GAMES AND ECONOMIC BEHAVIOR 1, 295–326 (1989)

Collective Dynamic Consistency in Repeated Games*

B. DOUGLAS BERNHEIM

Department of Finance, Northwestern University, Evanston, Illinois 60208

AND

DEBRAJ RAY

Indian Statistical Institute, New Delhi, 110 016, India

We formalize the notion of collective dynamic consistency for noncooperative repeated games. Intuitively, we require that an equilibrium not prescribe any course of action in any subgame that players would jointly wish to renegotiate, given the restriction that any alternative must itself be invulnerable to subsequent deviations and renegotiation. While the appropriate definition of collective dynamic consistency is clear for finitely repeated games, serious conceptual difficulties arise when games are repeated infinitely. We investigate several alternative solution concepts, and establish existence (under reasonably general conditions) for each.

1. INTRODUCTION

Selten's (1965, 1975) notion of subgame perfection requires dynamically consistent behavior on the part of each participant in a repeated game. Unfortunately, well-known "folk" theorems imply that, in general, the set of perfect equilibria is vast (see, e.g., Fudenberg and Maskin (1986) and Benoit and Krishna (1985) for discussions of infinite and finite horizon games, respectively). This observation has led most analysts to

* This work was supported by the National Science Foundation Grant SES-8404164 at the Institute for Mathematical Studies in the Social Sciences, Stanford University. We are also grateful for partial support from the Warshow endowment of Cornell University. We would like to thank Eric van Damme and Geir Asheim for helpful comments.

0899-8256/89 $3.00

impose additional refinements in specific applications. The most common practice is to consider "cooperative" equilibria—those that are Pareto efficient within the class of perfect equilibria. In imposing this further refinement, one implicitly assumes a kind of collective rationality. Specifically, if players bargain with complete information over possible outcomes subject to any set of constraints, one normally assumes that all feasible Pareto improvements are made. In noncooperative environments, the relevant constraints are incentive compatibility and individual dynamic consistency—the feasible set is defined by agreements that are self-enforcing both on and off the equilibrium path. Thus, it appears natural to study the most collusive perfect equilibria.

Unfortunately, this practice lacks internal consistency. Players typically have the opportunity to reexamine their self-enforcing agreement at any point during the course of play. Just as we require dynamically consistent behavior on the part of each individual, it is then essential to insist upon dynamic consistency at the *collective* level. For purposes of illustration, consider the one-shot game illustrated in Table I. There are two Nash equilibria (a_{12}, a_{22}) and (a_{13}, a_{23}). Since the first Pareto dominates the second, we would expect players to opt only for the first as a self-enforcing agreement. Now consider a single repetition of this game, and assume for simplicity that there is no discounting. In the two-stage game, there is a perfect equilibrium in which players cooperate (play (a_{11}, a_{21})) initially, and then play (a_{12}, a_{22}) in the terminal period, with any first-period deviation punished by reversion to (a_{13}, a_{23}). Yet each player knows that if either actually deviated in period 1, each would have an incentive post hoc to renegotiate the original agreement, and play (a_{12}, a_{22}) instead. If players actually have the opportunity to discuss strategies

TABLE I
ILLUSTRATION OF COLLECTIVE DYNAMIC CONSISTENCY

		Player II		
		a_{21}	a_{22}	a_{23}
Player I	a_{11}	3,3	0,4	0,0
	a_{12}	4,0	2,2	0,0
	a_{13}	0,0	0,0	1,1

at each stage, then it seems natural to rule out the cooperative equilibrium, on the grounds that it entails dynamically inconsistent behavior at the collective level.

Collective dynamic consistency might well rule out many equilibria commonly considered by applied game theorists. For example, grim strategies typically make all players strictly worse off relative to the equilibria which they support. On the other hand, collective dynamic consistency need not rule out cooperation: for instance, Rubinstein's (1979) notion of a strong perfect equilibrium is not vulnerable to renegotiation, and it is known that certain games admit cooperation as strong perfect equilibria. Unfortunately, Rubinstein's notion is more demanding than necessary for the purpose of imposing collective dynamic consistency, and as a result often fails to exist.

In this paper, we define and investigate several notions of collective dynamic consistency for repeated games. After describing an analytical framework and notation in Section 2, we consider finitely repeated games in Section 3. Intuitively, our concept requires that an equilibrium not prescribe any course of action in any subgame that players would jointly wish to renegotiate (given the restriction that any alternatives must themselves be invulnerable to subsequent renegotiation). This intuition suggests a recursive definition of the refinement. However, for infinitely repeated games, the recursive approach is invalid. Indeed, defining collective dynamic consistency becomes problematic. In Section 4, we discuss these difficulties, propose a specific notion of consistency, and demonstrate existence under relatively general conditions. In Section 5, we note that our formulation of consistency might rule out some apparently attractive equilibria, and it might fail to rule out some unattractive ones. We therefore propose two alternative refinements—*minimal consistency* and *simple consistency*—that yield more satisfactory results in the cases considered. We also establish existence under relatively general conditions. Section 6 contains a simple example, in which the requirement of collective dynamic consistency isolates interesting subsets of perfect equilibria. Section 7 describes the relationship between this paper and other work on renegotiation in noncooperative environments.

2. Preliminaries

2.1. *One-Shot Games*

Consider a one-shot simultaneous move game. The *player set* is $N = \{1, \ldots, n\}$. For $i \in N$, A_i is the *action set* of player i. This corresponds to player i's strategies in the one-shot game. Write $A \equiv X_{i \in N} A_i$. An *action* for player i is a choice $a_i \in A_i$. Write $a \equiv (a_1, \ldots, a_n)$ (so $a \in A$). The

payoff function of i is a real valued function π_i on A. The collection $G \equiv \{(A_i)_{i\in N}, (\pi_i)_{i\in N}\}$ is the *one-shot game*.

We make the following assumptions on G.

(A.1) A_i is compact for each $i \in N$

(A.2) $\pi_i: A \to \Re$ is continuous for each $i \in N$

(A.3) G has a Nash equilibrium in pure strategies.[1]

2.2. *Repeated Games*

We shall now *repeat* the game G for time periods $0, \ldots, T$. T is the horizon. It may be finite or infinite, respectively giving rise to a *finitely repeated game* or an *infinitely repeated game*.

Suppose that player i, $i \in N$, has a *discount factor* $\delta_i \in (0, 1)$. We denote by G^T the game formed by playing G during periods $0, \ldots, T$, and writing payoffs as a discounted sum of one-shot payoffs. More precisely, for any *action profile* $\alpha \equiv (a_t)_0^T$, where $a_t \in A$, the payoff to player i is evaluated by the expression

$$\Pi_i(\alpha) \equiv \sum_{t=0}^{T} \delta_i^t \pi_i(a_t).$$

Observe that $\Pi_i(\alpha)$ can be viewed as a real valued function on the space of all action profiles A^T, continuous in the topology of pointwise convergence.

A *t-history* h_t is defined for any $t = 1, \ldots, T$ as the sequence of previous actions. A typical t-history is of the form

$$h_t = (a_0, \ldots, a_{t-1}), \qquad t = 1, \ldots, T$$

Clearly, A^t is the set of all possible t-histories. A *history* h is a sequence of t-histories: $h \equiv (h_1, h_2, h_3, \ldots)$.

A *strategy* for player i is a sequence of functions $\psi_i \equiv \langle \psi_{it} \rangle_0^T$ such that

$$\psi_{i0} \in A_i$$

and

$$\psi_{it}: \quad A^t \to A_i, \qquad t \geq 1$$

[1] We are using the interpretation that the elements of A_i are actions involving no randomization. Of course, we could just as well think of these elements as mixed strategies, though the interpretation of this in a repeated situation is somewhat forced.

Let Ψ_i^T be the set of all strategies for i, and define $\Psi^T \equiv \bigtimes_{i \in N} \Psi_i^T$. Write $\psi \equiv (\psi_i)_{i \in N} \in \Psi^T$. We shall call ψ a *strategy profile*.

Given $\psi \in \Psi^T$, let $\alpha(\psi)$ be the action profile induced (in the obvious way) by ψ. More generally, define for any ψ and t-history h_t, $\alpha^t(\psi, h_t)$ as the action profile $(a_t, \ldots, a_T)$ starting from time t and induced by the t-history h_t and a subsequent application of ψ. When $T = \infty$, this can be identified easily with an action profile $\langle a_t' \rangle_0^\infty$ starting from $t = 0$, by putting $a_s' = a_{t+s}$ for all $s \geq 0$. When $T = \infty$, we shall omit the superscript T on Ψ^T.

Throughout, we will use the notation that for any n-tuple $(x_1, \ldots, x_n)$, x_{-i} denotes the $(n-1)$-tuple $(x_1, \ldots, x_{i-i}, x_{i+1}, \ldots, x_n)$.

3. Consistent Equilibria in Finitely Repeated Games

Our analysis of finitely repeated games makes use of some additional notation. For each α^t, denote discounted payoffs as

$$\Pi_i^t(\alpha^t) = \sum_{\tau=t}^{T} \delta^{\tau-t} \pi_i(a_\tau).$$

For each $\psi \in \Psi^T$, t such that $0 < t \leq T$, and t-history h_t, let

$$V_i(\psi) = \Pi_i^0(\alpha(\psi))$$

and

$$v_i^t(\psi, h_t) = \Pi_i^t(\alpha^t(\psi, h_t)).$$

The strategy profile $\psi^* \in \Psi^T$ is a *subgame perfect Nash equilibrium* (abbreviated PE) if for each i, t, and t-history h_t, we have

$$V_i(\psi^*) \geq V_i(\psi_i, \psi_{-i}^*)$$

and

$$v_i^t(\psi^*, h_t) \geq v_i^t(\psi_i, \psi_{-i}^*, h_t)$$

for all $\psi_i \in \Psi_i^T$.

We define a *consistent* equilibrium by induction on the number of stages in the game. For $T = 0$, let $\overline{\Psi}^0$ be the set of $\psi^* \in \Psi^0$ such that, for all i and $\psi_i \in \Psi_i^0$,

$$V_i(\psi^*) \geq V_i(\psi_i, \psi_{-i}^*).$$

$\overline{\Psi}^0$ is simply the set of Nash equilibria in the one-shot game. Next, define the consistent equilibrium set $\hat{\Psi}^0$ as $\psi^* \in \overline{\Psi}^0$ such that for all $\psi \in \overline{\Psi}^0$,

$$V_i(\psi^*) \geq V_i(\psi)$$

for some i. $\hat{\Psi}^0$ is simply the weak Pareto frontier of the Nash set. Intuitively, in a one-shot game players would always renegotiate to the Pareto frontier of the set of self-enforcing agreements. Clearly, assumptions (A.1) through (A.3) guarantee that $\hat{\Psi}^0$ is nonempty.

Now consider the t-stage game. Define the set $\overline{\Psi}^t$ as $\psi^* \in \Psi^t$ such that for all i and $\psi_i \in \Psi_i^t$,

$$V_i(\psi^*) \geq V_i(\psi_i, \psi^*_{-i})$$

and, for all h_1, $\langle \psi^*_\tau(\cdot, h_1)\rangle_{\tau=1}^t \in \hat{\Psi}^{t-1}$. Intuitively, we have already defined the consistent set $\hat{\Psi}^{t-1}$, and must require that all continuation strategy profiles lie in this set, or players would renegotiate upon reaching period 1. Finally, define the consistent set $\hat{\Psi}^t$ as $\psi^* \in \overline{\Psi}^t$ such that for all $\psi \in \overline{\Psi}^t$,

$$V_i(\psi^*) \geq V_i(\psi)$$

for some i. That is, we look at the Pareto frontier of $\overline{\Psi}^t$—otherwise, players would renegotiate in period 0. Once again, existence follows in a straightforward way from (A.1) through (A.3).

It is immediate from definitions that every consistent equilibrium is subgame perfect, but not vice versa. It is, for example, clear that the game considered in the introduction (Table I) has only one consistent equilibrium, in which no cooperation occurs.

The notion of a consistent equilibrium for finitely repeated games was developed in an earlier draft of this paper (Bernheim and Ray, 1985). The concept also appears under different names in van Damme (1987) and Farrell and Maskin (1987). Bernheim *et al.* (1987) noted that this notion may not be entirely satisfactory for games involving more than two players, in that we have restricted attention to renegotiation by the set of all players. Their notion of a perfectly coalition proof equilibrium coincides with consistency in two-player, finite horizon games. Bernheim and Whinston (1987) provided an example in which consistency isolates subgame perfect equilibria of particular interest (the equilibria are cyclical, despite the fact that the game also admits noncyclical subgame perfect equilibria). Bernheim and Ray (1987) described a more elaborate example along these lines in an economic context. More recently, Benoit and Krishna (1988) provided an analysis of the behavior of consistent equilibrium payoffs in an undiscounted, finitely repeated game, as the horizon

tends to infinity. (Their refinement is slightly different in the choice of the Pareto frontier.)

While the definition of collective dynamic consistency for finitely repeated games is not controversial (at least for games with two players), its extension to infinitely repeated games has proven problematic. In the remaining sections, we focus exclusively on infinitely repeated games.

4. Consistent Equilibria in Infinitely Repeated Games

In this section, we motivate and define the notion of collective dynamic consistency for infinitely repeated games. Our main result concerns the existence of consistent equilibria.

4.1. *Conceptual Issues and a Definition of Consistency*

For infinitely repeated games, the definition of consistency given in Section 3 is inapplicable. The reason is simple: there is no finite horizon from which one can perform a backward recursion to isolate consistent equilibria. Consequently, we must look for a nonrecursive definition of collective dynamic consistency.

The key motivation for our definition of collective dynamic consistency is the observation that, for infinitely repeated games, all subgames are identical. Any outcome that is possible in one subgame is also possible in every other subgame. Thus, an equilibrium implies a description of a set of outcomes that are possible in every subgame. At the beginning of each period, irrespective of history, players should have the option of collectively renouncing their prescribed strategies, and adopting the strategies prescribed for any other subgame. At a minimum, collective dynamic consistency should imply that players would never find it in their joint interests to exercise this option.

Below, we refer to this requirement as *internal consistency*. We also argue that internal consistency does not by itself capture the full implications of collective dynamic consistency. In particular, if one internally consistent equilibrium Pareto dominates another in every subgame, one would expect players to agree upon the first. This motivates a notion of *external consistency*. A *consistent* equilibrium is both internally and extremely consistent.

We formalize these notions as follows. Recall the description of a repeated game in Section 2, with $T = \infty$. Define, for each $i \in N$, $V_i: \Psi \to \Re$ by

$$V_i(\psi) \equiv \Pi_i(\alpha(\psi)), \tag{4.1}$$

and v_i: $\Psi \times A^t \to \Re$ by

$$v_i(\psi, h_t) \equiv \Pi_i(\alpha^t(\psi, h_t)), \qquad i \in N. \tag{4.2}$$

V_i is i's payoff function in the infinitely repeated game G^∞. Write $V \equiv (V_1, \ldots, V_n)$ and $v \equiv (v_1, \ldots, v_n)$.

A strategy profile $\psi^* \in \Psi$ is a *subgame perfect Nash equilibrium* if for each i and every t-history h_t, we have

$$V_i(\psi^*) \geq V_i(\psi_i, \psi^*_{-i}), \qquad \psi_i \in \Psi_i \tag{4.3}$$

and

$$v_i(\psi^*, h_t) \geq v_i(\psi_i, \psi^*_{-i}, h_t) \qquad \psi_i \in \Psi_i \tag{4.4}$$

for all $\psi_i \in \Psi_i$.

Let $E \equiv \{p \in \Re^n | p = V(\psi) \text{ for some PE } \psi\}$. Clearly, E is nonempty by (A.3).[2] Let Ξ be the power set of E. A set $P \in \Xi$ is *internally consistent* (IC) if P is nonempty and

(c.1) $p \in P$ implies that there is a PE ψ with $V(\psi) = p$ and with the property that for every t-history h_t,

$$v(\psi, h_t) \in P,$$

(c.2) for no $p, p' \in P$ is it the case that

$$p \gg p'.$$

Two observations about IC sets are worth making right away. First, IC sets rule out equilibria that support cooperative behavior in games by the threat of reverting to "grim" strategies in the case of a deviation. For example, the well-known method of supporting collusive outcomes in repeated oligopoly games by threatening to revert to one-shot Nash behavior is unacceptable under this approach. If collusive behavior *is* supportable to start with (by whatever means), and subsequently a deviation occurs, it is difficult to imagine that players who are communicating or negotiating at every stage will revert permanently to a Pareto inferior one-shot equilibrium when a "better" equilibrium is available. Of course, this may mean that the better equilibrium may not be an equilibrium to start with!

[2] Our only use of (A.3) is to guarantee nonemptiness.

Second, IC sets do not *necessarily* rule out all forms of cooperative behavior. One easy example is Rubinstein's (1979) method of supporting collusion in the prisoner's dilemma as a *strong* perfect equilibrium outcome. The set formed by collecting all the equilibrium payoffs (in the original game and in continuation games) under the strong equilibrium is indeed IC.

Nevertheless, an IC set may not capture *all* of our intuition regarding collective dynamic consistency. The simplest way to see this is to note that the *singleton* set consisting of the payoff vector arising from an (infinite) repetition of any one-shot equilibrium is always IC, despite the fact that *another* internally consistent equilibrium might be "superior" to it (e.g., an equilibrium formed by repeating a Pareto superior one-shot equilibrium).

We therefore need additional restrictions to capture the notion of collective dynamic consistency. Some additional notation is required. Let $\Pi \subseteq \Xi$ be the set of all IC sets. Clearly, Π is nonempty by (A.3) and the discussion in the preceding paragraph. Now suppose $P, P' \in \Pi$. We will say that P *directly dominates* P' (written P d P') if there is $p \in P$, $p' \in P'$ such that $p \gg p'$. (Note that in principle, it is quite possible that both P d P' and P' d P.)

This definition of dominance is motivated by the following considerations. Suppose that the players contemplate playing a perfect equilibrium, ψ', with its payoffs yielding some IC set, P'. Suppose further that there is another equilibrium ψ, the payoffs of which form an IC set P, such that for some $p \in P$ and $p' \in P'$, $p \gg p'$. By choosing ψ', players assert that in some subgame they will play strategies that yield payoffs p'. Yet if they ever reached this subgame, they might consider renouncing their strategies, choosing instead the subgame strategies from ψ that yield p. Since these new subgame strategies are internally consistent, players might have good reasons to believe that p is in fact achievable. An incentive to renounce the original strategies would then exist. Recognizing this in advance, players would realize that ψ' may not prescribe a credible course of action, or at the very least, that ψ' may be "threatened" in some subgame.

The preceding discussion suggests one possible notion of external consistency. Say that an IC set satisfies *strong consistency* if there is no other IC set that directly dominates it.

Unfortunately, there exist games with multiple IC sets, *none* of which satisfy strong consistency. This possibility is illustrated by Fig. 1. Let $A = (A_1, A_2)$ and $B = (B_1, B_2)$. Suppose that A and B are both IC sets. A d B and B d A, so neither set is strong consistent.

The diagram is only indicative of the problem, however, and does not constitute a concrete example. By restricting attention to pure strategies,

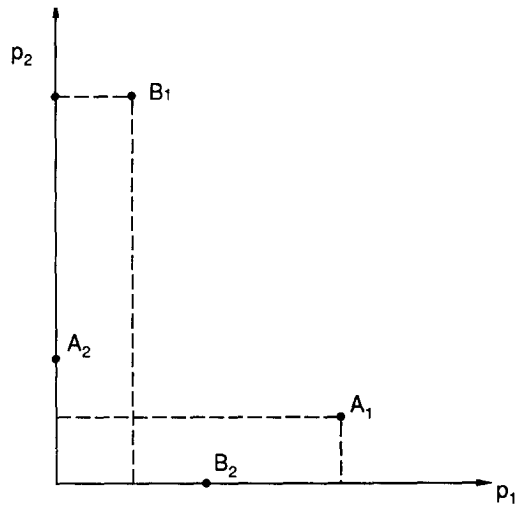

FIG. 1. Failure of strong consistency.

it is relatively easy to construct situations in which IC sets exist, but strong consistent sets do not. We provide the following example.

EXAMPLE 1. Consider the bimatrix game depicted in Table II. Each player chooses either a_1, b_1, a_2, or b_2. In each cell, I's payoff is listed first. Suppose that this game is repeated infinitely and that $\delta = \frac{1}{9}$. We will characterize the IC sets and show that, for sufficiently large F, none of them satisfy strong consistency. For the moment, we restrict attention to pure strategies (mixed strategies are considered in turn).

First, we argue that, for F sufficiently large, no off-diagonal element can occur along any pure strategy equilibrium path. This follows trivially from the observation that repeated play of b_2 guarantees each player a payoff of 0.

Now consider some perfect equilibrium. Suppose that (a_1, a_1) occurs along the equilibrium path. By deviating in the corresponding period, agent I can increase his payoff by 1. Thus, for this action pair to occur along the equilibrium path, it must be the case that

$$1 \leqq (v^c - v^p)/9,$$

or $v^c - v^p \geqq 9$, where v^c denotes I's payoff in the continuation game, and v^p denotes his payoff on the punishment path. Note that the upper bound on v^c is $(1 - \delta)^{-1}\, 8 = 9$ (we know that (b_2, a_1) can never occur) and that this bound is achieved *only* by repeating (a_1, a_1) in every period. Note also that $v^p \geqq 0$ (since I can play b_2) and that this bound can be achieved *only*

TABLE II
PAYOFF MATRIX FOR EXAMPLE 1

		Player II			
		a_1	b_1	a_2	b_2
Player I	a_1	8, 2	-F, -F	-F, -F	-F, 2
	b_1	-F, -F	2, 8	-F, -F	-F, 9
	a_2	-F, -F	-F, -F	0, 3	-F, 0
	b_2	9, -F	2, -F	0, -F	3, 0

by repeating (a_2, a_2). Thus, if (a_1, a_1) *ever* occurs on the equilibrium path, it also occurs in all subsequent periods, and any deviation is punished by permanent reversion to (a_2, a_2). It is trivial to check that this is, in fact, an IC equilibrium. The corresponding IC set, denoted A, is simply $\{(9, \frac{9}{4}), (0, \frac{27}{8})\}$. A completely symmetric argument applies for (b_1, b_1). Let B denote the corresponding IC set, $\{(\frac{9}{4}, 9), (\frac{27}{8}, 0)\}$. Note that A d B d A.

The only other candidates for perfect equilibrium paths either include one of these two IC sets (i.e., consist of a finite sequence of (a_2, a_2) and (b_2, b_2), followed by one of the equilibria described above) or simply consist of sequences of (a_2, a_2) and (b_2, b_2). In the latter case, let C denote a corresponding IC set. It is immediate that either A d C or B d C. We conclude that, within the class of pure strategy equilibria, there does not exist $P \in \Pi$ such that for all $P' \in \Pi$, P' does not dominate P.

Consideration of mixed strategies does not alter this result. We will briefly sketch the argument. For sufficiently large F, all equilibria entail actions that are "almost" pure in every period. Indeed, the outcome must lie on the diagonal with probability close to 1 (again, this follows from the fact that b_2 guarantees a payoff of 0). If we restrict attention to sequences of mixtures that are in the neighborhood of (a_2, a_2) or (b_2, b_2), then the resulting outcomes will be dominated by the pure strategy equilibria discussed above. Therefore, we need only consider equilibria that involve mixtures in the neighborhood of (a_1, a_1) (a symmetric argument applies to mixtures in the neighborhood of (b_1, b_1)).

Using an argument analogous to that employed for pure strategy equilibria, one can show the following: if the equilibrium prescribes a mixture around (a_1, a_1) in any period, it must prescribe a mixture in the neighbor-

hood of (a_1, a_1) in every subsequent period. Moreover, I's punishment must entail v_p close to 0.

It is certainly possible to construct a punishment with this property by using mixtures in the neighborhood of (a_2, a_2). However, unless the punishment also yields II a discounted payoff in excess of 9, it will be dominated by the pure strategy equilibrium, B. It follows that, for this equilibrium to escape domination, the punishment must employ a mixture around (b_1, b_1).

Since this punishment provides I with a discounted payoff near 0, and since the continuation payoff can never be below 0 in any period, I's payoff in the first period cannot exceed some small number $\varepsilon > 0$. Since the mixture is in the neighborhood of (b_1, b_1), I's total gains from deviating to b_2 permanently are close to 2. To discourage this deviation, one would have to provide a punishment path along which I's payoffs are roughly -18. This is clearly infeasible, since I can guarantee himself 0.

While this nonexistence result is discouraging, we would argue that, in any case, strong consistency is too strong. When players contemplate a joint deviation from one equilibrium to another, they must be convinced that the second equilibrium would actually prevail and that additional joint deviations would not subsequently occur. The mere fact that the second equilibrium is IC does not rule such deviations out. Note in particular that, in Fig. 1 (and in the example), both IC sets directly dominate each other. Suppose that the players contemplate playing equilibrium strategies ψ associated with the IC set A. Let ψ' be equilibrium strategies associated with the IC set B. For some subgame, players could make a Pareto improvement by shifting to the appropriate subgame strategies in ψ'. If players regard ψ' as a credible equilibrium, then they will not regard ψ as credible. Of course, the relationship between sets A and B is entirely symmetric, so if players regard ψ as a credible equilibrium, they will not regard ψ' as credible. Thus, there are two self-fulfilling sets of beliefs: either ψ is credible and ψ' is not, or ψ' is credible and ψ is not. Accordingly, we would regard both ψ and ψ' as consistent.

This discussion motivates a weaker requirement. We might call an IC set "consistent" if it directly dominates every IC set that directly dominates it. Unfortunately, even this definition is too strong to guarantee the existence of such sets. We provide the following example.

EXAMPLE 2. Consider the three-player game depicted in Table III. Player I chooses row, player II chooses column, and player III chooses box. Each player's action set is given by $\{a_1, b_1, c_1, a_2, b_2, c_2\}$. In each cell, I's payoff is listed first, II's second, and III's third. Suppose we repeat this game infinitely and that $\delta = \frac{1}{9}$.

This game is essentially the three-player counterpart to Example 1. There are three equilibria of interest. In each, (x_1, x_1, x_1) is repeated forever, and punishment of an opportunistic deviation results in perma-

nent reversion to (x_2, x_2, x_2), for $x = a$, b, and c. Let A, B, and C denote the associated IC sets. Note that A d C d B d A. Arguing as in Example 1, one can show that all other IC sets are dominated by either A, B, or C. We leave the details to the reader.

We would argue that, on theoretical grounds, even this weaker notion of consistency is too demanding. After all, it requires the candidate IC equilibrium to survive challenges from all other IC equilibria, while these other equilibria need only survive challenges from the candidate equilibrium. To put it another way, if none of the other IC equilibria satisfy the external consistency requirement, then none of them are credible alternatives. The candidate equilibrium would then be externally consistent.

It is helpful to study this problem in the context of Example 2. A, B, and C are IC sets, and the only dominance relations are A d B d C d A. Thus, the relationship between these three sets is completely symmetric, and none satisfies either notion of external consistency defined above.

Given the symmetry of this configuration, one has only two options: reject A, B, and C as inconsistent, or accept them all as consistent. The first option leads to a paradox. If none of the IC equilibria are externally consistent, then there are no credible alternatives to any one of them, in which case all of them must be externally consistent.

At first, it might appear that the second option (accepting all three as consistent) poses the same paradox in reverse. We resolve this problem as we did for the case depicted in Fig. 1: one need not take all three sets to be credible *simultaneously*. Suppose that we take the equilibrium corresponding to one IC set to be credible. In our view, every other IC set that is dominated (either directly or indirectly) by the original set is thereby rendered not credible. Thus, if we take A to be credible, we rule out B and C through domination. Players would then never deviate from the equilibrium corresponding to A in any subgame, as the desirable alternative (a subgame strategy from C) would not be considered viable. Likewise, the belief that either B or C is credible is also self-justifying. Accordingly, we would accept all three sets as consistent. Our theory is, however, silent on which of these sets would actually prevail.

These considerations motivate a third notion of external consistency. To formalize this notion, we require some additional notation. We will say that P *dominates* P' (written P d* P') if there are finitely many elements of Π, say $P_1, \ldots, P_m$, such that

$$P \text{ d } P_1 \text{ d } P_2 \text{ d } \cdots \text{ d } P_m \text{ d } P'.$$

P *does not dominate* P' if it is not true that P d* P'.

The following criterion embodies our notion of external consistency. $P \in \Xi$ is *externally consistent* (EC) if P is nonempty and

(c.3) For every $P' \in \Pi$ such that P' d* P, P d* P'.

TABLE III
PAYOFF MATRIX FOR EXAMPLE 2

		Player II					
	a_1	a_1	b_1	c_1	a_2	b_2	c_2
Player I	a_1	8,8,2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	b_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	c_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	a_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	b_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	c_2	9,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,0,-F

		Player II					
	b_1	a_1	b_1	c_1	a_2	b_2	c_2
Player I	a_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	b_1	-F,-F,-F	2,8,8	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,9,-F
	c_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	a_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	b_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	c_2	0,-F,-F	2,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,0,-F

		Player II					
	c_1	a_1	b_1	c_1	a_2	b_2	c_2
Player I	a_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	b_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	c_1	-F,-F,-F	-F,-F,-F	8,2,8	-F,-F,-F	-F,-F,-F	-F,2,-F
	a_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	b_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
	c_2	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,0,-F

If some $P \in \Xi$ is both internally and externally consistent (i.e., if it satisfies (c.1), (c.2), and (c.3)), then we will say that it is *consistent*. We will often refer to the strategy profile that supports any element of a consistent set as a *consistent equilibrium*.

TABLE III—*Continued*

a_2

Player I \ Player II	a_1	b_1	c_1	a_2	b_2	c_2
a_1	-F,-F, 2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
b_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
c_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
a_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	0,3,3	-F,-F,-F	-F,0,-F
b_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
c_2	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,0,-F

b_2

Player I \ Player II	a_1	b_1	c_1	a_2	b_2	c_2
a_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
b_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
c_1	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
a_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,0,-F
b_2	-F,-F,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F	3,0,3	-F,0,-F
c_2	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,-F,-F	0,0,-F

c_2

Player I \ Player II	a_1	b_1	c_1	a_2	b_2	c_2
a_1	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,0,0
b_1	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,0,0
c_1	-F,-F,0	-F,-F,0	-F,-F,9	-F,-F,0	-F,-F,0	-F,0,0
a_2	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,0,0
b_2	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,-F,0	-F,0,0
c_2	0,-F,0	0,-F,0	0,-F,0	0,-F,0	0,-F,0	3,3,0

It is worth noting that if P is consistent and $\hat{P}$ d* P, then $\hat{P}$ is also consistent (this is easily checked). If in particular P'' d P' d P and P is consistent, then P'' is also consistent (as is P'). This justifies our practice of describing P' as not credible on the grounds that P'' is a credible alternative.

4.2. *Existence of a Consistent Set*

Our main result is

THEOREM 1. *There exists a consistent set.*

Proof of the Theorem. For the proof, we need some additional notation and a number of lemmas.

Consider any $P \in \Pi$, and $p \in P$. Pick some PE strategy profile ψ which fulfills condition (c.1) for p. We will say that $\alpha \equiv \alpha(\psi)$ is an action profile *associated with* p. Let $\alpha \equiv (a_0, a_1, a_2, \ldots)$.

Define a *payoff sequence associated with* p, $\langle p_t \rangle_0^\infty$, by

$$p_t = \left(\sum_{\tau=t}^{\infty} \delta^{\tau-t} \pi_i(a_t) \right)_{i \in N}, \qquad t \geq 0,$$

where $(a_0, a_1, \ldots)$ is an action profile associated with p.

Finally, h defined from α is an *associated* history.

Next, we define *simple strategy profiles* (Abreu, 1988). Consider an $(n+1)$-tuple of action profiles $(\alpha(0), \ldots, \alpha(n))$. Informally, a simple strategy profile $\psi(\alpha(0), \ldots, \alpha(n))$ requires

(a) the playing of $\alpha(0)$ until some player deviates singly from $\alpha(0)$ by choosing a different action in some period,

(b) for any $j \in N$, the playing of $\alpha(j)$ should the jth player deviate singly from $\alpha(i)$, an ongoing previously specified action profile; continue with $\alpha(i)$ if no deviations from $\alpha(i)$ occur or if two or more players deviate simultaneously.[3]

LEMMA 1. *Let $\mathbf{C} \subseteq \Xi$ be the set of all nonempty closed subsets of E. Then, equipped with the Hausdorff metric, $\mathbf{C}$ is compact.*[4]

Proof. Note first that E is compact. A direct proof is easy (using (A.1) and (A.2)) though the reader may consult Abreu (1988). The compactness of $\mathbf{C}$ in the Hausdorff metric is a well-known consequence (see, e.g., Hildenbrand, 1974). Q.E.D.

Define $\overline{\Pi} \subseteq \Pi$ as the set of all IC sets that are closed (in the Euclidean metric). $\overline{\Pi}$ is nonempty, by (A.3). Obviously, $\overline{\Pi} \subseteq C$. Choose, for each $P \in \overline{\Pi}$ and $i \in N$,

$$p(i) \in \arg\min \{p_i \mid p \in P\}. \tag{4.5}$$

[3] This informal definition is taken directly from Abreu (1988). A formal definition is easy to construct, though tedious to write down. The interested reader may consult Abreu (1988).

[4] See, e.g., Hildenbrand (1974) for a definition of the Hausdorff metric.

Of course, $p(i)$ may not be unique but this does not matter. Denote by $\alpha(i)$ an action sequence associated with $p(i)$, for $i \in N$.

LEMMA 2. *Let $P \in \overline{\Pi}$, $p \in P$ and let α be an action profile associated with p. Then the simple strategy profile $\psi(\alpha, \alpha(1), \ldots, \alpha(n))$ is a PE.*

Proof. For notational ease, let $p(0) \equiv p$ and $\alpha(0) \equiv \alpha$. Now consider $\psi(\alpha(0), \alpha(1), \ldots, \alpha(n))$. By Abreu (1988, Proposition 1), $\psi(\alpha(0), \ldots, \alpha(n))$ is a PE iff for all $i = 0, \ldots, n$, $\psi(\alpha(i), \alpha(1), \ldots, \alpha(n))$ is a Nash equilibrium.

Define, for any i and $a_{-i} \in X_{\substack{j \in N \\ j \neq i}} A_j$,

$$D_i(a_{-i}) \equiv \operatorname*{Max}_{a_i \in A_i} \pi_i(a_i, a_{-i}). \tag{4.6}$$

This is the maximal one-shot payoff possible for i when others choose the tuple of actions a_{-i}. Now, for each $k = 0, \ldots, n$, $\alpha(k)$ is an action sequence associated with $p(k)$. So for each t and each i, there exists $p(t, i) \in P$ such that

$$\begin{aligned} \sum_{\tau=t}^{\infty} \delta_i^{\tau-t} \pi_i(a_t(k)) &\geq D_i(a_{t,-i}(k)) + \delta_i p_i(t, i) \\ &\geq D_i(a_{t,-i}(k)) + \delta_i p_i(i). \end{aligned} \tag{4.7}$$

The first inequality in (4.7) follows from the definition of an associated action sequence (and also using (c.1)), and the second follows from the definition of $p(i)$, $i \in N$.

But (4.7) immediately implies that for each i, $\psi(\alpha(i), \alpha(1), \ldots, \alpha(n))$ is a Nash equilibrium. So $\psi(\alpha(0), \alpha(1), \ldots, \alpha(n))$ is a PE. Q.E.D.

LEMMA 3. *$\overline{\Pi}$ is compact.*

Proof. Because $\overline{\Pi} \subseteq C$, it suffices to prove (given Lemma 1) that $\overline{\Pi}$ is closed. To this end, let P^q be a sequence in $\overline{\Pi}$ with $P^q \to P$ (in the Hausdorff metric). We are to show that $P \in \overline{\Pi}$. Clearly, because $\boldsymbol{C}$ is compact, P is closed. Moreover, it is easy to check that P satisfies (c.2). It remains to prove that P must satisfy (c.1). So fix any $p \in P$. We must find a PE ψ satisfying the conditions in (c.1).

For each P^q, we may choose $p^q(i)$, $i \in N$, as in (4.5) above, and their associated action profiles $\alpha^q(i)$, $i \in N$.

Now, given $p \in P$, we have (by Hausdorff convergence) a sequence $\langle p^q \rangle$, with $p^q \in P^q$ for all q and $p^q \to p$. For each q, let α^q be an action

profile associated with p^q. Next, use a diagonal argument to extract a subsequence of q (call it k) such that

$$P^k(i) \to p^*(i) \in P \tag{4.8}$$

$$\alpha^k(i) \to \alpha^*(i) \tag{4.9}$$

$$\alpha^k \to \alpha, \tag{4.10}$$

where the convergence in (4.9) and (4.10) is in the sense of pointwise convergence of sequences (we are using (A.1) here). The membership of the limit point $p^*(i)$ in P (see (4.8)) is a consequence of Hausdorff convergence.

Now, by (4.9) and (4.10) we have for each i, by using the continuity of discounted payoffs in the topology of pointwise convergence on action sequences (see (A.2) and the remarks in Section 2.2):

$$\begin{aligned} p_t^k(i) &\equiv \left(\sum_{\tau=t}^{\infty} \delta_j^{\tau-t}\pi_j(a_\tau^k(i))\right)_{j\in N} \to \left(\sum_{\tau=t}^{\infty} \delta_j^{\tau-t}\pi_j(a_\tau^*(i))\right)_{j\in N} \\ &\equiv p_t^*(i) \in P \end{aligned} \tag{4.11}$$

and similarly

$$\begin{aligned} p_t^k &\equiv \left(\sum_{\tau=t}^{\infty} \delta_j^{\tau-t}\pi_j(a_\tau^k)\right)_{j\in N} \to \left(\sum_{\tau=t}^{\infty} \delta_j^{\tau-t}\pi_j(a_\tau)\right)_{j\in N} \\ &\equiv p_t \in P. \end{aligned} \tag{4.12}$$

The L.H.S.'s of (4.11) and (4.12) are the tth terms of the associated payoff sequences of $p^k(i)$ and p^k, respectively. And the R.H.S.'s of (4.11) and (4.12) are the tth terms of the payoff sequences associated with $p^*(i)$ and p, respectively. By Hausdorff convergence and (c.1) applied to $\{p^k\}$, $p_t^*(i)$ and p_t belong to P for all t.

We are therefore done if we can show that the simple strategy profile $\psi(\alpha, \alpha^*(1), \ldots, \alpha^*(n))$ is a PE. Observe that by Lemma 2, we have for each k, t and j, because $\psi(\alpha^k, \alpha^k(1), \ldots, \alpha^k(n))$ is a PE,

$$\sum_{\tau=t}^{\infty} d_j^{\tau-t}\pi_j(a_\tau^k) \geq D_j(a_\tau^{k,-j}) + \delta_j p_j^k(j) \tag{4.13}$$

and additionally for each $i \in N$,

$$\sum_{\tau=t}^{\infty} d_j^{\tau-t}\pi_j(a_\tau^k(i)) \geq D_j(a_{\tau,-j}^k(i)) + \delta_j p_j^k(j). \tag{4.14}$$

Passing to the limit as $k \to \infty$ in (4.13) and (4.14), using the continuity of the L.H.S. in pointwise convergence and the continuity of $D_j(\cdot)$ for all j (the "maximum theorem," applying (A.2)), we have verified that $\psi(\alpha, \alpha^*(1), \ldots, \alpha^*(n))$ is a PE (Abreu, 1988, Proposition 1). This completes the proof. Q.E.D.

LEMMA 4. *For each* $P \in \overline{\Pi}$, *define* $L(P) = \{P' \in \overline{\Pi} | P \, d \, P'\}$. *Then* $L(P)$ *is open relative to* $\overline{\Pi}$.

Proof. Pick $P \in \overline{\Pi}$, and let $P' \in L(P)$. In particular, there is $p \in P$, $p' \in P'$ such that $p \gg p'$. Now there is $\varepsilon > 0$ such that for all $q \in B_\varepsilon(p')$ (the (Euclidean) open ball of radius ε around p'), we have $p \gg q$. Now, for this ε, construct an open ball of (Hausdorff) distance ε, $B_\varepsilon(P')$, around P'. It is immediate that if $P'' \in \overline{\Pi}$ and $P'' \in B_\varepsilon(P')$, P d P''. This establishes the lemma. Q.E.D.

LEMMA 5. *Let* $P \in \Pi$. *Then* closure$(P) \in \overline{\Pi}$.

Proof. Let $\bar{P}$ = closure(P), and define $p(i)$, $i \in N$ by (4.5). It is clear that $\bar{P}$ satisfies (c.2). We are to show that $\bar{P}$ satisfies (c.1). Pick $p \in \bar{P}$. If $p \in P$, we are done. If not, there is a sequence p^k in P such that $p^k \to p$.

Now consider $p(i)$, $i \in N$. If $p(i) \in P$, let $\alpha(i)$ be an associated action sequences. If for some i, $p(i) \notin P$, let p^m in P be chosen, $p^m \to p(i)$. Without loss of generality choose the sequence such that a corresponding sequence of associated action sequences, (α^m), converges (pointwise) (this can be done by noting that S is compact and using a diagonal argument). Call the pointwise limit $\alpha(i)$.

Return, now, to p and the sequence p^k that converges to it. Once again, choose the sequence so that (α^k), the corresponding sequence of action sequences, converges to some α. In a manner analogous to Lemma 2, it is easy to check that the simple strategy profile $\psi(\alpha^k, \alpha(1), \ldots, \alpha(n))$ is a PE for each k. One can then use a limiting argument (as in (4.13) and (4.14)) to conclude that $\psi(\alpha, \alpha(1), \ldots, \alpha(n))$ is a PE.

We are now home. By using arguments similar to that following (4.11) and (4.12) of Lemma 3, it is easy to verify that given p, $\psi(\alpha, \alpha(1), \ldots, \alpha(n))$ is a PE that does the job required by (c.1). Q.E.D.

LEMMA 6. *Suppose* $P, P' \in \Pi$. *If* P *dominates* P', *then* closure (P) *dominates* P'. *If* P *does not dominate* P', *then* P *does not dominate* closure (P').

Proof. Obvious. Q.E.D.

Proof of the Theorem. We shall show that a consistent set may be found in $\overline{\Pi}$ itself.

Suppose not. Then for each $P \in \overline{\Pi}$, there is $P'' \in \Pi$ such that (a) P'' d* P and (b) it is not true that P d* P''. By Lemmas 5 and 6, it must be the case that there is $P' \in \overline{\Pi}$ such that (a) P' d* P and (b) it is not true that P d* P'.

It follows, in particular, that

$$\bigcup_{P \in \overline{\Pi}} L(P) = \overline{\Pi}. \tag{4.15}$$

By Lemmas 3 and 4, and using the property of compactness, there are $P_1, \ldots, P_m \in \overline{\Pi}$ such that

$$\bigcup_{j=1}^{m} L(P_j) = \overline{\Pi}. \tag{4.16}$$

Pick P_1. By our supposition, there is $P_1' \in \overline{\Pi}$ such that P_1' d* P_1 and it is not true that P_1 d* P_1'. But $P_1' \in L(P_i)$ for some $i \neq 1$ (otherwise we have a contradiction). Now consider P_i. Again, there is $P_i' \in \overline{\Pi}$ such that P_i' dominates P_i but not vice versa. Continuing in this way, we get an infinite sequence of the form

$$\cdots P_k' \text{ d* } P_k \text{ d } P_i' \text{ d* } P_i \text{ d } P_1' \text{ d* } P_1.$$

But as there are only finitely many P_j's, there must be a repetition of at least one index. This is easily seen to yield a contradiction. Q.E.D.

5. Alternative Notions of Consistency in Infinitely Repeated Games

In this section, we discuss some conceptual problems with the notion of consistency described in Section 4. We then propose and analyze two alternative criterion—minimal consistency and simple consistency—that seem conceptually superior and perform better in the context of particular examples.

5.1. *Motivation for Alternative Refinements*

The central problem with consistency stems from the following observation: if P and P' are both internally consistent, and if *neither* P d P' *or* P' d P, then $P \cup P'$ is also internally consistent. The fact that two

TABLE IV
PAYOFF MATRIX FOR EXAMPLE 3

		Player II				
		a_1	b_1	a_2	b_2	c
Player I	a_1	16, 1	-F, -F	-F, -F	-F, -F	-F, 1
	b_1	-F, -F	$3,8+\varepsilon$	-F, -F	-F, -F	$-F, 9(1+\frac{\varepsilon}{8})$
	a_2	-F, -F	-F, -F	0, 8	-F, -F	-F, 0
	b_2	-F, -F	-F, -F	-F, -F	4, 0	-F, 0
	c	18, -F	3, -F	0, -F	0, -F	2, 4

unrelated sets can be joined together in this way causes us to retain some undesirable equilibria, and to reject some attractive ones. We illustrate these points via the following examples.

EXAMPLE 3. Consider the bimatrix game depicted in Table IV. Suppose that we repeat this game infinitely and that $\delta = \frac{1}{8}$. This is essentially an adaptation of Example 1. We have changed the numerical payoffs, and we have added a fifth choice, labeled c. Note that (c, c) is a pure strategy equilibrium for the static game.

For small $|\varepsilon|$, there are three equilibria of interest. The first two consist of repeating (x_1, x_1) forever and punishing opportunistic deviations by permanently reverting to (x_2, x_2), $x = a, b$. We will refer to the corresponding IC sets as A and B. The third equilibrium of interest consists of repeating (c, c) forever. We will refer to the corresponding IC set as C. Note that $A \cup C$ is also IC.[5] Henceforth, we will restrict attention to these four sets; arguing as in Example 1, one can show that this involves no loss of generality.

We illustrate the sets A, B, and C for $\varepsilon = 0$ in Fig. 2 (we depict normalized payoffs and adopt the convention that the point X_i represents the payoff vector associated with repeating the action pair (x_i, x_i),

[5] In fact, one could actually "mix" the equilibria associated with points in A and C. Specifically, choose any point in A, and, for the associated equilibrium, replace continuation strategies following any simultaneous deviation by two or more players with the equilibrium corresponding to C.

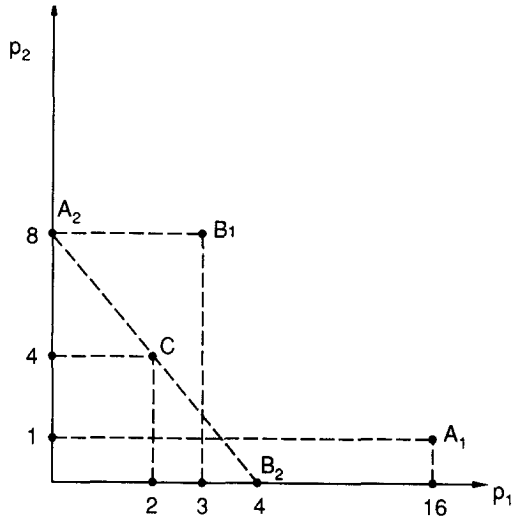

FIG. 2. IC sets for Example 3.

$X = A, B, i = 1, 2$, and similarly for the point C). Changing ε simply moves the point B_1 up or down.

Suppose initially that $\varepsilon < 0$. We have A d B d C, and A d B d $(A \cup C)$. Moreover, A is not dominated by any other IC set. Thus, only A is consistent.

Now suppose $\varepsilon > 0$. As before, A d B d C. However, we now have B d A as well. Thus, the improvement in B has rendered it externally consistent, as well as internally consistent. However, we also have $(A \cup C)$ d B d A. Thus, the set $A \cup C$ is now externally consistent *as well.* Our refinement therefore implies that repetitions of (c, c) can occur if and only if $\varepsilon > 0$.

This conclusion strikes us as odd. While raising ε certainly makes B more attractive, it does nothing for C. If anything, C becomes less desirable. When $\varepsilon > 0$, C survives, but not on its own merits. Rather, it is saved by an irrelevant association with A via $A \cup C$.

One might argue for inclusion of C on the following grounds. C is dominated only by B, and B is dominated by A. Thus, one can without inconsistency maintain the belief that both A and C are credible, while B is not. While this argument has some merit, it applies with equal force regardless of whether ε is greater or less than 0. It seems that one ought to either reject C in both cases or accept it in both cases. The refinement developed in this section always rules out C. One could always avoid ruling out C by adopting the second notion of external consistency discussed in Section 4. Unfortunately, for that notion, existence is problematic as we have already seen.

EXAMPLE 4. Consider the three-player game depicted in Table V. Players I and II chose from the set $\{a_1, b_1, a_2, b_2\}$ (rows and columns, respectively), while player III chose from $\{z_1, z_2\}$. Suppose that we repeat this game infinitely and that $\delta = \frac{1}{9}$. For simplicity, we confine our discussion to pure strategies.

Once again, this is an adaptation of Example 1. Indeed, if III chooses z_1, I and II's payoffs are exactly as in Example 1, and III always receives 0. On the other hand, if III chooses z_2, matters are quite different. Note that the action triplet (a_1, b_1, z_2) yields payoffs of (U, V, W). We will consider two different sets of values for these variables.

TABLE V
PAYOFF MATRIX FOR EXAMPLE 4

		Player II			
		a_1	b_1	a_2	b_2
z_1					
Player I	a_1	8,2,0	-F,-F,0	-F,-F,0	-F,2,0
	b_1	-F,-F,0	2,8,0	-F,-F,0	-F,9,0
	a_2	-F,-F,0	-F,-F,0	0,3,0	-F,0,0
	b_2	9,-F,0	2,-F,0	0,-F,0	3,0,0

		Player II			
		a_1	b_1	a_2	b_2
z_2					
Player I	a_1	-F,0,-F	U,V,W	-F,-F,-F	-F,-F,-F
	b_1	1,1,5	0,-F,-F	0,-F,-F	0,-F,-F
	a_2	-F,0,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F
	b_2	-F,0,-F	-F,-F,-F	-F,-F,-F	-F,-F,-F

The first case to consider is $U = V = W = -F$. There are three equilibria of interest. The first two consist of repeating (x_1, x_1, z_1) forever and punishing opportunistic deviations by reverting permanently to (x_2, x_2, z_1), $x = a, b$. We will refer to the corresponding IC sets as A and B. The third equilibrium consists of repeating (b_1, a_1, z_2) forever. We will refer to this IC set as C.

Note that we have A d B d A. Moreover, there is no dominance relationship between C and either A or B. Accordingly, A, B, and C are all consistent (incidentally, so are $A \cup C$ and $B \cup C$).

The second case to consider is $U = V = 1 - \varepsilon$, $W = 5 - \varepsilon$, for some small $\varepsilon > 0$. There is now a fourth equilibrium of interest, which consists of repeating (a_1, b_1, z_2) forever. Let D denote the corresponding IC set. Note that C d D, and it is *not* the case that D d C. Moreover, there is no dominance relationship between D and either A or B.

One would think that the addition of the "inferior" set D should have no impact on the attractiveness of A or B. Unfortunately, this is not the case. Note that $A \cup D$ and $B \cup D$ are also IC. Moreover, C d $(A \cup D)$ d B, C d $(B \cup D)$ d A, and neither A d* C or B d* C. Thus, in this second case, *only* C is consistent (even $A \cup C$ and $B \cup C$ are ruled out). Both A and B have been defeated via an apparently irrelevant association with D.

5.2. *Minimal Consistency*

We have seen in Examples 3 and 4 that our notion of consistency can rule out some desirable equilibria and fail to rule out undesirable ones. In the cases considered, the problem arises because the union of unrelated IC sets is also IC, as long as there is no dominance relationship between them. One possible solution is to demand that each candidate set stand or fall on the basis of its own merits, rather than on the basis of incidental associations with more or less meritorious sets. We pursue this possibility in the remainder of the current subsection.

We define the collection of *minimal internally consistent* (MIC) sets, Π^m, as follows. Let $\subset$ denote strict inclusion. Consider $P, P' \in \Pi$. Say that P is *properly included* in P' if closure$(P) \subset$ closure(P'). A set $P \in \Pi$ is *minimal internally consistent* if there is no collection of IC sets $\{P_\alpha\}$ such that P_α is properly included in P, for each α, and $\cup_\alpha$ closure$(P_\alpha) =$ closure(P). Let $\overline{\Pi}^m$ denote the set of minimal IC sets that are *closed*. Again, by (A.3), $\overline{\Pi}^m$ is clearly nonempty. For $P, P' \in \Pi^m$, we will say that P *m-dominates* P' (written P d^m P') if there are finitely many elements of Π^m, say $P_1, \ldots, P_m$, such that

$$P \text{ d } P_1 \text{ d } P_2 \text{ d } \cdots \text{ d } P_m \text{ d } P'.$$

P does not m-dominate P' if it is not true that $P\ d^m\ P'$. We now replace (c.3) with the following requirement. $P \in \Xi$ is *minimal externally consistent* (MEC) if P is nonempty and

(c.4) for every $P' \in \Pi^m$ such that $P'\ d^m\ P$, $P\ d^m\ P'$.

If some $P \in \Xi$ is both MIC and MEC, then we will say that it is a *minimal consistent* (MC) set.

Note that this refinement produces the desired results in the preceding examples. In Example 3, A is MC for $\varepsilon < 0$, and both A and B are MC for $\varepsilon > 0$. In Example 4, A, B, and C are all MC, even in the presence of the IC set D.

Our main result is

THEOREM 2. *There exists a minimal consistent set.*

We require three results in addition to Lemmas 1 through 5.

LEMMA 7. *Let* $P \in \Pi^m$. *Then* closure$(P) \in \overline{\Pi}^m$.

Proof. By Lemma 5, closure(P) is IC. Observing that closure(P) = closure(closure(P)), we are done. Q.E.D.

LEMMA 8. *Suppose* $P, P' \in \Pi^m$. *If* $P\ d^m\ P'$, *then* closure$(P)\ d^m\ P'$. *If P does not m-dominate P', then P does not m-dominate* closure(P').

Proof. Obvious. Q.E.D.

LEMMA 9. *Let* $P \in \overline{\Pi}$, *and* $p \in P$. *Then there exists* $P^* \in \overline{\Pi}^m$ *such that* $p \in P^*$.

Proof. Define $\Pi(p) = \{P' \in \overline{\Pi} | p \in P'\}$. Then $\Pi(p) \neq \varnothing$, because $P \in \Pi(p)$. Also, $\Pi(p)$ is a closed subset (in the Hausdorff metric) of $\overline{\Pi}$, and so by Lemma 3, $\Pi(p)$ is compact. Consider, now, any chain in $\Pi(p)$, that is, any subset Ω of $\Pi(p)$ totally ordered by $\supseteq$. Define $\overline{P} \equiv \cap_{P' \in \Omega} P'$. Because this is a nested intersection of nonempty compact sets, $\overline{P}$ is nonempty and compact. Moreover, $\overline{P} \subseteq P'$ for all $P' \in \Omega$. Also, note that for each integer $q \geq 1$, there is $P^q \in \Omega$ such that $h(\overline{P}, P^q) < 1/q$ where h is the Hausdorff distance). Therefore $P^q \to \overline{P}$, and so because $\Pi(p)$ is compact, $\overline{P} \in \Pi(p)$.

We have therefore shown that each totally ordered (by $\supseteq$) subset of $\Pi(p)$ has a lower bound in $\Pi(p)$. By Zorn's Lemma, there is $P^* \in \Pi(p)$ which is a minimal element under $\supseteq$; i.e., for no $P' \in \Pi(p)$ does $P' \subset P^*$ hold.

We claim that $P^* \in \overline{\Pi}^m$. Suppose not. Then closure$(P^*) = P^* = \cup_\alpha$ closure(P_α) for some collection $\{P_\alpha\}$, with P_α IC and properly included in P^* for all α. Because $p \in P^*$, $p \in$ closure$(P_\alpha) \equiv \overline{P}_\alpha$ for some α. By Lemma 5, $\overline{P}_\alpha \in \overline{\Pi}$, and by proper inclusion, $\overline{P}_\alpha \subset P^*$, which contradicts the defining property of P^*. This proves the lemma. Q.E.D.

Proof of the Theorem: We shall show that a minimal consistent set can be found in $\overline{\Pi}^m$ itself.

Suppose that the theorem is false. Then for each $P \in \overline{\Pi}^m$, there exists $P'' \in \Pi^m$ such that (a) $P''\ \mathrm{d}^m\ P$ and (b) *not* $P\ \mathrm{d}^m\ P''$. By Lemmas 7 and 8, there exists $P' \in \overline{\Pi}^m$ such that (a) $P'\ \mathrm{d}^m\ P$ and (b) *not* $P\ \mathrm{d}^m\ P'$.

By Lemma 9, if we have $P, P' \in \overline{\Pi}$ and $P'\ \mathrm{d}\ P$, then there exists $P^* \in \overline{\Pi}^m$ such that $P^*\ \mathrm{d}\ P$. Thus, under our supposition

$$\bigcup_{P \in \overline{\Pi}^m} L(P) = \overline{\Pi}.$$

Using Lemmas 3 and 4, it then follows that there exists $P_1, \ldots, P_m \in \overline{\Pi}^m$ such that

$$\bigcup_{j=1}^{m} L(P_j) = \overline{\Pi}.$$

Pick P_1, and recall that, by construction, $P_1 \in \overline{\Pi}^m$. Under our supposition, there exists $P_1' \in \overline{\Pi}^m$ such that $P_1'\ \mathrm{d}^m\ P_1$, but *not* $P_1\ \mathrm{d}^m\ P_1'$. We know that $P_1' \in L(P_i)$ for some $i \neq 1$ (otherwise we have an immediate contradiction). Now consider P_i, and recall that $P_i \in \overline{\Pi}^m$. Again, there is some $P_i' \in \overline{\Pi}^m$ such that $P_i'\ \mathrm{d}^m\ P_i$, but *not* $P_i\ \mathrm{d}^m\ P_i'$. Continuing this way, we get an infinite sequence of the form

$$\cdots P_k'\ \mathrm{d}^m\ P_k\ \mathrm{d}\ P_i'\ \mathrm{d}^m\ P_i\ \mathrm{d}\ P_1'\ \mathrm{d}^m\ P_1,$$

where $P_j, P_j' \in \overline{\Pi}^m$ for all j. But there are finitely many P_j''s, so there must be repetition of at least one index, l. For l, $P_l'\ \mathrm{d}^m\ P_l$, and $P_l\ \mathrm{d}^m\ P_l'$, which is a contradiction. Q.E.D.

5.3. *Simple Consistency*

The spirit of minimal consistency is that one wishes to delete irrelevant portions of IC sets, thereby reducing them to "essential units," before considering dominance relationships. Intuitively, an IC set can be divided up into several distinct pieces wherever it includes payoff vectors from several different perfect equilibria. This observation suggests an alternative approach: require each IC set to coincide with the set of payoffs achieved (in all subgames) by a single perfect equilibrium.

Unfortunately, this approach is problematic. Even with the requirement described above, internally consistent sets that correspond to different equilibria can be combined by modifying equilibrium strategies appropriately (see footnote 5 for an example).

To resolve this difficulty, we propose that attention be restricted to *simple equilibria,* which are perfect equilibria that involve *simple strategy profiles* (see Abreu, 1988). This involves no loss of strategic richness, in the sense that every equilibrium payoff can be supported as a simple equilibrium.

Define a *simple internally consistent* (SIC) set to be an IC set which corresponds to the payoffs achieved in all subgames for some simple equilibrium. Let $\Pi^s \subseteq \Pi$ be the set of all simple internally consistent sets. For $P, P' \in \Pi^s$, we define *s-dominance* (d^s) as follows: $P\ \mathrm{d}^s\ P'$ if there are finitely many elements of Π^s, say $P_1, \ldots, P_m$, such that

$$P\ \mathrm{d}\ P_1\ \mathrm{d}\ P_2\ \mathrm{d} \cdots \mathrm{d}\ P_m\ \mathrm{d}\ P'.$$

We now replace (c.3) with the following requirement. $P \in \Xi$ is *simple externally consistent* (SEC) if P is nonempty and

(c.5) for every $P' \in \Pi^s$ such that $P'\ \mathrm{d}^s\ P$, $P\ \mathrm{d}^s\ P'$.

If some $P \in \Xi$ is both SIC and SEC, then we will say that it is a *simple consistent* (SC) set. Note that simple consistency produces the desired results in Examples 3 and 4.

Following the lines of Theorem 2 and using a few additional arguments, one can prove (details omitted)

THEOREM 3. *A simple consistent set exists.*

One practical advantage of simple consistency is that it allows us to fully identify each payoff set satisfying collective dynamic consistency with a single equilibrium. In general, this interpretation does not apply either to consistent sets or to minimal consistent sets, except in the formalistic sense mentioned earlier (see footnote 5). We may therefore speak of a simple consistent equilibrium, rather than of a set of equilibrium payoffs.

6. THE INFINITELY REPEATED PRISONERS' DILEMMA

In this section, we apply our refinements to a familiar problem: the repeated prisoners' dilemma. Recent work by van Damme (1989) establishes that in this particular context, when players are sufficiently patient, any feasible and individually rational outcome can be sustained by means of an internally consistent equilibrium. Here, we demand external consistency, as well as internal consistency. Our object is to show that, for a robust set of parameter values, our refinements single out equilibria with interesting properties. Specifically, there is a range of discount factors

TABLE VI
THE PRISONERS' DILEMMA

		Player II: a_{21}	Player II: a_{22}
Player I	a_{11}	a,a	d,c
	a_{12}	c,d	b,b

strictly below unity for which all consistent equilibria are nonstationary (despite the fact that stationary perfect equilibria do exist), and only partially cooperative (hence, they are not strong perfect).

The static prisoners' dilemma is depicted in Table VI. As usual, we assume that $c > a > b > d > 0$. Let

$$\mu_1 = (b - d)(c - d)^{-1},$$

and

$$\mu_2 = \min\{(c - a)(a - d)^{-1}, (c - a)(2c - a - b)^{-1}\}.$$

Note that $\mu_2 < 1$. We will consider the class of games satisfying the following two inequalities:

$$c - a < a - d \tag{6.1}$$

and

$$\mu_1 < \mu_2. \tag{6.2}$$

These inequalities are satisfied for a nonempty and open set of parameters.[6] Throughout this discussion, we also restrict attention to pure strategy equilibria.

[6] For example, take $a = 6$, $b = 1$, $c = 10$, and $d = 0$. The inequalities continue to be satisfied for an open ball around these parameter values.

When $\mu_1 < \delta < \mu_2$, it is possible to show that all consistent equilibria are nonstationary. We will provide a brief outline of the argument, leaving details to the reader.

First, one shows that there exists an IC equilibrium which entails alternation between the off-diagonal corners and which "bootstraps" itself (in the sense that one punishes defections by restarting the equilibrium from the corner that is unfavorable to the defector). The second step consists of showing that there does not exist a perfect equilibria for which (a_{11}, a_{21}) occurs in any period (the incentive to deviate, $c - a$, exceeds the present value of any feasible punishment). Third, one shows that there is no IC set that dominates the alternating equilibrium described above. Having established that (a_{11}, a_{21}) can never occur, this is simply a matter of demonstrating that convex combinations of the payoffs associated with the remaining three outcomes cannot Pareto dominate the payoffs associated with the alternating equilibrium. It follows that the alternating equilibrium satisfies strong consistency. This equilibrium is therefore both consistent and minimal consistent. The fourth step is to note that the repeated static solution is dominated by the alternating solution. Since the latter is strong consistent, the former cannot be consistent. Finally, one notes that infinite repetition of (a_{12}, a_{21}) or (a_{11}, a_{22}) on the equilibrium path would not be individually rational. Taken together, the second, fourth, and fifth steps rule out all stationary possibilities.

The nonstationary, consistent equilibrium identified in the preceding paragraph entails alternation between the two off-diagonal corners. It is also straightforward to show that this outcome is not Pareto efficient—one can in general create a Pareto improvement by replacing the outcome in some appropriately chosen periods by (a_{11}, a_{21}). Thus, the consistent equilibrium is not strong perfect.

7. Related Work on Infinitely Repeated Games

We have already mentioned related papers on finitely repeated games in Section 3. For infinitely repeated games, several competing notions of collective dynamic consistency, or immunity to renegotiation, have been proposed. These appear in papers by Rubinstein (1979), Pearce (1987), Asheim (1989), and Farrell and Maskin (1989).

Rubinstein's (1979) requirement of strong perfection is excessive in two respects. First, Pareto efficiency in the space of all *feasible* outcomes should not be imposed as a precondition for collective dynamic consistency. Second, strong perfection requires that an equilibrium survive all conceivable deviations, most of which are unreasonable. This is unsatis-

factory at a conceptual level. Moreover, given the stringent requirements, it is hardly surprising that these equilibria often fail to exist.

Pearce (1987) also provides a notion of collective dynamic consistency that differs fundamentally from that considered here. In particular, he does not require internal consistency. He argues that, when attempting to renegotiate, players should compare their current payoffs to the worst payoffs in any subgame of the proposed alternative. In essence, Pearce insists that behavior must be history dependent—once the game has started, players can never collectively declare that past losses are sunk costs and start play over as if it was period 0. The lack of internal consistency implies that at some dates, players will not renegotiate to a Pareto superior equilibrium even if one is available *within all the constraints*. Pearce's work is therefore based on different conceptual premises.

Asheim (1989) develops a theory of renegotiation in repeated games using Greenberg's (1987) theory of social situations and compares his developments to our work at several points. He argues that his approach identifies a specific refinement, which he labels Pareto perfect Nash equilibrium. Moreover, he points out that this concept differs from the refinements proposed here. Specifically, he explores a different method of establishing external consistency. In addition, he regards time stationarity of the equilibrium set as an unnecessary imposition, even in a (stationary) repeated game.

The current paper is most closely related to the work of Farrell and Maskin (1989). Their weak renegotiation proof (WRP) concept coincides exactly with our requirement of consistency.[7] Strong renegotiation proofness (SRP) coincides with strong consistency. Their notion of relative strong renegotiation proofness (RSRP) is similar in some respects to consistency, but yields different results in specific games.

This final remark requires some explanation. One constructs an RSRP set as follows. Consider sets formed by taking the union of WRP sets. Restrict attention to sets that are maximal within the class of sets having the property that some WRP set lies entirely on the efficient boundary. Any WRP set lying entirely on the boundary of such a maximal set is called an RSRP set.

It is instructive to compare the performance of RSRP, consistency, and minimal consistency in specific games. For Example 3, A is the only RSRP set when $\varepsilon < 0$, but A, B, and C are RSRP when $\varepsilon > 0$. This outcome is similar to that obtained by imposing consistency. The one

[7] The notion of an IC/WRP equilibrium was developed simultaneously and independently by ourselves and by Joseph Farrell, although Farrell's original note on renegotiation (Farrell, 1983) predates the first draft of this paper (Bernheim and Ray, 1985).

distinction is that $A \cup C$, rather than C alone, is consistent when $\varepsilon > 0$. This distinction underscores the dependence of C on A. Note that both concepts have the undesirable property that C is excluded if and only if $\varepsilon < 0$, whereas minimal consistency and simple consistency exclude C in all cases.

For Example 4, C is the only RSRP set both in case 1 and in case 2. In contrast, A, B, and C are all consistent in case 1, while only C is consistent in case 2. Recall also that A, B, and C are all minimal consistent and simple consistent in both cases.

These examples illustrate the fact that there is no hierarchical relationship between consistency, RSRP and either MC or SC—none is a refinement of another. The relationship between MC and SC remains an open question.

REFERENCES

ABREU, D. (1988). "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica* **56,** 383–396.

ASHEIM, G. B. (1989). "Extending Renegotiation-Proofness to Infinite Horizon Games," mimeo, The Norwegian School of Economics and Business Administration.

BENOIT, J.-P., AND KRISHNA, V. (1985). "Finitely Repeated Games," *Econometrica* **53,** 905–922.

BENOIT, J.-P., AND KRISHNA, V. (1988). "Renegotiation in Finitely Repeated Games," Harvard Business School Working Paper No. 89-004.

BERNHEIM, B. D., PELEG, B., AND WHINSTON, M. (1987). "Coalition-Proof Nash Equilibria. I. Concepts," *J. Econ. Theory* **42,** 1–12.

BERNHEIM, B. D., AND RAY, D. (1985). "Pareto Perfect Nash Equilibria," mimeo, Stanford University.

BERNHEIM, B. D., AND RAY, D. (1987). "Collective Dynamic Consistency in Repeated Games," mimeo.

BERNHEIM, B. D., AND WHINSTON, M. (1987). "Coalition-Proof Nash Equilibria. II. Applications," *J. Econ. Theory* **42,** 13–29.

FARRELL, J. (1983). "Credible Repeated Game Equilibria," mimeo, MIT.

FARRELL, J., AND MASKIN, E. (1987). "Renegotiation in Repeated Games," mimeo, Harvard University.

FARRELL, J., AND MASKIN, E. (1989). "Renegotiation in Repeated Games," *Games Econ. Behav.* **1,** 327–360.

FUDENBERG, D., AND MASKIN, E. (1986). "The Folk Theorem in Repeated Games with Discounting and Incomplete Information," *Econometrica* **54,** 533–554.

GREENBERG, J. (1987). "The Theory of Social Situations," Manuscript, Haifa University.

HILDENBRAND, W. (1974). *Core and Equilibrium of a Large Economy*. Princeton, NJ: Princeton Univ. Press.

PEARCE, D. G. (1987). "Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation," mimeo, Yale University.

RUBINSTEIN, A. (1979). "Strong Perfect Equilibrium in Supergames," *Int. J. Game Theory* **9,** 1–12.

SELTEN, R. (1965). "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetraegheit," *Zeitschrift fuer die Gesamte Staatswissenschaft* **12,** 301–324.

SELTEN, R. (1975). "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *Int. J. Game Theory* **4,** 25–55.

VAN DAMME, E. (1987). *Stability and Perfection of Nash Equilibria,* Berlin/New York: Springer-Verlag.

VAN DAMME, E. (1989). "Renegotiation-Proof Equilibria in Repeated Prisoners' Dilemma," *J. Econ. Theory* **47,** 206–217.