JOURNAL OF ECONOMIC THEORY 49, 376–378 (1989)

Renegotiation-Proof Equilibrium: Reply

JOSEPH FARRELL

University of California, Berkeley, California 94720 and Hoover Institution, Stanford, California 94305

AND

ERIC MASKIN

Harvard University, Cambridge, Massachusetts 02138

Received June 6, 1988; revised November 8, 1988

We report a general theorem on sustaining cooperation in infinitely-repeated games when the possibility of renegotiation constrains punishments not to harm both players. This result generalizes observations recently reported by E. van Damme (Renegotiation-proof equilibria in repeated prisoner's dilemma, *J. Econ. Theory* **47** (1989), 206–217) for the Prisoner's Dilemma. *Journal of Economic Literature* Classification Number 026.

In an infinitely repeated game, many outcomes can be sustained as subgame-perfect equilibria that are not Nash equilibria of the one-shot game. However, some of the threatened "punishments" for deviation that sustain short-run cooperative behavior may seem implausible even though they are perfect equilibria, since they may involve punishing the innocent as well as the guilty. This is implausible if the players can communicate and "renegotiate" the continuation equilibrium. For example, it is a perfect equilibrium to sustain cooperation in the repeated prisoner's dilemma by the threat of never again cooperating after any player deviates from mutual cooperation; but if a player did cheat once, the other might propose to overlook it ("just this once") and continue cooperating.

In notes circulated in 1983–1984, one of us addressed this problem by defining the concept of a *renegotiation-proof equilibrium* in an infinitely-repeated game.[1] A *set* S of subgame-perfect equilibria is *weakly renegotiation-proof* if (i) all continuation equilibria of equilibria in S are also in S, and (ii) no equilibrium in S strictly Pareto dominates any other equi-

[1] In independent work, Bernheim and Ray [1] propose the same idea, which they call "partial Pareto perfection."

0022-0531/89 $3.00

librium in S. (For details, see Farrell and Maskin [2].) A *renegotiation-proof equilibrium* is a perfect equilibrium that is an element of some such set S.

Van Damme [3] has shown that, contrary to a claim made in early versions of those notes, cooperation in the repeated Prisoner's Dilemma is a renegotiation-proof equilibrium for large enough discount factors—indeed, for exactly the same discount factors as those for which it is a subgame-perfect equilibrium. Independently of his work, we have obtained stronger results on renegotiation-proof equilibria in repeated games. Here, we report one of our results, which generalizes van Damme's observation; the proof is in our working paper [2].

Our result characterizes the (discounted) average payoffs that can be achieved in renegotiation-proof equilibrium for large enough discount factors $\delta < 1$. The idea of the conditions in Theorem 1 is that the action-pair (a_1^i, a_2^i) is to be repeated some finite number of times in order to punish player i for a deviation, before reverting to cooperative behavior.

Before stating the Theorem, we introduce some notation. Let $a = (a_1, a_2)$ be a pair of actions. We write $g_i(a)$ for i's payoff from the action-pair a, and $c_i(a)$ for i's *cheating payoff* from the action-pair a: that is, $c_i(a)$ is i's payoff from his best response to his opponent j's move a_j.

THEOREM 1. *Let (v_1, v_2) be in V^*, the convex hull of the set of feasible payoffs from the one-shot game G. If there exist action-pairs $a^i = (a_1^i, a_2^i)$ $(i, j = 1, 2)$ in G such that $c_i(a^i) < v_i$, while $g_j(a^i) \geqslant v_j$ for $j \neq i$, then, for discount factors δ close enough to 1, there exists a renegotiation-proof equilibrium with average payoffs (v_1, v_2). Moreover, the sufficient conditions are necessary if the first inequality is made weak.*

Proof. See Farrell and Maskin [2]. ∎

Theorem 1 implies that mutual cooperation can be sustained in renegotiation-proof equilibrium in the prisoner's dilemma, since the action-pair (*cooperate*, *fink*) suffices for 1's punishment (a_1^1, a_2^1) and similarly the action-pair (*fink*, *cooperate*) will do for player 2's punishment. (Moreover, from the proof of the theorem it follows in this case that cooperation can be sustained for the same range of δ as if we did not insist on renegotiation-proof equilibrium.)

This does not mean, however, that renegotiation-proofness has no bite in general: in our working paper, we show that it strictly constrains the set of collusive outcomes that are equilibria in most games, including (for instance) infinitely-repeated Cournot duopoly.

REFERENCES

1. B. D. BERNHEIM AND D. RAY, Collective dynamic consistency in repeated games, mimeo, Northwestern University, 1987.
2. J. FARRELL AND E. S. MASKIN, "Renegotiation in Repeated Games," Working Paper 8759, Department of Economics, University of California, Berkeley, 1987.
3. E. VAN DAMME, Renegotiation-proof equilibria in repeated prisoners' dilemma, *J. Econ. Theory* **47** (1989), 206–217.