Tilburg University

The impossibility of stable renegotiation

van Damme, E.E.C.

Published in:
Economics Letters

Publication date:
1988

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
van Damme, E. E. C. (1988). The impossibility of stable renegotiation. *Economics Letters*, 26(4), 321-324.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

THE IMPOSSIBILITY OF STABLE RENEGOTIATION

Eric VAN DAMME *

Universität Bonn, D-5300 Bonn 1, FRG

Received 20 November 1987

This note describes an example of a repeated game which has a unique renegotiation-proof equilibrium [Farrell (1983)] that, however, is not stable in the sense of Kohlberg and Mertens (1986). Hence, the requirements of renegotiation-proofness and stability may be mutually inconsistent.

1. Introduction

The (perfect) folk theorem for finitely repeated games states that, for generic games with at least two Nash equilibria, every feasible and individually rational payoff can be approximated by subgame perfect equilibrium payoffs provided that the game is repeated sufficiently often [Benoit and Krishna (1986)]. Several authors have noted that this result hinges on the fact that the perfectness concepts allows self-lacerating threats of the type 'if we don't keep to the agreement, we'll shoot each other'. It has been argued that such threats (and, hence, the equilibria supported by them) are not really credible, and it has been shown that by excluding such threats one may not be able to sustain cooperation [see e.g., Farrell (1983), Farrell and Maskin (1987), Bernheim, Peleg and Whinston (1987), Bernheim and Ray (1987), Pearce (1987) and Van Damme (1986)].

Essentially the above authors argue that one should require some sort of collective dynamic consistency, i.e., that one should restrict attention to equilibria that in no subgame are Pareto dominated by other (viable) equilibrium. The motivation is that, if an equilibrium does not satisfy this requirement then, if a subgame is reached in which a Pareto superior equilibrium is available, all players will have an incentive to renegotiate to this better equilibrium and they will not continue with the original plan. Requiring renegotiation-proofness seems especially justified if players have the opportunity to communicate during the game, however, if the requirement is really compelling, then an application of Schelling's (1960) principle of tacit bargaining shows that players should accept the same concept also in the case in which no such communication is possible. (Especially in case there is a unique renegotiation-proof equilibrium, as in our example.)

A second (and maybe more fundamental) drawback of perfectness has been discussed in Kohlberg and Mertens (1986): not all perfect equilibria are self-enforcing. Essentially this is because of the fact that perfectness allows a player to explain any deviation from the equilibrium path by means of unintended small mistakes; the concept does not require a player to scrutinize the possibility that someone deviated intentionally. Hence, some equilibria may be supported only by incredible beliefs. Kohlberg and Mertens propose the concept of stable equilibria to eliminate such incredible equilibria

* Financial support from the Deutsche Forschungsgemeinschaft through the SFB 303 is gratefully acknowledged.

and it has been shown in Van Damme (1987) that requiring stability may also drastically reduce the set of payoffs obtainable by repetition.

This is not the place to discuss the relative merits of renegotiation-proofness versus stability; I just want to point out that in general one cannot satisfy both properties: the two-fold repetition of the game in section 2 has a unique renegotiation-proof equilibrium and this can be shown not to be stable. Furthermore, the example is generic, i.e., incompatibility remains for small perturbations in the data.

I conclude this section with an inductive definition of renegotiation-proof equilibria (henceforth *RPE*). This concept was first informally described by Farrell (1983). For two-person games, it coincides with the perfectly coalition-proof Nash equilibrium of Bernheim, Peleg and Whinston (1987). In Bernheim and Ray (1987) exactly the same concept is studied, there, however, it is called Pareto perfect equilibrium. Let G be a normal form game and let $G(T)$ be the T -fold repetition of G . In $G(T)$ players try to maximize the sum of their payoffs and they are always perfectly informed about the pure actions chosen by their opponents in the past. Then s is an *RPE* of $G(1)$ if it is a Nash equilibrium of G that is not (strictly) Pareto dominated by another equilibrium. Inductively, s is an *RPE* of $G(t+1)$ if

$$s \text{ is a Nash equilibrium of } G(t+1), \quad (1.1)$$

$$s \text{ continues with an } RPE \text{ of } G(t) \text{ after each period 1 history,} \quad (1.2)$$

and

$$\text{there does not exist } s' \text{ satisfying (1.1) and (1.2) that strictly Pareto dominates } s. \quad (1.3)$$

Note that these conditions imply that each *RPE* is a subgame perfect equilibrium.

2. The example

Consider the bimatrix game G from table 1 in which a and α are strictly dominated strategies.

Write $A = (a, \alpha)$, $B = (b, \beta)$ and $C = (c, \gamma)$. G has three equilibria, viz. B , C and a mixed one which yields each player an expected payoff of $4/5$, hence, the latter is not an *RPE* of $G(1)$. The two pure equilibria both are *RPE*. Note that these are strict equilibria (each player losses by deviating unilaterally) so they are stable in every conceivable sense. Next, consider the following strategy pair $\bar{s}$ in $G(2)$.

$$\begin{array}{lll} \text{at } t = 1 \text{ play } A; & \text{at } t = 2 \text{ play } B & \text{after } A, \\ & = & C \text{ otherwise.} \end{array} \quad (2.1)$$

Clearly, this profile satisfies the conditions (1.2) and (1.1). Furthermore, it results in a payoff of 14 for each player. I will show that any *RPE* s of $G(2)$ results in the same payoff as $\bar{s}$ (actually, that it

Table 1

	α	β	γ
a	10,13	-5,0	-5,0
b	11,-5	4,1	0,0
c	0,-5	0,0	1,4

yields the same path as $\bar{s}$) which will establish that all *RPE* lie in the same equilibrium component of $G(2)$. I will then show that this component is not stable.

Assume s is a subgame perfect equilibrium of $G(2)$ and write s_i^t for player i 's action at $t = 1$. Obviously, if $s_1^1(a) = 0$, then player 2 will choose $s_2^1(\alpha) = 0$. Furthermore, at $t = 2$ each player's payoff is at most 4 so that the payoffs in the overall game are bounded by (8,8) whenever $s_1^1(a) = 0$, hence, in this case s is Pareto dominated by $\bar{s}$ and s is not an *RPE*. By symmetry we must also have $s_2^1(\alpha) > 0$ in any *RPE* of $G(2)$. Therefore, assume $p = s_1^1(a) > 0$ and $q = s_2^1(\alpha) > 0$. Note that this implies that player 1's equilibrium payoff is equal to the payoff this player gets if he chooses a for sure at $t = 1$. Now, if $q < 1$, playing a at $t = 1$ always yields a total payoff strictly less than 14, so that an *RPE* with $q < 1$ should yield player 2 at least 14. However, this is impossible: if $q < 1$, then 2's equilibrium payoff is equal to the payoff resulting from playing β or γ for sure at $t = 1$, and this payoff is at most 8. Hence $q = 1$ in any *RPE*: Player 2 starts with α . If we would have $p < 1$ then player 1 should be indifferent between playing a for sure at $t = 1$ and playing b for sure at $t = 1$ and this is impossible given that the continuation payoff belongs to $\{4, 1, 4/5\}$ in a perfect equilibrium. Hence, $p = 1$ in any *RPE* and the play starts with A . This leaves only two possibilities for the *RPE* path, either (A, B) or (A, C) . As the latter is not a subgame perfect equilibrium path, only (A, B) is an *RPE* which establishes our first claim.

The formal proof that (A, B) is not stable is somewhat cumbersome and is referred to the appendix. However, an intuitive reason for why it is not self-enforcing can easily be given: it does not survive application of the 'equilibrium dominance' criterion of Cho and Kreps (1987). Player 1 is 'guaranteed' the payoff 14 if he does not deviate from (A, B) . If he, however, deviates to b at $t = 1$, then player 2 should conclude that 1 will not play c as b followed by c yields player 1 at most 12 and this is less than what this player gets in equilibrium. Furthermore, 1 will not play a in the second round as this is dominated by b . Hence, player 2 should conclude that 1 will again play b after a deviation, but then he should respond with β (his best response) and 1 will indeed gain from deviating. The equilibrium is not self-enforcing, i.e., it is not stable.

Appendix

I show that the equilibrium component supporting the path (A, B) is not stable; specifically, this component does not survive when dominated strategies or strategies that are inferior responses against this component are iteratively eliminated. Consider the normal form of $G(2)$. A typical pure strategy of player 1 is $(w, xyz)_1$ where w is the action at $t = 1$ and x (resp. y, z) is the response at $t = 2$ if the opponent has chosen α (resp. β, γ). Pure strategies of player 2 will be denoted by $(w, xyz)_2$. The equilibrium component generating the path (A, B) contains strategies $(a, byz)_1$ of player 1 (with y and z arbitrary) and the strategies $(\alpha, \beta yz)_2$ of player 2 where z is arbitrary and y assigns a large enough probability to γ . Given that player 1 chooses a strategy in this component, any strategy $(w, xyz)_2$ with $w \neq \alpha$ is an inferior response, (i.e., it is not a best response against any strategy of player 1 in the component). To verify instability, one may eliminate these, hence, one only retains strategies in which player 2 starts with α . Furthermore, any strategy $(b, cyz)_1$ is inferior against the component for player 1, and $(b, ayz)_1$ is dominated by $(b, byz)_1$. If we eliminate these, we only retain strategies in which 1 plays again b after a defection to b . Consequently, in the reduced game $(\alpha, xyz)_2$ with $y \neq \beta$ is dominated by $(\alpha, x\beta z)_2$. After having eliminated these dominated strategies, we are left with strategies in which 2 starts with α and responds to b with β . However, if 2 plays such a strategy, 1 can guarantee 15 by playing b twice which shows that no equilibrium of the original component remains in this reduced game. Proposition 6 of Kohlberg and Mertens (1986) implies that the component (A, B) is not stable.

References

- Benoit, J.P. and Krishna, 1986, Finitely repeated games, *Econometrica* 53, 905–922.
- Bernheim, B.D., B. Peleg and M.D. Whinston, 1987, Coalition-proof equilibria. I. Concepts, *Journal of Economic Theory* 42, 1–12.
- Bernheim, B.D. and D. Ray, 1987, Collective dynamic consistency in repeated games, Mimeo. IMSSS (Stanford University, Stanford, CA).
- Cho, J.K. and D. Kreps, 1987, Signalling games and stable equilibria, *Quarterly Journal of Economics* 102, no. 2, 179–221.
- Farrell, J., 1983, Credible repeated game equilibrium, Unpublished manuscript.
- Farrell, J. and E. Maskin, 1987, Renegotiation in repeated games, Mimeo. (Harvard University, Cambridge, MA).
- Kohlberg, E. and J.F. Mertens, 1986, On the strategic stability of equilibria, *Econometrica* 54, 1003–1037.
- Pearce, D., 1987, Renegotiation-proof equilibria: Collective rationality and intertemporal cooperation, Mimeo. (Yale University, New Haven, CT).
- Schelling, T.C., 1960, *The strategy of conflict* (Harvard University Press, Cambridge, MA).
- Van Damme, E.E.C., 1986, Renegotiation-proof equilibria in repeated prisoners' dilemma, Discussion paper no. A-84, SFB 303, Universität Bonn.
- Van Damme, E.E.C., 1987, Stable equilibria and forward induction, Discussion paper no. A-128, SFB 303, Universität Bonn.