

Working Paper 93-41
Economics Series 16
December 1993

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9875

ON THE POSSIBILITY OF STABLE RENEGOTIATION.
A NOTE

José Luis Ferreira*

Abstract

Van Damme (Oct 87) claims that renegotiation and stability may be mutually inconsistent concepts. This note shows that this claim may be incorrect and proposes a way to restore the compatibility of these two concepts that, in general, do not apply in the same context.

Key Words

Renegotiation-proof equilibrium, Pareto perfect equilibrium, stable set of equilibria.

* Departamento de Economía, Universidad Carlos III de Madrid. Correspondence to:
Departamento de Economía, Universidad Carlos III de Madrid, Calle Madrid, 126, 28903 Getafe
(Madrid), Spain.

1. INTRODUCTION

This note discusses the relation between the concepts of renegotiation-proof and stable equilibria. In particular, the question about existence of stable renegotiation is addressed. This discussion is motivated by an example with which van Damme claims that the answer to this question is negative. The example is a finite horizon game of two players, the definition of renegotiation-proof he uses is the Pareto perfect equilibrium (PPE, widely accepted for this class of games) and the definition of stability is the one given by Kohlberg and Mertens (1986) (see appendix). The concept of Pareto perfect equilibrium was first introduced by Farrell and Maskin (1989) and by Bernheim and Ray (1989). Here we present the definition using the notation in van Damme (Oct 87):

DEFINITION 1 Let G be a normal form game and let $G(T)$ be the T -fold repetition of G . Then s is a PPE of $G(1)$ if it is a Nash equilibrium (NE) of G that is not (strictly) Pareto dominated by another equilibrium. Inductively, s is a PPE of $G(t+1)$ if

- (1) s is a Nash equilibrium of $G(t+1)$
- (2) s continues with a PPE of $G(t)$ after each period 1 history, and
- (3) there does not exist s' satisfying (1) and (2) that strictly Pareto dominates s .

Now we can present van Damme's example: consider the bimatrix game G in figure 1, and let $G(2)$ be the 2-fold repetition of G .

FIGURE 1

Write $A=(a,\alpha)$, $B=(b,\beta)$ and $C=(c,\gamma)$ and consider the following strategy pair s^* in $G(2)$:

$$(*) \quad \text{at } t=1 \text{ play } A \quad \text{at } t=2 \text{ play } \begin{cases} B & \text{after } A \\ C & \text{otherwise} \end{cases}$$

Clearly s^* satisfies definition 1. This strategy results in a payoff of 14 for each player. Van-Damme shows that all PPEa of $G(2)$ yield the same path as s^* , and that this equilibrium component cannot be stable. The intuitive reason is as follows:

"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". (van Damme (Oct 87))

In another place, van Damme notes that

"stability is a pure non cooperative concept, hence, it requires that all aspects relevant to the situation be modeled by the rules of the game. This is important as it will turn out that stable equilibria depend crucially on modeling 'details'. Hence, stability is inconsistent with 'small words' arguments and if one cannot model all details, then stability should not be used as the solution concept."
(van Damme (Aug 87)).

However, in van Damme (Oct 87), he confronts the notion of stability with a definition of renegotiation-proof equilibrium, the Pareto perfect equilibrium (PPE), which, in his own words "seems especially justified if players have the opportunity to communicate during the game". In order to be able to compare both concepts in the same context, van Damme appeals to Schelling's (1960) principle of tacit bargaining to conclude that "if the requirement (of PPE) is really compelling, then ... 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)."

In this note I will try to show that the incompatibility between the two concepts may disappear once the following two claims are realized:

CLAIM 1. The principle of tacit bargaining does not apply in general when communication may take place.

CLAIM 2. In the particular cases when the principle seems to be applicable, the PPE may not be the "good" definition of a renegotiation-proof equilibrium if one considers stability

issues.

2. DISCUSSION OF CLAIM 1

If a definition of equilibrium is to be interpreted as a theory of how players would behave, it is clear that any suggested strategy profile that does not conform to that definition will not be obeyed: players will find some incentive to deviate. This is true for the Nash equilibrium and all its refinements. For instance, if players believe in the theory of perfect Nash equilibria, some player will unilaterally deviate from a non-perfect Nash equilibrium because he or she will be afraid of "trembling hand" moves by other players. However, the same cannot be said about the PPE if the game is to be played without communication. To see this, consider the game in figure 2:

FIGURE 2

Only (m,c) and (b,r) are PPE. If players believe in the theory of PPE, they will choose among these two pairs of strategies, but if they cannot communicate with each other, (t,l) cannot be discarded as an equilibrium. Even if they are fervent believers of the PPE and know which are the PPEa of the game, they have no way to move from (t,l) if this strategy is somehow proposed first: they need communication to go to either (m,c) or (b,r) . The knowledge of the theory of PPE is useful in general only if communication exists (that was

precisely the motivation for renegotiation-proof equilibria). Another examples of this kind may occur when coalitions are permitted to deviate (e.g. in the Coalition-proof Nash equilibrium or in the strong Nash equilibrium concepts): it may happen that a coalition is necessary to deviate from a "bad" equilibrium, but if several coalitions are possible, again communication is necessary to decide which one will be formed. For very particular examples, however it is true that communication is not necessary to apply the PPE. For example, in figure 3, the knowledge and acceptance of the theory of PPE make players simultaneously move from (t,l) to (b,r). This occurs, of course because there is only one place to go according to the theory.

FIGURE 3

Therefore, in general, we need communication to apply the definition of PPE as a theory that predicts players' behavior, but then, following van Damme, stability in the sense of Kohlberg and Mertens is not a relevant concept and we do not have any incompatibility: stability simply does not apply.

I will finish this section with two more examples to show how stability issues (and forward induction arguments) may be ruled out when communication may take place. The first one is the example given in van Damme (Oct 87), it will also be analyzed without communication (as van Damme did) in the discussion of claim 2 in the following section.

Let us consider, then, van Damme's example: if the PPE (s^*) is proposed and communication can take place, then we find that player 1 will not deviate: if he does deviate at the first stage (as in the discussion for stability in van Damme), at the beginning of the second stage he will find that player 2 will recall the first agreement (s^*) which implies that C will follow, since it is an equilibrium. Of course, player 1 may defend his new proposal B based on his last deviation, but since B is now viewed as a deviation from C, it has no chance to survive because not all players involved (read player 2) find it attractive.

As a last example consider the game of battle of sexes in figure 4.

FIGURE 4

If player 2 has the opportunity to "burn a dollar" before playing, it is well known that there is only one stable equilibrium in which player 2 does not burn the dollar and (t,l) follows. With communication, however, (b,r) is still plausible: if (b,r) is decided and player 2 burns the dollar to show that he will play aggressively afterwards (to induce (t,l)), at the beginning of the second stage he will hear from player 1 something like: "Ok, you burnt a dollar, so what? we planned to play now (b,r) and we shall do that way since it is an equilibrium; your deviation is worthless so you better follow the equilibrium path."

3. DISCUSSION OF CLAIM 2

For the cases in which the principle of tacit bargaining is applicable and no communication is necessary to have a PPE as a result (e.g. when only one such equilibrium exists as in the game in figure 3 above), it may well be the case that a PPE is not stable. An illustration of this is van Damme's example, but one can argue that if that game is to be played without communication and if players believe in the theories of both Renegotiation-proof and stable equilibria, then it is not clear why at the beginning they will choose s , the Pareto optimal within the subgame perfect equilibria. They are now aware of forward induction arguments and know that a deviation will occur, therefore, the PPE is not the "good" renegotiation-proof equilibrium in this case. Players should better concentrate their attention on the set of strategy profiles that are "optimal within the set of stable equilibria". The formal definition of this "Pareto Stable equilibrium", using van Damme's notation, is as follows:

DEFINITION 2. s is a Pareto stable equilibrium (PSE) of $G(1)$ if and only if it is a stable equilibrium (SE) of G that is not strictly dominated by any other equilibrium. Inductively,

s is a PSE of $G(t+1)$ iff

- (1) s is a SE of $G(t+1)$
- (2) s continues with a PSE of $G(t)$ after each period 1 stage history and
- (3) there does not exist s' satisfying (1) and (2) that strictly Pareto dominates

s .

REMARK 1. SE is the definition of stable equilibrium by Kohlberg and Mertens.

REMARK 2. Since SEa always exist when the sets of strategies are compact, the existence of $PSEa$ is immediate for those games with a compact set of outcomes in SEa .

REMARK 3. In van Damme's example, it is easy to check that the only two $PSEa$ are:

- (i) play (m,c) in the first period and (b,l) in the second after any history and
- (ii) play (b,l) in the first period and (m,c) in the second after any history.

5. FINAL COMMENTS

Pareto perfectness is a definition that is very generally applicable. Stability, however, is only meaningful in a very special kind of situations. To make the point simpler, consider van Damme's condition for stability (in the sense of forward induction) in two-player games:

"A solution concept S is consistent with forward induction on the class of generic 2-person games if $p \in S$ for any path p for which there exists a player i who by unilaterally deviating from p can enforce that a subgame is reached for which exactly one solution (according to S) yields this player more than p does and for which all other solutions yield this player less".

(van Damme (Aug 87)).

It is clear that, for this definition to be applicable (in the sense that it restricts the set of solutions), one needs a coincidence to happen ("... exactly one solution ..."). Furthermore, the cases in which both PPE and SE are applicable concepts are even more restricted: in

addition, one needs the coincidence that makes the principle of tacit bargaining adequate (as in the discussion of claim 2). As a result, only in a very particular set of games both concepts can be contrasted. It has been shown that, in this case, the definition of PPE can be modified to be a "reasonable" renegotiation-proof equilibrium concept when players accept the theory of stability (they interpret deviations from a Nash equilibrium, not as mistakes, but as something to be rationalized).

APPENDIX

Kohlberg and Mertens (1986):

"We will say that a set of equilibria is stable in a game G if it is minimal with respect to the following property:

PROPERTY (S): S is a closed set of Nash Equilibria of G satisfying: for any $\epsilon > 0$ there exists some $\delta_0 > 0$ such that for any completely mixed strategy vector $\gamma_1, \dots, \gamma_n$ (n players) and for any $\delta_1, \dots, \delta_n$, ($0 < \delta_i < \delta_0$), the perturbed game where every strategy s of player i is replaced by $(1-\delta_i)s + \delta_i\gamma_i$ has an equilibrium ϵ -closed to S .

REFERENCES

Bernheim, B.D. and D. Ray (1989): "Collective Dynamic Consistency in Repeated Games", *Games and Economic Behavior*, 1, 295-326.

Farrell, J. and E. Maskin (1989): "Renegotiation in Repeated Games", *Games and Economic Behavior*, 1, 327-360.

Kohlberg, E. and J. L. Mertens (1986): "On the Strategic Stability of Equilibria"; *Econometrica* 54, 1003-1037.

Schelling, T. C. (1960): "The Strategy of Conflict"; Harvard Univ. Press.

van Damme, E. (August 1987): "Stable Equilibria and Forward Induction"; Discussion Paper No. A-128, SFB 303 Universitat Bonn.

van Damme, E. (October 1987): "The Impossibility of Stable Renegotiation"; Discussion Paper No. A-136, SFB 303 Universitat Bonn.

Player 2

Player 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

Figure 1

Player 2

Player 1

	l	c	r
t	2	0	0
m	0	3	0
b	0	0	3

Figure 2

		Player 2	
		l	r
Player 1	t	2	0
	b	0	3

Figure 3

		Player 2	
		l	r
Player 1	t	3	0
	b	0	1

Figure 4