ELSEVIER

Economics Letters 47 (1995) 269–274

economics letters

# On the possibility of stable renegotiation

José Luis Ferreira

*Department of Economics, Universidad Carlos III, Madrid, Getafe 28903, Spain*

Received 16 May 1994; accepted 6 July 1994

**Abstract**

This paper discusses the relation between the concepts of stability and renegotiation. They have been shown to be incompatible in their standard definitions. Here we propose a way to restore their compatibility.

*Keywords:* Renegotiation; Stable equilibrium; Forward induction

*JEL classification:* C70; C72

## 1. Introduction

This paper discusses the relation between the concepts of renegotiation-proof and stable equilibria. In particular, the question about the existence of stable renegotiation is addressed. This discussion is motivated by an example (van Damme, 1987b) where the answer to this question seems to be negative. The example is a finite horizon game of two players, the definition of renegotiation-proof that van Damme 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 the appendix). The concept of Pareto-perfect equilibrium was first introduced by Farrell and Maskin (1989) and by Bernheim and Ray (1989). In this paper we propose a way to restore the existence of stable renegotiation in the appropriate context. As a result we can interpret the definitions in a way that provides an insight into how communication destroys forward induction arguments.

*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 $s$.



*SSDI* 0165-1765(94)00578-8

| | | Player 2 | | |
|---|---|---|---|---|
| | | α | β | γ |
| | a | 10, 13 | -5, 0 | -5, 0 |
| Player 1 | b | 11, -5 | 4, 1 | 0, 0 |
| | c | 0, -5 | 0, 0 | 1, 4 |

Fig. 1.

Now we can present the motivating example: consider the bimatrix game $G$ in Fig. 1, and let $G(2)$ be the two-fold repetition of $G$. Write $A=(a,\alpha)$, $B=(b,\beta)$ and $C=(c,\gamma)$ and consider the following strategy pair $s^*$ in $G(2)$:

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

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 has a payoff of 14 if he plays as in $(A, B)$. If he deviates to $b$ at $t=1$, then player 2 should conclude that 1 will not play $c$ because in this case ($b$ at $t=1$ and $c$ at $t=2$) player 1 gets at most 12, less than the 14 he gets in the equilibrium. Since we also have that $a$ is dominated by $b$, player 2 should conclude that 1 will again play $b$ after a deviation, but then his best response is $\beta$. This gives player 1 a payoff of 15 and he gains from deviating. The equilibrium is not stable.

Thus we have an incompatibility between the two concepts. The notion of PPE was meant to reflect players' interaction in situations with communication, while the concept of stability is purely non-cooperative. One can appeal to Schelling's (1960) principle of tacit bargaining to conclude that if the notion of PPE is compelling, then the players should accept it even in the case when no communication is possible. Now the two concepts can be confronted. This paper shows that the incompatibility between the two concepts may disappear once we take into account the following two claims:

*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 most of 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 Fig. 2. Only $(b, c)$ and $(c, r)$ 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, $(a, l)$ 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 are not clearly compelled to move from $(a, l)$ if this strategy is seen as the *status quo*: they need communication to go to either $(b, c)$ or $(c, r)$ and gain by this deviation. The knowledge of the theory of PPE is useful, in general, only if communication exists (that was precisely the motivation for renegotiation-proof equilibria). Other 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 strategy profile, 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

| | | Player 2 | | |
|---|---|---|---|---|
| | | l | c | r |
| | a | 2, 2 | 0, 0 | 0, 0 |
| Player 1 | b | 0, 0 | 3, 3 | 0, 0 |
| | c | 0, 0 | 0, 0 | 3, 3 |

Fig. 2.

Player 2

| | | l | r |
|---|---|---|---|
| Player 1 | t | 2, 2 | 0, 0 |
| | b | 0, 0 | 3, 3 |

Fig. 3.

communication is not necessary to apply the PPE. For example, in Fig. 3, the knowledge and acceptance of the theory of PPE make players move simultaneously from $(t, l)$ to $(b, r)$. The reason is that there is only one place to go according to the theory.

Therefore, in general, we need communication to apply the definition of PPE as a theory that predicts players' behavior. But then, stability in the sense of Kohlberg and Mertens (a purely non-cooperative concept), is not relevant and we do not have any incompatibility. Stability simply does not apply. To better see why, this section ends with two more examples to show how stability issues (and forward-induction arguments) may be ruled out when communication may take place.

Let us go back to our first example (Fig. 1). If the PPE ($s^*$ as defined above) 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), 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 find it attractive.

As a last example consider the game of battle of the sexes in Fig. 4. If player 2 has the opportunity to 'burn a dollar' before playing, it is well known that there is only one stable

Player 2

| | | l | r |
|---|---|---|---|
| Player 1 | t | 1, 3 | 0, 0 |
| | b | 0, 0 | 3, 1 |

Fig. 4.

equilibrium (which is also the only one that survives the process of iterative elimination of weakly dominated strategies) 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 it that way since it is an equilibrium; your deviation is worthless so you had 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 propose a PPE as a solution seems reasonable (e.g. when only one such equilibrium exists as in the game in Fig. 3), but then, this PPE may not be stable. An illustration of this is the example in Fig. 1. 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 chose $s$, the Pareto-optimal within the subgame perfect equilibria. They are now aware of the notion of stability and know that a deviation will occur. Therefore, the PPE is not the 'good' renegotiation-proof concept 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' is as follows:

*Definition 2.* The strategy profile $s$ 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 $B$ in the first period and either $B$ or $C$ in the second after any history and
(ii) play $C$ in the first period and either $B$ or $C$ in the second after any history.

## 4. Final comments

Pareto-perfectness is a definition that is very generally applicable. Stability, however, is only meaningful in a very special kind of situation. 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, 1987a).

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 can both concepts 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):
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 $\varepsilon > 0$ there exists some $\delta_0 > 0$ such that for any completely mixed strategy vector $\gamma_1, \ldots, \gamma_n$ ($n$ players) and for any $\delta_1, \ldots, \delta_n$, $(0 < \delta_i < \delta_0)$, the perturbed game where every strategy $s$ of player $i$ is replaced by $(1 - \delta_i)s + \delta_i \gamma_i$ has an equilibrium $\varepsilon$-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 University Press).

van Damme, E., 1987a, Stable equilibria and forward induction, Discussion Paper no. A-128, SFB 303, Universitat Bonn.

van Damme, E., 1987b, The impossibility of stable renegotiation, Discussion Paper no. A-136, SFB 303, Universitat Bonn.