

# A Communication-Proof Equilibrium Concept\*

JOSÉ LUIS FERREIRA

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

Received February 4, 1992; revised October 24, 1994

This paper proposes an equilibrium concept for the classes of environments in which players can communicate with each other but cannot make binding agreements. The communication-proof equilibrium is intended to be regarded as an extension of both coalition- and renegotiation-proof equilibria. Conceptual foundations for this particular definition are discussed as it is confronted with other definitions in these environments. *Journal of Economic Literature* Classification Number: C72. © 1996 Academic Press, Inc.

## 1. INTRODUCTION

In recent years, many authors have studied the role of communication in non-cooperative game theory and two approaches have been developed. The first explicitly models the process by which players communicate and includes it as a pre-play stage in which players interchange messages in a definite way before actually playing. In the second approach, this process is left unmodeled and intuitive arguments justify properties that one may expect from free communication. For instance, it is often assumed that coordination on mutually beneficial strategies occurs whenever this common interest is realized. This paper follows the second approach.

Bernheim, Peleg, and Whinston [3] (from now on BP&W) defined the coalition-proof Nash equilibrium (Coalition-PNE) for normal form games in the spirit of Nash equilibrium: since all players move simultaneously, they allow for a deviating coalition that takes as given the opponent coalition's strategy. But when a first coalition considers whether to deviate, it should know that a subcoalition may consider further deviations, and so on.

In multi-stage games, several definitions of renegotiation-proof equilibria (RPE) have been proposed. To illustrate the problem, consider the typical

\* The author acknowledges helpful comments from P. DeMarzo, A. Manelli, and especially E. Kalai. This paper is based on the second chapter of his Ph.D. thesis at Northwestern University.

way to achieve cooperation in the infinitely repeated prisoners' dilemma: both players agree to cooperate as long as this cooperative behavior is observed in the past and if at some point someone deviates, the Nash equilibrium in the stage game continues forever. But then a problem arises: if the cooperative behavior is sustainable, players will never agree to play the "grim" strategy, since they can renegotiate from it. But without punishments, deviations may not be prevented.

In finite horizon games, backward induction allows for a natural definition of renegotiation-proof equilibrium, the Pareto perfect equilibrium (PPE) (Bernheim and Ray [2]). Pearce [7], Farrell and Maskin [4], Bernheim and Ray [2], and Asheim [1] extended this concept to infinite horizon games in different ways. A drawback of these works is that only the coalition of all players can renegotiate. BP&W extended the Coalition-PNE to games in extensive form (the perfectly coalition-proof Nash equilibrium, PCoalition-PNE); but, as will be argued, PCoalition-PNE does not capture the idea of renegotiation.

The purpose of this paper is to define an equilibrium that may be regarded as both coalition- and renegotiation-proof; the communication-proof equilibrium (Com-PE). To understand this definition we can go back to the intuitive justification of Coalition-PNE by BP&W. A group of people arrive at an agreement and then must leave the room. The problem is that, when a player leaves, he may rightly suspect that the remaining players will change their actions, taking as given his own strategy. A Coalition-PNE is an arrangement that will not be changed regardless of the order of exit. In dynamic games (multi-stage or extensive form games) a difficulty arises: when considering a deviation, to take the opponents' strategies as given is a natural assumption only if those strategies and the deviation are played simultaneously. To modify the definition for extensive form games one needs to take into consideration not only the usual problems of time consistency but the renegotiation issues as well. In particular, if part of the deviation can be observed by some opponent (the natural case in multi-stage games), the possibility of a reaction by the opponents cannot be ruled out. Then the equilibrium must be immune to deviations by coalitions that not only take as fixed the actions of the complementary coalition in the current period, but also consider reactions by any coalition in future periods. Unlike the definition of PCoalition-PNE, our proposal of Com-PE reflects these points.

## 2. THE EQUILIBRIUM

Let  $\Gamma$  be an extensive form game with no chance moves with  $N = \{1, \dots\}$  as the set of players. A subset of  $N$  will denote a coalition. The (finite) set

of pure strategies for player  $i$  will be denoted by  $S_i$ . Strategy profiles for sets of players are denoted by  $S = \times_{i \in N} S_i$ ,  $S_c = \times_{i \in c} S_i$ , and  $S_{-c} = \times_{i \notin c} S_i$ , where  $C \subset N$  is a coalition and  $-C = N \setminus C$ . Their respective typical elements will be  $s_i, s, s_c$ , and  $s_{-c}$ . The vector  $(s_{-i}, s'_i)$  denotes the strategy profile  $s$  in which  $s_i$  has been replaced with  $s'_i$ . Let  $u^i: S \rightarrow \mathbb{R}$  be the outcome function for player  $i$  (it will be identified with his von Neumann–Morgenstern expected utility function). The game that  $s_c$  induces on  $\Gamma$  (i.e., the game that coalition  $C$  faces when  $s_{-c}$  is regarded as fixed) will be denoted by  $\Gamma | s_{-c}$ . For details see Peleg [8]. A subgame of  $\Gamma$  is indicated by  $g$  and the restriction of  $s$  in the subgame  $g$  by  $s(g)$ .

The number of stages in the game is the maximum number of nested subgames in it. All the actions between the beginning of a subgame and the beginning of the next subgame belong to the same stage. By  $s(t)$  we will denote the behavior that strategy  $s$  induces at stage  $t$ .

The set of histories when the game starts (stage 1) is given by  $H_1 = \{\emptyset\}$ . The set of feasible histories  $H$  for a  $\tau$  stage game is given by  $H = \bigcup_{t=1}^{\tau} H_t$ ;  $H_t$  is defined recursively by  $H_{t+1} = \{h_{t+1} \mid \exists h_t \in H_t \text{ s.t. } h_{t+1} \in \{h_t\} \times S(h_t)\}$ , where  $S = \times_{i=1}^n S_i$  and where  $S_i(h_t)$  is the set of actions for player  $i$  in stage  $t$  given  $h_t$ . Some more notation about histories include: the set of feasible histories following  $h$ ,  $H^h$ ; the subgame induced by history  $h$ ,  $g^h$ ; the strategy induced by  $s$  after history  $h$ ,  $s^h$ ; the stage at which history  $h$  is observed,  $t(h)$  (history  $h$  ends at stage  $t(h) - 1$ ), and the set of histories of length  $t(h) + 1$  that belong to  $H^h$ ,  $h_1$  (see Asheim [1] for more details).

The short name of an equilibrium will denote the set of those equilibria in a given game (e.g., Com-PE or Com-PE( $\Gamma$ )). In what follows, games are assumed to be of perfect recall so that only behavioral strategies will be considered.

To motivate the definition of Com-PE as an extension of Coalition-PNE to extensive form games in the same spirit as that in which subgame perfect equilibria are an extension of Nash equilibria, we introduce the following definitions.

**DEFINITION 2.1.** The strategy profile  $s^*$  is a *Nash equilibrium restricted to*  $T \subset S$ , NE( $T$ ), if  $s^* \in T$  and for all players  $i \in N$  and all strategies  $s_i \in S_i$  such that  $(s_i, s_{-i}^*) \in T$ ,  $u^i(s_i, s_{-i}^*) \leq u^i(s^*)$ .

**DEFINITION 2.2.** The strategy vector  $s^* \in S$  is a *subgame perfect equilibrium*, SPE, if, for any subgame  $g$ ,  $s^*(g)$  is a NE.

In a game with a finite number of stages, an alternative definition of SPE, using Definition 2.1 above, is as follows:

DEFINITION 2.2'. Let  $\Gamma$  be an extensive form game with  $t$  ( $t < \infty$ ) stages.

(i) If  $t = 1$ ,  $s^* \in S$  is a SPE' if  $s^*$  is a NE.

(ii) Assume that a SPE' has been defined for all games with  $r < t$  stages and consider a game with  $t$  stages; then  $s^* \in S$  is SPE' if  $s^*$  is a NE( $T$ ) where  $T = \{s \in S \mid s \text{ induces a SPE' in proper subgames of } \Gamma\}$ .

PROPOSITION 2.1. *The strategy profile  $s^*$  is a SPE if and only if it is a SPE'.*

The proof is straightforward (see Ferreira [5]).

DEFINITION 2.3. (i) In a single player game,  $s^*$  is a *coalition proof Nash equilibrium restricted to*  $T \subset S$ , Coalition-PNE( $T$ ), if  $s^* \in \arg \max_{s \in T} u^1(s)$ .

(ii) Assume that Coalition-PNE( $T$ ) has been defined for games with less than  $n$  players. Then, in an  $n$ -player game:

(a)  $s^* \in S$  is *self-enforcing restricted to*  $T$ , SE( $T$ ), if for any coalition  $C \neq N$ ,  $C \neq \emptyset$ ,  $s^*$  is Coalition-PNE( $T$ ) in the game  $\Gamma \mid s_{-c}^*$ ;

(b)  $s^* \in S$  is Coalition-PNE( $T$ ) if it is SE( $T$ ) and if there does not exist any other  $s \in T$  such that  $s$  is SE( $T$ ) and  $u^i(s) > u^i(s^*)$  for all  $i \in N$ .

If  $T = S$ , Coalition-PNE( $T$ ) is the definition of Coalition-PNE as given by BP&W.

DEFINITION 2.4. (i) In a single stage game  $\Gamma$ ,  $s^*$  is a *communication-proof equilibrium*, Com-PE, if it is a Coalition-PNE.

(ii) Let  $t > 1$  and assume that Com-PE has been defined for games with  $r < t$  stages. Then, in a game  $\Gamma$  with  $t$  stages,  $s^*$  is Com-PE if it is a Coalition-PNE( $S^1$ ), where  $S^1 = \{s \in S \mid s \text{ induces a Com-PE in proper subgames of } \Gamma\}$ .

Remark 2.1. If we replace Coalition-PNE with NE we obtain SPE and Com-PE is seen as a natural extension of Definition 2.2'.

Remark 2.2. The strategy  $s$  is a Com-PE if for every history  $h$  and every coalition  $C$ ,  $(s_c(t(h)), s(h_1))$  is a Com-PE in  $g^h \mid s_{-c}(t(h))$ . For example, at the beginning of any subgame (after history  $h$ ), and when the strategies by any coalition are fixed for the current period,  $s_c(t(h))$ , the restriction of  $s$  in this game is a Com-PE.

Proposition 2.2 studies the existence of Com-PEa for a special and important class of games.

PROPOSITION 2.2. *Let  $\Gamma$  be a perfect information finite game (the number of stages, players, and pure strategies is finite); then there exists a Com-PE.*

*Proof.* Perfect information implies that only one player moves at every reached stage of the game. The proof is by induction in the number of stages,  $t$ .

Let  $t = 1$ ; then a Com-PE exists by the finiteness of alternatives.

Let  $t > 1$  and assume that the proposition is true for games of  $s < t$  stages. W.l.o.g. let player one be the (only) player moving at the first stage in  $\Gamma$ . Choose a Com-PE  $s(k_i)$  in every subgame after  $k_i$  (one-stage histories) and let player one select his preferred action  $s_1$  in his first move if these  $s(k_1)$  are the continuation of the play. Construct the strategy profile  $s = \{s_1, \{s(k_i)\}\}$

If  $s$  is a Com-PE we are done. If not, there exists a coalition  $C(1)$  with a deviation  $s_{c(1)}^1$  satisfying the conditions in the definition. Let  $s^1 = (s_{-c(1)}, s_{c(1)}^1)$ . If  $s^1$  is not a Com-PE, construct  $s^2$  in a similar way and so on. If we never get Com-PE in this way, we have a sequence  $\{s^k\}_k$  where  $s^k$  is the result of a deviation from  $s^{k-1}$ .

By the finiteness of the set of players and pure strategies, the sequence either converges or cycles in the set of coalitions and in the set of strategies played with positive probability (the support of a strategy). In the second case, define a convergent subsequence and consider its limit  $\hat{s}$ . If it is not a Com-PE we have a sequence of Com-PEa in proper subgames that does not converge to a Com-PE. The only reason for this limit not to be a Com-PE is if in the limit we can form a deviating coalition that was not possible in the tail of the sequence. Since the support of the limit must be included in the support of the strategies in the tail of the sequence, it must be the case that this deviating coalition includes a player that, in the limit, is made indifferent among the choices that both the original strategy and the deviation induce *at the time* he plays. But in this case it must be that before the deviation, the path in  $\hat{s}$  did not pass through him, otherwise he is not improving in the deviation (and therefore, the path in the sequence did not pass through him either). We could have considered the strategies in that subgame to be fixed and still have had the sequence of Com-PE converging to a strategie that is not a Com-PE. Thus, the limit of this sequence is Com-PE in proper subgames.

We can always start with a selection of Com-PEa in subgames after  $k_i$  and with a sequence of deviations after the strategy profile  $s$  as constructed above so that the limit of the utility of player one in this sequence is the supremum of the utility levels he can get in strategy profiles that have Com-PE in proper subgames. Thus, the corresponding limit  $\hat{s}$  of the appropriate subsequence must be a Com-PE. This is because player one cannot have his utility increased in a deviation, so that he may not be part

of a deviating coalition, and because  $\hat{s}$  is a Com-PE in subgames so that no coalition without player one can deviate. Q.E.D.

Proposition 2.3 relates the concept of Com-PE with those of Pareto perfect equilibrium and perfectly coalition-proof Nash equilibrium, defined below.

DEFINITION 2.5 (BP&W). (i) In a single player, single-stage game  $\Gamma$ ,  $s^* \in S$  is a *perfectly coalition-proof Nash equilibrium* (PCoalition-PNE) if  $s^*$  maximizes  $u^i(s)$ .

(ii) Let  $(n, t) \neq (1, 1)$ . Assume that PCoalition-PNE has been defined for all games with  $m$  players and  $s$  stages, where  $(m, s) \leq (n, t)$  and  $(m, s) \neq (n, t)$ .

(a) For any game  $\Gamma$  with  $n$  players and  $t$  stages,  $s^* \in S$  is *perfectly self-enforcing* (PSE) if (1) for all  $C \subset N$ ,  $s_c^*$  is a PCoalition-PNE in the game  $\Gamma \upharpoonright s_{-c}^*$ , and (2) for any proper subgame of  $\Gamma$ ,  $g$ ,  $s^*(g)$  is a PCoalition-PNE in  $g$ .

(b) For any game  $\Gamma$  with  $n$  players and  $t$  stages,  $s^* \in S$  is *PCoalition-PNE* if it is PSE and if there does not exist another PSE strategy vector  $s \in S$  such that  $u^i(s) > u^i(s^*)$  for all  $i \in N$ .

DEFINITION 2.6 (Bernheim and Ray [2]). (i) In a single-stage game, a *Pareto perfect equilibrium* (PPE) is a Nash equilibrium (NE) that is not strictly Pareto-dominated by another NE.

(ii) Let  $t > 1$  and assume that PPE has been defined for all games with less than  $t$  stages. Let  $\Gamma$  be a  $t$ -stage game; then a strategy profile  $s$  is a PPE in  $\Gamma$  if

(a)  $s$  is a NE and  $s^h$  is a PPE of  $g^h$  for all  $h \in H$ ,  $h \neq \emptyset$ , and

(b) there is no profile  $x$  satisfying part (a) such that  $u_i(s) < u_i(x)$  for all  $i \in N$ .

PROPOSITION 2.3. *Let  $\Gamma$  be a finite horizon game, then*

(i) *Com-PE  $\subset$  SPE.*

(ii) *If  $\Gamma$  is a one-stage game ( $t = 1$ ), Com-PE = Coalition-PNE.*

(iii) *If  $\Gamma$  is a two-player game ( $n = 2$ ), Com-PE = PPE.*

(iv) *If  $t = 2$  PCoalition-PNE  $\subset$  Com-PE.*

(v) *Neither Com-PE  $\subset$  PCoalition-PNE nor PCoalition-PNE  $\subset$  Com-PE is satisfied.*

The reason we have (iv) is, basically, that both concepts coincide in one-stage games and that possible deviations according to Com-PE must satisfy a stronger condition than deviations according to PCoalition-NE. Since we are in two-stage games, deviations of deviations by arbitrary coalitions have no place and we have the inclusion. When these deviations may occur we have (v): a strategy profile satisfying the definition of PCoalition-PNE may not be Com-PE because there may be a deviation that induces Com-PEa in proper subgames but that does not induce PCoalition-PNEa (by (iv) Com-PE may be a larger set in some cases). Thus the deviation is viable according to Com-PE but not according to PCoalition-PNE.

*Proof.* (i) This follows from Proposition 2.1 and from the fact that Coalition-PNE  $\subset$  NE.

(ii) This follows from the definition of Com-PE.

(iii) Let us proceed by induction on the number of stages of the game. For  $t = 1$  it is trivial (Coalition-PNE reduces to a non-strictly Pareto dominated Nash equilibrium). Assume now that the proposition has been proven for  $s < t$ , and let  $\Gamma$  be a  $t$ -stage game. First prove PPE  $\subset$  Com-PE. Let  $s \in S$  be a PPE; by Definition 2.6 (ii) (a)  $s \in S^* = \{s \in S \mid s \text{ induces a PPE in proper subgames}\}$ ; by the induction hypothesis  $S^* = S' = \{s \in S \mid s \text{ induces a Com-PE in proper subgames}\}$ ; also  $s \in \text{NE}$ , which for  $n = 2$ , means that  $s$  is self-enforcing. These two facts and Definition 2.6 (ii) (b) show that  $s$  is Coalition-PNE restricted to  $S'$ . To prove Com-PE  $\subset$  PPE, let  $s$  be Coalition-PNE restricted to  $S'$ ;  $s \in S'$  implies that  $s \in S^*$  by induction hypothesis. By the definition of Com-PE we also have that

(\*) there does not exist any other  $s' \in S'$  such that  $s_j = s'_j$  and  $u_j(s') > u_j(s)$  for all  $i \in \{1, 2\}$  and  $j \neq i$ .

This means that  $s$  is a NE and Definition 2.6 (ii) (a) is satisfied; if not, there exists a player  $i \in \{1, 2\}$  and a strategy  $s'_i \in S_i$  such that  $u_i(s') = u_i(s'_i, s_j) > u_i(s)$ . But then  $s(1) = s'(1)$  (otherwise  $s(1)$  was not a NE in subgames) and  $s' \in S$ , which contradicts (\*). Finally, part (ii)(b) of Definition 2.6 is satisfied by the optimality of Coalition-PNE.

(iv) By induction on the number of players. For one-player games, the statement is trivial. Assume, then, that it is true for  $n - 1$  players and prove it for  $n$ . If strategy profile  $s^*$  is a PCoalition-PNE, then (1) for every coalition  $C \subset N$ ,  $C \neq N$ ,  $s_c^*$  is a PCoalition-PNE in  $\Gamma \mid s_{-c}^*$ , (2)  $s^*$  is a PCoalition-PNE in proper subgames of  $\Gamma$ , and (3) there does not exist a strategy profile  $s'$  that satisfies (1) and (2) such that every player in  $N$  strictly prefers  $s'$  to  $s^*$ . But (1) implies that (1') for every coalition  $C \subset N$ ,  $C \neq N$ ,  $s_c^*$  is a Com-PE in  $\Gamma \mid s_{-c}^*$  by the induction hypothesis; (2) implies that (2')  $s^*$  is a Com-PE in proper subgames of  $\Gamma$  because both concepts

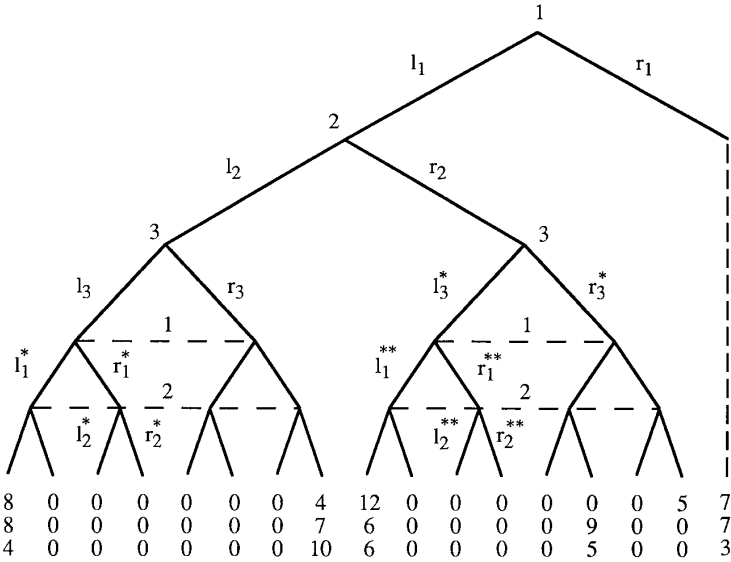


FIGURE 2.1

coincide in one-stage games. Now, (1') and (2') together imply that  $s_c^*$  is a Com-PE in  $\Gamma | s_{-c}^*(1)$ , which means that  $s_c^*$  is a coalition-PNE( $s^1$ ). This and (3) imply that  $s^*$  is a Coalition-PNE( $s^1$ ), which is the definition of Com-PE in two-stage games.

(v) Consider the game in Fig. 2.1. Among the SPEa of the game are the following:

- (a)  $(l_1, l_2, l_3, l_3^*, l_1^*, l_1^{**}, l_2^*, l_2^{**})$  and
- (b)  $(r_1, r_2, r_3, l_3^*, r_1^*, l_1^{**}, r_2^*, l_2^{**})$ .

The equilibrium in (a) is Com-PE but not PCoalition-PNE. To show that it is Com-PE see that no coalition of  $n$  players can deviate to a better strategy. If  $n=1$ , this is because (a) is SPE; if  $n=3$ , because the payoffs induced in every subgame are Pareto optimal, if  $n=2$ , because only the coalition formed by players 2 and 3 can find an outcome in which both are better off, namely the outcome (0, 9, 5), which is preferred to (8, 8, 4), but to obtain that outcome, the deviation to  $(r_2, r_3^*, r_2^{**})$  has to take place in the second stage of the game, but that deviation does not conform to a Com-PE in the subgame after  $r_2$ . Since that deviation is a PCoalition-PNE (it is SPE and optimal) in the game induced by player 1's strategy  $(l_1, l_1^*, l_1^{**})$ , then (a) is not a PCoalition-PNE.

Equilibrium (b) is PCoalition-PNE but not a Com-PE. It is not a Com-PE because it is Pareto dominated by (a). It is a PCoalition-PNE because it induces a PCoalition-PNE in every proper subgame, because it induces a PCoalition-PNE in games induced by player  $i$ 's strategies (since only subgame perfection and optimality is needed), and because it is not dominated by any other SPE. Q.E.D.

Let  $\Gamma$  be an extensive game form with perfect information. Peleg [8] shows, using a very involved proof, that for every profile of *linear* preferences (no indifferences), the set of SPE of  $\Gamma$  coincides with the set of PCoalition-PNE. Propositions 2.2 and 2.3 (i) show that the Com-PE satisfies the same property (important in the theory of voting). Thus we have the following.

**COROLLARY 2.1.** *Let  $\Gamma$  be a finite horizon perfect information game with linear preferences. Then  $\text{Com-PE}(\Gamma) = \text{SPE}(\Gamma)$ .*

Finally, the definition of Com-PE can be extended to the case of an infinite number of stages and players using the stable sets approach in Greenberg [6] and Asheim [1]. This definition, and the proposition showing its equivalence to the one given here for the finite case, can be found in Ferreira [5].

## REFERENCES

1. B. ASHEIM, Extending renegotiation-proofness to infinite horizon games, *Games Econ. Behav.* **3** (1991), 278–294.
2. B. D. BERNHEIM AND D. RAY, Collective dynamic consistency in repeated games, *Games Econ. Behav.* **1** (1989), 295–326.
3. B. D. BERNHEIM, B. PELEG, AND M. D. WHINSTON, Coalition-proof Nash equilibria I: Concepts, *J. Econ. Theory* **42** (1987), 1–12.
4. J. FARRELL AND E. MASKIN, Renegotiation in repeated games, *Games Econ. Behav.* **1** (1989), 327–360.
5. J. L. FERREIRA, “A Communication-Proof Equilibrium Concept,” Discussion Paper No. 896, CMSEMS, Northwestern University, 1990.
6. J. GREENBERG, Deriving strong and coalition-proof Nash equilibria from an abstract system, *J. Econ. Theory* **49** (1989), 195–202.
7. D. PEARCE, “Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation,” Cowles Foundation Discussion Paper No. 855, Yale University, 1987.
8. B. PELEG, On perfectly coalition-proof Nash equilibria, mimeo, Department of Mathematics, The Hebrew University, 1988.