

# **Universal Coalition-Proof Equilibrium: Concepts and Applications**

by

**Bhaskar Chakravorti**  
*Monitor Group*  
*Two Canal Park*  
*Cambridge, MA 02141*

and

**Charles M. Kahn**  
*Department of Finance*  
*University of Illinois at Urbana-Champaign*  
*Champaign, IL 61820*

Revised: 1994

We are grateful to George Mailath, Jacques Robert, Bill Sharkey, Sang-Seung Yi, and Bill Zame for helpful discussions on the subject, and to two anonymous referees for their insightful comments. The authors retain joint responsibility for all errors and (quite unlike the dastardly players whose exploits are chronicled in this paper) are not willing to unilaterally defect when complaints are made.

# UNIVERSAL COALITION-PROOF EQUILIBRIUM: CONCEPTS AND APPLICATIONS

## ABSTRACT

We propose a solution concept that characterizes stable agreements in games in which coalitions can make non-binding self-enforcing deviations from such agreements. We allow for "universal" coalition formation, i.e. the validity of a deviation is checked not only against further deviations by subsets of the deviating coalition but also against deviations agreed upon by some members of the deviating coalition convincing players from the complementary coalition to deviate by employing a very simple signaling argument. Consequently, the blocking device we introduce is weaker -- and less restrictive -- than that associated with Bernheim, Peleg and Whinston's (1987) concept of Coalition-proof Equilibrium (CPE). Universal coalition formation leads to a failure of an inductive approach to characterizing a solution; we show how the von Neumann-Morgenstern stable sets approach can be extended to derive an appropriate characterization. However, the solution concept we obtain has no logical inclusion relationship with CPE. It is a Nash Equilibrium refinement, which by no means is a foregone conclusion, since, a priori, the unilateral best-response property is not a necessary condition for stability (in the sense of von Neumann-Morgenstern stable sets) in the presence of universal coalition formation. The key aspect of our solution concept is a "lateral induction" condition, which is related to the idea of forwards induction in sequential games. We present a diverse set of economic applications where our solution concept outperforms CPE in providing plausible predictions; these include the analyses of hierarchies, standards-setting negotiations, divide-the-dollar problems, and renegotiation of contracts. In particular, we rationalize (i) a "pyramid", as opposed to vertical hierarchical structures in organizations, (ii) agreement on a neutral standard among competing firms with their own favorite product specifications, (iii) stable division rules with majority rule voting on the rule, and (iv) effort-inducing labor contracts which are stable despite the incentive among unions and management to renegotiate the contract.

*JEL Classification Number: C70*

## 1. INTRODUCTION

Consider a finite  $n$ -player game in which coalitions of players communicate prior to actual play and make non-binding agreements on strategy choices. We wish to know which agreements are stable in such environments. Following upon Aumann's (1959) "strong" Nash equilibrium, one answer to this question is given by Bernheim, et al.'s (1987) notion of Coalition-proof Nash equilibrium (CPE), a refinement of Nash equilibrium. But CPE is not without its critics. To borrow Kalai's remark (quoted in Greenberg (1989)): "The concept of CPE does not go far enough in its analysis of stability. When considering a deviating coalition, the validity of the deviation is checked only against further deviations of subcoalitions of the deviating coalition. (*We refer to this as the "nestedness" assumption.*) However, members of the deviating coalitions could also deviate by convincing other players (from the nondeviating coalition) to deviate provided they improve their payoff (*we refer to this as universal coalition formation*)."

The primary motivation of our paper is to address these concerns and to characterize stable agreements in the presence of universal coalition formation. We present a new solution concept called Universal Coalition-proof equilibrium (UCPE). We pose and resolve three issues that arise in this context: (i) Bernheim et al. (1987) pioneered an approach towards characterizing such solutions using induction on the number of players; this fails in the presence of universal coalition formation, since induction presumes nestedness of deviating coalitions. On the other hand, Greenberg (1989) pioneered an approach using von Neumann-Morgenstern stable sets; we show that this fails as well in the presence of universal coalition formation. (ii) We extend the stable set approach to arrive at a solution concept. However, this approach does not require a Nash-like best-response property for stability; hence, the containment of UCPE in the set of Nash equilibria is not a foregone conclusion, as is the case with CPE. We prove that such a containment is guaranteed by our definition. (iii) There are several classes of economic applications where CPE yields no predictions; we show that a remarkably simple "signaling" argument can overcome the nestedness restriction and lead to sharp predictions.

The key to arriving at our equilibrium is a "lateral induction" condition. Its intuition is derived from the idea of forwards induction, which plays a critical role in sequential games (see Kohlberg and Mertens (1986), Cho and Kreps (1987), Banks and Sobel (1987), etc.). There are, conceivably, several alternative ways in which informed players can signal to uninformed ones. The strength of the lateral induction criterion we propose is that it serves two critical roles simultaneously. First, it bridges the informational asymmetry that arises in the absence of the nestedness assumption. Second, as shall argue below, the condition ensures that every UCPE is also a Nash equilibrium, thereby validating its role as a non-cooperative solution concept.<sup>1</sup>

We establish the relationships of UCPE with other related solution concepts. We shall show that UCPE contains the set of strong Nash equilibria but has no necessary containment relationships with CPE, or with Greenberg's (1989, 1989a) notion of Coalitional Contingent Threats Equilibrium (CCTE).

As indicated earlier, UCPE is a refinement of Nash equilibrium. Unlike the case with CPE, this crucial containment property does not trivially follow from the definitions. A priori, we should not expect a UCPE to also be a Nash equilibrium. The characterization of UCPE is obtained by partitioning the space of agreements among players, using the logic of von Neumann and Morgenstern (1947) stable sets. In games with universal coalition formation, the Nash best-response property is not necessary for stability of an agreement in the von Neumann-Morgenstern sense. A Nash equilibrium corresponds to an  $n$ -player agreement that is not threatened by any one-player deviation. However, it is possible for an  $n$ -player agreement to be non-credibly threatened by a one-player deviation. The deviator may subsequently wish to

---

<sup>1</sup>Our analysis will be based on the same paradigm as that of Bernheim et al. In other words, actual payoff-relevant play occurs simultaneously in one shot. The stability of the strategies played is rationalized via an "extensive-form" involving coordinated defections discussed within coalitions and further defections from such defections. In terms of Bernheim et al.'s "room" analogy (1987, pp 5), we allow that players who leave the discussion room may be urged to return. Briefly, players meet in a room and discuss strategies and subsequently leave the room in an arbitrary order after a (non-binding) agreement on strategies is reached. When a player leaves the room the remaining players assume she will play as agreed unless she is invited to return to the room. Each player, regardless of the order of exit is concerned that the remaining players in the room may reach a new agreement either by themselves or by secretly calling up some of the players who have already left the room and urging them to return. Each player also knows that any such new agreement is also suspect if the next player to leave has the same concerns. Our objective is to find an agreement among the grand coalition such that regardless of the order of exit, the remaining players in the room have no incentive to form a new agreement.

form a coalition with other players and deviate from the initial deviation in a self-enforcing manner. Thus, the  $n$ -player agreement (which is non-Nash) is stable (in the von Neumann-Morgenstern sense).

As is the case for all other coalitional equilibrium concepts, there is no general existence theorem for UCPE (other than the assurance that it exists under conditions sufficient for existence of a strong equilibrium). Instead, we think that the case for UCPE is most strongly made by the application to problems in which CPE does not provide us with any plausible prediction about the likely course of events. In this spirit, we provide several applications. Using a variety of simple models, we show the following: (i) a "pyramid", as opposed to a vertical form of organization, is a stable hierarchical design in a firm; (ii) agreement on a neutral standard is possible among competing firms even if they have their own favored product specifications; (iii) there exists a stable method of dividing a dollar using majority rule voting on division rules; and (iv) effort-inducing labor contracts can be self-enforcing despite the incentive among management and unions to renegotiate. In each of these problems, CPE yields no predictions.

## 2.1 THE GAME

$N = \{1, \dots, n\}$  is a set of *players*.  $\mathcal{H}$  is the *set of non-empty subsets of  $N$* . For every  $H \in \mathcal{H}$ ,  $-H$  is the complement of  $H$  in  $N$ . In any set  $X$ , given  $x = (x_i)_{i \in N} \in X$ , for every  $H \in \mathcal{H}$ , let  $x_H = (x_i)_{i \in H}$  and for all  $i \in N$ ,  $x_{-i} = (x_j)_{j \in N \setminus \{i\}}$ .  $M = \times_{i \in N} M_i$  is a (finite) set of *joint moves* by the members of  $N$ , with  $M_i$  being the set of moves for  $i$ . Every  $i \in N$  has a *utility function*  $u_i: M \rightarrow \mathbb{R}$ . A *game*  $\Gamma$  is the triple  $\langle N, M, (u_i)_{i \in N} \rangle$ . An *agreement* is a pair  $[m, H] \in M \times \mathcal{H}$ , with  $\mathcal{A}$  denoting the set of all agreements.

An agreement is interpreted as a specification of (a) the moves to be taken by all parties to the agreement, for a given list of moves for all other players, and (b) the set of players forming the agreement. The objective is to model a one-shot simultaneous move (non-cooperative) game with the possibility of pre-play negotiation among coalitions of players. An agreement is simply a representation of the moves negotiated among the members of  $H$ , given the assumption that the play of those in  $-H$  is held fixed. This is the standard method of analyzing the incentive to defect from a status quo in a

non-cooperative context with simultaneous play.

## 2.2 A SOLUTION TO THE GAME: UCPE

The solution is based on the following reasoning: suppose, to start with, that all players in  $N$  agree on an  $n$ -tuple of moves, and ask whether any sub-coalition  $H$  would have the incentive to defect from the  $n$ -tuple, assuming that the remaining players play according to the last agreement that they have knowledge of. This follows from the principle (employed by all progenitors of our solution) that given knowledge of a particular candidate list of moves, players believe that other players will play according to this list unless they receive some new information. Consistency requires that the agreement made by the defecting sub-coalition must be subjected to the same test, and so should further defections from this defection, ...and so on. Once the players have contemplated all such agreements in a consistent manner, they make their moves simultaneously if it is determined that the candidate  $n$ -tuple is not threatened in a "credible" manner; and payoffs result. The remainder of this section formalizes this logic.

Define a binary relation  $\succ$  on  $\mathcal{A} \times \mathcal{A}$  as follows:  $[\hat{m}, H] \succ [m', J]$  if  $H \cap J \neq \emptyset$  and

$$(i) \forall i \in H, u_i(\hat{m}) > u_i(m'),$$

$$(ii) \hat{m}_{-H} = m'_{-H}, \text{ and}$$

$$(iii) \forall m \in M, \forall i \in HN, \forall j \in H \cap J, u_j(\hat{m}_H, m_{-H}) > u_j(m'_i, m_{-i}) \Leftrightarrow u_i(\hat{m}_H, m_{-H}) > u_i(m'_i, m_{-i}).$$

If  $[\hat{m}, H] \succ [m', J]$ , we say that  $[\hat{m}, H]$  *threatens*  $[m', J]$ . Moreover,  $[\hat{m}, H]$  and  $[m', J]$  are referred to, respectively, as the *source* and the *target* (of the threat). Members of the sets  $H \cap J$  and  $HN$  are referred to, respectively, as the *defectors* and the *recruits*.<sup>2</sup>

$[\hat{m}, H]$  threatens  $[m', J]$  if three conditions are satisfied. Both the defectors and the recruits must

---

<sup>2</sup>It is important to note that the set of defectors is required to be non-empty to initiate a threat; this follows from the presumption that each player assumes that the status quo is maintained unless informed otherwise. Thus, suppose a candidate "equilibrium", say  $m$ , is common knowledge, and it is threatened by  $[m', J]$  which, in turn, is threatened by  $[\hat{m}, H]$ . If  $H$  and  $J$  have no members in common, the agreement  $[m', J]$  is unobservable (and, therefore, cannot be a target) to those in  $H$ , who assume that the status quo, i.e. the "equilibrium" will be played.

be made better off than they were in the target agreement (condition (i)). This provides the incentive for the formation of the source agreement. The moves of all players not party to the source agreement are held fixed at the values specified by the target agreement (condition (ii)).

Condition (iii) is referred to as a *lateral induction* condition. If the defectors know what the profile of moves in target agreement is, they must convince the recruits that it is indeed  $m'$ . (Subsequently, we shall argue that defectors will always be fully informed.) When the target agreement is a candidate "equilibrium", of course, we presume that all the moves are common knowledge. However, all other agreements are out of equilibrium and each recruit would know only what he/she is playing in the status quo. The moves to be played by the recruits and the defectors in the source agreement are observable to everybody party to that agreement. The defectors make the following implicit "speech" (see Cho and Kreps (1987) for other examples of such "speeches") to each recruit  $i$ .

*"We want to convince you that the target agreement contains the profile  $m'_i$ . The only situation in which we would have an incentive to defect from the target is when we stand to benefit from the defection. If the target contains an  $N\{i\}$ -profile of moves for which you would be made worse off by joining the source agreement, then at least one defector would be made worse off as well. If there is any  $N\{i\}$ -profile of moves that makes us better off from the defection, then you would be better off as well. This coincidence of interests is captured as follows:  $\hat{m}_H$  are the moves that you can observe if the defection occurs, and  $m'_i$  is the move that you can observe in the absence of the defection. Hence, the fact that we are proposing the formation of the source agreement must convey a signal that we are truthfully revealing our information about the moves in the target agreement, since we have no incentive to lie."*<sup>3</sup>

Begin with an  $n$ -player agreement that is common knowledge and identify the agreements that threaten it. The defectors are fully informed at this point. Next, identify the agreements that threaten the second agreement. Again the defectors are fully informed. And so on. By inductively applying the lateral induction argument above, we can ensure that defectors will always be fully informed about the

---

<sup>3</sup>It is important to note that the condition presented is sufficient but not necessary to "signal" information. We focus on this condition because, as we shall show later, it has several other critical properties: it leads to the resulting solution concept being a Nash equilibrium refinement, and is extremely easy to use in applications.

moves profile in any target agreement in the hierarchy of agreements.

Our objective is to determine whether an  $n$ -player agreement is stable, i.e. it is never threatened or is threatened only by an agreement that is threatened by an agreement that is never threatened or ...and so on. This logic follows the von Neumann and Morgenstern stable sets approach adopted by Greenberg (1989), Kahn and Mookherjee (1989) and Dutta *et al* (1989). This approach is formalized as follows:

Consider three subsets of  $\mathcal{A}$ : the *good*,  $\mathcal{G}(\succ)$ , the *bad*,  $\mathcal{B}(\succ)$ , and the *ugly*,  $\mathcal{U}(\succ)$ , defined by:

$$\mathcal{B}(\succ) = \{[m', H] \in \mathcal{A} : \exists [m'', J] \in \mathcal{G}(\succ) \text{ such that } [m'', J] \succ [m', H]\}.$$

$$\mathcal{G}(\succ) = \{[m', H] \in \mathcal{A} : \text{if } \exists [m'', J] \in \mathcal{A} \text{ such that } [m'', J] \succ [m', H], \text{ then } [m'', J] \in \mathcal{B}(\succ)\}.$$

$$\mathcal{U}(\succ) = \mathcal{A} \setminus (\mathcal{G}(\succ) \cup \mathcal{B}(\succ)).$$

$\{\mathcal{G}(\succ), \mathcal{B}(\succ), \mathcal{U}(\succ)\}$  is said to be a  $(\succ)$ -*semi-stable* partition of  $\mathcal{A}$  if  $\mathcal{G}(\succ) \cap \mathcal{B}(\succ) = \emptyset$ ; and if, in addition,  $\mathcal{U}(\succ) = \emptyset$ , it is said to be a  $(\succ)$ -*stable* partition of  $\mathcal{A}$ .

To account for the possibility that these partitions are non-unique, define a *minimal*  $(\succ)$ -*semi-stable* partition  $\{\mathcal{G}^*(\succ), \mathcal{B}^*(\succ), \mathcal{U}^*(\succ)\}$  of  $\mathcal{A}$  as one that satisfies:  $\mathcal{G}^*(\succ) \subseteq \mathcal{G}(\succ)$  and  $\mathcal{B}^*(\succ) \subseteq \mathcal{B}(\succ)$  for every  $(\succ)$ -semi-stable partition  $\{\mathcal{G}(\succ), \mathcal{B}(\succ), \mathcal{U}(\succ)\}$  of  $\mathcal{A}$ .

Now we are in a position to define our solution concept.

An  $n$ -tuple  $m$  is a *Universal Coalition-proof Equilibrium* (UCPE) if  $[m, N] \in \mathcal{G}^*(\succ)$ .

Our solution concept says that an  $n$ -player agreement is stable if it is threatened only by strictly bad agreements. A weaker notion could also be defined as follows:

An  $n$ -tuple  $m$  is a *weak Universal Coalition-proof Equilibrium* (W-UCPE) if  $[m, N] \in \mathcal{G}^*(\succ) \cup \mathcal{U}^*(\succ)$ .

The weaker concept says that an  $n$ -player agreement is stable if it is not threatened by any strictly good agreements. In the following section we shall argue that the latter is not a plausible notion of stability in a non-cooperative context.<sup>4</sup>

In the appendix, we show that there is no guarantee of either a  $(\succ)$ -stable partition or of a unique

---

<sup>4</sup>Observe that the existence of a minimal semi-stable partition does not ensure existence of either UCPE or W-UCPE.



( $\succ$ )-semi-stable one. Hence, both the ugly set and minimal semi-stability are unavoidable. The former introduces some ambiguity in the recursive logic derived from stable set theory. The latter is intended to resolve any conflict arising in situations where an agreement is good under one partition and bad under another.<sup>5</sup>

To show existence of a minimal ( $\succ$ )-semi-stable partition, consider the following construction from Kahn and Mookherjee (1989). Define the sets  $\mathcal{G}_0^*$  and  $\mathcal{B}_0^*$  as follows:

$$\mathcal{G}_0^* = \{[m, H] \in \mathcal{A} : \exists \text{ no } [m', J] \in \mathcal{A} \text{ such that } [m', J] \succ [m, H]\}.$$

$$\mathcal{B}_0^* = \{[m, H] \in \mathcal{A} : \exists [m', J] \in \mathcal{G}_0^* \text{ such that } [m', J] \succ [m, H]\}.$$

Next, inductively define  $\mathcal{G}_z^*$ ,  $\mathcal{B}_z^*$  with  $z = 1, 2, \dots$  as follows:

$$\mathcal{G}_z^* = \{[m, H] \in \mathcal{A} : \text{if } [m', H'] \succ [m, H], \text{ then } [m', H'] \in \mathcal{B}_{z-1}^*\}.$$

$$\mathcal{B}_z^* = \{[m, H] \in \mathcal{A} : \exists [m', H'] \in \mathcal{G}_z^* \text{ such that } [m', H'] \succ [m, H]\}.$$

For all  $z$ ,  $\mathcal{G}_{z-1}^* \subseteq \mathcal{G}_z^*$  and  $\mathcal{B}_{z-1}^* \subseteq \mathcal{B}_z^*$ , by the definitions given above.

Define  $\mathcal{G}^* = \bigcup_{z=0}^{\infty} \mathcal{G}_z^*$  and  $\mathcal{B}^* = \bigcup_{z=0}^{\infty} \mathcal{B}_z^*$ . Observe that if  $[m, H] \in \mathcal{G}^*$  and there exists  $[m', H'] \in \mathcal{A}$  such that

$[m', H'] \succ [m, H]$ , then there exists  $z$  such that  $[m, H] \in \mathcal{G}_z^*$  and  $[m', H'] \in \mathcal{B}_{z-1}^*$ . Hence,  $[m', H'] \in \mathcal{B}^*$ .

Conversely, if  $[m, H] \in \mathcal{B}^*$ , then there exists  $z$  such that  $[m, H] \in \mathcal{B}_z^*$  and  $[m', H'] \in \mathcal{G}_z^*$  such that  $[m', H'] \succ [m, H]$ . Hence,  $[m', H'] \in \mathcal{G}^*$ .  $\mathcal{G}^*$  and  $\mathcal{B}^*$  satisfy the definition of good and bad sets respectively.

In addition, we claim that  $\mathcal{G}^* \cap \mathcal{B}^* = \emptyset$ .

Suppose otherwise, i.e.  $[m, H] \in \mathcal{G}^* \cap \mathcal{B}^*$ . By construction, for some  $z$ ,  $[m, H] \in \mathcal{G}_z^* \cap \mathcal{B}_z^*$ .  $[m, H] \in \mathcal{B}_z^*$  implies that there exists  $[m', H'] \in \mathcal{G}_z^*$  such that  $[m', H'] \succ [m, H]$ . Since  $[m, H] \in \mathcal{G}_z^*$  as well,  $[m', H'] \in \mathcal{B}_{z-1}^*$ . Hence, there exists  $[m'', H''] \in \mathcal{G}_{z-1}^*$  such that  $[m'', H''] \succ [m', H']$ . Since  $[m', H'] \in \mathcal{G}_z^*$ , and  $[m'', H''] \succ [m', H']$ ,  $[m'', H''] \in \mathcal{B}_{z-1}^*$ . Repeating this argument, we conclude that there exists  $[m, H] \in \mathcal{G}_0^* \cap \mathcal{B}_0^*$ . This is in contradiction with the definitions of  $\mathcal{G}_0^*$  and  $\mathcal{B}_0^*$ .

Finally, define  $\mathcal{U}^*$  as the complement of  $\mathcal{G}^* \cup \mathcal{B}^*$  in  $\mathcal{A}$ . Note that every agreement in  $\mathcal{G}_0^*$  is a good

<sup>5</sup>Note that there will never be a situation where a unique stable partition into good and bad sets exists, and yet the minimal semi-stable partition labels everything as ugly. Uniqueness of a stable partition, ensures that there will never be another partition that is not stable; hence, the trivial semi-stable partition cannot occur.

agreement for any ( $\succ$ )-semi-stable partition and, therefore, every agreement in  $\mathcal{G}_o^*$  is a bad agreement for any ( $\succ$ )-semi-stable partition. By applying an induction argument for all  $\mathcal{G}_z^*$  and  $\mathcal{B}_z^*$ , we have just shown the existence of a partition of  $\mathcal{A}$  which satisfies minimal semi-stability.

### 3. A SOLUTION UNDER NESTEDNESS: CPE

A stronger blocking device is used in Bernheim, et al. (1987), Greenberg (1989), Kahn and Mookherjee (1989), and Chakravorti (1993). This can be expressed as follows:

Define a second binary relation  $\succ\triangleright$  on  $\mathcal{A} \times \mathcal{A}$  as follows:  $[\hat{m}, H] \succ\triangleright [m', J]$  if

$$(i) [\hat{m}, H] \succ [m', J]$$

$$(ii) H \subseteq J.$$

If  $[\hat{m}, H] \succ\triangleright [m', J]$ , then  $[\hat{m}, H]$  *trumps*  $[m', J]$ . The trumping relation requires that the threat relation be met and a nestedness assumption (i.e. condition (ii) above) be satisfied. Note that the lateral induction condition is trivially satisfied in the case where  $H \subseteq J$ . Hence, the nestedness assumption ensures that this condition is automatically met.

Define  $\mathcal{G}(\succ\triangleright)$  and  $\mathcal{B}(\succ\triangleright)$  and  $\mathcal{U}(\succ\triangleright)$  as the good, bad and ugly subsets of  $\mathcal{A}$  by using trumping instead of a threat as a blocking device. Define  $(\succ\triangleright)$ -stability and  $(\succ\triangleright)$ -semi-stability and minimal  $(\succ\triangleright)$ -stability analogously.

A solution concept is derived in the following manner. First, show that there exists a unique  $(\succ\triangleright)$ -stable partition of  $\mathcal{A}$ . Second,  $m$  is a solution to the game if  $[m, N]$  is a good agreement.

**PROPOSITION 0:** (Greenberg (1989), Kahn and Mookherjee (1989)):  $\mathcal{A}$  admits a unique  $(\succ\triangleright)$ -stable partition.<sup>6</sup>

This facilitates an alternative to the inductive definition of Bernheim, et al. (1987)<sup>7</sup>. Formally,

An  $n$ -tuple  $m$  is a (nested) Coalition-proof Nash Equilibrium (CPE) if  $[m, N] \in \mathcal{G}(\succ\triangleright)$ .

<sup>6</sup>Finiteness of the game is necessary for existence and uniqueness.

<sup>7</sup>The induction being on the number of players.

Under nestedness, there are two methods of characterizing a solution. Without the assumption, neither method applies: the inductive approach fails since induction on the number of players assumes that only nested deviations occur; the stability approach fails since neither uniqueness nor the existence of a stable partition is guaranteed (see the appendix). Hence, the necessity of minimal ( $\succ$ )-semi-stability.

#### 4. RELATIONSHIP WITH OTHER SOLUTION CONCEPTS

In this section, we shall explore the relationship of UCPE with other solution concepts that have been proposed to apply to the same class of problems. These include CPE, an equilibrium concept based on Coalitional Contingent Threats (Greenberg (1989, 1989a)) and Nash equilibrium. Clearly, the set of strong Nash equilibria (Aumann (1959)) are contained in the strong UCPE set. Hence, UCPE exist when the former exists. A general proof of existence of UCPE would face the same difficulties as those associated with CPE and other coalitional concepts. First, we shall give some additional definitions.

##### 4.1 ADDITIONAL DEFINITIONS

We consider an alternative equilibrium concept based on the notion of Coalitional Contingent Threats, proposed by Greenberg (1989, 1989a) to characterize stable agreements in the presence of universal coalition formation. This concept is applicable when all negotiations are made publicly.

Define a binary relation  $\succ$  on  $\mathcal{A} \times \mathcal{A}$  as follows:  $[\hat{m}, H] \succ [m', J]$  if

(i)  $\forall i \in H, u_i(\hat{m}) > u_i(m')$ , and

(ii)  $\hat{m}_{-H} = m'_{-H}$ .

If  $[\hat{m}, H] \succ [m', J]$ , we say that  $[m, H]$  is an *objection* to  $[m', J]$ . Observe that an objection

involves precisely the same requirements as a threat except for the lateral induction criterion. The latter is not used since negotiations are assumed to be made in public. The underlying process is as follows (see Greenberg (1989, 1989a)). An  $n$ -player agreement  $[m, N]$  is on the table. A coalition  $J$  may openly declare that it objects to the agreement and will adopt  $m'_J$  instead provided the remaining players play  $m_{-J}$ . Another coalition, say  $H$ , can then object to the agreement  $[m' \equiv (m'_J, m_{-J}), J]$  by threatening to play  $\hat{m}_H$  provided the remaining players play  $m'_{-H}$ . This process continues until no coalition objects to a proposed profile, taking into account the possible reaction of other coalitions.

Define the sets  $\mathcal{G}^*(\succ)$ ,  $\mathcal{B}^*(\succ)$  and  $\mathcal{U}^*(\succ)$  such that  $\{\mathcal{G}^*(\succ), \mathcal{B}^*(\succ), \mathcal{U}^*(\succ)\}$  is a minimal  $(\succ)$ -semi-stable partition of  $\mathcal{A}$ . The construction of this partition is analogous to that given for the relation  $\succ$ .

An  $n$ -tuple  $m$  is a *Coalitional Contingent Threat Equilibrium (CCTE)* if  $[m, N] \in \mathcal{G}^*(\succ)$ ;  $m$  is a *weak Coalitional Contingent Threat Equilibrium (W-CCTE)* if  $[m, N] \in \mathcal{G}^*(\succ) \cup \mathcal{U}^*(\succ)$ .

In general, CCTE will not be a Nash equilibrium, which is defined as follows:

An  $n$ -tuple  $m$  is a *Nash Equilibrium* if  $\exists$  no  $[m', \{i\}] \in \mathcal{A}$  such that  $[m', \{i\}] \succ [m, N]$ .

A priori, only CPE satisfies the Nash criterion by definition, since the condition given above follows from the condition that  $[m, N] \in \mathcal{G}(\succ)$ .

## 4.2 PROPOSITIONS

The points made by the examples below do not rely on any non-genericity in the payoffs. The first proposition shows that even though  $\succ$  is weaker than  $\succ$ , this does not translate into a relationship between the corresponding solution concepts.

We begin with some notation which will be useful in making the arguments that follow. For any  $H, J \in$

$\mathcal{H}$  and  $m', m'' \in M$  define  $X[m'_H, m''_i; J] = \{m \in M: \forall j \in J, u_j(m'_H, m_{-H}) > u_j(m''_i, m_{-i})\}$ . By the definition, the following is true:

**LEMMA 1:**  $[\hat{m}, H] \succ [m', J]$  if and only if  $H \cap J \neq \emptyset$ , and

- (i)  $\forall i \in H, u_i(\hat{m}) > u_i(m')$ ,
- (ii)  $\hat{m}_{-H} = m'_{-H}$ , and
- (iii)  $\forall i \in H \cap J, X[\hat{m}_H, m'_i; H] = X[\hat{m}_H, m'_i; H \cap J]$ .

**PROPOSITION 1:** Neither UCPE nor W-UCPE is logically related to CPE.

Proof: To show that there are no logical containment relationships between either UCPE or W-UCPE and CPE, consider the following examples. The first example is provided for the sake of completeness; the applications in Section 5 also demonstrate the non-containment of UCPE in CPE.

**EXAMPLE 1** (To show that UCPE (or W-UCPE) is not a subset of CPE):

*[Insert Figure 1 here]*

Consider the game given above. Player 1 chooses from  $\{T, B\}$ , 2 chooses from  $\{L, R\}$  and 3 chooses from  $\{\ell, r\}$ . This example will show that  $[(B, R, \ell), \{1, 2, 3\}] \in \mathcal{B}(\succ)$  and  $[(B, R, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$ .

A Nash equilibrium of the game is  $(B, R, \ell)$ . It is not a CPE since  $[(T, L, \ell), \{1, 2\}] \succ [(B, R, \ell), \{1, 2, 3\}]$ . Check that  $[(T, L, \ell), \{1, 2\}] \in \mathcal{G}(\succ)$ . We claim, however, that  $[(B, R, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$ . The argument is in several steps.

(I)  $[(T, L, \ell), \{1, 2\}] \succ [(B, R, \ell), \{1, 2, 3\}]$  implies that  $[(T, L, \ell), \{1, 2\}] \succ [(B, R, \ell), \{1, 2, 3\}]$ .

(II) Next, we claim that  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, r), \ell; \{1, 3\}] = \{(B, L, \ell), (B, L, r), (T, L, \ell), (T, L, r)\} = X[(B, r), \ell; \{1\}]$ . By Lemma 1,  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$ .

Since  $[(B, L, r), \{1, 3\}]$  is not threatened by any other agreement, it is in  $\mathcal{G}^*(\succ)$ . Hence,  $[(T, L,$

$\ell$ ),  $\{1, 2\}\} \in \mathcal{G}^*(\succ)$ . Since  $[(T, L, \ell), \{1, 2\}]$  is the unique agreement that threatens  $[(B, R, \ell), \{1, 2, 3\}]$ , the latter is in  $\mathcal{G}^*(\succ)$ .

**EXAMPLE 2** (To show that CPE is not a subset of UCPE (or W-UCPE)):

*[Insert Figure 2 here]*

This game has player 3 choosing from  $\{\ell, c, r\}$ . The remaining players' strategy spaces are the same as in the previous game. This example will show that  $[(T, L, c), \{1, 2, 3\}] \in \mathcal{G}(\succ)$  and  $[(T, L, c), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$ .

Consider the Nash equilibrium  $(T, L, c)$ . It is also a CPE. This may be checked as follows. The only agreement that trumps  $[(T, L, c), \{1, 2, 3\}]$  is  $[(B, R, \ell), \{1, 2, 3\}]$ . However,  $[(T, L, \ell), \{1, 2\}] \succ [(B, R, \ell), \{1, 2, 3\}]$ .  $[(T, L, \ell), \{1, 2\}] \in \mathcal{G}(\succ)$  since it is not trumped by any agreements.

We claim that  $[(T, L, c), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$ . The argument is as follows:

- (I)  $[(T, L, \ell), \{1, 2\}] \succ [(B, R, \ell), \{1, 2, 3\}]$  implies that  $[(T, L, \ell), \{1, 2\}] \succ [(B, R, \ell), \{1, 2, 3\}]$ .
- (II) Next, we claim that  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, r), \ell; \{1, 3\}] = \{(B, L, \ell), (B, L, r), (T, L, \ell), (T, L, r), (B, L, c), (T, L, c)\} = X[(B, r), \ell; \{1\}]$ . By Lemma 1,  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$ .

Since  $[(B, L, r), \{1, 3\}]$  is not threatened by any other agreement, it is in  $\mathcal{G}^*(\succ)$ . Hence,  $[(T, L, \ell), \{1, 2\}] \in \mathcal{G}^*(\succ)$ . Since  $[(T, L, \ell), \{1, 2\}]$  is the only agreement that threatens  $[(B, R, \ell), \{1, 2, 3\}]$ , the latter is in  $\mathcal{G}^*(\succ)$ .  $[(B, R, \ell), \{1, 2, 3\}] \succ [(T, L, c), \{1, 2, 3\}]$  implies that  $[(T, L, c), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$ . ■

The second proposition shows that even though  $\succ$  is stronger than  $\succcurlyeq$ , this does not translate into a relationship between the corresponding solution concepts.

**PROPOSITION 2:** *Neither UCPE nor W-UCPE is logically related to either CCTE or W-CCTE.*

Proof: Consider the following examples.

**EXAMPLE 3:** (To show that UCPE is not contained in CCTE.)

*[Insert Figure 3 here]*

Consider the game given above. Player 1 chooses from  $\{T, B\}$ , player 2 chooses from  $\{L, C, R\}$  and player 3 chooses from  $\{\ell, r\}$ . The example shows that  $[(B, C, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$  and  $[(B, C, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ\!\succ)$ .

First, we show that  $[(B, C, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ)$ . The argument is in several steps:

(I)  $[(T, L, \ell), \{1, 2\}] \succ\!\succ [(B, C, \ell), \{1, 2, 3\}]$ , which implies that  $[(T, L, \ell), \{1, 2\}] \succ [(B, C, \ell), \{1, 2, 3\}]$ .

(II) We claim that  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, r), \ell; \{1, 3\}] = \{(B, L, \ell), (B, L, r), (T, L, \ell), (T, L, r)\} = X[(B, r), \ell; \{1\}]$ . By Lemma 1,  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$ .

Since  $[(B, L, r), \{1, 3\}]$  is not threatened by any other agreement, it is in  $\mathcal{G}^*(\succ)$ . Hence,  $[(T, L, \ell), \{1, 2\}] \in \mathcal{G}^*(\succ)$ . Since  $[(T, L, \ell), \{1, 2\}]$  is the only agreement that threatens  $[(B, C, \ell), \{1, 2, 3\}]$ , the latter is in  $\mathcal{G}^*(\succ)$ .

Next, we show that  $[(B, C, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ\!\succ)$ . First, observe that  $[(T, L, \ell), \{1, 2\}] \succ [(B, C, \ell), \{1, 2, 3\}]$  implies that  $[(T, L, \ell), \{1, 2\}] \succ\!\succ [(B, C, \ell), \{1, 2, 3\}]$ . Moreover,  $[(B, L, r), \{1, 3\}] \succ [(T, L, \ell), \{1, 2\}]$  implies that  $[(B, L, r), \{1, 3\}] \succ\!\succ [(T, L, \ell), \{1, 2\}]$ . Finally, observe that  $[(B, R, r), \{2\}] \succ\!\succ [(B, L, r), \{1, 3\}]$ .

Since there is no agreement that constitutes an objection to  $[(B, R, r), \{2\}]$ , the latter is in  $\mathcal{G}^*(\succ\!\succ)$ .  $[(B, R, r), \{2\}] \succ\!\succ [(B, L, r), \{1, 3\}]$  implies that  $[(B, L, r), \{1, 3\}] \in \mathcal{G}^*(\succ\!\succ)$ . Since  $[(B, L, r), \{1, 3\}]$  is the only agreement that is an objection to  $[(T, L, \ell), \{1, 2\}]$ , the latter is in  $\mathcal{G}^*(\succ\!\succ)$ .  $[(T, L, \ell), \{1, 2\}] \succ\!\succ [(B, C, \ell), \{1, 2, 3\}]$  implies that  $[(B, C, \ell), \{1, 2, 3\}] \in \mathcal{G}^*(\succ\!\succ)$ .

Next, consider the following example.

**EXAMPLE 4:** (To show that CCTE is not contained in UCPE.)

*[Insert Figure 4 here]*

In the game above, Player 1 chooses from  $\{T, B\}$  and player 2 chooses from  $\{L, R\}$ . This example shows

that  $[(T, L), \{1, 2\}] \in \mathcal{G}^*(\Rightarrow)$  and  $[(T, L), \{1, 2\}] \in \mathcal{B}^*(\succ)$ .

$[(T, R), \{2\}] \gg [(T, L), \{1, 2\}]$  and  $[(B, R), \{1\}] \gg [(T, R), \{2\}]$ . Also,  $[(B, L), \{1\}] \gg [(T, L), \{1, 2\}]$  and  $[(B, R), \{2\}] \gg [(B, L), \{1\}]$ . Since there is no agreement that constitutes an objection to  $[(B, R), \{1\}]$ , the latter is in  $\mathcal{G}^*(\Rightarrow)$ . Hence,  $[(B, R), \{1\}] \gg [(T, R), \{2\}]$  implies that  $[(T, R), \{2\}] \in \mathcal{B}^*(\Rightarrow)$ . Moreover, since there is no agreement that constitutes an objection to  $[(B, R), \{2\}]$ , the latter is in  $\mathcal{G}^*(\Rightarrow)$ . Hence,  $[(B, R), \{2\}] \gg [(B, L), \{1\}]$  implies that  $[(B, L), \{1\}] \in \mathcal{B}^*(\Rightarrow)$ . Given that the set of agreements that are objections to  $[(T, L), \{1, 2\}]$  is  $\{[(T, R), \{2\}], [(B, L), \{1\}]\} \subseteq \mathcal{B}^*(\Rightarrow)$ , we have  $[(T, L), \{1, 2\}] \in \mathcal{G}^*(\Rightarrow)$ .

Next, observe that  $[(T, R), \{2\}] \succ [(T, L), \{1, 2\}]$  and  $[(B, L), \{1\}] \succ [(T, L), \{1, 2\}]$ . However, there is no agreement that threatens either  $[(T, R), \{2\}]$  or  $[(B, L), \{1\}]$ . Hence,  $\{[(T, R), \{2\}], [(B, L), \{1\}]\} \subseteq \mathcal{G}^*(\succ)$ , which implies that  $[(T, L), \{1, 2\}] \in \mathcal{B}^*(\succ)$  since the set of agreements that threaten it is  $\{[(T, R), \{2\}], [(B, L), \{1\}]\}$ . ■

The examples show why the relationship between  $\gg$ ,  $\succ$  and  $\succ\rightarrow$  do not carry over to the corresponding solution concepts. When we strengthen a dominance relation the effect on the set of equilibria is ambiguous. On the one hand, candidates for equilibrium are now dominated by fewer defections, thereby increasing the number of equilibria. On the other hand, defections are now dominated by fewer counter-defections, thereby increasing the number of defections and reducing the number of equilibria.

The last example also shows that a CCTE need not be a Nash equilibrium. This might be expected of a UCPE as well. However, we have the following result.

**PROPOSITION 3:** *A UCPE is also a Nash equilibrium.*

Proof: Suppose  $m$  is a UCPE and is not a Nash equilibrium. Then there exists  $i \in N$  such that  $[m' = (m'_1, m'_2), \{i\}] \succ [m, N]$ . By definition of UCPE,  $[m, N] \in \mathcal{G}^*(\succ)$ . Hence,  $[m', \{i\}] \in \mathcal{B}^*(\succ)$  and, therefore, there must exist an agreement  $[m'' = (m''_{H \cup \{i\}}, m''_{-(H \cup \{i\})}), \{\{i\} \cup H\}] \in \mathcal{G}^*(\succ)$  such that  $[m'', \{\{i\} \cup H\}] \succ [m', \{i\}]$ . We shall consider two cases:

Case (i):  $\text{There exists } j \in H \text{ such that } u_j(m) > u_j(m'')$  This corresponds to the situation in which some recruit(s) prefer the initial  $n$ -tuple  $m$  to the  $n$ -tuple arising from the last defection,  $m''$ . By Lemma 1,



we must have  $X[m''_{H \cup \{i\}}, m_j; H \cup \{i\}] = X[m''_{H \cup \{i\}}, m_j; \{i\}]$ , i.e.  $u_i(m) > u_i(m'')$ . We have a contradiction with the requirement that  $[m'', \{\{i\} \cup H\}] \succ [(m'_1, m_{-1}), \{i\}] \succ [m, N]$  must imply  $u_i(m'') > u_i(m'_1, m_{-1}) > u_i(m)$ .

Case (ii): There exists no  $j \in H$  such that  $u_j(m) > u_j(m'')$  This corresponds to the situation in which no

recruit prefers the initial  $n$ -tuple  $m$  to the post-defection  $n$ -tuple  $m''$ . Consider two sub-cases:

Case (ii)-a: Suppose that for all  $j \in H$ ,  $u_j(m) < u_j(m'')$ .  $[m'', \{\{i\} \cup H\}] \succ [(m'_1, m_{-1}), \{i\}] \succ [m, N]$  implies that  $u_i(m'') > u_i(m)$ . Given that for all  $j \in H$ ,  $u_j(m'') > u_j(m)$ , we have  $[m'', \{\{i\} \cup H\}] \succ [m, N]$ . This implies that  $[m'', \{\{i\} \cup H\}] \succ [m, N]$  and is in contradiction with the assumption that  $[m, N] \in \mathcal{E}^*(\succ)$ .

Case (ii)-b: Suppose there exists  $j \in H$  such that  $u_j(m) = u_j(m'')$ .  $[m'', \{\{i\} \cup H\}] \succ [(m'_1, m_{-1}), \{i\}] \succ [m, N]$  implies that  $u_i(m'') > u_i(m)$ . By Lemma 1, we must have  $X[m''_{H \cup \{i\}}, m_j; H \cup \{i\}] = X[m''_{H \cup \{i\}}, m_j; \{i\}]$ , i.e.  $u_k(m) > u_k(m'')$  for all  $k \in H$ . Thus, we have a contradiction. ■

The intuition underlying the proof above is the following: the only way in which a UCPE, say  $m$ , could exist without satisfying the Nash best response property is if a player's unilateral defection could be held in check by a further defection in which the player enters into a good agreement with some other players. However, to convince these latter players to join in the agreement which results in  $m''$ , the player's preference for the post-defection  $n$ -tuple  $m''$  over  $m$  should be mirrored in a similar preference for  $m''$  by the recruits as well. This would help in convincing the recruits that even if the alternative to  $m''$  is  $m$ , they would have nothing to worry about. However, if the recruits and the original defecting player prefer  $m''$  over  $m$ , then this coalition could upset  $m$  directly; given that the agreement forged by the coalition is good, it would eliminate  $m$  as a candidate for UCPE.

The weaker notion W-UCPE does not, however, have the Nash equilibrium property. This can be seen from the following proposition. This makes the W-UCPE less desirable as a characterization of stability in a non-cooperative context.

**PROPOSITION 4:** *There is no logical relationship between W-UCPE and Nash equilibrium.*

Proof: From example 2, we know that a Nash equilibrium is not necessarily W-UCPE. To see that a W-UCPE is

not necessarily a Nash equilibrium, consider the following example.

**Example 5:** (To show that W-UCPE is not contained in NE.)

*[Insert Figure 5 here]*

Consider the game above. Player 1 chooses from  $\{T, B\}$ , 2 chooses from  $\{L, R\}$  and 3 chooses from  $\{\ell, \rho\}$ . We claim that  $[(T, L, \ell), \{1, 2, 3\}] \in \mathcal{U}^*(\succ)$  whereas  $(T, L, \ell)$  is not a Nash equilibrium. We need the following lemmata.

**LEMMA 2:**  $\{[(B, R, \ell), \{1, 2\}], [(B, L, \rho), \{2, 3\}], [(T, L, \ell), \{1, 3\}]\} \subseteq \mathcal{U}^*(\succ)$ .

**Proof of Lemma 2:** We shall show that the relation  $\succ$  induces a cycle in the set of agreements  $\{[(B, R, \ell), \{1, 2\}], [(B, L, \rho), \{2, 3\}], [(T, L, \ell), \{1, 3\}]\}$ . The argument proceeds in several steps.

(I) We claim that  $[(B, L, \rho), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(L, \rho), \ell; \{2, 3\}] = \{(B, L, \ell), (B, L, \rho), (B, R, \ell), (B, R, \rho)\} = X[(L, \rho), \ell; \{2\}]$ . By Lemma 1,  $[(B, L, \rho), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}]$ .

(II) Next, we claim that  $[(T, L, \ell), \{1, 3\}] \succ [(B, L, \rho), \{2, 3\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(T, \ell), B; \{1, 3\}] = \{(T, L, \ell), (B, L, \ell), (T, L, \rho), (B, L, \rho)\} = X[(T, \ell), B; \{3\}]$ . By Lemma 1,  $[(T, L, \ell), \{1, 3\}] \succ [(B, L, \rho), \{2, 3\}]$ .

(III) Finally, we claim that  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, R), L; \{1, 2\}] = \{(T, L, \ell), (B, L, \ell), (T, R, \ell), (B, R, \ell)\} = X[(B, R), L; \{1\}]$ . By Lemma 1,  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}]$ .

The steps (I)-(III) have generated a cycle of threatened agreements since  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}] \succ [(B, L, \rho), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}] \succ \dots$

We claim that  $[(B, R, \ell), \{1, 2\}] \in \mathcal{U}^*(\succ)$ . Suppose otherwise. There are two cases to be examined.

Suppose  $[(B, R, \ell), \{1, 2\}] \in \mathcal{G}^*(\succ)$ , in which case it is threatened by an agreement in  $\mathcal{G}^*(\succ)$ . However, there is no agreement other than  $[(B, L, \rho), \{2, 3\}]$  that threatens  $[(B, R, \ell), \{1, 2\}]$ . Hence,  $[(B, L, \rho), \{2, 3\}] \in \mathcal{G}^*(\succ)$ . Then  $[(T, L, \ell), \{1, 3\}] \in \mathcal{G}^*(\succ)$  since  $[(T, L, \ell), \{1, 3\}] \succ [(B, L, \rho), \{2, 3\}]$ . If  $[(T, L, \ell), \{1, 3\}] \in \mathcal{G}^*(\succ)$ , we must have  $[(B, R, \ell), \{1, 2\}] \in \mathcal{G}^*(\succ)$  since

there is no agreement other than  $[(B, R, \ell), \{1, 2\}]$  that threatens  $[(T, L, \ell), \{1, 3\}]$ . Hence, we have a contradiction.

Suppose that  $[(B, R, \ell), \{1, 2\}] \in \mathcal{G}^*(\succ)$ , in which case  $[(B, L, \rho), \{2, 3\}] \in \mathcal{G}^*(\succ)$ , since  $[(B, L, \rho), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}]$ . Thus,  $[(T, L, \ell), \{1, 3\}] \in \mathcal{G}^*(\succ)$  since  $[(T, L, \ell), \{1, 3\}]$  is the only agreement that threatens  $[(B, L, \rho), \{2, 3\}]$ . If  $[(T, L, \ell), \{1, 3\}] \in \mathcal{G}^*(\succ)$ , we must have  $[(B, R, \ell), \{1, 2\}] \in \mathcal{G}^*(\succ)$  since  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}]$ . Hence, we have a contradiction.

An analogous argument can be given to show that  $[(T, L, \ell), \{1, 3\}] \in \mathcal{U}^*(\succ)$  and  $[(B, L, \rho), \{2, 3\}] \in \mathcal{U}^*(\succ)$ . ■

**LEMMA 3:**  $[(T, R, \ell), \{2\}] \in \mathcal{U}^*(\succ)$ .

Proof of Lemma 3: First, we claim that  $[(B, R, \ell), \{1, 2\}] \succ [(T, R, \ell), \{2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, R), T; \{1, 2\}] = \{(T, L, \ell), (B, L, \ell), (T, R, \ell), (B, R, \ell), (T, R, \rho), (B, R, \rho)\} = X[(B, R), T; \{2\}]$ . By Lemma 1,  $[(B, R, \ell), \{1, 2\}] \succ [(T, R, \ell), \{2\}]$ .

Observe that neither  $[(B, L, \rho), \{1, 2, 3\}]$  nor  $[(B, L, \rho), \{2, 3\}]$  threatens  $[(T, R, \ell), \{2\}]$ . In each case, though conditions (i) and (ii) for a threat are met, condition (iii) is not satisfied. The unique agreement that threatens  $[(T, R, \ell), \{2\}]$  is  $[(B, R, \ell), \{1, 2\}]$ .  $[(B, R, \ell), \{1, 2\}] \in \mathcal{U}^*(\succ)$  (from Lemma 2) implies that  $[(T, R, \ell), \{2\}] \in \mathcal{U}^*(\succ)$ . ■

To check that  $[(T, L, \ell), \{1, 2, 3\}] \in \mathcal{U}^*(\succ)$ , observe that the only agreements that threaten it are  $[(B, R, \ell), \{1, 2\}]$  and  $[(T, R, \ell), \{2\}]$ . By Lemma 2 and Lemma 3,  $\{[(B, R, \ell), \{1, 2\}], [(T, R, \ell), \{2\}]\} \subseteq \mathcal{U}^*(\succ)$ . ■

## 5. APPLICATIONS

This section presents applications of the concept presented earlier. The applications focus on

classes of economic problems in which UCPE identifies plausible outcomes while its competitor, CPE, provides no solution. In keeping with the earlier material, we focus exclusively on pure strategies. Although they are presented as games with fairly limited strategy spaces, each example can be generalized. The key insights are obtained with three players; hence, to highlight these insights instead of the details of a general model, we limit ourselves to three agent problems.

### 6.1 STABLE MANAGEMENT HIERARCHIES WITH MORAL HAZARD

Consider a simple hierarchical model of an organization, where the top management and middle and lower level employees have divergent agendas; we shall refer to the top management agenda as the "public agenda" (such as customer focus) and the middle and lower level employees' agenda as the "private agenda" (such as pursuing leisure-time activities on the job). As pointed out by Tirole (1990), the study of stable hierarchies involves a careful understanding of the role of collusion between various levels in the hierarchy. In the game to be developed below we shall formally examine such collusion possibilities; in addition, we will treat compensation schemes as exogenously determined. Endogenizing the compensation schemes and determining "optimal" schemes will not affect the results.

There are three players: *principal* ( $P$ ), *supervisor* ( $S$ ) and *agent* ( $A$ ). There are two potential management schemes that  $P$  can adopt:  $M_P = \{\text{"Manage" (MAN), "Delegate" (DEL)}\}$ .  $S$  can choose from the following three actions:  $M_S = \{\text{"Work For P's Agenda" (PUB), "Manage" (MAN), "Work For Private Agenda" (PVT)}\}$ .  $A$  can choose from two actions:  $M_A = \{\text{"Work For P's Agenda" (PUB), "Work For Private Agenda" (PVT)}\}$ . The firm's output level can take one of three values: "High" ( $h$ ), "Low" ( $l$ ) or 0, with  $h > l > 0$ .  $P$  receives this output. The act of "PVT" is more generally interpreted as "shirking" -- that is taking an action not in  $P$ 's interest, and, in fact, it generates high payoffs to those who shirk if they can do so unsupervised. The act of "DEL" yields utility to  $P$ , since, *ceteris paribus*, it is assumed that it is costly for  $P$  to manage.

The firm's output is generated by a production function  $f: M_P \times M_S \times M_A \rightarrow \{h, l, 0\}$  defined by:

$$f(\text{MAN}, \text{PUB}, \text{PUB}) = h.$$

$$f(MAN, MAN, PUB) = f(MAN, PVT, PUB) = f(MAN, PUB, PVT) = f(DEL, MAN, PUB) = \ell.$$

$$f(DEL, PVT, PUB) = f(DEL, PUB, PUB) = f(DEL, PVT, PVT) = f(DEL, MAN, PVT) = f(DEL, PUB, PVT) = f(MAN, MAN, PVT) = f(MAN, PVT, PVT) = 0.$$

The production function has the following natural properties. A manager and a single public-agenda worker produce a low output. An additional manager makes no contribution, if there is a manager already. A manager with two public agenda workers yield a high output. The production of non-zero output requires at least one manager and one public agenda worker.

$P$  receives the output and pays wages and bonuses as necessary; in addition, he derives utility from delegating to  $S$ , denoted by a payoff of  $d > 0$ . If there is positive output, then anyone who works for the public agenda receives wages  $w > 0$  with  $w < \ell$ ,  $w < 0.5h$ . If output is positive, anyone who works for a private agenda receives 0; in case output is 0, and there is a manager, anyone who works for a private agenda is penalized by an amount  $p > 0$ . If there is no manager, then anyone who works for the public agenda receives 0 and anyone who works for the private agenda unilaterally gets a benefit  $s > 0$ , and anyone who works for a private agenda jointly with someone else gets benefit  $k > s$ . There are "network externalities" in working for a private agenda (this is consistent with situations in which leisure is more enjoyable when there is company).

Finally, if  $P$  delegates to  $S$  and if the latter manages, then  $S$  receives a bonus equal to a fraction  $\beta$  times the output, with  $\beta\ell > w$  and  $k > \beta\ell > s$ .  $S$  receives no reward for managing if  $P$  has decided to manage, since  $S$ 's choice generates no additional output. When someone works for a private agenda and no one manages,  $P$  incurs a cost  $c > d$ . Also,  $h - 2w > d > \beta\ell$

The three players may communicate how they plan to act in the game, however, they cannot sign binding contracts. For the sake of convenience, the payoffs can be displayed as follows ( $P$  chooses the rows,  $S$  chooses columns and  $A$  chooses the matrix):

*[Insert Figure 9 here]*

The question to be addressed here is: what management scheme will be adopted? There are two Nash equilibria:  $(MAN, PUB, PUB)$  and  $(DEL, MAN, PUB)$ . These equilibria correspond, respectively, to a "pyramid" and a "linear" hierarchical structure. An analysis of the problem using CPE yields the

following conclusion: no matter how high a wage  $w$  and bonus  $\beta$ , the principal pays, provided these payments satisfy: (i)  $\beta l > w$ , (ii)  $k > \beta l > s$  and (iii)  $h - 2w > d > \beta l$ , there is no stable outcome in this game -- neither hierarchical structure is expected to arise. This can be seen as follows.

Observe that  $[(DEL, MAN, PUB), \{P, S\}] \succ [(MAN, PUB, PUB), \{P, S, A\}]$  and  $[(DEL, MAN, PUB), \{P, S\}] \in \mathcal{S}(\succ)$ . Also,  $[(DEL, PVT, PVT), \{S, A\}] \succ [(DEL, MAN, PUB), \{P, S, A\}]$  and  $[(DEL, PVT, PVT), \{S, A\}] \in \mathcal{S}(\succ)$ .

However, this game admits a UCPE, i.e.  $(MAN, PUB, PUB)$ . Observe that  $[(DEL, PVT, PVT), \{S, A\}] \succ [(DEL, MAN, PUB), \{P, S\}] \succ [(MAN, PUB, PUB), \{P, S, A\}]$ . Check that the lateral induction condition is satisfied. Also  $[(DEL, PVT, PVT), \{S, A\}] \in \mathcal{S}^*(\succ)$ , and there is no other agreement that threatens  $[(DEL, MAN, PUB), \{P, S\}]$ . The only other agreement that threatens  $[(MAN, PUB, PUB), \{P, S, A\}]$  is  $[(DEL, MAN, PUB), \{P, S, A\}]$ . But then, by the argument above  $[(DEL, MAN, PUB), \{P, S, A\}] \in \mathcal{S}^*(\succ)$ . Hence,  $[(MAN, PUB, PUB), \{P, S, A\}] \in \mathcal{S}^*(\succ)$ .

To sum up, UCPE captures an important feature of management planning which CPE misses entirely due to its definitional constraints: when the cat is away, the mice will play; hence, the cat should not be away.<sup>8</sup>

## 6.2 STANDARDS-SETTING NEGOTIATIONS

There are three firms 1, 2 and 3 that compete for a given market worth  $\Pi$  in profits in the absence of any cost saving technologies. The firms must choose product specifications which save costs. If a particular specification is adopted as the industry standard, the aggregate cost saving over the entire market is equal to  $\epsilon$ . This surplus could be shared in different ways depending on which specification is adopted and is outlined below. If all three choose the same product specification, then it is the

---

<sup>8</sup>Several major corporations in the U.S. are moving in the direction of a pyramid structure, where the corporate headquarters are moved from a distant location, and brought closer to the employees, thereby eliminating several layers of middle management. The most prominent example of this is IBM's recent announcement of abandonment of their "house on the hill" in Armont, NY, and locating top management near the workforce.

industry standard, and the market (and  $\Pi$ ) is equally shared. If any two firms choose the same specification then it is the industry standard and each firm gets half of the market (and  $\Pi$ ) due to network externalities among consumers. If the firms choose three different specifications, they lose the entire market to a powerful foreign competitor. Each firm's objective is to maximize its profits.

The set of possible product specifications,  $X$ , is given by  $X = \{x^1, x^2, x^3, x^0\}$ . For each firm  $i$ ,  $x^i$  is the "favorite" product specification which has been privately developed in  $i$ 's R&D lab, with  $x^i \neq x^j$  if  $i \neq j$ . Firm  $i$  owns the cost saving technology for producing specification  $x^i$ . If  $i$  licenses its technology out to the others, it must mutually agree on a licensing fee cum inducement scheme with firms to which licenses are granted, denoted  $\delta(i) = (\delta^1(i), \delta^2(i), \delta^3(i))$  drawn from the set  $\Delta(i) \equiv \{\delta(i): \delta^1(i) + \delta^2(i) + \delta^3(i) = 1; \delta^1(i) > 0\}$ . The choice of  $\delta(i)$  gives  $i$  the right to retain a strictly positive share of the cost saving (the licensing fee),  $\delta^i(i)\epsilon$ , while the others who service the market using  $x^i$ , say  $j$ ,  $k$ , share  $(1 - \delta^i(i))\epsilon$  in the proportions  $\delta^j(i)$  and  $\delta^k(i)$ . The specification  $x^0$  is in the public domain. If the latter emerges as the industry standard, the surplus  $\epsilon$  is automatically shared equally by all three firms. Any firm which chooses  $x^i$  must obtain the license from firm  $i$  to produce the product and earn profits. If firm  $i$  licenses  $x^i$ , both firms  $j$  and  $k$  can choose  $x^i$ .

Now consider a simultaneous-move version game to analyze the situation above. Each firm  $i$  chooses a specification, with the choice denoted  $x_i \in X$  and makes a decision as to whether or not it will license out its technology; in case  $i$  does decide to license the technology out, it proposes a rule drawn from  $\Delta(i)$  for sharing the surplus  $\epsilon$ . Otherwise, it proposes a rule drawn from  $\Delta(j)$  or  $\Delta(k)$ . Firm  $i$ 's choice of a sharing rule is denoted  $\delta_i$ . Note that the subscript on  $x$  identifies the firm that is choosing the strategy, while a superscript on  $x$  identifies the firm whose favorite specification has been chosen. Thus, the strategy space for firm  $i$  is  $M_i = X \times \Delta(1) \cup \Delta(2) \cup \Delta(3)$ .

Given a profile of moves  $m = (x_i, \delta_i)_{i=1,2,3} \in M$ , define

$$S(m) = \{i \in \{1, 2, 3\}: \text{either (i) } \exists j \neq i, k \text{ such that } x_i = x_j = x^k, \text{ and } \delta_k = \delta_i = \delta_j \in \Delta(k), \\ \text{or (ii) } \exists j \neq i \text{ such that } x_i = x_j = x^0\}.$$

$S(m)$  is the set of firms that will produce according to the industry standard, and earn positive profits. Also, define  $\hat{k}(S(m)) = \{k \in \{1, 2, 3\}: \exists i, j \in S(m) \text{ with } x_i = x_j = x^k\}$ . The latter is the firm whose

product specification is the industry standard. Note that  $S(m)$  and  $\hat{k}(S(m))$  could both be empty.

The payoffs are specified by distinguishing between the following cases:

*Partition of  $M$ :*

Case 1:  $S(m) \neq \emptyset$ , and  $\hat{k}(S(m)) = \emptyset$ .

Case 2:  $S(m) \neq \emptyset$ , and  $\hat{k}(S(m)) \neq \emptyset$ .

Case 3:  $S(m) = \emptyset$ , and  $\hat{k}(S(m)) = \emptyset$ .

*Payoffs:*

$$\text{Case 1: } u_i(m) = \begin{cases} \Pi/|S(m)| + \epsilon/3 & \text{if } i \in S(m) \\ \epsilon/3 & \text{otherwise} \end{cases}$$

$$\text{Case 2: } u_i(m) = \begin{cases} \Pi/|S(m)| + (1-\delta^i(j))\epsilon/|S(m)\hat{k}(S(m))| & \text{if } i \in S(m) \setminus \hat{k}(S(m)), j \in \hat{k}(S(m)) \\ \Pi/|S(m)| + \delta^i(i)\epsilon & \text{if } i \in S(m) \cap \hat{k}(S(m)) \\ \delta^i(i)\epsilon & \text{if } i \in \hat{k}(S(m)) \setminus S(m) \\ 0 & \text{otherwise} \end{cases}$$

Case 3:  $\forall i \in \{1, 2, 3\}, u_i(m) = 0$ .

In a Nash equilibrium in which positive profits are earned, either Case 1 or Case 2 must hold with  $|S(m)| = 3$ , i.e. all three choose the same  $x \in X$  and agree on the industry standard. However, no such equilibrium is a CPE since any two-firm coalition would upset it in a self-enforcing way. Suppose the initial equilibrium is  $m = (x_i, \delta_i)_{i=1,2,3}$  with  $x_i = x^0$  for all  $i$ . Observe that  $[(x^1, \bar{\delta}_1), (x^1, \bar{\delta}_2), (x_3, \delta_3), \{1, 2\}] \succ [m, \{1, 2, 3\}]$ , provided  $1/3 < \bar{\delta}_1^1 < 2/3$  and  $\bar{\delta}_1 = \bar{\delta}_2 \in \Delta(1)$ . Check that  $[(x^1, \bar{\delta}_1), (x^1, \bar{\delta}_2), (x_3, \delta_3), \{1, 2\}] \in \mathcal{S} (\succ)$ . Hence,  $m$  is not CPE. Next, suppose, without loss of generality, that the initial equilibrium is  $m = (x_i, \delta_i)_{i=1,2,3}$  with  $x_i = x^3$  and  $\delta_i = \delta(3) \in \Delta(3)$  for all  $i$ . Observe that, since  $\delta^3(3) > 0$ , there exists  $\bar{\delta} = \bar{\delta}_1 = \bar{\delta}_2 \in \Delta(1)$  such that  $[(x^1, \bar{\delta}_1), (x^1, \bar{\delta}_2), (x_3, \delta_3), \{1, 2\}] \succ [m, \{1, 2, 3\}]$ . Check that  $[(x^1, \bar{\delta}_1), (x^1, \bar{\delta}_2), (x_3, \delta_3), \{1, 2\}] \in \mathcal{S} (\succ)$ . Hence,  $m$  is not CPE. The argument is identical for any other Nash equilibrium in which positive profits are earned.

However, this problem does admit a stable outcome with positive profits if UCPE is used as the



predictor. To argue that UCPE exist in this problem, consider the Nash equilibrium,  $m = ((x^0, \delta_1), (x^0, \delta_2), (x^0, \delta_3))$ . Clearly, given any  $\delta_1^1$  such that  $1/3 < \delta_1^1 < 2/3$ , and  $\delta_1 = \delta_2 \in \Delta(1)$ , we have  $[(x^1, \delta_1), (x^1, \delta_2), (x_3, \delta_3)], \{1, 2\}] \succ [m, \{1, 2, 3\}]$ . However, since  $\delta_1^1 > 0$ , there exists  $\hat{\delta} = \hat{\delta}_2 = \hat{\delta}_3 \in \Delta(2)$  such that  $[(x^1, \delta_1), (x^2, \hat{\delta}_2), (x^2, \hat{\delta}_3)], \{2, 3\}] \succ [(x^1, \delta_1), (x^1, \delta_2), (x_3, \delta_3)], \{1, 2\}]$ . Check that the lateral induction condition holds; i.e. 3 can be convinced by 2 to join the latter agreement, which in turn is Pareto efficient among all agreements involving  $\{2, 3\}$ . Furthermore, 1 cannot be convinced in a similar way to create any further defections. Hence,  $[(x^1, \delta_1), (x^2, \hat{\delta}_2), (x^2, \hat{\delta}_3)], \{2, 3\}] \in \mathcal{S}^*(\succ)$ . The argument is similar for any agreement that threatens  $[m, \{1, 2, 3\}]$ . Hence,  $[m, \{1, 2, 3\}] \in \mathcal{S}^*(\succ)$ .<sup>9</sup>

### 6.3 THE DIVIDE-THE-DOLLAR GAME

The example presented in this sub-section is similar to the standard-setting problem posed above. However, it plays an important role in the development of the theory of coalition-proof equilibrium. Virtually the same example is used in Bernheim et al. (1987), pp. 8 to show non-existence of CPE, even in mixed strategies. In contrast, we shall show that UCPE exists (in pure strategies) in this game. Moreover, the game represents a natural division problem with much wider implications.

Consider a three player (A, B, C) game of divide-the-dollar, where the allocation is decided by majority. Each player simultaneously announces a strictly positive share for herself and a residual share for at most one other player. If two or more players propose the same allocation, then that division is implemented; if all disagree, the dollar is discarded. Each player's objective is to maximize expected share of the dollar. This game has no CPE (see Bernheim, et al. (1987) for the argument).

---

<sup>9</sup>The issue of permitting coalition-formation in the context of standards setting has recently been discussed in the context of auctioning the radio spectrum for a new generation of wireless telecommunications called PCS (Personal Communications Services). It has been suggested that prior to the auction various potential bidders should coordinate on a single standard for deploying PCS, given that there are several competing mutually incompatible alternatives. Our argument suggests that such pre-play communication among coalitions is likely to yield socially beneficial results.

Consider the Nash equilibrium in which  $A$  and  $B$  propose  $(0.5, 0.5, 0)$  and  $C$  proposes, say,  $(0, 0.4, 0.6)$ . This outcome is threatened by the possible joint defection by  $A$  and  $C$  to a proposal  $(0.6, 0, 0.4)$ . Note that both are made better off and the new proposal is Pareto efficient for  $A$  and  $C$ , and is a Nash equilibrium for the two-person game consisting of  $A$  and  $C$ . Therefore, the original proposal is not CPE.

Consider, for example, a second defection coordinated by  $C$  and  $B$  to  $(0, 0.55, 0.45)$ .  $C$  can convince  $B$  to participate in the second defection by arguing that no matter what the status quo was, if it were one that makes  $B$  worse off by defecting to  $(0, 0.55, 0.45)$ , then it would make  $C$  worse off as well. Thus, the very fact that  $C$  has agreed to defect to  $(0, 0.55, 0.45)$  would act as a signal to  $B$  that the defection is worthwhile. The argument is easily made in this particular case since  $B$  is unambiguously better off by participating in the agreement with  $C$ . Finally, observe that the last defection is Pareto efficient for  $B$  and  $C$ , and neither can coordinate a defection with  $A$ , since  $A$ 's knowledge of the status quo was the agreement on  $(0.6, 0, 0.4)$ .  $A$  cannot be completely convinced that the status quo is different -- both  $B$  and  $C$  would benefit by bluffing and pretending that the status quo was *not*  $(0.6, 0, 0.4)$ . Any defection with  $A$  that is attractive to  $B$  or  $C$  would make  $A$  worse off if the status quo were indeed  $(0.6, 0, 0.4)$ .

Thus, we would argue that the  $B, C$  defection is a credible threat to the  $A, C$  defection. A similar argument holds for *any*  $A, C$  or  $B, C$  defection from  $(0.5, 0.5, 0)$ ; hence the latter is a UCPE outcome.

#### 6.4 RENEGOTIATION OF CONTRACTS

There are three players: *Management* ( $P$ ), *Labor Union* ( $U$ ) and *Workers* ( $W$ ). The  $U$ 's payoff function is an isotone transformation of  $W$ 's payoff function. Each player has one of two actions.  $P$  and  $U$  must choose from an "(ex ante) efficient contract" ( $EFF$ ) and an "(ex ante) inefficient contract" ( $IN$ ).  $W$  must choose whether to "Work" ( $WOR$ ) or "Shirk" ( $SHK$ ).

$P$  and  $U$  must mutually agree to a contract, otherwise the payoffs are  $(0, 0, 0)$ .  $P$ 's objective is to induce  $W$  to choose  $WOR$ . For a given level of effort, the efficient contract is mutually beneficial. But it induces shirking, while the inefficient contract induces effort. This issue has motivated the recent

literature on contract renegotiation (Green and Laffont (1987), Fudenberg and Tirole (1990), Hart and Moore (1988), Ma (1991), etc.). There could be several reasons for the effort-inducing contract to be ex ante inefficient. The inefficiencies may arise from the fact that risk is imposed on a risk-averse agent or from various aspects of the structuring, monitoring and bonding costs associated with the writing of effort-inducing contracts. An example of possible payoffs consistent with this problem is given below in Figure 10. In this example,  $U$  and  $W$  have identical payoffs.

*[Insert Figure 10 here]*

The question is the following: given the potential to renegotiate the contract, is it possible to establish a contract that induces effort?

The game has two Nash equilibria:  $(EFF, EFF, SHK)$  and  $(IN, IN, WOR)$ . Clearly,  $(IN, IN, WOR)$  is not CPE since  $P$  and  $U$  will renegotiate the inefficient contract, and will not unilaterally defect after renegotiation.

If  $U$  and  $W$  were merged into one player, CPE predicts that an effort inducing contract is feasible; however, if the two are independent players, such a contract is infeasible. Consider the two-player game below (Figure 11) in which  $P$  plays with  $UW$  (the merger of the union and the workers) and the latter chooses both the contract and the level of effort. Given the coincidence of interests of workers and the union, this divergence of predictions in the two versions is somewhat troubling.

*[Insert Figure 11]*

With coincident interests between the players  $U$  and  $W$ , the lateral induction property is trivially satisfied.<sup>10</sup> Thus, it is easy to see that the UCPE of the two-player game coincides with the UCPE of the three-player game;  $UCPE = \{(EFF, EFF, SHK), (IN, IN, WOR)\}$ .

## 7. CONCLUDING REMARKS

---

<sup>10</sup>The fact that UCPE accounts for non-nested coalition formation between parties with coincident interests should be seen as a general strength of this concept.

We have proposed a solution concept that identifies stable outcomes in games with pre-play negotiation among coalitions without binding agreements or public observability of actual strategy choice. This solution generalizes the ideas underlying CPE by relaxing the assumption that deviations are made by nested coalitions. Moreover, we show that our solution is a Nash equilibrium refinement -- thereby making it consistent with the fundamental non-cooperative nature of the problem. Both of these objectives are achieved by a single condition known as "lateral induction". We also show that our concept makes plausible predictions in specific economic applications where CPE fails.

The general characterization is necessarily complex. It cannot rely on the inductive approach to defining such concepts (as adopted by Bernheim et al.) since induction on the number of players presumes nestedness; also, we cannot directly apply the von Neumann Morgenstern stability approach (as adopted by Greenberg (1990)) because of the non-existence and non-uniqueness of stable partitions. Moreover, the lateral induction criterion adds to the complexity of the definition. However, the basic intuition underlying the concept is a simple one. In particular, in a variety of applications such as the ones given here, UCPE are relatively easy to compute.

Unfortunately, as with other coalitional equilibrium concepts, there is no general existence theorem for UCPE, other than the ones available for strong equilibrium. As with the other coalitional concepts, the interest in ours derives from the fact that it provides plausible predictions about outcomes in several interesting economic contexts -- where UCPE do indeed exist.

We close with some remarks about the relationship with Chakravorti and Sharkey (1993), where the question of non-nested coalition formation is also addressed. In this paper, we were motivated by the observation that defectors will convince recruits to form a defection by signaling their superior information to the latter. In Chakravorti and Sharkey (1993), there is no signaling between defectors and recruits. Deviations are made consistent with whatever information players are presumed to have at any stage in the game. Hence, in contrast to the complete information generated at each level of defection via the lateral induction condition, in the other paper the information is asymmetric within defecting coalitions.

## References

- AUMANN, R. (1959): "Acceptable Points in General Cooperative  $n$ -Person Games," *Ann. Math. Stud.*, **40**, 287-324.
- BANKS, J. AND J. SOBEL (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, **55**, 647-662.
- BERNHEIM, D., B. PELEG AND M. WHINSTON (1987): "Coalition-proof Nash Equilibria, I: Concepts," *J. Econ. Theory*, **42**, 1-12.
- CHAKRAVORTI, B. AND W. SHARKEY (1993): "Consistency, Un-Common Knowledge, and Coalition-Proofness," mimeo, Bellcore.
- CHAKRAVORTI, B. (1993): "Dynamic Public Goods Provision with Coalitional Manipulation", *Journal of Public Economics*, forthcoming.
- CHO, I. AND D. KREPS (1987): "Signaling Games and Stable Equilibria," *Quart. J. Econ.*, **102**, 179-221.
- DUTTA, B., D. RAY, K. SENGUPTA, AND R. VOHRA (1989): "A Consistent Bargaining Set," *J. Econ. Theory*, **49**, 93-112.
- FUDENBERG, D. AND J. TIROLE (1990): "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, **58**, 1279-319.
- GREEN, J. AND J.J. LAFFONT (1987): "Renegotiation and the Form of Efficient Contracts," mimeo Harvard.
- GREENBERG, J. (1989): "Deriving Strong and Coalition-proof Nash Equilibrium from an Abstract System," *J. Econ. Theory*, **49**, 195-202.
- GREENBERG, J. (1989a): *The Theory of Social Situations*, Cambridge, UK, Cambridge University Press.
- HART, O. AND J. MOORE (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, **56**, 755-85.
- KAHN, C. AND D. MOOKHERJEE (1989): "The Good, the Bad and the Ugly: Coalition-proof Equilibrium in Games with Infinite Strategy Spaces," *Games and Econ. Behav.*
- KOHLBERG, E. AND J. MERTENS (1986): "On the Strategic Stability of Equilibria," *Econometrica*, **54**, 1003-38.
- MA, C.-T. (1991): "Adverse Selection in Dynamic Moral Hazard," *Quart. J. Econ.*, **106**.
- TIROLE, J. (1990): "Collusion and the Theory of Organizations," forthcoming in *Advances in Economic Theory*, 6th World Congress of the Econometric Society, J.J. Laffont (ed.) Cambridge University Press.

## APPENDIX

### NON-EXISTENCE OF A UNIQUE $\succ$ -(SEMI)-STABLE PARTITION

Note that the following arguments refer to Lemma 1 which is stated in Section 4.

**PROPOSITION A.1 :** *In general,  $\mathcal{A}$  does not admit a  $\succ$ -stable partition.*

Proof: The proof is by way of an example. We shall present a game for which  $\mathcal{U}(\succ) \neq \emptyset$ .

**Example 6:** Consider the following game. Player 1 chooses from  $\{T, B\}$ , 2 chooses from  $\{L, R\}$  and 3 chooses from  $\{\ell, r\}$ .

*[Insert Figure 6 here]*

We shall show that the relation  $\succ$  induces a cycle in the set of agreements  $\{[(B, R, \ell), \{1, 2\}], [(B, L, r), \{2, 3\}], [(T, L, \ell), \{1, 3\}]\}$ . The argument proceeds in several steps.

(I) We claim that  $[(B, L, r), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(L, r), \ell; \{2, 3\}] = \{(B, L, \ell), (B, L, r), (B, R, \ell), (B, R, r)\} = X[(L, r), \ell; \{2\}]$ . By Lemma 1,  $[(B, L, r), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}]$ .

(II) We claim that  $[(T, L, \ell), \{1, 3\}] \succ [(B, L, r), \{2, 3\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(T, \ell), B; \{1, 3\}] = \{(T, L, \ell), (B, L, \ell), (T, L, r), (B, L, r)\} = X[(T, \ell), B; \{3\}]$ . By Lemma 1,  $[(T, L, \ell), \{1, 3\}] \succ [(B, L, r), \{2, 3\}]$ .

(III) We claim that  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, R), L; \{1, 2\}] = \{(T, L, \ell), (B, L, \ell), (T, R, \ell), (B, R, \ell)\} = X[(B, R), L; \{1\}]$ . By Lemma 1,  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}]$ .

The steps (I)-(III) have generated a cycle of threatened agreements since  $[(B, R, \ell), \{1, 2\}] \succ [(T, L, \ell), \{1, 3\}] \succ [(B, L, r), \{2, 3\}] \succ [(B, R, \ell), \{1, 2\}] \succ \dots$

We claim that  $[(B, R, \ell), \{1, 2\}] \in \mathcal{U}(\succ)$ . Suppose otherwise. There are two cases to be examined.

Suppose  $[(B, R, \ell), \{1, 2\}] \in \mathcal{B}(\succ)$ , in which case it is threatened by an agreement in  $\mathcal{S}(\succ)$ . However, there is no agreement other than  $[(B, L, r), \{2, 3\}]$  that threatens  $[(B, R, \ell), \{1, 2\}]$ .

Hence,  $[(B, L, \rho), \{2, 3\}] \in \mathcal{G}(\succ)$ . Then  $[(T, L, \iota), \{1, 3\}] \in \mathcal{B}(\succ)$  since  $[(T, L, \iota), \{1, 3\}] \succ [(B, L, \rho), \{2, 3\}]$ . If  $[(T, L, \iota), \{1, 3\}] \in \mathcal{B}(\succ)$ , we must have  $[(B, R, \iota), \{1, 2\}] \in \mathcal{G}(\succ)$  since there is no agreement other than  $[(B, R, \iota), \{1, 2\}]$  that threatens  $[(T, L, \iota), \{1, 3\}]$ . Hence, we have a contradiction.

Suppose that  $[(B, R, \iota), \{1, 2\}] \in \mathcal{G}(\succ)$ , in which case  $[(B, L, \rho), \{2, 3\}] \in \mathcal{B}(\succ)$ , since  $[(B, L, \rho), \{2, 3\}] \succ [(B, R, \iota), \{1, 2\}]$ . Thus,  $[(T, L, \iota), \{1, 3\}] \in \mathcal{G}(\succ)$  since  $[(T, L, \iota), \{1, 3\}]$  is the only agreement that threatens  $[(B, L, \rho), \{2, 3\}]$ . If  $[(T, L, \iota), \{1, 3\}] \in \mathcal{G}(\succ)$ , we must have  $[(B, R, \iota), \{1, 2\}] \in \mathcal{B}(\succ)$  since  $[(B, R, \iota), \{1, 2\}] \succ [(T, L, \iota), \{1, 3\}]$ . Hence, we have a contradiction. ■

**PROPOSITION A.2:** *In general,  $\mathcal{A}$  does not admit a unique  $(\succ)$ -semi-stable partition.*

Proof: The proof is by way of an example. We shall present a game such that a cycle is generated as in the previous example. Each agreement in the cycle can be defined as both good and bad relative to corresponding re-definitions of the agreements in the cycle that threaten it.

**Example 7:** Consider the following game. Player 1 chooses from  $\{T, B\}$ , 2 chooses from  $\{L, R\}$ , 3 chooses from  $\{\iota, \rho\}$  and 4 chooses from  $\{U, D\}$ .

*[Insert Figure 7 here]*

We shall show that the relation  $\succ$  induces a cycle of the form discussed above in the set  $\{[(B, R, \iota), U], \{1, 2\}], [(B, L, \rho, U), \{2, 3\}], [(B, L, \rho, D), \{3, 4\}], [(T, L, \iota, U), \{1, 3, 4\}]\}$ . The argument proceeds in several steps.

(I) We claim that  $[(B, L, \rho, U), \{2, 3\}] \succ [(B, R, \iota, U), \{1, 2\}]$ . Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(L, \rho), \iota; \{2, 3\}] = \{(B, L, \iota, U), (B, L, \rho, U), (B, R, \iota, U), (B, R, \rho, U)\} = X[(L, \rho), \iota; \{2\}]$ . By Lemma 1,  $[(B, L, \rho, U), \{2, 3\}] \succ [(B, R, \iota, U), \{1, 2\}]$ .

(II) We claim that  $[(B, L, \rho, D), \{3, 4\}] \succ [(B, L, \rho, U), \{2, 3\}]$ . Conditions (i) and (ii) for a threat met. To check for condition (iii), verify that  $X[(\rho, D), U; \{3, 4\}] = \{(B, L, \iota, U), (B, L, \rho, U), (B, L, \iota, D), (B, L, \rho, D)\} = X[(\rho, D), U; \{3\}]$ . By Lemma 1,  $[(B, L, \rho, D), \{3, 4\}] \succ [(B, L, \rho, U), \{2, 3\}]$ .

$U), \{2, 3\}$ ].

(III) We claim that  $[(T, L, \ell, U), \{1, 3, 4\}] \succ [(B, L, r, D), \{3, 4\}]$ , Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(T, \ell, U), B; \{1, 3, 4\}] = \{(T, L, \ell, U), (B, L, \ell, U), (T, L, r, U), (B, L, r, U), (T, L, r, D), (B, L, r, D), (T, L, \ell, D), (B, L, \ell, D)\} = X[(T, \ell, U), B; \{3, 4\}]$ . Therefore,  $[(T, L, \ell, U), \{1, 3, 4\}] \succ [(B, L, r, D), \{3, 4\}]$ .

(IV) We claim that  $[(B, R, \ell, U), \{1, 2\}] \succ [(T, L, \ell, U), \{1, 3, 4\}]$ , Conditions (i) and (ii) for a threat are met. To check for condition (iii), verify that  $X[(B, R), L; \{1, 2\}] = \{(T, L, \ell, U), (B, L, \ell, U), (T, R, \ell, U), (B, R, \ell, U), (B, L, \ell, D), (B, R, \ell, D)\} = X[(B, R), L; \{1\}]$ . By Lemma 1,  $[(B, R, \ell, U), \{1, 2\}] \succ [(T, L, \ell, U), \{1, 3, 4\}]$ .

The steps (I)-(IV) have generated a cycle of trumping agreements since  $((B, R, \ell, U), \{1, 2\}) \succ ((T, L, \ell, U), \{1, 3, 4\}) \succ ((B, L, r, D), \{3, 4\}) \succ ((B, L, r, U), \{2, 3\}) \succ ((B, R, \ell, U), \{1, 2\}) \succ \dots$

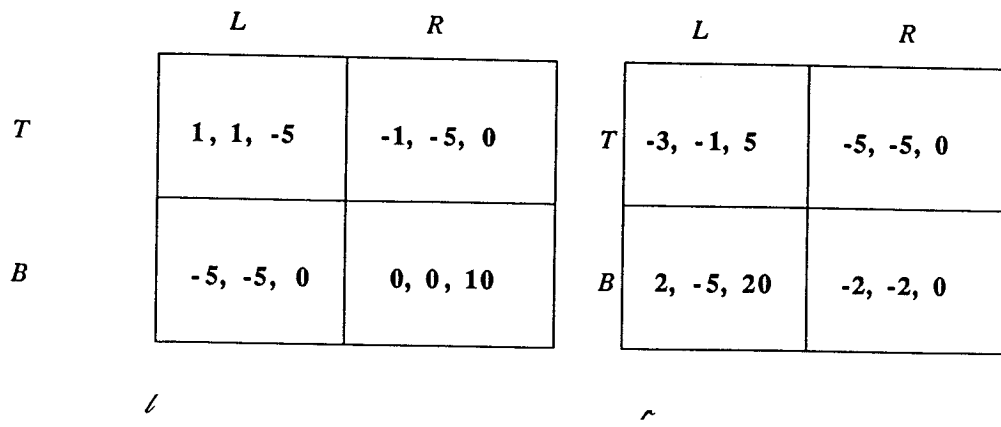
Also, check that for each one of the agreements in this cycle, there is only one agreement that threatens it. The cycle generated above has the following structure:

*[Insert Figure 8 here]*

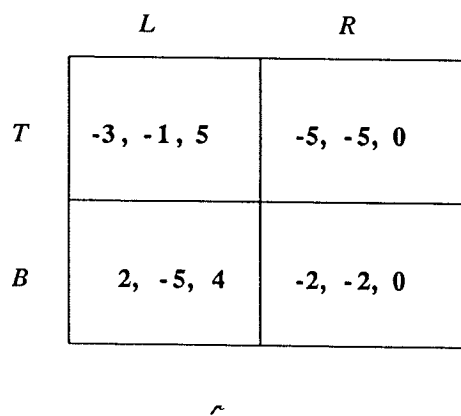
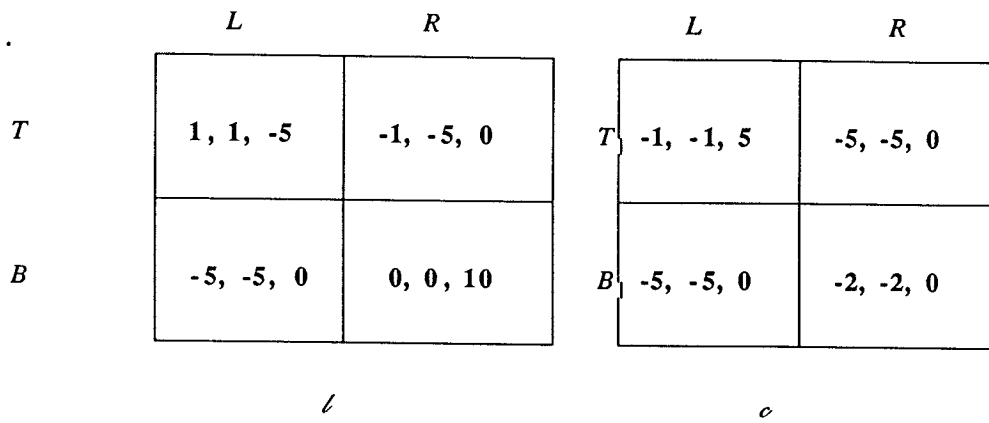
$\{A, B, C, D\}$  is such that  $A \succ B \succ C \succ D \succ A \succ \dots$ . B is the only agreement that threatens A. C is the only agreement that threatens B. D is the only agreement that threatens C and A is the only agreement that threatens D.

Suppose that  $A \in \mathcal{S}(\succ)$ . Then  $D \in \mathcal{B}(\succ)$ ,  $C \in \mathcal{S}(\succ)$  and  $B \in \mathcal{B}(\succ)$ . Alternatively, suppose  $A \in \mathcal{B}(\succ)$ . Then  $D \in \mathcal{S}(\succ)$ ,  $C \in \mathcal{B}(\succ)$  and  $B \in \mathcal{S}(\succ)$ . Both the partitions of  $\{A, B, C, D\}$  are admissible. ■





*Figure 1*



*Figure 2*

|          | <i>L</i>  | <i>C</i>  | <i>R</i> |
|----------|-----------|-----------|----------|
| <i>T</i> | 1, 1, -5  | -1, -5, 0 | 0, 0, 0  |
| <i>B</i> | -5, -5, 0 | 0, 0, 10  | 0, 0, 0  |

|          | <i>L</i>  | <i>C</i>  | <i>R</i> |
|----------|-----------|-----------|----------|
| <i>T</i> | -3, -1, 5 | -5, -5, 0 | 0, 0, 0  |
| <i>B</i> | 2, -5, 20 | -2, -2, 0 | 0, 0, 0  |

*l*
*r*

*Figure 3.*

|          | <i>L</i> | <i>R</i> |
|----------|----------|----------|
| <i>T</i> | 0, 5     | 0, 6     |
| <i>B</i> | 6, 0     | 1, 1     |

*Figure 4.*

|          |            |             |
|----------|------------|-------------|
|          | <i>L</i>   | <i>R</i>    |
| <i>T</i> | 10, 10, 10 | -3, 10.5, 4 |
| <i>B</i> | 9, 9, 6    | 11, 11, 5   |

|          |          |            |
|----------|----------|------------|
|          | <i>L</i> | <i>R</i>   |
| <i>T</i> | 9, 8, 3  | -4, -5, 0  |
| <i>B</i> | 8, 12, 7 | -2, -2, 11 |

*l*
*r*

*Figure 5.*

|          |            |           |
|----------|------------|-----------|
|          | <i>L</i>   | <i>R</i>  |
| <i>T</i> | 10, 10, 10 | -3, 9, 4  |
| <i>B</i> | 9, 9, 6    | 11, 11, 5 |

|          |          |            |
|----------|----------|------------|
|          | <i>L</i> | <i>R</i>   |
| <i>T</i> | 9, 8, 3  | -4, -5, 0  |
| <i>B</i> | 8, 12, 7 | -2, -2, 11 |

*l*
*r*

*Figure 6.*

*U*

|          | <i>L</i>       | <i>R</i>      |          | <i>L</i>    | <i>R</i>       |
|----------|----------------|---------------|----------|-------------|----------------|
| <i>T</i> | 10, 10, 10, 10 | -3, 9, 4, -3  | <i>T</i> | 9, 8, 3, 9  | -4, -5, 0, -4  |
| <i>B</i> | 9, 9, 6, 9     | 11, 11, 5, 11 | <i>B</i> | 8, 12, 7, 8 | -2, -2, 11, -2 |

*D*

|          | <i>L</i>      | <i>R</i>      |          | <i>L</i>      | <i>R</i>     |
|----------|---------------|---------------|----------|---------------|--------------|
| <i>T</i> | 12, 13, -1, 9 | 9, 13, -1, 9  | <i>T</i> | 12, 10, -1, 9 | 9, 10, -1, 9 |
| <i>B</i> | 9, 12, -1, 9  | 10, 13, -1, 9 | <i>B</i> | 9, 10, 9, 9.5 | 9, 10, -1, 9 |

Figure 7.

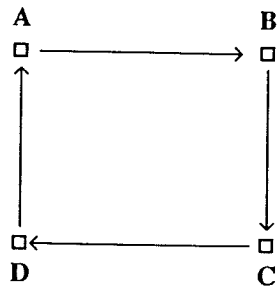


Figure 8.

|            | <i>PUB</i>     | <i>MAN</i>                       | <i>PVT</i>    |
|------------|----------------|----------------------------------|---------------|
| <i>MAN</i> | $h - 2w, w, w$ | $l - w, 0, w$                    | $l - w, 0, w$ |
| <i>DEL</i> | $d, 0, 0$      | $l - \beta l - w + d, \beta l w$ | $d - c, s, 0$ |

*PUB*

|            | <i>PUB</i>    | <i>MAN</i> | <i>PVT</i>    |
|------------|---------------|------------|---------------|
| <i>MAN</i> | $l - w, w, 0$ | $0, 0, -p$ | $0, -p, -p$   |
| <i>DEL</i> | $d - c, 0, s$ | $d, 0, -p$ | $d - c, k, k$ |

*PVT*

Figure 9

|            |            |           |            |            |           |
|------------|------------|-----------|------------|------------|-----------|
|            | <i>EFF</i> | <i>IN</i> |            | <i>EFF</i> | <i>IN</i> |
| <i>EFF</i> | 10, 10, 10 | 0, 0, 0   | <i>EFF</i> | 2, 15, 15  | 0, 0, 0   |
| <i>IN</i>  | 0, 0, 0    | 9, 9, 9   | <i>IN</i>  | 0, 0, 0    | 1, 3, 3   |
|            | <i>WOR</i> |           |            | <i>SHK</i> |           |

FIGURE 10

|            |                |               |                |               |
|------------|----------------|---------------|----------------|---------------|
|            | <i>WOR/EFF</i> | <i>WOR/IN</i> | <i>SHK/EFF</i> | <i>SHK/IN</i> |
| <i>EFF</i> | 10, 10         | 0, 0          | 2, 15          | 0, 0          |
| <i>IN</i>  | 0, 0           | 9, 9          | 0, 0           | 1, 3          |

FIGURE 11