Coalition-Proof Equilibrium*

Diego Moreno[†]

*Departamento de Economía, Universidad Carlos III de Madrid, 28903 Getafe
(Madrid), Spain*

and

John Wooders[‡]

*Department of Economics, University of Arizona, McClelland Hall, Tucson,
Arizona 85721*

We characterize the agreements that the players of a noncooperative game may reach when they can communicate prior to play, but they cannot reach binding agreements: A *coalition-proof equilibrium* is a correlated strategy from which no coalition has an improving and self-enforcing deviation. We show that any correlated strategy whose support is contained in the set of actions that survive the iterated elimination of strictly dominated strategies and weakly Pareto dominates every other correlated strategy whose support is contained in that set, is a coalition-proof equilibrium. Consequently, the unique equilibrium of a dominance solvable game is coalition-proof. *Journal of Economic Literature* Classification Numbers: C72, D82. © 1996 Academic Press, Inc.

INTRODUCTION

When the players of a noncooperative game have the opportunity to communicate prior to play, they will try to reach an agreement to coordinate their actions in a mutually beneficial way. The aim of this paper is to characterize the set of agreements that the players may reach. Since we consider situations where agreements are nonbinding, only those agreements that are not subject to viable (i.e., *self-enforcing*) deviations are of

interest. As preplay communication allows the players to correlate their play, we take the set of all correlated strategies as the space of feasible agreements. We characterize the set of coalition-proof equilibria as the set of agreements from which no coalition has a self-enforcing deviation making all its members better off.

Admitting correlated strategies as feasible agreements alters the set of coalition-proof equilibria of a game in a fundamental way (viz., no inclusion relationship between the notion of coalition-proofness that we propose and others previously introduced is to be found). In fact, there are games where the only plausible agreements are correlated (and not mixed) agreements. We provide examples with this feature and we show that the notion of coalition-proof equilibrium that we propose identifies these agreements. Unfortunately, as with other notions of coalition proofness previously introduced, the existence of an equilibrium cannot be guaranteed. We are able to establish, however, that if there is a correlated strategy which (i) has a support contained in the set of actions that survive the iterated elimination of strictly dominated strategies, and (ii) weakly Pareto dominates every other correlated strategy whose support is contained in that set, then this strategy is a coalition-proof equilibrium. Consequently, the unique equilibrium of a dominance solvable game is coalition-proof.

Other authors have explored the implications of preplay communication when agreements are mixed strategy profiles. Aumann (1959) introduced the notion of strong Nash equilibrium, which requires that an agreement not be subject to an improving deviation by any coalition of players. This requirement is too strong, since agreements must be resistant to deviations which are not themselves resistant to further deviations. Recognizing this problem, Bernheim, Peleg, and Whinston (1987) (henceforth referred to as BPW) introduced the notion of coalition-proof Nash equilibrium (CPNE), which requires only that an agreement be immune to improving deviations which are self-enforcing. A deviation is self-enforcing if there is no further self-enforcing and improving deviation available to a proper subcoalition of players. This notion of self-enforceability provides a useful means of distinguishing coalitional deviations that are viable from those that are not resistant to further deviations. Only viable deviations can upset potential agreements. A deficiency of CPNE, however, is that it does not allow players to agree to correlate their play.

Although the possibility that players correlate their actions when given the opportunity to communicate was recognized as early as in Luce and Raiffa (1957), only recently did Einy and Peleg (1995) (E & P) introduce a concept of coalition-proof communication equilibrium. The difference between E & P's notion and ours can be better understood if we assume that correlated agreements are carried out with the assistance of a media-

tor. The mediator selects an action profile according to the agreement and then makes a (private and nonbinding) recommendation of an action to each player.

E & P consider situations where the players may plan deviations only after receiving recommendations. In our framework, however, players plan deviations before receiving recommendations, and no further communication is possible after recommendations are issued. This difference manifests itself most clearly in two-person games where an agreement is coalition-proof in our sense only if it is Pareto-efficient within the set of correlated equilibria, while an agreement that is coalition-proof in E & P's sense need not be. We provide an example with this feature in Section 4. The second difference is that in our framework deviations may involve the members of a coalition jointly ''misreporting'' their types, while this possibility is not considered by E & P's notion. In Section 4 these differences are discussed in detail.

Ray (1996) proposes a notion of coalition-proof correlated equilibrium in which the players' possibilities of correlating their play are limited by an exogenously given *correlation device*, and he shows that there are coalition-proof equilibria which cannot be attained by means of *direct devices* (i.e., devices in which players' messages are their action spaces). This finding raises the question whether allowing nondirect devices might alter the coalition-proof correlated equilibria (CPCE) of a game when, as in our definition, players' possibilities of correlating their play are not exogenously given. We do not have a general answer to this question. For games which satisfy the sufficient conditions we provide for existence of a unique CPCE, however, the set of equilibria we identify as coalition-proof is the same regardless of whether or not nondirect devices are available to the players.

As the following example illustrates, correlated play naturally arises when communication is possible (and regardless of whether or not players have access to a correlation device). Therefore one should take the set of correlated strategies as the set of feasible agreements, and one must consider deviations that involve correlated play by members of a deviating coalition.

Three-Player Matching Pennies Game (*TPMPG*). Three players each simultaneously choose heads or tails. If all three faces match, then players 1 and 2 each win a penny while player 3 loses two pennies. Otherwise, player 3 wins two pennies while players 1 and 2 each lose a penny.

The matrix representation of this game is given in Table I. This game has two pure-strategy and one mixed-strategy Nash equilibria: one pure-

TABLE I
The Three-Player Matching Pennies Game

	H_3		T_3	
	H_2	T_2	T_2	H_2
H_1	$1, 1, -2$	$-1, -1, 2$	$-1, -1, 2$	$-1, -1, 2$
T_1	$-1, -1, 2$	$-1, -1, 2$	$-1, -1, 2$	$1, 1, -2$

strategy equilibrium consists of players 1 and 2 each choosing heads (tails) and player 3 choosing tails (heads). In the mixed strategy equilibrium each player chooses heads with probability $\frac{1}{2}$.

The game does not have a CPNE, as each of the Nash equilibria is upset by a deviation of the coalition of players 1 and 2; in the pure-strategy Nash equilibrium where players 1 and 2 both choose heads, they each obtain a payoff of -1. By jointly deviating (both choosing tails instead) players 1 and 2 each obtain a payoff of 1. This deviation is self-enforcing as players 1 and 2 each obtain their highest possible payoffs and therefore neither player can improve by a further unilateral deviation. (A symmetric argument shows that the other pure-strategy Nash equilibrium is not a CPNE either.) In the mixed-strategy Nash equilibrium, players 1 and 2 each obtain an expected payoff of $-\frac{1}{2}$. This equilibrium is not a CPNE as players 1 and 2 can jointly deviate (both choosing heads instead) and obtain a payoff of zero. This deviation is self-enforcing, since given that player 3 chooses heads or tails with equal probability, neither player can obtain more than zero by a further deviation. Since a CPNE must be a Nash equilibrium, this game has no CPNE.

Nevertheless, the game does have an agreement that is resistant to improving deviations. This agreement is the correlated strategy where with probability $\frac{1}{2}$ players 1 and 2 both choose heads and with probability $\frac{1}{2}$ both choose tails, and player 3 chooses heads or tails with equal probability. Under this agreement each player has an expected payoff of zero. No single player can deviate and improve upon this agreement: neither player 1 nor player 2 can benefit by unilaterally deviating, as they both lose a penny whenever their faces do not match. Neither does player 3 benefit from deviating: given the probability distribution over the moves of players 1 and 2, he is indifferent between heads and tails. Moreover, since the interests of players 1 and 2 are completely opposed to those of player 3, no coalition involving player 3 can improve upon the given agreement. Finally, given player 3's strategy, players 1 and 2 obtain at most a payoff of zero, and therefore they cannot benefit by deviating. Hence, no coalition can gain by deviating from the agreement.

TABLE II
An Entry Game

	Enter	Not enter
Enter	$-2, -2$	$1, -1$
Not enter	$-1, 1$	$0, 0$

Notice that the agreement described above is *not* a mixed strategy profile and so it cannot possibly be a CPNE. As we shall see, however, when we expand the space of agreements to include all the correlated strategies, this agreement is the unique coalition-proof equilibrium of the game. (See Moreno and Wooders, 1995, for an experimental study of this game.)

The possibility of players correlating their play arises even when communication is limited. Consider, for instance, the game described in Table II which is related to a class of games discussed in Farrell (1987); in this game two identical firms must simultaneously decide whether to enter a market which is a natural monopoly. This game has three Nash equilibria: (Enter, Not enter), (Not enter, Enter), and a mixed-strategy Nash equilibrium where each firm enters the market with probability $\frac{1}{2}$. Each of these Nash equilibria is also a CPNE.

Although the mixed Nash equilibrium is a CPNE, it is not resistant to improving deviations given the possibility of preplay communication. The firms can improve by augmenting the game with a round of cheap talk. In the game with cheap talk each firm simultaneously and publicly announces whether it intends to "Enter" or "Not enter" the market. Following the announcements each firm makes its choice.

Suppose the firms agree to play the following Nash (and subgame perfect) equilibrium of the game with cheap talk. Each firm announces "Enter" with probability $\frac{3}{4}$. If the profile of announcements is either (Enter, Not enter) or (Not enter, Enter), then each firm plays its announcement. Otherwise, each firm plays "Enter" with probability $\frac{1}{2}$. This equilibrium yields an expected payoff for each firm of $-\frac{5}{16}$ while in the mixed Nash equilibrium of the original game each firm has an expected payoff of only $-\frac{1}{2}$.

Preplay communication has enabled the firms to correlate their play. In this Nash equilibrium of the cheap talk game the firms effectively play the correlated strategy (of the original game) given in Table III. This joint probability distribution is not the product of its marginal distributions and therefore cannot be obtained from a mixed strategy profile of the game

TABLE III
The Correlated Strategy Induced by "Cheap Talk"

	Enter	Not enter
Enter	$\frac{5}{32}$	$\frac{11}{32}$
Not enter	$\frac{11}{32}$	$\frac{5}{32}$

without communication. This "correlated deviation" from the mixed strategy equilibrium makes both firms better off. Moreover, it is a self-enforcing deviation since it is a correlated equilibrium of the original game.

Expanding the set of feasible agreements from the mixed strategies (as in CPNE) to the set of correlated strategies does not lead simply to an expansion of the set of coalition-proof agreements. In the *Three-Player Matching Pennies* game we found a coalition-proof agreement where no CPNE existed. In the entry game we found a CPNE that was not coalition-proof. Thus, there is no inclusion between the set of CPNE and the set of equilibria that are coalition-proof in our sense.

In our framework the primitives are a set of feasible agreements and the concepts of feasible deviation and of self-enforcing deviation by a coalition from a given agreement. The set of feasible deviations by a coalition from a given agreement is the set of all correlated strategies that the coalition can induce when the complementary coalition behaves according to the given agreement and when the members of the coalition correlate their play. The definition of a self-enforcing deviation is recursive. For a coalition of a single player any feasible deviation is self-enforcing. For coalitions of more than one player, a deviation is self-enforcing if it is feasible and if there is no further self-enforcing and improving deviation by one of its proper subcoalitions. With these concepts, our notion of coalition-proofness is easily formulated; an agreement is coalition-proof if no coalition (not even the grand coalition) has a self-enforcing deviation that makes all its members better off.

Our notion of a self-enforcing deviation coincides with that implicit in the concept of CPNE. The difference between our notion of coalition-proofness and CPNE is only that we take the set of correlated strategies as the space of feasible agreements. For games of complete information, if feasible agreements are mixed strategies then our definition of coalition-proofness coincides with CPNE. (This is established in Appendix B.) In some situations it may be natural to restrict the space of feasible agreements (e.g., if communication is limited) or to limit the possibilities of players to form deviations. The framework we propose easily accommodates these kinds of changes.

In fact, our existence results are easily modified to provide conditions for the existence of a CPNE; namely, any mixed strategy profile whose support is contained in the set of action profiles that survive iterated elimination of strictly dominated strategies and which weakly Pareto-dominates any other mixed strategy profile whose support is contained in this set is a CPNE. For games with strategic complementarities, Milgrom and Roberts (1994) have independently obtained analogous results.

The paper is organized as follows: in Section 1 we discuss our framework and define our notion of equilibrium for games of complete information. In Section 2 we establish conditions for existence of these equilibria and show by means of an example that an equilibrium does not always exist. In Section 3 we extend the concept of coalition-proofness to games of incomplete information. Of course, the notion of coalition-proof equilibrium for games of incomplete information reduces to that formulated for games of complete information when every player has a single type. We present separately the notion of coalition-proofness for games of complete information, as the notion's simplicity in this context facilitates the discussion and because we want to stress the fact that our notion of coalition-proofness can be formulated without resorting to games of incomplete information. In Section 4 we compare our notion of coalition-proof equilibrium and E & P's notion of coalition-proof communication equilibrium, and we present some concluding remarks.

1. GAMES OF COMPLETE INFORMATION

A game in strategic form Γ is defined as

$$\Gamma = \left(N, (A_i)_{i \in N}, (u_i)_{i \in N} \right),$$

where N is the set of players, and for each $i \in N$, A_i is player i's set of actions (or pure strategies) and u_i is player i's utility (payoff) function, a real-valued function on $A = \Pi_{i \in N} A_i$. Assume that N and A are nonempty and finite. For any finite set Z, denote by ΔZ the set of probability distributions over Z. In particular, denote by ΔA the set of probability distributions over A, and refer to its members as *correlated strategies*. Given a correlated strategy μ, player i's expected utility when players' actions are selected according to μ is

$$U_i(\mu) = \sum_{a \in A} \mu(a) u_i(a).$$

A coalition of players S is a member of 2^N. When S consists of a single player $i \in N$, we write it as "i" rather than the more cumbersome $\{i\}$. For

each $S \in 2^N$, $S \neq \varnothing$, denote by A_S the set $\prod_{i \in S} A_i$. Given $a \in A$, we write $a = (a_S, a_{-S})$, where $a_S \in A_S$ and $a_{-S} \in A_{-S}$. If $S = N$, then $(a_S, a_{-S}) = a_S = a$.

Coalition-Proof Correlated Equilibrium

We conceive of communication and play as proceeding in two stages. In the first stage players communicate, reaching an agreement and possibly planning deviations from the agreement. Given an agreement $\mu \in \Delta A$, the players implement it with the assistance of a mediator who recommends the action profile $a \in A$ with probability $\mu(a)$. In the second stage, each player privately receives his component of the recommendation and then chooses an action. (No further communication occurs in this stage.)

A deviation by a coalition is a plan for its members to correlate their play in a way different from that prescribed by the agreement. We take a broad view of the ability of coalitions to plan deviations; for every different profile of recommendations received by its members, a deviating coalition may plan a different correlated strategy. Therefore, a deviation for a coalition S is a mapping from the set A_S of profiles of recommendations for its members to the set ΔA_S of probability distributions over the set of the coalition's action profiles.

Given an agreement $\overline{\mu}$, if a coalition S plans to deviate according to η_S: $A_S \to \Delta A_S$, and if the members of the complement of S play their part of the agreement, i.e., they obey their recommendations, then the induced probability distribution over action profiles for the grand coalition is given for each $a \in A$ by

$$\mu(a) = \sum_{\alpha_S \in A_S} \overline{\mu}(\alpha_S, a_{-S}) \eta_S(a_S \mid \alpha_S).$$

It will be convenient to define the feasible deviations for coalition S as those correlated strategies $\mu \in A$ which the coalition can induce, rather than as mappings from A_S to ΔA_S. Thus, a correlated strategy is a feasible deviation by coalition S from a given agreement if the members of S, using some plan to correlate their play, can induce the correlated strategy when each member of the complementary coalition obeys his recommendation.

DEFINITION 1.1. Let $\overline{\mu} \in \Delta A$ and $S \in 2^N$, $S \neq \varnothing$. We say that $\mu \in \Delta A$ is a feasible deviation by coalition S from $\overline{\mu}$ if there is an η_S: $A_S \to \Delta A_S$, such that for all $a \in A$ we have $\mu(a) = \sum_{\alpha_S \in A_S} \overline{\mu}(\alpha_S, a_{-S}) \eta_S(a_S \mid \alpha_S)$.

We illustrate our definition of a feasible deviation by describing a procedure that can be thought of as mimicking the process by which players select agreements and plan deviations. Given an agreement $\overline{\mu}$,

suppose that the mediator implementing $\bar{\mu}$ mails to each player a sealed envelope containing the player's recommendation. A coalition S deviates from $\bar{\mu}$ by employing a new mediator to which each member of S sends the (unopened) envelop it received from the mediator implementing $\bar{\mu}$. The new mediator opens the envelops, reads the recommendations α_S, and then selects a new profile of recommendations according to the correlated strategy $\eta_S(\alpha_S)$. The mediator then mails to each player $i \in S$ a sealed envelope containing his recommended action. When each player opens his envelope and obeys the recommendation it contains, the induced correlated strategy is given by the equation in Definition 1.1.

Given a coalition $S \in 2^N$, $S \neq \varnothing$, and an agreement $\mu \in \Delta A$, let $D(\mu, S)$ denote the set of feasible deviations by coalition S from μ; note that $\mu \in D(\mu, S)$, since a coalition always has the trivial "deviation" consisting of each member of the coalition obeying his own recommendation. Also note that for every $\mu \in \Delta A$, we have $D(\mu, N) = \Delta A$. A correlated equilibrium is a correlated strategy from which no individual has a feasible improving deviation.

Correlated Equilibrium.   A correlated strategy μ is a *correlated equilibrium* if no individual $i \in N$, has a feasible deviation $\tilde{\mu} \in D(\mu, i)$, such that $U_i(\tilde{\mu}) > U_i(\mu)$.

The definition of strong Nash equilibrium suggests the following definition of strong correlated equilibrium[1]: a strong correlated equilibrium is a correlated strategy from which no coalition has a deviation which makes every member of the coalition better off.

DEFINITION 1.2.   A correlated strategy $\mu \in \Delta A$ is a *strong correlated equilibrium* if no coalition $S \in 2^N$, $S \neq \varnothing$, has a feasible deviation $\tilde{\mu} \in D(\mu, S)$, such that for each $i \in S$, we have $U_i(\tilde{\mu}) > U_i(\mu)$.

The agreement described in the introduction for the *Three-Player Matching Pennies* game is, for example, the unique strong correlated equilibrium of that game. Like strong Nash equilibrium, the notion of strong correlated equilibrium is too strong. A strong correlated equilibrium must be resistant to *any* feasible deviation by *any* coalition. In particular, it must be resistant to deviations which are not themselves resistant to further deviations. Consider, for example, the *Prisoners' Dilemma* game described in Table IV. This game has a unique correlated equilibrium where (D, D) is played with probability one. This correlated equilibrium is not a strong correlated equilibrium since the correlated strategy $\tilde{\mu}$ consisting of playing (C, C) with probability one is a feasible deviation which makes both players better off. Since a strong correlated equilibrium must be a corre-

---

TABLE IV
A Prisoners' Dilemma Game

	C	D
C	1, 1	−1, 2
D	2, −1	0, 0

lated equilibrium, this game has no strong correlated equilibrium. Notice, however, that either player can unilaterally deviate from $\tilde{\mu}$ and increase his payoff. Hence $\tilde{\mu}$ should not undermine an agreement to play (D, D) with probability one.

In order to be able to distinguish those deviations that are viable from those that are not (and which therefore should not upset an agreement as coalition-proof) we introduce the notion of self-enforcing deviation; a correlated strategy μ is a self-enforcing deviation by coalition S from correlated strategy $\overline{\mu}$ if μ is a feasible deviation and if no proper subcoalition of S has a further self-enforcing and improving deviation. This notion of self-enforceability is identical to the one implicit in the concept of CPNE.

DEFINITION 1.3. Let $\overline{\mu} \in \Delta A$ and $S \in 2^N$, $S \neq \emptyset$. The *set of self-enforcing deviations by coalition S from* $\overline{\mu}$, SED($\overline{\mu}, S$), is defined, recursively, as follows:

(i) If $|S| = 1$, then SED($\overline{\mu}, S$) = $D(\overline{\mu}, S)$;

(ii) If $|S| > 1$, then SED($\overline{\mu}, S$) = $\{\mu \in D(\overline{\mu}, S) \mid \nexists [R \in 2^S \setminus S, R \neq \emptyset, \tilde{\mu} \in$ SED(μ, R)] such that $\forall i \in R: U_i(\tilde{\mu}) > U_i(\mu)\}$.

Since a coalition consisting of a single player has no proper (nonempty) subcoalitions, any feasible deviation by a one-player coalition is self-enforcing. With this notion of a self-enforcing deviation, a coalition-proof correlated equilibrium is defined to be a correlated strategy from which no coalition has a self-enforcing and improving deviation.

DEFINITION 1.4. A correlated strategy μ is a *coalition-proof correlated equilibrium* (CPCE) if no coalition $S \in 2^N$, $S \neq \emptyset$, has a deviation $\tilde{\mu} \in$ SED(μ, S), such that for each $i \in S$, we have $U_i(\tilde{\mu}) > U_i(\mu)$.

It is clear that a strong correlated equilibrium is a coalition-proof correlated equilibrium, which in turn is a correlated equilibrium. For two-player games the set of coalition-proof correlated equilibria is the set of correlated equilibria which are not strongly Pareto-dominated by other correlated equilibria (i.e., μ is a CPCE if it is a correlated equilibrium, and

there is no other correlated equilibrium μ' such that $U_i(\mu') > U_i(\mu)$ for each $i \in N$). Thus, for two-player games, the set of coalition-proof correlated equilibria is nonempty. Although existence of a CPCE cannot be guaranteed in general games, in the next section we identify conditions under which a CPCE exists.

2. EXISTENCE OF CPCE AND CPNE

In this section we show that a CPCE (CPNE) exists whenever there is a correlated (mixed) strategy whose support is contained in the set of action profiles that survive iterated elimination of strictly dominated strategies and which weakly Pareto-dominates every other correlated (mixed) strategy whose support is contained in this set. First, we define formally the notion of strict dominance.

DEFINITION 2.1.   Let $B = \prod_{i \in N} B_i \subset A$. An action $\bar{a}_j \in B_j$ is said to be *strictly dominated* in B, if there is $\sigma_j \in \Delta B_j$ such that for each $a_{-j} \in B_{-j}$,

$$\sum_{a_j \in B_j} \sigma_j(a_j) u_j(a_j, a_{-j}) > u_j(\bar{a}_j, a_{-j}).$$

Note that if $\bar{a}_j$ is strictly dominated in B, then it is also strictly dominated in $A_j \times B_{-j}$. The set of action profiles that survive iterated elimination of strictly dominated strategies, which we write as A^∞, is now easily defined.

DEFINITION 2.2.   The *set A^∞ of action profiles that survive the iterated elimination of strictly dominated strategies* is defined by $A^\infty = \prod_{i \in N} A_i^\infty$, where each $A_i^\infty = \bigcap_{n=0}^\infty A_i^n$, A_i^n is the set of actions that are not strictly dominated in $A^{n-1} = \prod_{i \in N} A_i^{n-1}$, and $A_i^0 = A_i$.

The following proposition establishes that if μ is a correlated strategy whose support is contained in A^∞, then the support of every self-enforcing deviation from μ by a coalition of more than one player is also contained in A^∞. For each $\mu \in \Delta A$ and $S \in 2^N$, $S \neq \varnothing$, write $A_S^+(\mu)$ for the set $\{a_S \in A_S \mid \mu(a_S, a_{-S}) > 0 \text{ for some } a_{-S} \in A_{-S}\}$, and write $A^+(\mu)$ for the set $A_N^+(\mu)$.

PROPOSITION.   *Let $\mu \in \Delta A$ be such that $A^+(\mu) \subset A^\infty$, and let $S \in 2^N$ be a coalition of more than one player. If $\tilde{\mu} \in \mathrm{SED}(\mu, S)$, then $A^+(\tilde{\mu}) \subset A^\infty$.*

Proof. Let S and $\mu \in \Delta A$ be as in the proposition, and let $\tilde{\mu} \in \mathrm{SED}(\mu, S)$. By the definition of feasible deviation (Definition 1.1) $A^+(\tilde{\mu}) \subset A_S \times A_{-S}^\infty$. We show that in fact $A^+(\tilde{\mu}) \subset A^\infty$. Suppose by way of contradiction that $A_S^+(\tilde{\mu})$ is not contained in A_S^∞. Let n^* be the largest n

such that $A_S^+(\tilde{\mu}) \subset A_S^n$. Hence there is $j \in S$ and $\bar{a}_j \in A_j^+(\tilde{\mu})$ such that $\bar{a}_j$ is strictly dominated in A^{n*}. Thus $\bar{a}_j$ is also strictly dominated in $A_j \times A_{-j}^{n*}$; i.e., there is $\sigma_j \in \Delta A_j$ such that for each $a_{-j} \in A_{-j}^{n*}$,

$$\sum_{a_j \in A_j} \sigma_j(a_j) u_j(a_j, a_{-j}) > u_j(\bar{a}_j, a_{-j}). \qquad (*)$$

Consider the deviation μ' by player j (a proper subcoalition of S) from $\tilde{\mu}$, where player j chooses an action according to σ_j when recommended $\bar{a}_j$ and takes the recommended action otherwise. Formally, the deviation η_j is defined as follows: for each $a_j \in A_j$ such that $a_j \neq \bar{a}_j$, let $\eta_j(a_j \mid \alpha_j) = 1$ if $a_j = \alpha_j$, and $\eta_j(a_j \mid \alpha_j) = 0$ if $a_j \neq \alpha_j$; for $\alpha_j = \bar{a}_j$, let $\eta_j(a_j \mid \alpha_j) = \sigma_j(a_j)$. Again by the definition of feasible deviation $A^+(\mu') \subset A_j \times A_{-j}^{n*}$; then

$$U_j(\mu') = \sum_{a \in A} \mu'(a) u_j(a) = \sum_{a \in A_j \times A_{-j}^{n*}} \left(\sum_{\alpha_j \in A_j} \tilde{\mu}(\alpha_j, a_{-j}) \eta_j(a_j \mid \alpha_j) \right) u_j(a).$$

Substituting η_j as defined above we have

$$U_j(\mu') = \sum_{a \in (A_j \setminus \{\bar{a}_j\}) \times A_{-j}^{n*}} \tilde{\mu}(a) u_j(a)$$

$$+ \sum_{a \in \{\bar{a}_j\} \times A_{-j}^{n*}} \tilde{\mu}(\bar{a}_j, a_{-j}) \left(\sum_{\alpha_j \in A_j} \sigma_j(\alpha_j) u_j(\alpha_j, a_{-j}) \right).$$

Since $\tilde{\mu}(\bar{a}_j, a_{-j}) > 0$ for some $a_{-j} \in A_{-j}^{n*}$, Eq. $(*)$ implies

$$U_j(\mu') > \sum_{a \in (A_j \setminus \{\bar{a}_j\}) \times A_{-j}^{n*}} \tilde{\mu}(a) u_j(a) + \sum_{a \in \{\bar{a}_j\} \times A_{-j}^{n*}} \tilde{\mu}(\bar{a}_j, a_{-j}) u_j(\bar{a}_j, a_{-j});$$

i.e.,

$$U_j(\mu') > \sum_{a \in A} \tilde{\mu}(a) u_j(a) = U_j(\tilde{\mu}).$$

Hence μ' is an improving and self-enforcing deviation from $\tilde{\mu}$ by player j (recall that every feasible deviation by a single player is self-enforcing). Thus, $\tilde{\mu}$ is not a self-enforcing deviation by S from μ; i.e., $\tilde{\mu} \notin \text{SED}(\mu, S)$. This contradiction establishes that $A^+(\tilde{\mu}) \subset A^\infty$. ∎

The following corollary establishes that if a correlated strategy μ whose support is contained in A^∞ weakly Pareto-dominates every other correlated strategy $\tilde{\mu}$ whose support is contained in A^∞ (i.e., $U_i(\mu) \geq U_i(\tilde{\mu})$, for each $i \in N$), then μ is a CPCE.

COROLLARY. *Let* $\mu \in \Delta A$ *be such that* $A^+(\mu) \subset A^\infty$ *and such that it weakly Pareto-dominates every other* $\tilde{\mu} \in \Delta A$ *for which* $A^+(\tilde{\mu}) \subset A^\infty$. *Then* μ *is a CPCE.*

Proof. Let μ be as in the corollary. We show that μ is a CPCE. It is easy to show that no single player has a feasible and improving deviation from μ. If a player j has an improving deviation $\tilde{\mu}$ from μ, then he also has an improving deviation $\tilde{\mu}'$ from μ such that $A^+(\tilde{\mu}') \subset A^\infty$. Since μ weakly Pareto-dominates every correlated strategy whose support is contained in A^∞, such a deviation $\tilde{\mu}'$ cannot exist. Moreover, neither does a coalition of more than one player have a self-enforcing and improving deviation, since the support of every self-enforcing deviation from μ by such a coalition is contained in A^∞ by the proposition above. Hence μ is a CPCE.   ∎

In Appendix B we show that the set of CPNE of a game can be characterized as the set of mixed strategies from which no coalition has a self-enforcing deviation which makes all its members better off. The proposition above is easily modified to show that if σ is a mixed strategy profile whose support is contained in A^∞, then any self-enforcing mixed deviation from σ by a coalition of more than one player also has its support contained in A^∞. Thus, the corollary above establishes conditions under which a CPNE exists; whenever there is a mixed strategy profile whose support is contained in A^∞ and which weakly Pareto-dominates every other mixed strategy whose support is contained in A^∞, then this strategy is a CPNE.

In fact, the existence of a correlated strategy μ whose support is contained in A^∞ and which weakly Pareto-dominates every other correlated strategy whose support is also contained in A^∞ implies the existence of an action profile $a \in A^\infty$ which weakly Pareto-dominates every action profile in A^∞ (i.e., such that for each $i \in N$ and each $a' \in A^\infty$, $u_i(a) \geq u_i(a')$). This action profile is therefore a (pure strategy) coalition-proof correlated (and Nash) equilibrium. Thus, the conditions that guarantee the existence of a CPCE also imply existence of a CPNE.

An obvious implication of our corollary is that the unique equilibrium of a *dominance solvable* game (i.e., a game for which the set A^∞ is a singleton) is the unique CPCE (and CPNE) of the game. Also if a correlated strategy μ whose support is contained in A^∞ *strongly* Pareto-dominates every other strategy whose support is contained in A^∞, then μ is the unique CPCE of the game (as any other correlated equilibrium will be upset by the deviation of the grand coalition in which players ignore their recommendations and play according to μ).

It is worth noticing that the equilibria characterized by our corollary are *strong* in the sense that any improving deviation by a coalition of players is

"upset" by a further deviation by a single player. Milgrom and Roberts (1994) refer to such equilibria as *strongly coalition-proof*. They provide conditions that guarantee the existence of these equilibria in games with strategic complementarities. Specifically, they show that if a game with strategic complementarities has a unique Nash equilibrium, then this equilibrium is the unique strongly coalition-proof equilibrium of the game (Theorem 1); if each player's payoff function is increasing (respectively, decreasing) in the other players' strategies, then the maximal (respectively, minimal) Nash equilibrium is the unique strongly coalition-proof equilibrium (Theorem 2). Milgrom and Roberts establish their results using Tarski's fixed point theorem, and they do not rely on dominance arguments.

For games with finite strategy spaces, Milgrom and Roberts's results are implied by our corollary: if a game with strategic complementaries has a unique Nash equilibrium, then it is dominance solvable[2]; hence, this equilibrium is the unique (strongly) coalition-proof Nash equilibrium. Moreover, if each player's payoff function is increasing (respectively, decreasing) in the other players' strategies, then the maximal (respectively, minimal) Nash equilibrium weakly Pareto-dominates every other strategy whose support is contained in A^∞, and therefore our Corollary implies that this equilibrium is a (strongly) coalition-proof Nash equilibrium. Of course, the equilibria identified by these conditions are also (strongly) coalition-proof correlated equilibria.

A Game Where a CPCE Does Not Exist

Unfortunately, as the following example shows, there are games with more than two players with no coalition-proof correlated equilibria. Consider the three-player game given in Table V, taken from Einy and Peleg, where player 1 chooses the row, player 2 chooses the column, and player 3 chooses the matrix.

TABLE V
A Game Where a CPCE Does Not Exist

	c_1		c_2	
	b_1	b_2	b_1	b_2
a_1	3, 2, 0	0, 0, 0	3, 2, 0	0, 3, 2
a_2	2, 0, 3	2, 0, 3	0, 0, 0	0, 3, 2

[2] See Milgrom and Shannon's (1994) Theorem 12.

We show that there does not exist a coalition-proof correlated equilibrium of this game. Let $\overline{\mu}$ be an arbitrary correlated equilibrium and suppose that player 1 has the lowest payoff of the three players. Then $U_3(\overline{\mu}) \leq \frac{13}{5}$. (This can be proven by maximizing player 3's utility over the set of correlated equilibria μ satisfying $U_1(\mu) \leq \max\{U_2(\mu), U_3(\mu)\}$.) Moreover, $U_1(\overline{\mu}) \leq \frac{13}{5}$ since player 1 has the lowest payoff. Now consider the following deviation from $\overline{\mu}$ by players 1 and 3: player 1 chooses the bottom row and player 3 chooses the left matrix. This deviation is improving as players 1 and 3 now receive payoffs of 2 and 3, respectively. To demonstrate that $\overline{\mu}$ is not a coalition-proof correlated equilibrium we need only show that this deviation is self-enforcing. Clearly player 3 does not deviate further as he now obtains 3, his highest possible payoff. It can be shown that player 1 obtains at most $\frac{5}{3}$ by deviating further and choosing the top row.[3] (The details of this calculation are in Appendix A.) Thus, $\overline{\mu}$ is not a coalition-proof correlated equilibrium as players 1 and 3 have a self-enforcing and improving deviation.

There was no loss of generality in assuming that player 1 has the lowest payoff. If player 2 has the lowest payoff, then there is a self-enforcing and improving deviation by players 2 and 1. If player 3 has the lowest payoff, then there is a self-enforcing and improving deviation by players 3 and 2. Since any correlated equilibrium has a self-enforcing deviation by two players which makes both players better off, this game has no coalition-proof correlated equilibrium. (This game does not have a CPNE either.)

3. GAMES OF INCOMPLETE INFORMATION

In this section we extend our notion of coalition proofness to games of incomplete information. A (finite) game of incomplete information (or Bayesian game) G is defined by

$$G = \left(N, (T_i)_{i \in N}, (A_i)_{i \in N}, (p_i)_{i \in N}, (u_i)_{i \in N}\right),$$

where N is the set of players, and for each $i \in N$, T_i is the set of possible types for players i, A_i is player i's action set, $p_i: T_i \to \Delta T_{-i}$ is player i's prior probability distribution over the set of type profiles for the other players in the game ($T_{-i} = \prod_{j \in N \setminus \{i\}} T_j$), and $u_i: T \times A \to \mathbb{R}$ is player i's

---

[3] Following the deviation by players 1 and 3, player 1 is choosing the bottom row with probability one. Hence, when considering a further deviation by player 1 there is no loss of generality in restricting attention to the deviation where he chooses the top row with probability one. If this deviation does not make him better off, then no deviation does.

utility (payoff) function ($A = \prod_{i \in N} A_i$, $T = \prod_{i \in N} T_i$). We assume that the sets N, A, and T are nonempty and finite. For every coalition of players $S \in 2^N$, $S \neq \varnothing$, we denote by T_S the set $\prod_{i \in S} T_i$.

A *correlated strategy* is a function $\mu: T \to \Delta A$. We let C denote the set of all correlated strategies. Given $\mu \in C$, if each player reports his type truthfully and obeys his recommendation, then player i's expected payoff when he is of type $t_i \in T_i$ is

$$U_i(\mu \mid t_i) = \sum_{t_{-i} \in T_{-i}} \sum_{a \in A} p_i(t_{-i} \mid t_i) \mu(a \mid t) u_i(a, t).$$

Notice that in order for the players to play according to a correlated strategy, information about the players' types must be *revealed* so that an action profile can be selected according to the probability distribution specified by the given correlated strategy. We therefore must allow deviations by a coalition in which the players reveal a type profile different from their true one, as well as deviations where the players take actions different from those recommended. In the conceptual framework of mediation, the members of a coalition can deviate from a correlated strategy $\overline{\mu}$ by *misreporting* their type profile to the mediator or by *disobeying* the mediator's recommendations.

Intuitively, a deviation can be conceived of as follows: a coalition S carries out a deviation by employing a new mediator who *represents* the coalition with the mediator implementing μ and with whom the members of S communicate. Each member of S reports his type to this mediator who then (1) selects according to some $f_S: T_S \to \Delta T_S$ a type profile for the coalition (which he reports to the mediator implementing μ) and, upon receiving from the mediator implementing μ the recommendations for the members of S, (2) selects according to some $\eta_S: T_S \times T_S \times A_S \to \Delta A_S$ an action profile (which he recommends to the coalition members). The action profile recommended by the new mediator depends upon the type profile reported to it, the type profile it reported to the mediator implementing μ, and the actions recommended by the mediator implementing μ. This deviation generates a new correlated strategy which can be calculated from f_S and η_S according to the formula given in Definition 3.1.

DEFINITION 3.1. Let $\overline{\mu} \in C$ and $S \in 2^N$, $S \neq \varnothing$. A correlated strategy μ is a *feasible deviation by coalition S from $\overline{\mu}$* if there are $f_S: T_S \to \Delta T_S$ and $\eta_S: T_S \times T_S \times A_S \to \Delta A_S$, such that for each $t \in T$ and each $a \in A$,

$$\mu(a \mid t) = \sum_{\tau_S \in T_S} \sum_{\alpha_S \in A_S} f_S(\tau_S \mid t_S) \overline{\mu}(\alpha_S, a_{-S} \mid \tau_S, t_{-S}) \eta_S(a_S \mid \tau_S, t_S, \alpha_S).$$

The set of feasible deviations by coalition S from a correlated strategy is the set of correlated strategies that the coalition can induce by means of

some f_S and η_S. Given $\mu \in C$ and $S \in 2^N$, $S \neq \varnothing$, denote by $D(\mu, S)$ the set of all feasible deviations by coalition S from correlated strategy μ. As in Section 1, for every $\mu \in C$ and $S \in 2^N$, we have $\mu \in D(\mu, S)$ and $D(\mu, N) = C$.

For expositional ease, we explicitly introduce a concept of Pareto-dominance: a correlated strategy $\tilde{\mu}$ *Pareto-dominates* another correlated strategy μ *for coalition S* if no member of S is worse off under $\tilde{\mu}$ than under μ for any type profile and if for at least one type profile every member of S is better off under $\tilde{\mu}$ than under μ.

DEFINITION 3.2.   Let $S \in 2^N$, $S \neq \varnothing$, and let $\mu, \tilde{\mu} \in C$. We say that $\tilde{\mu}$ *Pareto-dominates μ for coalition S* (or that $\tilde{\mu}$ *Pareto S-dominates* μ) if (3.1) and (3.2) below are satisfied.

$$\text{For each } i \in S \text{ and each } t_i \in T_i: U_i(\tilde{\mu} \,|\, t_i) \geq U_i(\mu \,|\, t_i). \qquad (3.1)$$

$$\text{For each } i \in S \text{ and some } \tilde{t}_i \in T_i: U_i(\tilde{\mu} \,|\, \tilde{t}_i) > U_i(\mu \,|\, \tilde{t}_i). \qquad (3.2)$$

In our framework, the notion of Pareto dominance used determines whether a deviation is an improvement for a coalition. Consequently, alternative notions of Pareto dominance will lead to different notions of coalition-proof communication equilibrium. There are two alternative notions worth considering.

We say that $\tilde{\mu}$ *weakly Pareto S-dominates* μ if no member of S is worse off under $\tilde{\mu}$ than under μ for any of his types (i.e., if (3.1) is satisfied), and if at least one member of S is better off under $\tilde{\mu}$ than under μ for one of his types (i.e., if (3.2) is satisfied for some $i \in S$ rather than for all $i \in S$). The notion of weak Pareto dominance does not seem appropriate; an agreement will be ruled out if a coalition has a self-enforcing deviation which makes only a proper subset of its members better off, even though there are not clear incentives for such a coalition to form.

We say that $\tilde{\mu}$ *strongly Pareto S-dominates* μ if each member of S is better off under $\tilde{\mu}$ than under μ for each of his types (i.e., if the inequalities (3.1) are satisfied with strict inequality). Strong Pareto dominance is sometimes too strong. For example, if the utility function of some player is constant for one of his types, then there is no deviation which is improving for this player. Using strong Pareto dominance rules out the possibility of this player participating in any deviation.

It is easy to see that a correlated strategy μ is a *communication equilibrium* if no single player $i \in N$ has a feasible deviation which Pareto i-dominates μ.[4] In the spirit of the notion of strong Nash equilibrium, a

---

[4] See Forges (1986) or Myerson (1986) for a definition of communication equilibrium.

strong communication equilibrium can be defined as follows: A correlated strategy μ is a *strong communication equilibrium* if no coalition S has a feasible deviation which Pareto S-dominates μ. We only want to require, however, that an agreement not be Pareto-dominated by self-enforceability deviations. The notion of self-enforceability we define is identical to that introduced in Section 1.

DEFINITION 3.3.   Let $\overline{\mu} \in C$ and $S \in 2^N$, $S \neq \varnothing$. The *set of self-enforcing deviations by coalition S from $\overline{\mu}$*, $\mathrm{SED}(\overline{\mu}, S)$, is defined, recursively, as follows:

(i) If $|S| = 1$, then $\mathrm{SED}(\overline{\mu}, S) = D(\overline{\mu}, S)$;

(ii) If $|S| > 1$, then $\mathrm{SED}(\overline{\mu}, S) = \{\mu \in D(\overline{\mu}, S) \mid \nexists [R \in 2^S \setminus S, R \neq \varnothing, \tilde{\mu} \in \mathrm{SED}(\mu, R)]$ such that $\tilde{\mu}$ Pareto R-dominates $\mu\}$.

With this notion of self-enforceability, a coalition-proof communication equilibrium is defined to be any correlated strategy μ from which no coalition S has a self-enforcing deviation which Pareto S-dominates μ.

DEFINITION 3.4.   A correlated strategy μ is a *coalition-proof communication equilibrium* (CPCE) if no coalition $S \in 2^N$, $S \neq \varnothing$, has a deviation $\tilde{\mu} \in \mathrm{SED}(\mu, S)$ such that $\tilde{\mu}$ Pareto S-dominates μ.

When the set of type profiles T is a singleton, the concepts of strong and coalition-proof *communication* equilibrium reduce to, respectively, strong and coalition-proof *correlated* equilibrium. Note that a strong communication equilibrium is a coalition-proof communication equilibrium, which in turn is a communication equilibrium.

In two-player Bayesian games, the set of coalition-proof communication equilibria consists of the communication equilibria that are not Pareto N-dominated by any other communication equilibrium (i.e., the set of *interim efficient* communication equilibria).[5] Hence, for two-player Bayesian games a CPCE always exists. As established by the example in Section 1, games with more than two players need not have a CPCE.

4. DISCUSSION

In this section we discuss the relation of CPCE to Einy and Peleg's notion of coalition-proof communication equilibrium (which we denote by $\mathrm{CPCE}_{\mathrm{EP}}$), and we present some concluding remarks.

In CPCE deviations are evaluated *prior* to the players receiving recommendations; a deviation is improving if it makes each member of the

---

[5] See Holmstrom and Myerson (1983).

TABLE VI
The Game of Chicken

	L	R
T	6, 6	2, 7
B	7, 2	0, 0

TABLE VII
Chicken: A Correlated Equilibrium

	L	R
T	1/3	1/3
B	1/3	0

deviating coalition better off, conditional on his type, for at least one of his types and no worse off for any of his types. In contrast, in $CPCE_{EP}$ deviations are considered *after* players receive recommendations; a deviation is improving if it makes each member of the deviating coalition better off, conditional on *both* his type and his recommendation, for each combination of types and recommendations that occur with positive probability. Consequently, for two person games, while a CPCE must be interim efficient a $CPCE_{EP}$ need not be.

This is illustrated by the game *Chicken* given in Table VI. Table VII describes a correlated equilibrium of *Chicken* which yields an expected payoff of 5 for each player. This correlated strategy is not a CPCE as the grand coalition has the self-enforcing deviation given in Table VIII, which yields an expected payoff of 5.25 for each player. (This deviation is self-enforcing since it is a correlated equilibrium and therefore is immune to further deviations by a single player.)

TABLE VIII
Chicken: A Deviation by the Grand Coalition

	L	R
T	1/2	1/4
B	1/4	0

Nonetheless, the correlated strategy given in Table VII is a CPCE$_{EP}$. In this game each player has only a single type; therefore, for a deviation to be improving in Einy and Peleg's sense, it must make each player better off, conditional on his recommendation, for each of his possible recommendations. Consider player 1 given the recommendation B. His expected payoff conditional on his recommendation is 7. Since 7 is player 1's highest possible payoff, no coalition involving player 1 can improve upon this strategy.[6]

One interpretation of E & P's framework is that players have the opportunity to communicate only after each player has received his recommendation. Thus, when determining whether or not an agreement is a CPCE$_{EP}$, the agreement is elevated to the position of a *status quo* agreement. It is required to be resistant to deviations following recommendations, but it is not confronted with alternative agreements which are improving at the stage prior to each player receiving his recommendation. If players have the opportunity to discuss their play prior to receiving recommendations, however, they will exhaust the opportunities for improvements at this stage. For the game *Chicken*, if the players must decide whether to play the strategy in Table VII or that of Table VIII, they should choose the latter as this strategy gives a higher expected payoff to each player, and it is also resistant to further deviations.

The second fundamental way in which the notions of coalition proofness differ is that Einy and Peleg do not admit the possibility that members of a coalition jointly ''misreport'' their types. A CPCE$_{EP}$ must be a communication equilibrium, and so a CPCE$_{EP}$ is immune to deviations where a *single* player misreports his type and disobeys his recommendation. However, in Einy and Peleg's framework, at the stage where deviations are considered, the players are assumed to have already truthfully reported their types. Thus, deviations may not involve the members of a coalition jointly misreporting their types, or involve one member of a coalition misreporting his type and another member of the coalition disobeying his recommendation. An example of a CPCE$_{EP}$ which fails to be immune to this latter kind of deviation is illustrated in the game of incomplete information below. The game is the same as the *Three-Player Matching Pennies* game (see Table I), except that player 1's moves have now become his types. This game is given in Table IX below. Player 1 now has two possible types $\{H_1, T_1\}$ and no actions, while players 2 and 3 both have a singleton

---

[6] It can be shown that there is no improving deviation upon $\bar{\mu}$ in E & P's sense even with the weaker requirement that a deviation makes each member of the deviating coalition better off for at least one recommendation and at least as well off for all recommendations.

TABLE IX
An Incomplete Information Version of the TPMPG

	$t_1 = H_1$		$t_1 = T_1$	
	H_3	T_3	T_3	H_3
H_2	$1, 1, -2$	$-1, -1, 2$	$-1, -1, 2$	$-1, -1, 2$
T_2	$-1, -1, 2$	$-1, -1, 2$	$-1, -1, 2$	$1, 1, -2$

type set and their action sets remain, respectively, $\{H_2, T_2\}$ and $\{H_3, T_3\}$. Assume that the priors of players 2 and 3 over player 1's types are, respectively, $p_2(H_1) = p_3(H_1) = \frac{1}{2}$.

The correlated strategy μ given by $\mu(H_2, T_3 \mid H_1) = 1$ and $\mu(T_2, H_3 \mid T_1) = 1$, is a communication equilibrium of the game which yields expected payoffs of $U_1(\mu \mid H_1) = U_1(\mu \mid T_1) = -1$, $U_2(\mu) = -1$, and $U_3(\mu) = 2$. It is also a CPCE_{EP}; in E & P's framework, a deviation by a coalition is a mapping from the set of type and action (recommendation) profiles for the coalition to probability distributions over the coalition's set of action profiles. The coalition $\{1, 2\}$ has no improving deviation since, if player 1 is of type H_1, then player 3 moves T_3 with probability one and players 1 and 2 have a payoff of -1 regardless of the action taken by player 2. By the same argument, the coalition cannot improve if player 1 is of type T_1. No coalition involving player 3 has an improving deviation as the interests of players 1 and 2 are completely opposed to the interests of player 3. That no single player has an improving deviation follows from the fact that μ is a communication equilibrium.

In contrast, μ is not a CPCE of the game. Consider the deviation by the coalition $\{1, 2\}$, where player 1 reports T_1 when his type is H_1 and he reports H_1 when his type is T_1, and where player 2 moves H_2 when recommended T_2 and moves T_2 when recommended H_2. This deviation results in the correlated strategy $\tilde{\mu}$ given by $\tilde{\mu}(H_2, H_3 \mid H_1) = \tilde{\mu}(T_2, T_3 \mid T_1) = 1$, which yields expected payoffs of $U_1(\mu \mid H_1) = U_1(\mu \mid T_1) = 1$ and $U_2(\mu) = 1$. The deviation makes both players better off and is also self-enforcing (as both players attain their maximum possible payoff). Hence μ is not a CPCE.

Note that even if players can communicate only following the receipt of recommendations, CPCE_{EP} assumes a certain myopia on the part of player 1. Consider again the CPCE_{EP} of the *Three-Player Matching Pennies* game, where $\mu(H_2, T_3 \mid H_1) = 1$ and $\mu(T_2, H_3 \mid T_1) = 1$. If player 1 is of type H_1 and if he anticipates the opportunity to communicate following

player 2's receipt of his recommendation, then player 1 should report type T_1 and, at the communication stage, suggest to player 2 that he should move H_2. Player 2 should follow player 1's suggestion given that his interests are coincident with player 1's.

This game has a unique CPCE (which is also a CPCE$_{EP}$), where player 3 moves H_3 with probability $\frac{1}{2}$ regardless of player 1's type, and player 2 moves H_2 when player 1's type is H_1 and moves T_2 when player 1's type is T_1. This is essentially the same agreement predicted for the complete information version of the game. In fact, given that the interests of players 1 and 2 are coincident and opposed to those of player 3, this seems the only reasonable outcome.

For the game of *Chicken* and the incomplete information version of the *Three-Player Matching Pennies* game we have found correlated strategies which are CPCE$_{EP}$ and which are not CPCE. For the *Coordination/Defection* game in Appendix A we find a correlated strategy which is a CPCE and which is not a CPCE$_{EP}$. Thus, there is no inclusion relation between these two notions.

We conclude by emphasizing our findings. First, we show that when players can communicate they will reach correlated agreements. For example, in the *Three-Player Matching Pennies* game the only intuitive agreement is a correlated (and not mixed) agreement. Second, we offer a natural definition of coalition-proof equilibrium when correlated agreements are possible, and we show that no inclusion relationship between this new notion and CPNE is to be found. (Consequently, the notion of coalition proofness is sensitive to the possibility of correlated agreements.) And third, we obtain conditions under which a coalition-proof equilibrium exists.

APPENDIX A

In this appendix we present three examples. The first example is a game that has no coalition-proof correlated equilibrium. The second example is the *Three-Player Matching Pennies* game; we show that the correlated strategy described in the Introduction is the unique coalition-proof correlated equilibrium (and the unique strong correlated equilibrium) of the game. The third example is a game with a CPCE which is not a CPCE$_{EP}$.

A Game with No Coalition Proof Correlated Equilibrium

We show that the game described in Table V has no coalition-proof correlated equilibrium. A correlated strategy for this game is a vector

$\mu = (\mu_{ijk})_{i,j,k \in \{1,2\}}$, where $\mu_{ijk} \geq 0$ denotes the probability that players 1, 2, and 3 are recommended, respectively, actions a_i, b_j, and c_k.

If μ is a correlated equilibrium, then it satisfies the system of inequalities (I) given by

$$(I.a_1) \qquad \mu_{111} - 2\mu_{121} + 3\mu_{112} \geq 0$$

$$(I.a_2) \qquad -\mu_{211} + 2\mu_{221} - 3\mu_{212} \geq 0$$

$$(I.b_1) \qquad 2\mu_{111} - \mu_{112} - 3\mu_{212} \geq 0$$

$$(I.b_2) \qquad -2\mu_{121} + \mu_{122} + 3\mu_{222} \geq 0$$

$$(I.c_1) \qquad -2\mu_{121} + 3\mu_{211} + \mu_{221} \geq 0$$

$$(I.c_2) \qquad 2\mu_{122} - 3\mu_{212} - \mu_{222} \geq 0$$

We show that for each correlated equilibrium there is a coalition of two players which has an improving and self-enforcing deviation. Therefore, since a coalition-proof correlated equilibrium must be a correlated equilibrium, the set of CPCE of this game is empty.

Let $\overline{\mu}$ be an arbitrary correlated equilibrium and suppose that player 1 has the lowest payoff of the three players. We show that the coalition of players 1 and 3 has a self-enforcing and improving deviation. If player 1 has the lowest payoff in a correlated equilibrium, then player 3's payoff is no larger than $\frac{13}{5}$, which is the value of the solution to the linear programming problem

$$\max_{\mu \in \Delta A} U_3(\mu) \qquad \text{subject to } (I), U_1(\mu) \leq U_2(\mu), U_1(\mu) \leq U_3(\mu).$$

We also have $U_1(\overline{\mu}) \leq \frac{5}{3}$ since player 1 has the lowest payoff.

Consider the deviation $\tilde{\mu}$ induced by players 1 and 3 playing (a_2, c_1) with probability one for each profile of recommendations. (Then $\tilde{\mu}_{211} = \mu_{111} + \mu_{211} + \mu_{112} + \mu_{212}$, $\tilde{\mu}_{221} = \mu_{121} + \mu_{221} + \mu_{122} + \mu_{222}$, and $\tilde{\mu}_{ijk} = 0$ otherwise.) Given this deviation, players 1 and 3 obtain payoffs of, respectively, 2 and 3, regardless of player 2's action. Hence, $U_1(\tilde{\mu}) = 2 > U_1(\overline{\mu})$ and $U_3(\tilde{\mu}) = 3 > U_3(\overline{\mu})$ and so $\tilde{\mu}$ is an improving deviation for $\{1, 3\}$.

We now show that $\tilde{\mu}$ is self-enforcing. Clearly player 3 does not have a further improving deviation as he obtains his highest possible payoff. Player 1 has an improving deviation if the expected payoff of deviating to a_1, which is $3\tilde{\mu}_{211}$, is greater than $U_1(\tilde{\mu}) = 2$ (his expected payoff when he follows a recommendation to play a_2). However, this payoff is not larger than $\frac{5}{3}$, which is the value of the solution to the linear programming

problem

$$\max_{\mu \in \Delta A} 3(\mu_{111} + \mu_{211} + \mu_{112} + \mu_{212})$$

$$\text{subject to } (I), U_1(\mu) \leq U_2(\mu), U_1(\mu) \leq U_3(\mu).$$

The value of the solution to this problem is the maximum payoff that player 1 can obtain by a further deviation to a_1 from the correlated strategy $\tilde{\mu}$ given that the original agreement μ was a correlated equilibrium in which player 1 had the lowest payoff. Hence, player 1 has no further improving deviation.

There was no loss of generality in assuming that player 1 has the lowest payoff. Given the symmetry of this game, we can construct the following self-enforcing and improving deviations in each case: If player 2 has the lowest payoff, then players 1 and 2 deviate to $\{a_1, b_1\}$. If player 3 has the lowest payoff, then players 2 and 3 deviate to $\{b_2, c_2\}$. Therefore, this game has no coalition-proof correlated equilibrium.

Three-Player Matching Pennies

In the Introduction we demonstrated that the correlated strategy μ^* given in Table X below is a strong correlated equilibrium of the *Three-Player Matching Pennies* game. We now establish that μ^* is the unique coalition-proof correlated equilibrium of this game. (A strong correlated equilibrium is also a coalition-proof correlated equilibrium; therefore μ^* is also the unique strong correlated equilibrium.) Let μ be any correlated strategy. We reduce notation by writing μ_{xyz} for the probability $\mu(x_1, y_2, z_3)$, where $(x_1, y_2, z_3) \in \{H_1, T_1\} \times \{H_2, T_2\} \times \{H_3, T_3\}$; e.g., we write μ_{TTH} for $\mu(T_1, T_2, H_3)$. If μ is a correlated equilibrium, then it must

TABLE X
The CPCE of the TPMPG

	H_3		T_3	
	H_2	T_2	T_2	H_2
H_1	$\frac{1}{4}$	0	$\frac{1}{4}$	0
T_1	0	$\frac{1}{4}$	0	$\frac{1}{4}$

satisfy the system of inequalities (1) given by

$$2\mu_{HHH} - 2\mu_{HTT} \geq 0 \qquad (1.H_1)$$

$$-2\mu_{THH} + 2\mu_{TTT} \geq 0 \qquad (1.T_1)$$

$$2\mu_{HHH} - 2\mu_{THT} \geq 0 \qquad (1.H_2)$$

$$-2\mu_{HTH} + 2\mu_{TTT} \geq 0 \qquad (1.T_2)$$

$$-4\mu_{HHH} + 4\mu_{TTH} \geq 0 \qquad (1.H_3)$$

$$4\mu_{HHT} - 4\mu_{TTT} \geq 0. \qquad (1.T_3)$$

Note that $(1.H_3)$ implies $\mu_{TTH} \geq \mu_{HHH}$ and $(1.T_3)$ implies $\mu_{HHT} \geq \mu_{TTT}$. Hence player 3's payoff,

$$U_3(\mu) = 2(-\mu_{HHH} + \mu_{HTH} + \mu_{THH} + \mu_{TTH}$$
$$+ \mu_{HHT} + \mu_{HTT} + \mu_{THT} - \mu_{TTT}),$$

satisfies $U_3(\mu) \geq 0$. Since for each $(x_1, y_2, z_3) \in \{H_1, T_1\} \times \{H_2, T_2\} \times \{H_3, T_3\}$, $u_1(x_1, y_2, z_3) + u_2(x_1, y_2, z_3) + u_3(x_1, y_2, z_3) = 0$, we have $U_1(\mu) = U_2(\mu) \leq 0$. We now establish the following result.

CLAIM. *If μ is a CPCE, then $U_1(\mu) = U_2(\mu) = 0$, and $\mu_{HHH} + \mu_{HTH} + \mu_{THH} + \mu_{TTH} = \frac{1}{2}$.*

Proof. Let μ be a coalition-proof correlated equilibrium. We have shown that in any correlated equilibrium $U_1(\mu) = U_2(\mu) \leq 0$. Suppose by way of contradiction that $U_1(\mu) = U_2(\mu) < 0$. Consider the deviation where players 1 and 2 play (H_1, H_2) with probability $\frac{1}{2}$ and (T_1, T_2) with probability $\frac{1}{2}$, regardless of their recommendations. This deviation induces the correlated strategy $\overline{\mu}$ given in Table XI, where $\lambda = \mu_{HHH} + \mu_{HTH} +$

TABLE XI
A Deviation from the CPCE of the TPMPG

	H_3		T_3	
	H_2	T_2	H_2	T_2
H_1	$\dfrac{\lambda}{2}$	0	$\dfrac{1-\lambda}{2}$	0
T_1	0	$\dfrac{\lambda}{2}$	0	$\dfrac{1-\lambda}{2}$

$\mu_{THH} + \mu_{TTH}$. This deviation is improving since $U_1(\overline{\mu}) = U_2(\overline{\mu}) = 0$. It is also self-enforcing since a further deviation by either player 1 or 2 makes the player strictly worse off. The existence of such a deviation contradicts that μ is a CPCE. Thus, we must have $U_1(\mu) = U_2(\mu) = 0$; i.e.,

$$\mu_{HHH} - \mu_{HTH} - \mu_{THH} - \mu_{TTH} - \mu_{HHT} - \mu_{HTT} - \mu_{THT} + \mu_{TTT} = 0. \tag{2}$$

We now show that $\lambda = \frac{1}{2}$. Suppose that $\lambda > \frac{1}{2}$. Then the deviation by players 1 and 2, where, regardless of their recommendation, they move (H_1, H_2) with probability one is improving (players 1 and 2 each have an expected payoff of $\lambda + (1 - \lambda) > 0$) and self-enforcing, contradicting that μ is a CPCE. The case $\lambda < \frac{1}{2}$ is symmetric. Therefore $\lambda = \frac{1}{2}$; i.e.,

$$\mu_{HHH} + \mu_{HTH} + \mu_{THH} + \mu_{TTH} = \tfrac{1}{2}. \quad \blacksquare \tag{3}$$

Finally, we show that if μ is a CPCE, then $\mu_{HHH} = \mu_{HHT} = \mu_{TTT} = \mu_{TTH} = \frac{1}{4}$. As μ is a correlated strategy, we have

$$\mu_{HHH} + \mu_{HTH} + \mu_{THH} + \mu_{TTH} + \mu_{HHT} + \mu_{HTT} + \mu_{THT} + \mu_{TTT} = 1. \tag{4}$$

Adding (2) and (4) we get

$$\mu_{HHH} + \mu_{TTT} = \tfrac{1}{2}. \tag{$*$}$$

Also $(1.H_3)$ and $(1.T_3)$ yield $\mu_{TTH} \geq \mu_{HHH}$ and $\mu_{HHH} \geq \mu_{TTT}$. Adding these two inequalities and noticing (4), we get

$$\mu_{TTH} + \mu_{HHT} = \tfrac{1}{2}. \tag{$**$}$$

Thus, (4) implies $\mu_{HTH} = \mu_{THH} = \mu_{HTT} = \mu_{THT} = 0$. Substituting in (3) we get

$$\mu_{HHH} + \mu_{TTH} = \tfrac{1}{2}. \tag{$***$}$$

Subtracting $(***)$ from $(*)$ we get $\mu_{TTT} - \mu_{TTH} = 0$; i.e., $\mu_{TTT} = \mu_{TTH}$. Substituting in $(**)$ and subtracting $(*)$, we have $\mu_{HHT} - \mu_{HHH} = 0$; i.e., $\mu_{HHT} = \mu_{HHH}$. Using $(1.H_3)$ and $(1.T_3)$ again implies

$$\mu_{HHH} = \mu_{HHT} \geq \mu_{TTT} = \mu_{TTH} \geq \mu_{HHH}.$$

Hence $\mu_{HHH} = \mu_{HHT} = \mu_{TTT} = \mu_{TTH} = \frac{1}{4}$. $\quad \blacksquare$

The Coordination/Defection Game

In the *Coordination/Defection* game there are three players and each player has four actions: L (left), C (center), R (right), and D (defect).

Players' payoffs are given for each $(a_1, a_2, a_3) \in \{L, C, R, D\}^3$ by

$$u(a_1, a_2, a_3) = \begin{cases} (2, 0, 0) & \text{if } (a_1, a_2, a_3) = (L, L, L) \\ (0, 2, 0) & \text{if } (a_1, a_2, a_3) = (C, C, C) \\ (0, 0, 2) & \text{if } (a_1, a_2, a_3) = (R, R, R) \\ (-2, 1, 1) & \text{if } (a_1, a_2, a_3) = (L, D, D) \\ (1, -2, 1) & \text{if } (a_1, a_2, a_3) = (D, C, D) \\ (1, 1, -2) & \text{if } (a_1, a_2, a_3) = (D, D, C) \\ (0, 0, 0) & \text{otherwise}. \end{cases}$$

Let μ be a correlated strategy for this game, and write $\mu_{a_1 a_2 a_3}$ for the probability of action profile $(a_1, a_2, a_3) \in \{L, C, R, D\}^3$. We will show that the correlated strategy $\overline{\mu}$ satisfying $\overline{\mu}_{CRL} = 1$ is a CPCE, but it is not a CPCE_{EP}.

First, we show that $\overline{\mu}$ is a CPCE by proving that any improving deviation from $\overline{\mu}$ is vulnerable to a further self-enforcing and improving deviation. It is easy to see that no single player or two-player coalition has an improving deviation from $\overline{\mu}$. Let $\tilde{\mu}$ be an improving deviation from $\overline{\mu}$ by the grand coalition; i.e., $U_i(\tilde{\mu}) > 0$, $i = 1, 2, 3$. Hence $U_1(\tilde{\mu}) + U_2(\tilde{\mu}) + U_3(\tilde{\mu}) > 0$, and therefore

$$\tilde{\mu}_{LLL} + \tilde{\mu}_{CCC} + \tilde{\mu}_{RRR} > 0.$$

We show that $\tilde{\mu}$ is not a self-enforcing deviation (i.e., that there is a coalition which has a further deviation which is improving and self-enforcing).

Without loss of generality, assume that $\tilde{\mu}_{CCC} > 0$. If $\tilde{\mu}$ is not a correlated equilibrium, then there is a single player with an improving and self-enforcing deviation. If $\tilde{\mu}$ is a correlated equilibrium, then consider the deviation by the coalition of players 1 and 3 in which each chooses D when each is recommended C, and otherwise each chooses the action he is recommended. This deviation induces the correlated strategy $\tilde{\mu}'$ given for each $a_2 \in \{L, C, R, D\}$ by $\tilde{\mu}'_{Ca_2C} = 0$, $\tilde{\mu}'_{Da_2D} = \tilde{\mu}_{Ca_2C} + \tilde{\mu}_{Da_2D}$, and $\tilde{\mu}'_{a_1 a_2 a_3} = \tilde{\mu}_{a_1 a_2 a_3}$ if $(a_1, a_3) \notin \{(D, D), (C, C)\}$. Since $\tilde{\mu}_{CCC} > 0$, this deviation is improving. Moreover, it can be seen that because $\tilde{\mu}$ is a correlated equilibrium, neither player 1 nor player 3 has a further unilateral improving deviation from $\tilde{\mu}'$, and therefore $\tilde{\mu}'$ is a self-enforcing deviation. Thus, $\overline{\mu}$ is a CPCE_{MW}.

In order to show that $\overline{\mu}$ is not a CPCE_{EP}, we note the grand coalition has an *internally consistent improvement upon* $\overline{\mu}$, given by the correlated strategy μ' satisfying $\mu'_{LLL} = \mu'_{CCC} = \mu'_{RRR} = \frac{1}{3}$—see Einy and Peleg

(1995). Informally, E & P's notion of CPCE considers the deviation μ' to be self-enforcing because each player has some recommendation such that a further deviation is not improving. For example, if after the deviation to μ' player 1 (respectively, player 2 or player 3) is recommended L (respectively, C or R), then his expected payoff, conditional on his recommendation, is 2, his highest possible payoff. Thus, no coalition of players has a further deviation which makes each member of the coalition better off for each of his possible recommendations.

APPENDIX B

In this appendix we prove Proposition B.1 which characterizes the set of coalition-proof Nash equilibria as the set of mixed strategies from which no coalition has a self-enforcing deviation which makes all its members better off.

For $S \in 2^N$, $S \neq \varnothing$, let Σ_S denote the set of probability distributions σ_S over $A_S = \Pi_{i \in S} A_i$, satisfying $\sigma_S(a_S) = \Pi_{i \in S} \sigma_i(a_i)$ for all $a_S \in A_S$, where $\sigma_i(a_i) = \Sigma_{a_{S \setminus \{i\}} \in A_{S \setminus \{i\}}} \sigma_S(a_{S \setminus \{i\}}, a_i)$ is the marginal distribution of σ_S over A_i. Write Σ for the set Σ_N and refer to its members as *mixed strategies*. If $|N| > 1$, then Σ is a proper subset of ΔA. Given $\sigma \in \Sigma$ and $S \in 2^N$, we denote by σ_S the marginal distribution of σ over A_S (i.e., $\forall a_S \in A_S$: $\sigma_S(a_S) = \Sigma_{\alpha_{-S} \in A_{-S}} \sigma(a_S, \alpha_{-S})$). Here a mixed strategy is a probability distribution over A. A mixed strategy $\sigma \in \Sigma$ has an equivalent and more conventional representation as a strategy profile, $(\sigma_1, \ldots, \sigma_n)$.

Given an agreement $\bar{\sigma} \in \Sigma$, define the set of feasible *mixed* deviations by coalition S from $\bar{\sigma}$ as those mixed strategies that are obtained when each player i, $i \in S$, randomizes independently according to some $\tilde{\sigma}_i$, while each player j, $j \in N \setminus S$, follows the agreement and randomizes according to $\bar{\sigma}_j$. In other words, σ is a feasible deviation from $\bar{\sigma}$ by coalition S if σ can be written as a mixed strategy profile $((\tilde{\sigma}_i)_{i \in S}, (\bar{\sigma}_j)_{j \in N \setminus S})$, where $(\tilde{\sigma}_i)_{i \in S}$ is some mixed strategy profile for members of S. This is established formally in Definition B.1.

DEFINITION B.1. Let $\bar{\sigma} \in \Sigma$ and $S \in 2^N$, $S \neq \varnothing$. We say that $\sigma \in \Sigma$ is a *feasible mixed deviation by coalition S from $\bar{\sigma}$* if there is a $\tilde{\sigma}_S \in \Sigma_S$, such that for all $a \in A$, we have $\sigma(a) = \tilde{\sigma}_S(a_S) \bar{\sigma}_{-S}(a_{-S})$.

Let $D_M(\bar{\sigma}, S)$ denote the set of feasible mixed deviations by coalition S from $\bar{\sigma}$. It is clear that a mixed strategy is a Nash equilibrium if no single player has a feasible mixed deviation which makes him better off. A mixed strategy is a strong Nash equilibrium if no coalition has a feasible and improving mixed deviation.

The definition of a self-enforcing mixed deviation is obtained by replac-
ing in Definition 1.3 the set of deviations with the set of mixed deviations.
Hence, a mixed strategy σ is a self-enforcing mixed deviation by coalition
S from $\bar{\sigma}$ if σ is a feasible mixed deviation and if no proper subcoalition
of S has a further self-enforcing and improving mixed deviation from σ.

DEFINITION B.2.   Let $\bar{\sigma} \in \Sigma$ and $S \in 2^N$, $S \neq \varnothing$. The *set of self-en-
forcing mixed deviations by coalition S from $\bar{\sigma}$*, $\mathrm{SED}_M(\bar{\sigma}, S)$, is defined,
recursively, as

(i) If $|S| = 1$, then $\mathrm{SED}_M(\bar{\sigma}, S) = D_M(\bar{\sigma}, S)$;

(ii) If $|S| > 1$, then $\mathrm{SED}_M(\bar{\sigma}, S) = \{\sigma \in D_M(\bar{\sigma}, S) \mid \nexists\, [R \in 2^S \setminus S,$
$R \neq \varnothing, \tilde{\sigma} \in \mathrm{SED}_M(\sigma, R)]$ such that $\forall i \in R: U_i(\tilde{\sigma}) > U_i(\sigma)\}$.

Using the notions of feasible and of self-enforcing deviation by a
coalition from a mixed strategy, we define the notion of coalition-proof
Nash equilibrium as follows.

DEFINITION B.3.   Let $\Gamma = (N, (A_i)_{i \in N}, (u_i)_{i \in N})$ be a game in strategic
form. A strategy profile $\sigma \in \Sigma$ is a CPNE′ if no coalition $S \in 2^N$, $S \neq \varnothing$,
has a mixed deviation $\tilde{\sigma} \in \mathrm{SED}_M(\sigma, S)$ such that for each $i \in S$, we have
$U_i(\tilde{\sigma}) > U_i(\sigma)$.

Definition B.4 below formalizes the concept of CPNE as defined by
Bernheim, Peleg, and Whinston (1987). For convenience, the notion of
CPNE is cast in terms of *mixed strategies* (members of Σ) instead of
strategy profiles (members of $\prod_{i=1}^n \Sigma_i$). We abuse notation sometimes by
writing a mixed strategy $\sigma \in \Sigma$ as (σ_S, σ_{-S}), where $\sigma_S \in \Sigma_S$.
Let $\Gamma = (N, (A_i)_{i \in N}, (u_i)_{i \in N})$ be a game in strategic form. Given
$\bar{\sigma} \in \Sigma$ and $S \in 2^N \setminus N$, $S \neq \varnothing$, we write $\Gamma / \bar{\sigma}_{-S}$ for the game
$(S, (A_i)_{i \in S}, (\bar{u}_i)_{i \in S})$, where for each $i \in S$ and $a_S \in A_S$ we have

$$\bar{u}_i(a_S) = \sum_{\alpha_{-S} \in A_{-S}} \bar{\sigma}_{-S}(\alpha_{-S}) u_i(a_S, \alpha_{-S}).$$

For $S = N$, define $\Gamma / \bar{\sigma}_{-S} = \Gamma$. The definition of CPNE given by BPW is
recursive.

DEFINITION B.4.   Let $\Gamma = (N, (A_i)_{i \in N}, (u_i)_{i \in N})$ be a game in strategic
form.

(i) If $|N| = 1$, then $\sigma_1 \in \Sigma_1$ is a CPNE if for every $\tilde{\sigma}_1 \in \Sigma_1$: $U_1(\sigma_1)$
$\geq U_1(\tilde{\sigma}_1)$.

(ii) Assume that CPNE has been defined for games with fewer than n
players, and let Γ be a game such that $|N| = n$.

(a) A mixed strategy $\sigma \in \Sigma$ is *self-enforcing* if for every $S \in 2^N \setminus N$,
$S \neq \varnothing$, σ_S is a CPNE of Γ / σ_{-S}.

(b) A mixed strategy $\sigma \in \Sigma$ is a CPNE of Γ if it is self-enforcing, and if there is no other self-enforcing mixed strategy $\tilde{\sigma}$ such that for every $i \in N$: $U_i(\tilde{\sigma}) > U_i(\sigma)$.

For every game in strategic form Γ let $\text{CPNE}'(\Gamma)$ and $\text{CPNE}(\Gamma)$ represent the sets of mixed strategies satisfying, respectively, Definitions B.3 and B.4. Also, we denote by $\text{SE}(\Gamma)$ the set of all self-enforcing mixed strategies of Γ. For each $\sigma \in \Sigma$, and each $S \in 2^N$, $S \neq \varnothing$, we write $\text{SED}_M^\Gamma(\sigma, S)$ for the set of self-enforcing mixed deviations from σ by coalition S in the game Γ and we denote by $U_i^\Gamma(\sigma)$ the expected utility of player i given mixed strategy σ in the game Γ. Proposition B.1 can now be stated as follows.

PROPOSITION B.1. *For every game Γ in strategic form we have* $\text{CPNE}(\Gamma)$ $= \text{CPNE}'(\Gamma)$.

Before proving the proposition, we establish two lemmas.

LEMMA B.1. *For each Γ, each $\overline{\sigma} \in \Sigma$, and each $S \in 2^N$, $S \neq \varnothing$, we have that $\sigma = (\sigma_S, \overline{\sigma}_{-S}) \in \text{SED}_M^\Gamma(\overline{\sigma}, S)$ if and only if $\sigma_S \in \text{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S)$.*

Proof. We prove the lemma by induction on the number of players in S. Let Γ be a strategic form game, $\overline{\sigma} \in \Sigma$ and $S \in 2^N$.

(i) If $S = \{i\}$, then $\text{SED}_M^\Gamma(\overline{\sigma}, S) = D_M^\Gamma(\overline{\sigma}, S) = \Sigma_S \times \{\overline{\sigma}_{-S}\}$ and $\text{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S) = D_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S) = \Sigma_S$. Therefore, $\sigma = (\sigma_S, \overline{\sigma}_{-S}) \in \text{SED}_M^\Gamma(\overline{\sigma}, S)$ if and only if $\sigma_S \in \text{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S)$.

(ii) Assume Lemma B.1 holds for $|S| < k$. We show that it holds for $|S| = k$.

Step 1. If $\sigma_S \in \text{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S)$ then $\sigma = (\sigma_S, \overline{\sigma}_{-S}) \in \text{SED}_M^\Gamma(\overline{\sigma}, S)$.

Let $\sigma = (\sigma_S, \overline{\sigma}_{-S}) \notin \text{SED}_M^\Gamma(\overline{\sigma}, S)$. Then there are $R \in 2^S \setminus S$, $R \neq \varnothing$, and $\tilde{\sigma} = (\tilde{\sigma}_R, \sigma_{S \setminus R}, \overline{\sigma}_{-S}) \in \text{SED}_M^\Gamma((\sigma_S, \overline{\sigma}_{-S}), R)$ such that for each $i \in R$ we have $U_i^\Gamma(\tilde{\sigma}) > U_i^\Gamma(\sigma)$. Since $|R| < k$, the induction hypothesis yields $\tilde{\sigma}_R \in \text{SED}_M^{\Gamma/(\overline{\sigma}_{-S}, \sigma_{S \setminus R})}(\sigma_R, R)$. Noticing that $\Gamma/(\overline{\sigma}_{-S}, \sigma_{S \setminus R}) \equiv (\Gamma/\overline{\sigma}_{-S})/\sigma_{S \setminus R}$, it also yields $(\tilde{\sigma}_R, \sigma_{S \setminus R}) \in \text{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\sigma_S, R)$. Moreover, for each $i \in R$ we have $U_i^{\Gamma/\overline{\sigma}_{-S}}(\tilde{\sigma}_R, \sigma_{S \setminus R}) = U_i^\Gamma(\tilde{\sigma}) > U_i^\Gamma(\sigma) = U_i^{\Gamma/\overline{\sigma}_{-S}}(\sigma_S)$. Hence, $\sigma_S \notin \text{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S)$.

Step 2. If $(\sigma_S, \overline{\sigma}_{-S}) \in \mathrm{SED}_M^{\Gamma}(\overline{\sigma}, S)$ then $\sigma_S \in \mathrm{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S)$.

Let $\sigma_S \notin \mathrm{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\overline{\sigma}_S, S)$. Then there are $R \in 2^S \setminus S$, $R \neq \varnothing$, and $\tilde{\sigma}_S = (\tilde{\sigma}_R, \sigma_{S \setminus R}) \in \mathrm{SED}_M^{\Gamma/\overline{\sigma}_{-S}}(\sigma_S, R)$ such that for each $i \in R$ we have $U_i^{\Gamma/\overline{\sigma}_{-S}}(\tilde{\sigma}_S) > U_i^{\Gamma/\overline{\sigma}_{-S}}(\sigma_S)$. The induction hypothesis implies that $\tilde{\sigma}_R \in \mathrm{SED}_M^{(\Gamma/\overline{\sigma}_{-S})/\sigma_{S \setminus R}}(\sigma_R, R)$. Since $(\Gamma/\overline{\sigma}_{-S})/\sigma_{S \setminus R} \equiv \Gamma/(\overline{\sigma}_{-S}, \sigma_{S \setminus R})$, it also implies that $\tilde{\sigma} = (\tilde{\sigma}_R, \sigma_{S \setminus R}, \overline{\sigma}_{-S}) \in \mathrm{SED}_M^{\Gamma}((\sigma_S, \overline{\sigma}_{-S}), R)$. Furthermore, for each $i \in R$ we have $U_i^{\Gamma}(\tilde{\sigma}) = U_i^{\Gamma/\overline{\sigma}_{-S}}(\tilde{\sigma}_S) > U_i^{\Gamma/\overline{\sigma}_{-S}}(\sigma_S) = U_i^{\Gamma}(\sigma_S, \overline{\sigma}_{-S})$. Therefore $(\sigma_S, \overline{\sigma}_{-S}) \notin \mathrm{SED}_M^{\Gamma}(\overline{\sigma}, S)$. ∎

LEMMA B.2. *Let* Γ *be a game in strategic form. For each* $\sigma \in \mathrm{CPNE}'(\Gamma)$ *and each* $S \in 2^N$, $S \neq \varnothing$, *we have* $\sigma_S \in \mathrm{CPNE}'(\Gamma/\sigma_{-S})$.

Proof. Let $\sigma \in \Sigma$ be such that $\sigma_S \notin \mathrm{CPNE}'(\Gamma/\sigma_{-S})$ for some $S \in 2^N$, $S \neq \varnothing$. We show that $\sigma \notin \mathrm{CPNE}'(\Gamma)$.

Since $\sigma_S \notin \mathrm{CPNE}'(\Gamma/\sigma_{-S})$, then there are $S' \in 2^S$, $S' \neq \varnothing$, and $\tilde{\sigma}_S = (\tilde{\sigma}_{S'}, \sigma_{S \setminus S'}) \in \mathrm{SED}_M^{\Gamma/\sigma_{-S}}(\sigma_S, S')$ such that for each $i \in S'$, we have $U_i^{\Gamma/\sigma_{-S}}(\tilde{\sigma}_S) > U_i^{\Gamma/\sigma_{-S}}(\sigma_S)$. By Lemma B.1, $(\tilde{\sigma}_{S'}, \sigma_{S \setminus S'}) \in \mathrm{SED}_M^{\Gamma/\sigma_{-S}}(\sigma_S, S')$ implies that $\tilde{\sigma} = (\tilde{\sigma}_{S'}, \sigma_{S \setminus S'}, \sigma_{-S}) \in \mathrm{SED}_M^{\Gamma}(\sigma, S')$. Moreover, for each $i \in S'$ we have $U_i^{\Gamma}(\tilde{\sigma}) = U_i^{\Gamma/\sigma_{-S}}(\tilde{\sigma}_S) > U_i^{\Gamma/\sigma_{-S}}(\sigma_S) = U_i^{\Gamma}(\sigma)$. Hence $\sigma \notin \mathrm{CPNE}'(\Gamma)$. ∎

Proof of Proposition B.1. We prove the proposition by induction on the number of players.

(i) If $|N| = 1$ then Proposition B.1 clearly holds as for each $\sigma_1 \in \Sigma_1$, we have $\mathrm{SED}_M^{\Gamma}(\sigma_1, \{1\}) = \Sigma_1$.

(ii) Assume that Proposition B.1 is satisfied for games with fewer than n players. We need to show that it holds for Γ with $|N| = n$.

Step 1. If $\sigma \in \mathrm{CPNE}(\Gamma)$ then $\sigma \in \mathrm{CPNE}'(\Gamma)$.

Let $\sigma \notin \mathrm{CPNE}'(\Gamma)$. Then there is a $S \in 2^N$, $S \neq \varnothing$, and a $\tilde{\sigma} = (\tilde{\sigma}_S, \sigma_{-S}) \in \mathrm{SED}_M^{\Gamma}(\sigma, S)$ such that for each $i \in S$, we have $U_i^{\Gamma}(\tilde{\sigma}) > U_i^{\Gamma}(\sigma)$.

Case a: $S \neq N$. Since $\tilde{\sigma} \in \mathrm{SED}_M^{\Gamma}(\sigma, S)$, by Lemma B.1 $\tilde{\sigma}_S \in \mathrm{SED}_M^{\Gamma/\sigma_{-S}}(\sigma_S, S)$. Moreover, $U_i^{\Gamma}(\tilde{\sigma}) = U_i^{\Gamma/\sigma_{-S}}(\tilde{\sigma}_S) U_i^{\Gamma/\sigma_{-S}}(\sigma_S) = U_i^{\Gamma}(\sigma)$ for each $i \in S$. Hence, $\sigma_S \notin \mathrm{CPNE}'(\Gamma/\sigma_{-S}) = \mathrm{CPNE}(\Gamma/\sigma_{-S})$, where the equality follows from the induction hypothesis and that the game Γ/σ_{-S} has less than n players. Therefore, $\sigma \notin \mathrm{CPNE}(\Gamma)$.

Case b: $S = N$. Assume without loss of generality that $\not\exists\ \hat{\sigma} \in \mathrm{SED}_M^{\Gamma}(\sigma, N)$ such that for each $i \in N$, $U_i^{\Gamma}(\hat{\sigma}) > U_i^{\Gamma}(\tilde{\sigma})$. Then $\tilde{\sigma} \in$

CPNE$'(\Gamma)$ and so by Lemma B.2, $\sigma_S \in \text{CPNE}'(\Gamma/\sigma_{-S}) = \text{CPNE}(\Gamma/\sigma_{-S})$ $\forall S \in 2^N \setminus N$, $S \neq \varnothing$, where the equality follows from the induction hypothesis. Thus, $\tilde{\sigma} \in \text{SE}(\Gamma)$ and for each $i \in N$, $U_i^{\Gamma}(\tilde{\sigma}) > U_i^{\Gamma}(\sigma)$. Therefore, $\sigma \notin \text{CPNE}(\Gamma)$.

Step 2. If $\sigma \in \text{CPNE}'(\Gamma)$ then $\sigma \in \text{CPNE}(\Gamma)$.

Let $\sigma \notin \text{CPNE}(\Gamma)$. If $\sigma \notin \text{SE}(\Gamma)$, then there is a $S \in 2^N \setminus N$, $S \neq \varnothing$, such that $\sigma_S \notin \text{CPNE}(\Gamma/\sigma_{-S}) = \text{CPNE}'(\Gamma/\sigma_{-S})$, where the equality follows from the induction hypothesis. But Lemma B.2 and $\sigma_S \notin$ CPNE$'(\Gamma/\sigma_{-S})$ imply that $\sigma \notin \text{CPNE}'(\Gamma)$.

If $\sigma \in \text{SE}(\Gamma)$, then there is a $\tilde{\sigma} \in \text{SE}(\Gamma)$ such that for each $i \in N$, we have $U_i^{\Gamma}(\tilde{\sigma}) > U_i^{\Gamma}(\sigma)$. We show that $\tilde{\sigma} \in \text{SED}_M^{\Gamma}(\sigma, N)$, thereby proving that $\sigma \notin \text{CPNE}'(\Gamma)$.

Suppose to the contrary that $\tilde{\sigma} \notin \text{SED}_M^{\Gamma}(\sigma, N)$. Since $D_M^{\Gamma}(\sigma, N) = \Sigma$ (any deviation by the grand coalition is feasible), then there must be a $S \in 2^N \setminus N$, $S \neq \varnothing$, and a $\hat{\sigma} = (\hat{\sigma}_S, \tilde{\sigma}_{-S}) \in \text{SED}_M^{\Gamma}(\tilde{\sigma}, S)$ such that for each $i \in S$: $U_i^{\Gamma}(\hat{\sigma}) > U_i^{\Gamma}(\tilde{\sigma})$. Since $(\hat{\sigma}_S, \tilde{\sigma}_{-S}) \in \text{SED}_M^{\Gamma}(\tilde{\sigma}, S)$, Lemma B.1 yields $\hat{\sigma}_S \in \text{SED}_M^{\Gamma/\tilde{\sigma}_{-S}}(\tilde{\sigma}_S, S)$. Moreover, for each $i \in S$ we have $U_i^{\Gamma/\tilde{\sigma}_{-S}}(\hat{\sigma}_S) = U_i^{\Gamma}(\hat{\sigma}) > U_i^{\Gamma}(\tilde{\sigma}) = U_i^{\Gamma/\tilde{\sigma}_{-S}}(\tilde{\sigma}_S)$. Hence $\tilde{\sigma}_S \notin \text{CPNE}'(\Gamma/\tilde{\sigma}_{-S}) = \text{CPNE}(\Gamma/\tilde{\sigma}_{-S})$, where the equality follows from the induction hypothesis. Therefore, $\tilde{\sigma} \notin \text{SE}(\Gamma)$. This contradiction establishes the proposition. ∎

REFERENCES

Aumann, R. (1959). "Acceptable Points in General Cooperative *n*-Person Games," in *Contributions to the Theory of Games IV*, Princeton: Princeton Univ. Press.

Aumann, R. (1974). "Subjectivity and Correlation in Randomized Strategies," *J. Math. Econ.* **1**, 67–96.

Bernheim, B., Peleg, B., and Whinston, M. (1987). "Coalition-Proof Nash Equilibria I: Concepts," *J. Econ. Theory* **42**, 1–12.

Einy, E., and Peleg, B. (1995). "Coalition-Proof Communication Equilibria," in *Social Choice, Welfare, and Ethics* (W. A. Burnett, H. Moulin, M. Sales, and N. J. Schofield, Eds.), Chap. 13. Cambridge: Cambridge Univ. Press.

Farrell, J. (1987). "Cheap Talk, Coordination, and Entry," *Rand J. Econ.* **18**, 34–39.

Forges, F. (1986). "An Approach to Communication Equilibrium, *Econometrica* **54**, 1375–1385.

Holmstrom, B., and Myerson, R. (1983). "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica* **53**, 1799–1819.

Luce, R., and Raiffa, D. (1957). *Games and Decisions*, New York: Wiley.

Milgrom, P., and Roberts, J. (1994). "Strongly Coalition-Proof Equilibria in Games with Strategic Complementarities," Stanford mimeo.

Milgrom, P., and Shannon, C. (1994). "Monotone Comparative Statics," *Econometrica* **62**, 157–180.

Moulin, H. (1981). *Games Theory for the Social Sciences*, Studies in Game Theory and Mathematical Economics, New York: New York Univ. Press.

Moreno, D., and Wooders, J. (1993). ''Coalition-Proof Equilibrium,'' Discussion Paper 93-7, University of Arizona.

Moreno, D., and Wooders, J. (1995). ''An Experimental Study of Communication and Cooperation in Noncooperative Games,'' Working Paper 95-30, Universidad Carlos III.

Myerson, R. (1986). ''Multistage Games with Communication,'' *Econometrica* **54**, 323−358.

Myerson, R. (1991). *Game Theory: Analysis of Conflict*, Cambridge, MA: Harvard Univ. Press.

Ray, I. (1996). ''Coalition-Proof Correlated Equilibrium: A Definition,'' *Games Econ. Behav.* **17**, 56−79.