All-Stage strong correlated equilbrium

Heller, Yuval

Tel-Aviv University

11 March 2008

All-Stage Strong Correlated Equilibrium

Yuval Heller

School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel. Phone: 972-3-640-5386. Fax: 972-3-640-9357.

Abstract

A strong correlated equilibrium is a strategy profile that is immune to joint deviations. Different notions of strong correlated equilibria were defined in the literature. One major difference among those definitions is the stage in which coalitions can plan a joint deviation: before (*ex-ante*) or after (*ex-post*) the deviating players receive their part of the correlated profile. In this paper we show that an *ex-ante* strong correlated equilibrium is immune to deviations at all stages (assuming that deviating coalitions are allowed to use new correlating devices). Thus the set of *ex-ante* strong correlated equilibria of Moreno & Wooders (1996) is included in all other sets of strong correlated equilibria.

Key words: coalition-proofness, strong correlated equilibrium, common knowledge, incomplete information, noncooperative games. JEL classification: C72, D82.

1 Introduction

The ability of players to communicate prior to the play, influences the set of self-enforcing outcomes of a non-cooperative game. The communication allows the players to correlate their play, and to implement a correlated strategy profile as a feasible non-biding agreement. For such an agreement to be self-enforcing, it has to be stable against "plausible" coalitional deviations. Two

Email address: helleryu@post.tau.ac.il (Yuval Heller).

[1] This work is in partial fulfillment of the requirements for the Ph.D. in mathematics at Tel-Aviv University. I would like to thank Eilon Solan for his careful supervision, for the continuous help he offered, and for many insightful discussions. My deep gratitude is also given to to Ehud Lehrer and Yaron Azrieli for many useful comments, discussions and ideas.

notions in the literature describe such self-enforcing agreements: a *strong correlated equilibrium* is a profile that is stable against *all* coalitional deviations, and a *coalition-proof correlated equilibrium* is a profile that is stable against *self-enforcing* coalitional deviations (a deviation is self-enforcing if no sub-coalition has further self-enforcing and improving deviation).

Each notion has a few alternative definitions. One major difference among them, is the stage in which coalitions can plan a deviation from a correlated agreement. Assume that the correlated agreement is implemented by a mediator who privately recommends each player what to play. The definitions in Milgrom & Roberts (1996), Moreno & Wooders (1996), and Ray (1996) are *ex-ante* definitions: In their framework, players may plan deviations before receiving the recommendations, and no further communication is possible after recommendations are issued. The definitions in Einy & Peleg (1995), Ray (1998) and Bloch & Dutta (2007) are *ex-post* [2] definitions: In their framework, players may plan deviations only after receiving the recommendations.

However, in some frameworks coalitions can plan deviations at all stages. One example for such framework is an extended game with *cheap-talk* [3] In such a framework, the players can "mimic" a mediator, and implement a large set of strong correlated equilibria as strong Nash equilibria in the extended game (Heller, 2008). A coalition can plan a deviation in the early phases of the cheap-talk when no player has received his recommendation yet (*ex-ante* stage), in the late phases when all players have received their recommendations (*ex-post* stage), or in an intermediate stage when some of the players know their recommendations.

A natural question is whether any of the existing notions is appropriate to such frameworks, or whether new definitions are needed. In this paper we prove that the existing *ex-ante* strong correlated equilibrium (à la Moreno & Wooders) is resistant to deviations at all stages. The result is based on three assumptions about the communication framework (which hold in the cheap-talk framework):

(1) A deviating coalition can use new correlating devices (play a joint *correlated* deviation).
(2) When a coalition decides to deviate, that decision is common knowledge

[2] Referred to as *"interim"* in some of the referred papers.
[3] Cheap-talk is pre-play, unmediated, non-biding, non-verifiable communication among the players. For a good nontechnical introduction to some of the main issues of cheap-talk, see the survey of Farrel & Rabin (1996).

among its members.

(3) The players share a common prior about the possible states of the world in an incomplete information model à la Aumann (1987).

An immediate corollary is that the set of *ex-ante* strong correlated equilibria is included in all other sets of strong correlated equilibria, as defined in the literature referred above. One could hope that similar results might be obtained for the coalition-proof notions. However, in Section 5 we demonstrate that the *ex-ante* coalition-proof notion is not appropriate to frameworks in which coalitions can plan deviations at all stages. In Section 6 we discuss different approaches for coalitional stability, present the different notions of strong, and coalition-proof equilibria, and discuss the implications of our result.

The paper is organized as follows: Section 2 presents our model and the main result. The main result is demonstrated in Section 3, and proven in Section 4. We deal with the coalition-proof notion in Section 5, and discuss the the implications of our result in Section 6.

2 Model and Definitions

2.1 Preliminary Definitions

A game in strategic form G is defined as: $G = \left(N, (A^i)_{i\in N}, (u^i)_{i\in N}\right)$, where N is the finite (and non-empty) set of players. For each $i \in N$, A_i is player i's finite (and non-empty) set of actions (or pure strategies), and u^i is player i's utility (payoff) function, a real-valued function on $A = \prod_{i\in N} A^i$. The multi-linear extension of u^i to $\Delta(A)$ is still denoted by u^i. A member of $\Delta(A)$ is called a (correlated) strategy profile. A coalition S is a non-empty member of 2^N. For simplicity of notation, the coalition $\{i\}$ is denoted i. Given a coalition S, let $A^S = \prod_{i\in S} A^i$, and let $-S = \{i \in N \mid i \notin S\}$ denote the complementary coalition. A member of $\Delta(A^S)$ is called a (correlated) S-*strategy profile*. Given $q \in \Delta(A)$ and $a^S \in A^S$, we define $q_{|S} \subseteq \Delta(A^S)$ to be $q_{|S}(a^S) = \sum_{a^{-S}\in A^{-S}} q(a^S, a^{-S})$, and for simplicity we omit the subscript: $q(a^S) = q_{|S}(a^S)$. Given a^S s.t. $q(a^S) > 0$, we define $q(a^{-S}|a^S) = \frac{q(a^S,a^{-S})}{q(a^S)}$.

2.2 An Intuitive Description of Our Framework

Assume that the players of a game G (which will be played tomorrow), have agreed to play a correlated strategy profile $q \in \Delta(A)$. The players implement q with the assistance of a mediator who chooses the action profile $a \in A$ with probability $q(a)$. Throughout the day, the mediator calls each player i and privately gives him his recommendation: $a^i \in A^i$. The players do not necessarily know the order in which the mediator calls the players, or which players have already been called. During the day, the players can communicate, share information about their recommendations, and plan coalitional deviations from the agreement. If all the members of a coalition agree to use a deviation (each, with his own posterior information, believes that the deviation is profitable), then it is implemented with the assistance of a new mediator who receives (at the end of the day) the recommendations of the deviating players, and gives each of them a new recommendation. In the next day, each player simultaneously chooses an action in G. The profile q is an *all-stage strong correlated equilibrium* if, for every calling order and every stage, there is no coalition with a profitable deviation, as will be formally defined in the next subsection.

2.3 All-stage Strong Correlated Equilibrium

We model the situation of incomplete information of the players, by a state space that describes all possible states of the world (Aumann, 1987).

Definition 1 Given a game G, let a *state space* be *a 4-tuple:* $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i\in N})$*:*

(1) The 3-tuple (Ω, B, μ) is a probability space.
(2) The $(\tilde{a}^i)_{i\in N}$ are random variables in (Ω, B, μ), where $\tilde{a}^i : \Omega \to A^i$.

We interpret $(\Omega, \mathcal{B})$ as the space of all possible states of the world, μ as the common prior (for the states of the world) for all the players, and $a^i(\omega)$ as the recommendation of player i in the state ω.

Remark 2 The state of the world $\omega \in \Omega$ includes a full description of all parameters that may be object to uncertainty on the part of any player in G. We assume that all the players share a common prior about the states of the world. The justification of this assumption is discussed in Aumann (1987).

Given a non-null event $E \in \mathcal{B}$ and a coalition $S \in N$, let $\tilde{a}^S(E) \in \Delta(A^S)$ be the posterior distribution of $\tilde{a}^S = (\tilde{a}^i)_{i \in S}$ conditioned on the event E. In order to model the situation where an agreement is implemented, we require that the prior distribution $\tilde{a}^S(\Omega)$ to be equal to the agreement distribution.

Definition 3 Given a game G *and* $q \in \Delta(A)$*, we say that* a state space $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i \in N})$ is a *consistent state space* if $\forall a \in A$, $\tilde{a}^N(\Omega) = q(a)$.

A (joint) deviation of a coalition S is a random variable (in Ω) that is independent of $\tilde{a}^{-S}$ (given $\tilde{a}^S$).

Definition 4 Given a game G*,* a coalition S and a state space $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i \in N})$, let $\tilde{d^S} = \left(\tilde{d^i}\right)_{i \in S}$ be a *deviation of* S*,* where the $\left(\tilde{d^i}\right)_{i \in S}$ are random variables in Ω: $\tilde{d^i} : \Omega \to A^i$*,* and the random variables $\tilde{d^S}$ and $\tilde{a}^{-S}$ are conditionally independent given $\tilde{a}^S$.

The interpretation is the following: If the players of S agree to use the deviation $\tilde{d^S}$, they implement it with the help of a new mediator. The mediator receives S-part of the recommendation profile, but he does not receive any information about the recommendations of the non-deviating players. Thus, the new recommendations he sends to the deviating players may depend only on $\tilde{a}^S$ (and not on $\tilde{a}^{-S}$).

When the members of a coalition S consider to implement a joint deviation, they are in a situation of incomplete information: each player may know his recommendation, and may have additional private information acquired when communicating with the other deviating players. We assume that the deviating players have no direct information about the recommendations of the non-deviating players. We model this by the following definition of a consistent information structure.

Definition 5 Given a game G, a coalition S, an agreement $q \in \Delta(A)$ and a consistent state space $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i \in N})$, let *an information structure* (of S) be a $|S|$-tuple of partitions of Ω $(\mathcal{F}^i)_{i \in S}$, whose join ($\bigwedge_{i \in S} \mathcal{F}^i$, the coarsest common refinement of $(\mathcal{F}^i)_{i \in S}$) consists of non-null events. We say that $(\mathcal{F}^i)_{i \in S}$ is a *consistent information structure*, if $\forall \omega \in \Omega$, $\forall i \in S$, $\forall a \in A$, $\tilde{a}^N\left(F^i(\omega)\right)(a) = \tilde{a}^S\left(F^i(\omega)\right)\left(a^S\right) \cdot q(a^{-S} \mid a^S)$.

We interpret $\mathcal{F}^i$ as the information partition of player i; that is, if the true state of the world is $\omega \in \Omega$ then player i is informed of that element $F^i(\omega)$ of

$\mathcal{F}^i$ that contains ω.

When each player considers whether the implementation of a deviation is profitable to him, he compares his conditional expected payoff when playing the original agreement and when implementing the deviation. A player agrees to deviate, only if the latter conditional expectation is larger. We now formally define the conditional expected payoffs of each player in each state of the world ω, when following the agreement and when implementing a deviation.

Definition 6 Given a game G, an agreement $q \in \Delta(A)$, a coalition S, a player $i \in S$, a consistent state space $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i \in N})$, a deviation (of S) $\tilde{d^S}$, and a consistent information structure $(\mathcal{F}^i)_{i \in S}$, let the *conditional expected payoffs of player* i in $\omega \in \Omega$ be:

- The conditional expected payoff when all the players follow the agreement:

$$u_f^i(\omega) = \int_{F^i(\omega)} u^i\left(\tilde{a}^N(\omega)\right) d\mu$$

- The conditional expected payoff when the members of S deviate (by implementing $\tilde{d^S}$), and the players in $-S$ follow the agreement:

$$u_d^i(\omega) = \int_{F^i(\omega)} u^i\left(\left(\tilde{d^S}, \tilde{a}^{-S}\right)(\omega)\right) d\mu$$

If the players in S decide to implement a deviation in some state $\omega \in \Omega$, then it is common knowledge (in ω) that each player expects to earn if the deviation is implemented (conditioned on his information). Formally (Aumann, 1976):

Definition 7 Given a coalition S, a state space $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i \in N})$, an information structure $(\mathcal{F}^i)_{i \in S}$, and $\omega \in \Omega$, an event $E \in \mathcal{B}$ is *common knowledge* at ω if E includes that member of the meet $\mathcal{F}^{meet} = \bigwedge\limits_{i \in S} \mathcal{F}^i$ that contains ω.

We define a profitable deviation (with respect to a consistent information structure of a coalition S), as a deviation that it is common knowledge (in some state of the world), that each deviating player expects to earn more if the deviation is implemented.

Definition 8 Given a game G, an agreement $q \in \Delta(A)$, a coalition $S \subseteq N$ and a consistent state space $(\Omega, \mathcal{B}, \mu, (\tilde{a}^i)_{i \in N})$, we say that a deviation (of S) $\tilde{d^S}$ is a *profitable deviation*, if there exists a consistent information structure $(\mathcal{F}^i)_{i \in S}$ and a state $\omega_0 \in \Omega$ such that it is common knowledge in ω_0 that $\forall i \in S$, $u_d^i(\omega) > u_f^i(\omega)$. In that case, we say that $\tilde{d^S}$ *is* a *profitable deviation with respect to the information structure* $(\mathcal{F}^i)_{i \in S}$.

We can now define an all-stage strong correlated equilibrium as a strategy profile, from which no coalition has a profitable deviation.

Definition 9 A strategy profile $q \in \Delta(A)$ is an *all-stage strong correlated equilibrium* if no coalition $S \subseteq N$ has a profitable deviation.

2.4 Main Result

A profile is an *ex-ante* strong correlated equilibrium, if no coalition has a profitable deviation at the *ex-ante* stage, when the players have no information about the recommendations.

Definition 10 A strategy profile $q \in \Delta(A)$ is an *ex-ante strong correlated equilibrium* if no coalition $S \subseteq N$ has a profitable deviation w.r.t. to the *ex-ante* information structure $(\mathcal{F}^i)_{i \in S}$ that satisfies $\forall i, \mathcal{F}^i = \Omega$.

One can verify that our definition is equivalent to the definition of Moreno & Wooders (1996). It is obvious that an all-stage strong correlated equilibrium is also an *ex-ante* strong correlated equilibrium. Our main result shows that the converse is also true, and thus the two notions coincide.

Theorem 11 *A correlated profile $q \in \triangle(A)$ is an ex-ante strong correlated equilibrium if and only if it is an all-stage strong correlated equilibrium.*

3 An Example of the Main Result

In the following example we present an *ex-ante* strong correlated equilibrium in a 3-player game, and a specific deviation that is considered by the grand coalition at some intermediate stage. At first look, one may think that this deviation is profitable to all the players (conditioned on their posterior information at that stage), but a more thorough analysis reveals that this is not true. The analysis in this example gives the intuition for our use of a model of incomplete information *à* la Aumann, and for our main result.

Table 1 presents the matrix representation of a 3-player game, where player 1 chooses the row, player 2 chooses the column, and player 3 chooses the matrix.

Let q be the profile: $\left(\frac{1}{4}(a_1, b_1, c_1), \frac{1}{4}(a_2, b_1, c_1), \frac{1}{4}(a_1, b_2, c_1), \frac{1}{4}(a_1, b_1, c_2)\right)$, with an expected payoff of 10 to each player. Observe that q is an *ex-ante*

Table 1
A 3-Player Game With An Ex-Ante Strong Correlated Equilibrium

	c_1			c_2			c_3		
	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
a_1	10,10,10	5, 20,5	0,0,0	5,5,20	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0
a_2	20,5,5	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0
a_3	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	7,11,12

strong correlated equilibrium:

- No single player has a unilateral deviation (q is a correlated equilibrium).
- No coalition of two players has a profitable deviation. (Their uncertainty about the recommendation of the third player prevents them from being able to earn together more than 20 by a joint deviation.)
- The grand coalition cannot earn more than a total payoff of 30.

Now, consider an intermediate stage in which player 1 has received a recommendation a_1, player 2 has received a recommendation a_2, and player 3 has not received his recommendation yet. Each player does not know whether the other players have received their recommendations. At first look, the implementation of the deviation $\tilde{d}^S(\cdot) = (a_3, b_3, c_3)$, which gives a payoff of $(7, 11, 12)$, may look profitable for all the players:

- Conditioned on his recommendation (a_1), player 1 has an expected payoff of $6\frac{2}{3}$, and thus $\tilde{d}^S$ is profitable to him. The same is true for player 2.
- Player 3 does not know his recommendation. His *ex-ante* expected payoff is 10, and he would earn a payoff of 12 by implementing $\tilde{d}^S$.

However, a more thorough analysis of player 3's information, reveals that d^S is unprofitable for him. Player 1 can only earn from $\tilde{d}^S$ if he has received a recommendation a_1. Thus, if player 1 agrees to implement $\tilde{d}^S$, then it is common knowledge that he has received a_1. The expected payoff of players 2 and 3, conditioned on that player 2 has received a_1, is $11\frac{2}{3}$. Thus, if player 2 agrees to implement $\tilde{d}^S$, then he must have more information: that his recommendation is a_2. Therefore player 3 knows that if the other players agree to implement $\tilde{d}^S$, then their part of the recommendation profile is (a_1, a_2). Conditioned on that, his expected payoff is 15, and thus $\tilde{d}^S$ is unprofitable for him.

4 A Proof of the Main Result

In this Section we prove our main result. As discussed earlier, one direction is straightforward, and we have to prove only the other direction:

Theorem 12 *Every ex-ante strong correlated equilibrium is an all-stage strong correlated equilibrium.*

In other words: If a profitable deviation from an agreement $q \in \triangle(A)$ exists, then there also exists a profitable *ex-ante* deviation from q.

PROOF. Let $q \in \Delta(A)$ be a strategy profile (the agreement) that is not an all-stage strong correlated equilibrium in a game G with a consistent state space $(\Omega, \mathcal{B}, \mu, (a^i)_{i \in N})$. Thus there exists a coalition $S \subseteq N$ with a profitable deviation $\tilde{d}^S$ w.r.t. a consistent information structure $(\mathcal{F}^i)_{i \in S}$. This implies that there is a state $\omega_0 \in \Omega$, such that it is common knowledge in ω_0 that $\forall i,\ u^i_d(\omega) > u^i_f(\omega)$, i.e., $F^{meet}(\omega_0) \subseteq \left\{\omega \mid u^i_d(\omega) > u^i_f(\omega)\right\}$. For each deviating player $i \in S$, write $F^{meet} = F^{meet}(\omega_0) = \bigcup_j F^i_j$ where the F^i_j are disjoint members of $\mathcal{F}^i$, and let $\omega^i_j \in F^i_j$ be a state in F^i_j. We now construct an *ex-ante* profitable deviation $\tilde{d}^S_e$ (w.r.t to the *ex-ante* information structure $(\mathcal{F}^i_e)_{i \in S}$, which satisfies $\forall i,\ \mathcal{F}^i_e = \Omega$): $\tilde{d}^S_e(\omega) = \begin{cases} \tilde{d}^S(\omega) & \omega \in F^{meet} \\ \tilde{a}^S(\omega) & \omega \notin F^{meet} \end{cases}$

Observe that $\tilde{d}^S_e$ and $\tilde{a}^{-S}$ are conditionally independent given $\tilde{a}^S$, thus $\tilde{d}^S_e$ is a well defined deviation. Let $u^i_{d_e}(\omega)$, $u^i_{f_e}(\omega)$ be the conditional utilities of the players w.r.t. $(\mathcal{F}^i_e)_{i \in S}$. We finish the proof by showing that $\tilde{d}^S_e$ is profitable, i.e: $\forall i,\ u^i_{d_e}(\omega) > u^i_{f_e}(\omega)$. Let $i \in S$ be a deviating player.

$$
\begin{aligned}
u^i_{d_e}(\omega) - u^i_{f_e}(\omega) &= \int_{F^i_e(\omega)} \left(u^i\left(\left(\tilde{d}^S_e, \tilde{a}^{-S}\right)(\omega)\right) - u^i\left(\tilde{a}^N(\omega)\right)\right) d\mu && (1) \\
&= \int_{\Omega} \left(u^i\left(\left(\tilde{d}^S_e, \tilde{a}^{-S}\right)(\omega)\right) - u^i\left(\tilde{a}^N(\omega)\right)\right) d\mu && (2) \\
&= \int_{F^{meet}} \left(u^i\left(\left(\tilde{d}^S_e, \tilde{a}^{-S}\right)(\omega)\right) - u^i\left(\tilde{a}^N(\omega)\right)\right) d\mu && (3) \\
&= \int_{F^{meet}} \left(u^i\left(\left(\tilde{d}^S, \tilde{a}^{-S}\right)(\omega)\right) - u^i\left(\tilde{a}^N(\omega)\right)\right) d\mu && (4) \\
&= \sum_j \int_{F^i_j} \left(u^i\left(\left(\tilde{d}^S, \tilde{a}^{-S}\right)(\omega)\right) - u^i\left(\tilde{a}^N(\omega)\right)\right) d\mu && (5) \\
&= \sum_j u^i_d(\omega^i_j) - u^i_f(\omega^i_j) > 0 && (6)
\end{aligned}
$$

Equation 2 is due to the equality $F_e^i(\omega) = \Omega$, (3) holds since $\tilde{d}_e^S = \tilde{a}^S$ outside F^{meet}, (4) holds since $\tilde{d}_e^S = \tilde{d}^S$in F^{meet}, (5) follows from $F^{meet} = \dot{\bigcup_j} F_j^i$, and the last inequality is implied by $F^{meet} \subseteq \left\{\omega \mid u_d^i(\omega) > u_f^i(\omega)\right\}$. **QED**.

5 Coalition-Proof Correlated Equilibria

In the last Section we have shown that an *ex-ante* strong correlated equilibrium *à* la Moreno & Wooders is also appropriate to frameworks in which players can plan deviations at all stages. A natural question is whether a similar result holds for their notion of coalition-proof correlated equilibrium. [4] In this Section we show that the answer is negative, by presenting an example (adapted from Bloch & Dutta, 2007), in which there is an *ex-ante* coalition-proof correlated equilibrium that is not a self enforcing agreement in a framework in which communication is possible at all stages. Table 2 presents a two-player game and an *ex-ante* coalition-proof correlated equilibrium.

Table 2
A Two-Player Game and an *Ex-ante* Coalition-Proof Correlated Equilibrium

	b_1	b_2	b_3
a_1	6,6	-2,0	0,7
a_2	2,2	2,2	0,0
a_3	0,0	0,0	3,3

	b_1	b_2	b_3
a_1	1/2	0	0
a_2	1/4	1/4	0
a_3	0	0	0

We first show that the profile presented in table 2 is an *ex-ante* coalition-proof equilibrium. First, observe that the profile is a correlated equilibrium. Moreno & Wooders (1996) have proved that in a two-player game, every correlated profile that is not Pareto-dominated by another correlated equilibrium is a coalition-proof correlated equilibrium. The profile gives each player a payoff of 4. Thus we prove that it is an *ex-ante* coalition-proof correlated equilibrium, by showing that any correlated equilibrium q gives player 1 a payoff of at most 4. Let $x = q(a_1, b_1)$. Observe that $q(a_2, b_1) \geq x/2$ because otherwise player 2 would have a profitable deviation: playing b_3 when recommended b_1. This implies $q(a_2, b_2) \geq x/2$, because otherwise player 1 would have a profitable deviation (playing a_1 when recommended a_2). Thus the payoff of q conditioned

[4] Recall (Moreno & Wooders, 1996) that an *ex-ante* coalition-proof correlated equilibrium is a strategy profile from which no coalition has a self-enforcing and improving *ex-ante* deviation. An *ex-ante* deviation is self enforcing, if no proper sub-coalition has a further self-enforcing and improving *ex-ante* deviation.

on that the recommendation profile is in $A = ((a_1, b_1), (a_2, b_1), (a_2, b_2))$ is at most 4, and because the payoff of q conditioned on that the recommendation profile is not in A is at most 3, then the total payoff of q is at most 4.

We now explain why this profile is not a self-enforcing agreement in a framework in which the players can also plan deviations at the *ex-post* stage. Assume that the players have agreed to play the profile, and player 1 has received a recommendation a_2. In that case, he can communicate with player 1 at the *ex-post* stage, tell him that he received a_2 (and thus if the players follow the recommendation profile they would get a payoff of 2), and suggest a joint deviation: playing (a_3, b_3). As player 1 has no incentive to lie, player 2 would believe him, and they would both play (a_3, b_3). This *ex-post* deviation is self-enforcing because (a_3, b_3) is a Nash equilibrium.

Observe that the same deviation is not self-enforcing in the *ex-ante* stage. If the players agree at the *ex-ante* stage to implement a deviation that changes (a_2, b_1) into (a_3, b_3), then player 2 would have a profitable sub-deviation: playing b_3 when recommended b_1. Similarly, if they agree to implement a deviation that changes (a_2, b_2) into (a_3, b_3), then player 1 would have a profitable sub-deviation: playing a_1 when recommended a_2.

6 Discussion

6.1 Approaches for Coalitional Stability

In this paper we deal with self-enforcing agreements in (one-shot) games in environments where players can freely discuss their strategies before the play. Such agreements have to be stable against coalitional deviations. A few notions in the literature present different approaches for coalitional stability.

The first approach, is the *Pareto dominance refinement*, in which the set of Nash equilibria is refined by restricting attention to its efficient frontier. This approach is popular in applications due to its advantages: existence (in all games) and the simplicity of its use. However, when there are more than 2 players, it ignores the ability of coalitions other than the grand coalition to privately agree upon a joint deviation. [5]

[5] As discussed in Bernheim et al. (1987) and Yi (1999). The latter paper presents a set of conditions that if satisfied, the two notions of Pareto dominance refinement and coalition-proof equilibrium coincide.

Another approach is to explicitly model the procedure of communication as an extended-form game that specifies how messages are interchanged (see, e.g.: Farrel & Saloner, 1988; Rabin, 1994; Aumann & Hart, 2003). However, the results are sensitive to the exact procedure employed, and usually strong restrictions have to be be made to isolate the desired outcome.

A different approach is the *farsighted coalitional stability* (alternative variations are discussed in Greenberg, 1989, 1990; Chwe, 1994; Mariotti, 1997; Xue, 1998, 2000).[6] These notions focus on environments where deviations are *public*. At each stage coalitions propose deviations from the current *status-quo* outcome, until nobody wishes to deviate further. The set of possible final outcomes is defined using stable sets à la Von-Neumann & Morgenstern (1953). This approach is less appropriate when coalitions can *privately* plan deviations.[7]

6.2 Strong and Coalition-Proof Equilibria

In this paper we focus on the approach of coalition-proof and strong equilibria. *A Nash equilibrium is strong* (Aumann, 1959) if no coalition, taking the actions of its complement as given, can cooperatively deviate in a way that benefits all of its members. The main drawback of this notion, is that it exists in only a relatively small set of games.[8] Bernheim et al., 1987, has presented a wider refinement of Nash equilibrium, which exists in more games: *a coalition-proof Nash equilibrium.* A Nash equilibrium is coalition-proof if no coalition has a profitable self-enforcing deviation. A deviation is self-enforcing if no proper sub-coalition has a further self-enforcing and improving deviation.[9] The notion of coalition-proof equilibrium has been useful in a variety

[6] Also called *negotiation-proof equilibrium* and *full coalitional equilibrium.*

[7] Xue (2000, Table 1) presents an example for the difference between a negotiation-proof equilibrium and a coalition-proof Nash equilibrium. Observe that the negotiation-proof equilibrium in this example (the profile (U, L, A)) is not a plausible outcome if the coalition $(\{1, 2\})$ can privately deviate.

[8] Examples for games where strong Nash equilibria exist are congestion games (Holzman & Law-Yone, 1997); games where the preferences satisfying independence of irrelevant choices, anonymity, and partial rivalry (Konishi et al., 1997a); and games where the core of the cooperative game derived from the original normal form game, is non-empty (see Konishi et al., 1997b, and the references within). The latter paper also presents conditions for the equivalence of strong and coalition-proof Nash equilibria (independence of irrelevant choices and population monotonicity property).

[9] Observe that only members of the deviating coalition may contemplate deviations from the deviation. This rules out the possibility that members of the deviating

of applied contexts, such as: *menu* auctions (Bernheim & Whinston, 1986), oligopolies (Bernheim & Whinston, 1987; Thoron, 1998; Delgado & Moreno, 2004; Delgado, 2006), and common agency games (Konishi et al., 1999).

These notions focus on environments where coalitions can *privately* communicate before the play, and plan a joint deviation. However, they ignore the fact that the same private communication allow the players to correlate their moves. This deficiency is overcome by the notions of strong and coalition-proof correlated equilibria. *A correlated equilibrium is strong* if no coalition has a (possibly correlated) joint deviation that benefits all of its members. The close connection between strong correlated equilibrium and private pre-play communication is demonstrated by:

- The result in Heller (2008), which shows that any "punishable"[10] (*ex-ante*) strong correlated equilibrium is a strong Nash equilibrium in an extended game with *cheap-talk.*[11]
- The example in Moreno & Wooders (1998) of an *ex-ante* strong correlated equilibrium that is the only plausible outcome of the game (with pre-play communication), as experimentally demonstrated in the referred paper.

6.3 Relations among Different Notions of Strong Correlated Equilibria

A deficiency of the notion of strong correlated equilibrium, is that there are six different variants of it in the literature: three *ex-ante* notions and three *ex-post* notions. In this subsection we present those notions, the relations among them, and the implications of our result.

Notions of *ex-ante* strong correlated equilibria have been presented in Moreno & Wooders (1996), Ray (1996) and Milgrom & Roberts (1996). As discussed in Section 2, our *ex-ant*e definition is equivalent to the definition of Moreno & Wooders. In the framework of Ray, deviating coalitions are not allowed to

---

coalition might form a pact to deviate further with someone not included in this coalition. This limitation has been criticized, especially by the literature that deals with the farsighted coalitional stability approach (described earlier).

[10] Loosely speaking, a strategy profile is punishable if it Pareto-dominates another strategy profile, even when the deviating players do a joint scheme.

[11] The implementation presented in Heller (2008) is only as a $\lfloor n/2 \rfloor$-strong correlated equilibrium (an equilibrium that is resistant to deviations of coalitions with less than $n/2$ players). If one assumes that the players are computationally restricted and "one-way" functions exist, then the implementation can be as a strong Nash equilibrium (as discussed in Lepinski et al., 2004 and Abraham et al., 2006).

construct new correlating devices, and are limited to use only uncorrelated deviations.[12] In the framework of Milgrom & Roberts only some of the coalitions can communicate and coordinate deviations. In both cases the sets of feasible deviations is included in our set of deviations, and thus our set of *ex-ante* strong correlated equilibria is included in the other sets of equilibria.

An *ex-post* strong correlated equilibrium can be defined in our framework, as a profile which is resistant to deviations at the *ex-post* stage when each player knows his recommendation (i.e., no coalition $S \subseteq N$ has a profitable deviation w.r.t. to an *ex-post* information structure $(\mathcal{F}^i)_{i \in S}$, in which: $\forall \omega \in \Omega$, $\forall i \in S$, $\exists a^i \in A^i$ s.t. $\tilde{a}^i\left(F^i(\omega)\right)(a^i) = 1$).

Notions of *ex-post* strong correlated equilibria have been presented in Einy & Peleg (1995), Ray (1998) and Bloch & Dutta (2007). In the framework of Einy & Peleg, a deviating coalition can only use deviations that improve the conditional utilities of all deviating players for *all possible* recommendation profiles.[13] In the framework of Ray, a coalition S can only use *pure* deviations (functions $d^S : A^S \to A^S$). In the framework of Bloch & Dutta, a coalition S can only use deviations that are implemented iff the recommendation profile a^S is included in some set $E^S \subseteq A^S$ which satisfies:

(1) If $a^S \in E^S$, each player earns from implementing the deviation.
(2) If $a^S \notin E^S$, at least one player looses from implementing the deviation.

It can be shown that those conditions imply the existence of a profitable deviation in our framework. (The information structure is such that each deviator would know his recommendation and whether $a^S(\omega) \in E^S$.) Thus our set of *ex-post* strong correlated equilibria is included in the other sets of equilibria.

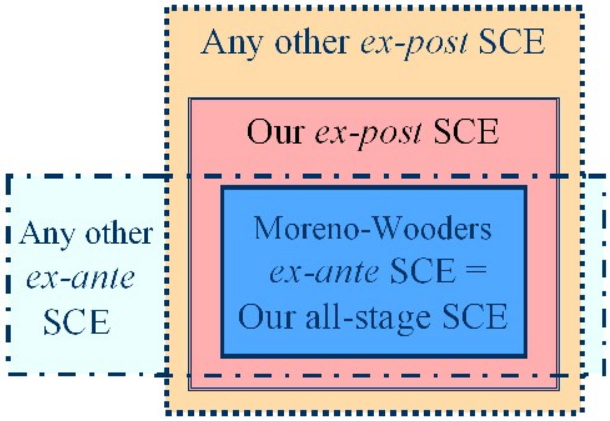
The main result of this paper, reveals inclusion relations among the different notions of strong correlated equilibria, which described in Fig. 1.[14] Thus, Moreno & Wooders *ex-ante* notion is much more robust than originally pre-

---

[12] In Ray's setup, the mediator can send an indirect signal to each player (holding more information than the recommendation itself). In that case, the uncorrelated deviation is a function from the set of S-part of the signals to the set of uncorrelated S-strategy profiles. In our framework, in which coalitions can use new correlating devices, any *ex-ante* strong correlated equilibrium that can be implemented by indirect signals, can also be implemented by a direct correlating device.
[13] In our notations, it is equivalent to requiring that $\forall i \in S$, $u^i_d(\omega) > u^i_f(\omega)$ in every $\omega \in \Omega$, and not only in every $\omega \in F^{meet}(\omega_0)$.
[14] See Moreno & Wooders (1996, Section 4) for an example of an *ex-post* strong correlated equilibrium that is not an *ex-ante* equilibrium.

sented: It is an appropriate notion not only for frameworks where players can only communicate before receiving the recommendations of the correlated agreement, but to any pre-play communication framework (i.e., an *ex-ante* strong correlated equilibrium is resistant to deviations at all stages).

Figure 1. Relations among Different Notions of Strong Correlated Equilibria (SCE)



6.4 Coalition-proof Correlated Equilibria

A correlated equilibrium is coalition-proof if no coalition has a (possibly correlated) profitable self-enforcing deviation. Again, a deficiency of the notion of strong correlated equilibrium, is that there are six different variants of it in the literature (3 *ex-ante* and 3 *ex-post*).[15] It is possible to extend our model of incomplete information, and define a notion of *all-stage coalition-proof correlated equilibrium* (by using an appropriate notion of consistent refinements of information structures). However, the example in Section 5 shows that this notion does not coincide with the *ex-ante* coalition-proof notion, nor that there is any inclusion relations among the different coalition-proof notions.[16] Thus, the notion of coalition-proof correlated equilibrium is not robust: it is sensitive to the exact communication structure.

[15] Conditions for the existence of strong and coalition-proof equilibria in games are discussed in Moreno & Wooders (1996), Milgrom & Roberts (1996), Ray (1996), and Bloch & Dutta (2007).

[16] The example in Section 5 presents an *ex-ante* coalition-proof correlated equilibrium that is not an all-stage equilibrium. Based on this, it is possible to construct a 3-player game with an all-stage coalition-proof correlated equilibrium that is not an *ex-ante* equilibrium, in which the coalition $\{1, 2\}$ would have a deviation that induce a similar situation to that described in table 2. Examples in the literature referred above, show that there are no inclusion relations with the *ex-post* notions as well.

6.5 Extensions of Our Result

(1) *Bayesian games:* Moreno & Wooders (1996) present a notion of *ex-ante* strong communication equilibrium in Bayesian games. Our result can be extended to this setup as well, to show that an *ex-ante* strong communication equilibrium is resistant to deviations at all stages.
(2) *k-strong equilibria*: In Heller (2008) an *ex-ante* notion of k-strong correlated equilibrium is defined as a strategy profile that is resistant to all coalitional deviations of up to k players. Our result can be directly extended to this notion as well: An *ex-ante* k-strong correlated equilibrium is resistant to deviations of up to k players at all stages.

References

[1] Abraham I., Dolev D., Gonen R., Halpern J., 2006. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. Proceedings of the 25th annual ACM symposium on principles of distributed computing, 53-62.

[2] Aumann, R.,1959. Acceptable Points in General cooperative n-person Games, in Kuhn H.W., Luce R.D. (Eds.), Contributions to the Theory of Games IV. Princeton University Press, NJ, 287-324.

[3] Aumann, R., 1976. Agreeing to disagree. Ann. Statist. 4 (no. 6), 1236-1239.

[4] Aumann, R., 1987. Correlated equilibrium as an expression of Bayesian rationality, Econometrica 55, 1-18.

[5] Aumann R., Hart, S., 2003. Long cheap talk. Econometrica 71 (6), 1619-1660.

[6] Bernheim, B.D., Whinston, M., 1986. Menu auctions, resource allocation, and economic influence, Quart. J. Econom. 101, 1-31.

[7] Bernheim, B.D., Peleg, B., Whinston, M., 1987. Coalition-proof Nash equilibria - I. concepts. J. Econ. Theory 42, 1-12.

[8] Bernheim, B.D., Whinston, M., 1987. Coalition-proof Nash equilibria - II. applications. J. Econ. Theory 42, 13-29.

[9] Bloch F., Dutta B., 2007. Correlated equilibria, incomplete information and coalitional deviations. mimeo (based on The Warwick Economics Research Paper Series - paper 763).

[10] Chwe, MS-Y, 1994. Farsighted coalitional stability. J. Econ. Theory 63, 299-325.

[11] Delgado J., Moreno D., 2004. Coalition-proof supply function equilibria in Oligopoly. J. Econ. Theory, 114 231-254.

[12] Einy, E., Peleg B., 1995. Coalition-proof communication equilibria, in Social Choice, Welfare & Ethics (W. Barnett, H. Moulin, M. Salles and N. Schofield, eds.), Cambridge, New-York and Melbourne: Cambridge University Press.

[13] Farrel J., Rabin M., 1996. Cheap-talk, J. Econ. Perspect. 10, 103-118.

[14] Farrel, J., Saloner, G., 1988. Coordination through committees and markets. RAND J. Econ. 19, 235-252.

[15] Greenberg, J., 1989. Deriving strong and coalition-prof Nash equilibrium from an abstract system. J. Econ. Theory 49, 195-202.

[16] Greenberg, J., 1990. The theory of social situations. Cambridge University Press.

[17] Heller Y., 2008, A Minority-proof cheap-talk protocol, mimeo, http://www.tau.ac.il/~helleryu/minority.pdf

[18] Holzman R., Law-Yone N., 1997. Strong equilibrium in congestion games, Games Econ. Behav. 21, 85-101.

[19] Konishi, H., Le Breton, M.,Weber, S., 1997a. Equilibria in a model with partial rivalry. J. Econ. Theory 72, 225-237.

[20] Konishi, H., Le Breton, M.,Weber, S., 1997b. Equivalence of strong and coalition-proof Nash equilibria in games without spillovers. Econ. Theory 9, 97-113.

[21] Konishi, H., Le Breton, M.,Weber, S., 1999. On coalition-proof Nash Equilibria in common agency games. J. Econ. Theory 85, 122-139.

[22] Lepinski, M., Micali, S., Peikert C., Shelat A., 2004. Completely fair SFE and coalition-Safe cheap talk. Proceedings of the 23rd annual ACM symposium on Principles of distributed computing, 1-10.

[23] Mariotti, M., 1997. A theory of agreements in strategic form games. J. Econ. Theory 74, 196-217.

[24] Milgrom, P., Roberts, J., 1996. Coalition-proofness and correlation with arbitrary communication possibilities, Games Econ. Behav. 17, 113-128.

[25] Moreno, D., Wooders, J., 1996. Coalition-proof equilibrium, Games Econ. Behav. 17, 80-113.

[26] Moreno, D., Wooders, J., 1998. An experimental study of communication and coordination in non-cooperative Games, Games Econ. Behav. 24, 47-76.

[27] Rabin, M., 1994. A model of pre-game communication. J. Econ. Theory 63, 370-391.

[28] Ray I., 1996. Coalition-proof correlated equilibrium: a definition, Games Econ. Behav. 17, 56-79.

[29] Ray, I., 1998. Correlated equilibrium as a stable standard of behavior, Rev. Econ. Design 3, 257-269

[30] Thoron, S., 1998. Formation of a Coalition-Proof Stable Cartel, Economics Can. J. Econ. 31, 63-76.

[31] Von-Neumann, J., Morgenstern O., 1953. Theory of games and economic behavior. 3rd ed., Princeton University Press, NJ.

[32] Xue, L., 1998. Coalitional stability under perfect foresight. Econ. Theory 11, 603-627.

[33] Xue, L., 2000. Negotiation-proof Nash equilibrium. Int. J. Game Theory 29, 339-357.

[34] Yi SS., 1999. On the coalition-proofness of the Pareto frontier of the set of Nash equilibria, Games Econ. Behav. 26 353-364.