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.

Email: helleryu@post.tau.ac.il

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. Thus the set of *ex-ante* strong correlated equilibria of Moreno & Wooders (Games Econ. Behav. 17 (1996), 80-113) is included in all other sets of strong correlated equilibria.

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

[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. I would also like to express my deep gratitude to to Ehud Lehrer and Yaron Azrieli for many useful comments, discussions and ideas.

1 Introduction

The ability of players to communicate prior to playing a game, 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-binding agreement. For such an agreement to be self-enforcing, it has to be stable against "plausible" coalitional deviations. Two 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. For a coalition of a single player any deviation is self-enforcing. For coalitions of more than one player, a deviation is self-enforcing if there is no further self-enforcing and improving deviation by one of its proper sub-coalitions.

Each notion has a few alternative definitions. One major difference between 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 [25,26,29] are *ex-ante* definitions: players may plan deviations before receiving the recommendations, and no further communication is allowed after recommendations are issued. The definitions in [9,13,30] are *ex-post* [2] definitions: 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], where the players can "mimic" a mediator, and implement a large set of strong correlated equilibria as strong Nash equilibria in the extended game ([18]). A coalition can plan a deviation in the early phases of the cheap-talk when no player has yet received his recommendation (*ex-ante* stage), in the late phases when all players have received their recommendations (*ex-post* stage), or in an intermediate stage when each player has some partial information about his recommendation (as described in subsection 2.2).

A natural question is whether any of the existing notions is appropriate to such frameworks, or whether new definitions are needed. Our main result

[2] Referred to as *"interim"* in some of these papers.
[3] Cheap-talk is pre-play, unmediated, non-binding, non-verifiable communication among players. For a good nontechnical introduction to some of the main issues of cheap-talk, see the survey in [14].

shows that the existing *ex-ante* strong correlated equilibrium à la Moreno & Wooders ([26]) is resistant to deviations at all stages. The result is based on three assumptions (which hold in the cheap-talk framework):

(1) A deviating coalition can use new correlation devices.
(2) When a coalition decides to deviate, that decision is common knowledge 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 ([4]).

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 the main result.

The paper is organized as follows: Section 2 presents the model and the main result. The 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 the 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, 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 A is called an action profile, and 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 The Framework

The framework consists of players who implement a correlated strategy profile (an agreement) $q \in \Delta(A)$ with the assistance of a mediator. The mediator performs a private lottery, and chooses each action profile $a \in A$ with probability $q(a)$. Then he reveals to each player his recommendation by the following revealing process. He makes another private lottery, possibly with a distribution that is unknown to the players, and chooses accordingly the order in which he reveals the recommendations. At each stage, he privately informs one of the players a part of his recommendation, until at the end of this process, each player knows his recommendation. During this process, the players can communicate, share some of the information they have acquired so far about their recommendations, and plan coalitional deviations from the agreement. A player agrees to deviate, if given his own information, he believes that the deviation is profitable for him. If at some stage of the process, all the members of a coalition agree to use a joint deviation, then it is implemented with the assistance of a new mediator. The new mediator receives the recommendations of the deviating players at the end of the revealing process, and gives each of them a new recommendation. The profile q is an *all-stage strong correlated equilibrium* if, for every revealing process, and for every stage of such a process, there is no coalition with a profitable deviation.

We would like to emphasize the following points:

(1) The mediator may reveal the recommendations in parts. For example, if the possible actions of a player are $\{a, b, c\}$,the mediator may reveal him first that his recommendation is not c, and only at a later stage reveal him that it is a.
(2) The framework allows to model a broad variety of revealing processes, including:
   - *Ex-ante* process ([25,26,29]) - The mediator simultaneously reveals the recommendations at the end of the process, with no further communication among the players. Communication is only allowed before the players receive any information about their recommendations.
   - *Ex-post* process ([9,13,30] - The mediator simultaneously reveals the recommendations to all the players at the beginning of the process, and

players can only communicate afterwords.
- Polite Cheap-talk process ([18]) - The players acquire their recommendations consecutively by some common known order.
- Unknown revealing process - The players do not know the order of the revealing process.

(3) An all-stage strong correlated equilibrium is required to be resistant against deviations in every revealing process. The main result shows that this is equivalent to the seemingly weaker requirement of resistance against deviations only in an *ex-ante* process.

(4) It is assumed that:
- The decision to implement a deviation is common knowledge in the deviating coalition.
- If a deviation is implemented, then all the deviating players follow it: report their true recommendations to the new mediator, follow the new recommendations, and avoid the implementation of sub-deviations.

2.3 All-stage Strong Correlated Equilibrium

The broad variety of revealing processes is modeled by an incomplete information model *à* la Aumann ([4]). A *state space* is a probability space, $(\Omega, \mathcal{B}, \mu)$, which describes all parameters that may be the object of uncertainty on the part of the players. We interpret Ω as the space of all possible states of the world, $\mathcal{B}$ as the σ-algebra of all measurable events, and μ as the common prior. [4]

Given a non-null event $E \in \mathcal{B}$ and a random variable $\mathbf{x} : \Omega \to X$ (where X is a finite set), let $\mathbf{x}(E) \in \Delta(X)$ denote the posterior distribution of $\mathbf{x}$ conditioned on the event E. The implementation of an agreement by a mediator is modeled by a random variable $\mathbf{a} : \Omega \to A$, which satisfies that the prior distribution $\mathbf{a}(\Omega)$ is equal to the agreement distribution.

Definition 1 Let G be a game, $q \in \Delta(A)$ an agreement*,* and $(\Omega, \mathcal{B}, \mu)$ a state space. A *recommendation profile* is a random variable $\mathbf{a} = (\mathbf{a}^i)_{i \in N} : \Omega \to A$ that satisfies: $\mathbf{a}(\Omega) = q$.

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

Definition 2 Let G be a game, $q \in \Delta(A)$ an agreement*,* $S \subseteq N$ a coalition,

[4] The justification of the common prior assumption is discussed in [4, Section 5].

$(\Omega, \mathcal{B}, \mu)$ a state space, and $\mathbf{a} : \Omega \to A$ a recommendation profile. A *deviation* (of S) is a random variable $\mathbf{d}^S = (\mathbf{d}^i)_{i\in S} : \Omega \to A^S$ that is conditionally independent of $\mathbf{a}^{-S}$ given $\mathbf{a}^S$.

The interpretation is the following: If the players of S agree to use deviation $\mathbf{d}^S$, they implement it with the assistance of a new mediator. The mediator receives the 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 $\mathbf{a}^S$, but not on $\mathbf{a}^{-S}$.

When the members of a coalition S consider the implementation of 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 3 Let G be a game, $q \in \Delta(A)$ an agreement, $S \subseteq N$ a coalition, $(\Omega, \mathcal{B}, \mu)$ a state space, and $\mathbf{a} : \Omega \to A$ a recommendation profile. *An information structure* (of S) is 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 $\forall \omega \in \Omega$, $\forall i \in S$, $\forall a \in A$, $\mathbf{a}\left(F^i(\omega)\right)(a) = \mathbf{a}^S\left(F^i(\omega)\right)\left(a^S\right) \cdot q(a^{-S} \mid a^S)$.

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. Formally, let G be a game, $q \in \Delta(A)$ an agreement, $S \subseteq N$ a coalition, $i \in S$ a player, $(\Omega, \mathcal{B}, \mu)$ a state space, $\mathbf{a} : \Omega \to A$ a recommendation profile, $\mathbf{d}^S : \Omega \to A^S$ a deviation, and $(\mathcal{F}^i)_{i\in S}$ a consistent information structure. The *conditional expected payoffs of player* i in $\omega \in \Omega$ are:

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

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

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

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

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. Formally ([3]):

Definition 4 Let G be a game, $S \subseteq N$ a coalition, $(\Omega, \mathcal{B}, \mu)$ a state space, $(\mathcal{F}^i)_{i\in S}$ an information structure, and $\omega \in \Omega$ a state. 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 5 Let G be a game. $q \in \Delta(A)$ an agreement, $S \subseteq N$ a coalition, $i \in S$ a player, $(\Omega, \mathcal{B}, \mu)$ a state space, and $\mathbf{a} : \Omega \to A$ a recommendation profile. A deviation (of S) $\mathbf{d}^S$ is *profitable*, 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 $\mathbf{d}^S$ *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 6 Let G be a game. 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 7 Let G be a game and $(\Omega, \mathcal{B}, \mu)$ a state space. A strategy profile $q \in \Delta(A)$ is an *ex-ante strong correlated equilibrium* if no coalition $S \subseteq N$ has a profitable deviation with respect to the *ex-ante* information structure $(\mathcal{F}^i)_{i \in S}$ that satisfies $\forall i, \mathcal{F}^i = \Omega$.

One can verify that Def. 7 is equivalent to the definition of ([26]). The definition immediately implies that an all-stage strong correlated equilibrium is also an *ex-ante* strong correlated equilibrium. The main result shows that the converse is also true, and thus the two notions coincide.

Theorem 8 *A correlated profile 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 glance, 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 the case. The analysis in this example provides the intuition for the use of a model of incomplete information *à* la Aumann ([4]), and for the 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.

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

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* strong correlated equilibrium:

- The profile q is a correlated equilibrium, and thus no player has a unilateral profitable deviation.
- No coalition of two players has a profitable deviation, because their uncertainty about the recommendation of the third player prevents them from earning 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. No player knows whether the other players have received their recommendations. At first look, the implementation of the deviation $\mathbf{d}(\cdot) = (a_3, b_3, c_3)$, 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 $\mathbf{d}$ 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 $\mathbf{d}$.

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

4 The Proof of the Main Result

In this Section we prove the main result. As discussed earlier, one direction immediately follows from the definitions, and we have to prove only the other direction:

Theorem 9 *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 an agreement that is not an all-stage strong correlated equilibrium in a game G. *Let* $(\Omega, \mathcal{B}, \mu)$ be the state space, and $\mathbf{a} : \Omega \to A$ the recommendation profile. Thus there exists a coalition $S \subseteq N$ with a profitable deviation $\mathbf{d}^S : \Omega \to A^S$ 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 $F^{meet} = F^{meet}(\omega_0) = \dot{\bigcup_j} F^i_j$ 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 $\mathbf{d}^S_e$ with respect to the *ex-ante* information structure $(\mathcal{F}^i_e)_{i \in S}$, which satisfies $\forall i,\ \mathcal{F}^i_e = \Omega$: $\mathbf{d}^S_e(\omega) = \begin{cases} \mathbf{d}^S(\omega) & \omega \in F^{meet} \\ \mathbf{a}^S(\omega) & \omega \notin F^{meet} \end{cases}$.

Observe that $\mathbf{d}^S_e$ and $\mathbf{a}^{-S}$ are conditionally independent given $\mathbf{a}^S$, thus $\mathbf{d}^S_e$ 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 $\mathbf{d}^S_e$ is profitable, i.e: $\forall i,\ u^i_{d_e}(\omega) > u^i_{f_e}(\omega)$. Let $i \in S$ be a deviating player.

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

Equation 2 is due to the equality $F^i_e(\omega) = \Omega$, (3) holds since $\mathbf{d}^S_e = \mathbf{a}^{-S}$ outside F^{meet}, (4) holds since $\mathbf{d}^S_e = \mathbf{d}^S$ in F^{meet}, (5) follows from $F^{meet} = \dot{\bigcup_j} F^i_j$, and the last inequality is implied by $F^{meet} \subseteq \left\{\omega \mid u^i_d(\omega) > u^i_f(\omega)\right\}$. **QED**.

5 Coalition-Proof Correlated Equilibria

In the previous Section we have shown that an *ex-ante* strong correlated equilibrium is also appropriate to frameworks in which players can plan deviations at all stages. A natural question is whether a similar result holds for the notion of coalition-proof correlated equilibrium. [5] In this Section we show that the answer is negative, by presenting an example, adapted from [9], 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. [26] shows 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 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 frame-

[5] Recall ([26]) 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. For a coalition of a single player any *ex-ante* deviation is self-enforcing. For coalitions of more than one player, an *ex-ante* deviation is self-enforcing if there is no further self-enforcing and improving *ex-ante* deviation by one of its proper sub-coalitions.

work 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

We deal with self-enforcing agreements in (one-shot) games in environments where players can freely discuss their strategies before the play starts. 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. [6]

Another approach is to explicitly model the procedure of communication as an extended-form game that specifies how messages are interchanged (e.g.: [5,15,28]). However, the results are sensitive to the exact procedure employed, and usually strong restrictions have to be made to isolate the desired outcome.

[6] As discussed in [6,35]. [35] presents a set of conditions that if satisfied, the two notions of Pareto dominance refinement and coalition-proof equilibrium coincide.

A different approach is the *farsighted coalitional stability.* Alternative variations are discussed in: [10,16,17,24,33,34]. [7] 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 ([32]). This approach is less appropriate when coalitions can *privately* plan deviations. [8]

6.2 Strong and Coalition-Proof Equilibria

A Nash equilibrium is strong ([2]) if no coalition, taking the actions of its complement as given, has an uncorrelated deviation that benefits all of its members. The main drawback of this notion, is that it exists in only a relatively small set of games. [9] [6] presents 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 uncorrelated deviation. For a coalition of a single player any deviation is self-enforcing. For coalitions of more than one player, a deviation is self-enforcing if there is no further self-enforcing and improving deviation by one of its proper sub-coalitions. [10] The notion of coalition-proof equilibrium has been useful in a variety of applied contexts, such as: *menu* auctions ([7]), oligopolies ([8,11,12,31]), and common agency games ([22]).

These notions focus on environments where coalitions can *privately* commu-

[7] Also called *negotiation-proof equilibrium* and *full coalitional equilibrium.*
[8] [34, Section 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.
[9] Examples for games where strong Nash equilibria exist are congestion games ([19]); games where the preferences satisfying independence of irrelevant choices, anonymity, and partial rivalry ([20]); and games where the core of the cooperative game derived from the original normal form game, is non-empty (see [21], and the references within). Conditions for the equivalence of strong and coalition-proof Nash equilibria are presented in [21] (games with population monotonicity property) and in [22] (common agency games).
[10] Observe that only members of the deviating coalition may contemplate deviations from the deviation. This rules out the possibility that members of the deviating coalition might form a pact to deviate further with someone not included in this coalition. This limitation has been criticized, especially in the literature that deals with the farsighted coalitional stability approach (described earlier).

nicate before the play starts, 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 [18], which shows that any "punishable"[11] *ex-ante* strong correlated equilibrium is a strong Nash equilibrium in an extended game with *cheap-talk*.[12]
- The example in [27] of an *ex-ante* strong correlated equilibrium that is the only plausible outcome of a 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 the main result.

Notions of *ex-ante* strong correlated equilibria have been presented in [26,29,25]. As discussed in Section 2, our *ex-ant*e definition is equivalent to the definition in [26]. In [29] deviating coalitions are not allowed to construct new correlation devices, and are limited to use only uncorrelated deviations.[13] In [25] 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

[11] Loosely speaking, a strategy profile is punishable if it Pareto-dominates another strategy profile, even when the deviating players do a joint scheme.
[12] The implementation presented in [18] is only as a $\lfloor n/2 \rfloor$-strong Nash 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 (see [1,23]).
[13] In [29]'s setup, the mediator can send an indirect signal to each player, which holds more information than the recommendation itself. In that case, the uncorrelated deviation is a function from the set of the S-part of the signals to the set of uncorrelated S-strategy profiles. In our framework, in which coalitions can use new correlation devices, any *ex-ante* strong correlated equilibrium that can be implemented by indirect signals, can also be implemented by a direct correlation device.

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 with respect to 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. $\mathbf{a}^i\left(F^i(\omega)\right)(a^i) = 1$).

Notions of *ex-post* strong correlated equilibria have been presented in [13,30,9]. In [13] a deviating coalition can only use deviations that improve the conditional utilities of all deviating players for *all possible* recommendation profiles. [14] In [30] a coalition S can only use *pure* deviations (functions $d^S : A^S \to A^S$). In [9], a coalition S can only use deviations that are implemented if 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 with respect to an *ex-post* information structure. [15] Thus our set of *ex-post* strong correlated equilibria is included in the other sets of equilibria.

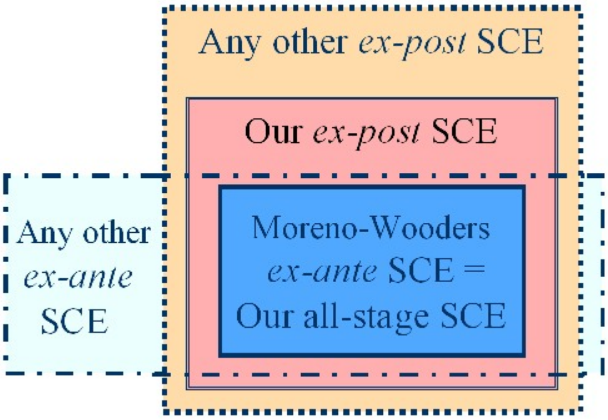
The main result reveals inclusion relations among the different notions of strong correlated equilibria, which described in Fig. 1. [16] Thus, [26]'s *ex-ante* notion is much more robust than originally presented: It is an appropriate notion not only for frameworks where players can only communicate before receiving the recommendations of the correlated agreement, but to any preplay mediation and communication framework.

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

[14] 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 $\omega \in F^{meet}(\omega_0)$.
[15] The information structure is such that each deviator would know his recommendation and whether $\mathbf{a}^S(\omega) \in E^S$.
[16] See [26, Section 4] for an example of an *ex-post* strong correlated equilibrium that is not an *ex-ante* equilibrium.

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



literature (3 *ex-ante* and 3 *ex-post*).[17] It is possible to extend the 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.[18] Thus, the notion of coalition-proof correlated equilibrium is not robust: it is sensitive to the exact communication structure.

6.5 Extensions of the Main Result

(1) *Bayesian games:* [26] presents a notion of *ex-ante* strong communication equilibrium in Bayesian games. The main 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 [18] 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. The main 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.

[17] Conditions for the existence of strong and coalition-proof correlated equilibria in games are discussed in [9,25,26,29].
[18] 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* coaliton-proof equilibrium, in which the coalition $\{1, 2\}$ would have a deviation that induces a similar situation to that described in table 2. The examples in [9,26,29] show that there are no inclusion relations with the *ex-post* notions as well.

References

[1] I. Abraham, D. Dolev, R. Gonen , J. Halpern, 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 (2006), 53-62.

[2] R.J. Aumann, Acceptable points in general cooperative n-person games, in: H.W. Kuhn , R.D. Luce (Eds.), Contributions to the Theory of Games IV, Princeton University Press, Princeton, 1959, pp. 287-324.

[3] R.J. Aumann, Agreeing to disagree, Ann. Statist. 4 (1976), 1236-1239.

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

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

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

[7] B.D. Bernheim, M. Whinston, 1986. Menu auctions, resource allocation, and economic influence, Quart. J. Econ. 101 (1986), 1-31.

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

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

[10] M.S.Y. Chwe, Farsighted coalitional stability. J. Econ. Theory 63 (1994), 299-325.

[11] J. Delgado, Coalition-proof supply function equilibria under capacity constraints, Econ. Theory 29 (2006), 219-229.

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

[13] E. Einy, B. Peleg, Coalition-proof communication equilibria, in: W. Barnett, H. Moulin, M. Salles, N. Schofield (Eds.), Social Choice, Welfare & Ethics, Cambridge University Press, Cambridge, 1995, pp. 289-300.

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

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

[16] J. Greenberg, Deriving strong and coalition-proof Nash equilibrium from an abstract system, J. Econ. Theory 49 (1989), 195-202.

[17] J. Greenberg, The Theory of Social Situations, Cambridge University Press, Cambridge, 1990.

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

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

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

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

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

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

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

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

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

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

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

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

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

[31] S. Thoron, Formation of a coalition-proof stable cartel, Can. J. Econ. 31 (1998), 63-76.

[32] J. von-Neumann, O. Morgenstern, Theory of games and economic behavior (3rd ed.), Princeton University Press, Princeton, 1953.

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

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

[35] S.S. Yi, On the coalition-proofness of the Pareto frontier of the set of Nash equilibria, Games Econ. Behav. 26 (1999), 353-364.