COALITIONAL RATIONALITY

Attila Ambrus*

Working paper
11/22/2002

Abstract

This paper investigates the implications of groups or coalitions of players acting in their collective interest in non-cooperative normal form games. It is assumed that players are unable to make binding agreements, and pre-play communication is neither precluded nor assumed. The main idea is that each member of a coalition will confine play to a subset of their strategies if it is in their mutual interest to do so. This leads to an iterative procedure of restricting players' beliefs and action choices in the game. The procedure defines a non-cooperative solution concept, the set of coalitionally rationalizable strategies. The solution set is a refinement of the the set of rationalizable strategies. In contrast to equilibrium based solution concepts, it is always nonempty, despite the fact that every coalition is simultaneously allowed to "make agreements". It is also robust to the order in which agreements are made. Relations to other non-cooperative solution concepts, refinements of the basic concept and various applications are also offered.

————————————————

1 Introduction

Many non-cooperative games contain elements of both conflicting interests among individual players (competition) and common interests among subgroups of players (cooperation). A natural question which arises is to what extent players can incorporate a recognition of common interest into their individual plays. In particular, different groups of players, that is *coalitions* may be aware of common interests in a game. We will assume that players are unable to make binding agreements, and we do not require that players are able to communicate with one another. Our objective is to investigate what restrictions on coalitional behavior may be deduced in this setting.

We present a theory to address these questions in finite normal-form games, although the principles we propose can be applied to more general settings. The scenario we have in mind is one in which players make their moves secretly and independently. We rely on four main assumptions to support our claims.

The central premise of our theory is that players utilize a distinct reasoning procedure when they formulate conjectures, which we call *coalitional rationality*. The key feature of this method of reasoning is that the type of implicit agreements players consider is restricting their play to a subset of their strategy sets. Another way to look at these agreements is that players agree upon *not* to play certain strategies, rather than upon exactly which strategies to play. The intuition behind coalitional rationality is that, whenever it is of mutual interest for a group of players to avoid certain strategies, individual members will make an implicit agreement not to play them and formulate their conjectures accordingly (they expect similar reasoning from others in the group). This excercise is in the same spirit as *iterated deletion of never best response strategies*: excluding the play of certain strategies. By being "of mutual interest for the group" we mean that every group member always (for every possible expectation) expects a higher payoff if the implicit agreement is made than if he instead chooses to play a strategy outside the agreement. If we associate conjectures with the payoffs that best response strategies yield, then we require an agreement to be such that players in the group prefer any conjecture compatible with the agreement to any for which a strategy outside the agreement is a best response.

There are two more basic assumptions that we rely on. The first is that players are *Bayesian decision makers*, they base their strategy choices on conjectures on other players' actions. Second, we assume that there *cannot be binding agreements* among players.

On the other hand, we do not require players' conjectures to be correct. Therefore, like rationalizability (Bernheim (84) and Pearce (84)) we present an *out-of-equilibrium* theory. Finally, we *do not assume pre-play communication.*

It suffices that the conclusions players draw about the actions of others are based purely on *public information* – that is, the payoffs of the game. We claim though that pre-play communication does not invalidate our results.

Since the set of coalitionally rationalizable strategies is an out-of-equilibrium concept, its purpose is to confidently rule out some outcomes of the game, as opposed to giving a sharp prediction. We do not claim that all outcomes in the solution set are equally likely, since players, in addition to coalitional reasoning, might have other type of considerations when assessing what strategies are plausible in the given game. But we claim that if players' strategic reasoning includes thinking about what strategies groups of players could agree upon playing, then we can rule out any outcomes which are inconsistent with coalitional rationality.

In Section 2 we provide a variety of motivating examples illustrating agreements that are of mutual interest to members of some coalition in the game. Section 3 establishes the technical foundations of the theory. We begin by defining an iterative procedure which formalizes the following intuitive line of thought. Essentially, by requiring every mutually beneficial implicit agreement of every coalition to be made, the support of players'conjectures is restricted to a subset of the original strategy set. In this new restricted game we again require players to make every mutually beneficial implicit agreement of every coalition, and so on. This leads to a sequence of implicit agreements in which certain strategies are agreed not to be played. The solution set we propose is the set of outcomes surviving this iterative procedure. We call this the *set of coalitionally rationalizable strategies*. We prove that the agreements we consider are compatible with each other at any stage of the iterative procedure and that the set of coalitionally rationalizable strategies is nonempty. The result is particularly surprising given that we allow every coalition to simultaneously make agreements and consequently leave open the possibility of a player making conflicting agreements as a member of two (or more) coalitions. Solution concepts which allow coalitions to make agreements simultaneously, like strong Nash equilibrium (see Aumann(59)) or coalition-proof equilibrium (see Bernheim, Peleg and Whinston (87)) typically suffer from incompatibility of agreements, which can give rise to empty solution sets in games of economic interest. An advantage of the type of coalitional agreements that we propose is that they are flexible enough to be compatible with each other, unlike agreements in which coalitions fix a unique strategy for every member.

The iterative procedure we provide resembles iterative removal of strategies that are never best responses. In section 3 we prove that indeed all strategies that are deleted in iterative removal of strictly dominated strategies are deleted in the procedure we define, making our solution set a refinement of rationalizability. A related result of section 3 establishes a further parallel between the two procedures by proving that the order in which coalitions make implicit agreements is irrelevant to our results in the same manner that the order of deletions is inconsequential in iterative removal of strictly dominated strategies.

3

Requiring coalitions to play inside agreements that are mutually advantageous is similar to requiring individual players not to play strategies that are never best responses.

2 Motivating examples

To motivate the ideas behind coalitional rationality, we start with the simplest possible example and gradually increase the level of complexity to illustrate additional points.

The game of Figure 1 is an asymmetric coordination game. Strategies A2 and B2 are rationalizable for player 1 and player 2, and they even constitute a Nash equilibrium. Still, coordinating on playing (A1,B1) is mutually beneficial for the players in the sense that it yields a higher payoff to both of them than any other payoff in the game. If players think in terms of common gains, they can only reasonably expect each other to play A1 and B1.

	B1	B2
A1	2,2	0,0
A2	0,0	1,1

Figure 1

The game of Figure 2 shows that even if there is no unique outcome that is the best for all players, there might be room to coordinate play. Restricting play to $(A1, A2) \times (B1, B2)$ is mutually beneficial for players 1 and 2 in the sense that no matter what outcome is played in that set, player 1 gets a higher payoff than anything he can get if he plays A3 and player 2 gets a higher payoff than anything he can get if he plays B3.

	B1	B2	B3
A1	4,3	2,2	0,0
A2	2,2	3,4	0,0
A3	0,0	0,0	1,1

Figure 2

The game of Figure 3 demonstrates that even if the coalition of all players doesnot have any opportunity to coordinate play in a way that is mutually advantegous for everyone, there might be a subgroup of players to whom coordinating their play is mutually beneficial.

4

	C1			C2	
	B1	B2		B1	B2
A1	4,4,0	0,0,1	A1	3,3,1	0,2,0
A2	3,0,1	2,2,0	A2	2,0,3	2,2,3

Figure 3

In this game no matter what (mixed) strategy player 3 plays, coordinating play to (A1,B1) is always mutually beneficial to players 1 and 2, in the sense that it yields a higher expected payoff than the payoff from any other pair of strategies they can play. Another point we can make here is that player 3, knowing that players 1 and 2 coordinate on playing A1 and B1 is better off playing C2. This suggests that strategy profiles that can be consistent with coalitional reasoning should be derived in an iterative manner. For more on this, consider the next game.

The last example we present below is a four-player version of a classical positive externality situation. "Farmer 1" and "Farmer 2" are two fruit farmers, having three pure strategy choices. Growing apples, growing peaches or not growing anything the given year. "Bee keeper 1" is a bee keeper operating next to "Farmer 1", and "Bee keeper 2" is a bee keeper operating next to "Farmer 2". The bee-keepers both have three pure strategies: keeping "apple bees", keeping "peach bees", or going out of business. Growing fruit and keeping bees next to each other inflict positive externalities on each other, raising each others' productivity. We assume that the externality is so strong that growing any kind of fruit is profitable for the farmers if and only if the bee-keeper operating next door is in business. Similarly, bee-keeping is profitable if and only if the farmer operating next door grows some kind of fruit.

The farmers sell their products to the same local market and therefore they are direct competitors. Competition affects farmers payoffs more if they produce the same kind of fruit. The type of bee next door does not affect profitability of fruit growing, but apple bees yield more honey and therefore more payoff to the bee-keeper if the farmer next door grows apples, while peach bees yield more payoff to the bee-keeper if the farmer next door grows peaches. We will call a farmer "weak" if the bee-keeper next to him is out of business and "strong" if the bee-keeper next to him is in business.

A farmer's payoff is 0 if he does not grow any fruits. His payoff is -1 if he grows some kind of fruit and the bee-keeper next door is out of business. Farmer 1's payoff is $8 - x$ and $7 - x$ respectively if he grows apples and peaches and the bee-keeper next door is in business, where x represents the severeness of the competition. Let x be 0 if the other farmer does not grow any fruits or he grows a different kind of fruit than the first farmer, and the other farmer is "weak". Let x be 2 if the other farmer is "weak" and he grows the same kind of fruit as the first farmer, or if the other farmer is "strong" and he grows different kind of fruit as the first farmer. Finally, let x be 3 if the other farmer is "strong" and he grows the same kind of fruit as the first farmer. Farmer 2's payoff is $7 - x$

5

and $8 - x$ respectively if he grows apples and peaches and the bee-keeper next door is in business, where x is defined the same way as above.

A bee-keeper's payoff is 0 if he is out of business. His payoff is -1 if he keeps some kind of bees and the farmer next door does not grow fruits. A bee-keeper's payoff is 3 if he keeps apple bees and the farmer next door grows apples, 1 if he keeps apple bees and the farmer next door grows peaches, 2 if he keeps peach bees and the farmer next door grows peaches, and 1 if he keeps peach bees and the farmer next door grows apples.

Payoffs are summarized in the tables below.

apple	-1
peach	-1
nothing	0

Table 1:
A farmer's payoffs if bee-keeper next door is out of business, rows represent the type of fruit that the farmer grows.

x	apple/s	apple/w	peach/s	peach/w	nothing
apple	5	6	6	8	8
peach	5	7	4	5	7
nothing	0	0	0	0	0

Table 2:
Farmer 1's payoffs if bee keeper next door is in business, rows represent the type of fruit that the farmer grows, columns represent the type of fruit the rival farmer grows, and whether the rival farmer is weak or strong.

x	apple/s	apple/w	peach/s	peach/w	nothing
apple	4	5	5	7	7
peach	6	8	5	6	8
nothing	0	0	0	0	0

Table 3:
Farmer 2's payoffs if the bee keeper next door is in business, rows represent the type of fruit that the farmer grows, columns represent the type of fruit the rival farmer grows, and whether the rival farmer is weak or strong.

x	apple	peach	nothing
apple bee	3	1	-1
peach bee	1	2	-1
out of business	0	0	0

Table 4:
A bee-keepers payoffs, rows represent the bee-keeper's action, columns represent the action of the farmer next door.

In this game every strategy is rationalizable. There are many Nash equilibria. One is the profile in which both farmers grow nothing and both bee-keepers are out of business. Another one is in which Farmer 1 grows peaches, Bee-keeper 1 keeps peach bees, Farmer 2 grows apples and Bee-keeper 2 keeps apple bees. A third one is in which Farmer 1 grows apples, Bee-keeper 1 keeps apple bees, Farmer 2 grows peaches and Bee-keeper 2 keeps peach bees. There is no Pareto-dominant Nash equilibrium. Still, there is a unique natural solution in this game, given by the following reasoning procedure. Independently of what strategies Farmer 2 and Bee-keeper 2 play, Farmer 1 and Bee-keeper 1 are always better off if they do not choose "not growing any fruits" and "going out of business" respectively, then if at least one of them chooses it. Farmer 1's and Bee-keeper 1's interests coincide in engaging in some kind of economic activity, so we expect them to make an implicit agreement to do so. We expect a similar implicit agreement between Farmer 2 and Bee-keeper 2, to engage in some kind of economic activity. Because of the above, the only candidates for reasonable conjectures in this game are those which allocate probability 0 to either of the farmers playing "not growing anything", and probability 0 to either of the bee-keepers playing "going out of business". Now note that if both farmers grow some kind of fruit and both bee-keepers keep some kind of bees, then the unique best payoff Farmer 1 and Farmer 2 can get is when Farmer 1 grows apples and Farmer 2 grows peaches, independently of what type of bees the bee-keepers keep. Therefore we expect them to make an implicit agreement to do so, and therefore the only candidates for reasonable conjectures in this game those which allocate probability 1 to Farmer 1 growing apples and Farmer 2 growing peaches. And given this restriction on conjectures, a rational Bee-keeper 1 keeps apple bees and a rational Bee-keeper 2 keeps peach bees.

Through a series of belief restrictions we solved the game. The belief restrictions we made in each step had the property that they were *mutually advantageous* for the players whose actions were restricted (henceforth referred to as a coalition), for every possible conjecture on the strategy of players whose actions were not restricted. By mutually advantageous we mean that the expected payoff of everyone in the coalition is higher if the restriction is made than if the restriction is not made and he plays outside it. In the next section we formally define the concept of mutually advantegous agreement for a group of players, which is the central building block for the solution concept that this paper offers.

The sequence of restrictions has an iterative structure. The implicit agreement between the two farmers that Farmer 1 grows apples and Farmer 2 grows peaches is only mutually advantageous if it is known that both of them grow either apples or peaches, so after the first round of implicit agreements. In this example, there are three rounds of agreements. In the first round, there is an agrement between Farmer 1 and Bee-keeper 1, and an agreement between Farmer 2 and Bee-keeper 2. In the second round, there is an agreement between Farmer 1 and Farmer 2. Finally, in the last round there are two "single player agreements", one by Bee-keeper 1 and one by Bee-keeper 2. In the next section we introduce an iterative procedure based on the concept of supported restrictions and call the solution set obtained by the procedure the set of coalitionally

rationalizable strategies. In the above example the solution consists of a single profile. In general the belief restrictions we consider give a set-valued solution concept.

Observe that in this example although in each step different coalitions of players make agreements and in some steps there are multiple coalitions making agreements at the same time, the agreements are compatible with each other, and each step results in a nonempty set of pure strategies on which candidates for reasonable conjectures are concentrated. We prove that this observation generalizes to finite games and that the iterative procedure above gives a nonempty solution set.

3   Restricting beliefs and coalitional rationalizability

Let $G = (I, S, u)$ be a normal form game, where $I = \{1, ..., n\}$ is the set of players, $S = \underset{i \in I}{\times} S_i$, is the set of strategies, and $u = \underset{i \in I}{\times} u_i$, $u_i : S \to R \ \forall \ i \in I$ are the payoff functions. We assume that S_i is finite for every $i \in I$. Let $S_{-i} = \underset{j \in I/\{i\}}{\times} S_j \ \forall \ i \in I$ and let $S_{-J} = \underset{j \in I/J}{\times} S_j \ \forall \ J \subset I$. Similarly, for a generic $s \in S$, let $s_{-i} = \underset{j \in I/\{i\}}{\times} s_j \ \forall \ i \in I$ and let $s_{-J} = \underset{j \in I/J}{\times} s_j \ \forall \ J \subset I$.

We will refer to nonempty groups of players (J such that $J \subset I$ and $J \neq \emptyset$) as *coalitions*.

We assume that players are Bayesian decision makers and we allow them to form *correlated conjectures* concerning other players' moves. Given this assumption, a strategy is a never best response if and only if it is strictly dominated, therefore from this point on we use these terms interchangeably. We note that requiring conjectures to be *independent* (to be product probability distributions over the strategy space) does not change the qualitative results in the paper.

Let Ω_{-i} be the set of probability distributions over S_{-i}, representing the set of possible conjectures player i can have concerning other players' moves. For every $J \subset I$, $i \in J$ and $f_{-i} \in \Omega_{-i}$ let f_{-i}^{-J} be the marginal distribution of f_{-i} over S_{-J}. Let Ω^{-J} be the set of probability distributions over S_{-J}, representing the set of marginal beliefs over the moves of players in $-J$.

We will compare expectations of players under different conjectures. For every $f_{-i} \in \Omega_{-i}$ and $s_i \in S_i$ let $u_i(s_i, f_{-i}) = \underset{t_{-i}:t_{-i} \in S_{-i}}{\sum} u_i(s_i, t_{-i}) \cdot f_{-i}(t_{-i})$ denote the expected payoff of player i if he has conjecture f_{-i} and plays pure strategy s_i.

Since players are Bayesian decision makers, we employ the concept of best response. For every $f_{-i} \in \Omega_{-i}$ let $BR_i(f_{-i}) = \{s_i \mid s_i \in S_i, u_i(s_i, f_{-i}) \geq u_i(t_i, f_{-i}) \ \forall \ t_i \in S_i\}$, the set of pure strategy best responses player i has against conjecture f_{-i}. For any B such that $B \subset S$, $B \neq \emptyset$ and $B = \underset{i \in I}{\times} B_i$, let

$\Omega^*_{-i}(B_i) = \{f_{-i} \mid f_{-i} \in \Omega_{-i}, \exists \, b_i \text{ such that } b_i \in B_i \text{ and } b_i \in BR_i(f_{-i})\}$. In words, $\Omega^*_{-i}(B_i)$ is the set of beliefs that player i has against which there is a best response in B_i.

Let $\widehat{u}_i(f_{-i}) = u_i(b_i, f_{-i})$ for any $b_i \in BR_i(f_{-i})$. Then $\widehat{u}_i(f_{-i})$ is the expected payoff of a player if he has conjecture f_{-i} and plays a best response to his conjecture. That means $\widehat{u}_i(f_{-i})$ is the expected payoff of a rational player if he has conjecture f_{-i}.

We will consider support restrictions on conjectures. For any A such that $A \subset S$ and $A = \underset{i \in I}{\times} A_i$, let $\Omega_{-i}(A) = \{f_{-i} \mid \text{supp} f_{-i} \subset A_{-i}\}$, the set of conjectures player i can have that are concentrated on A_{-i} (the set of conjectures which are consistent with player i believing that other players play inside A).

We assume that players coordinate on restricting their play to a subset of the strategy space whenever it is unambiguously in the interest of every player in the group to do so. The following definition identifies the restrictions which we consider to be unambigously in the interest of some group of players.

Let J be a set of players and B be such that $B \subset S$, $B \neq \emptyset$ and $B = \underset{i \in I}{\times} B_i$ (a product subset of the strategy space).

Definition: B is a *supported restriction* from S by J if

1) $B_i = S_i \; \forall \, i \notin J$, and

2) $\forall \, j \in J, f_{-j} \in \Omega^*_{-j}(S_j / B_j)$ it holds that

$\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j}) \; \forall \, g_{-j}$ such that $g_{-j} \in \Omega_{-j}(B)$ and $g^{-J}_{-i} = f^{-J}_{-i}$.

The first condition in the definition of supported restriction requires that only the strategies of those players who are members of the given group are restricted. The second condition requires that for any player in the coalition, any belief against which he has a best response strategy outside the agreement yields a strictly lower expected payoff than any belief that is consistent with other players in the coalition keeping the agreement, holding the marginal expectation concerning the strategies of players outside the coalition fixed. To state it simply, for a restriction to be supported, we require that if the restriction is not made, every player in the coalition either plays inside B, or wishes that the restriction was made, because he expects a lower payoff than any payoff he could expect if the agreement was made.

The reason we fix conjectures concerning the play of players outside the coalition on the two sides of the inequality is that since players make their moves secretly, the strategy choice of players outside the coalition cannot be made contingent on whether players in the coalition play inside B or not. The other players may or may not beleive that the restriction is made by the coalition, but they do not have a chance to find out. On the other hand, we require the condition to hold for any conjecture concerning the play of players outside the coalition. For this reason, we consider these agreements unambiguously in the interest of every player in the coalition: no matter what conjectures different members of the coalition have concerning the play of outsiders, it is always true that if they play outside B, they expect a lower payoff than the minimum payoff they could get if they all played inside the restriction.

Note that condition 2 cannot hold if there is a player j in J and a conjecture f_{-j} which is concentrated on B_{-j} and against which j has a best response in S_{-j}/B_{-j}. That means that if B is a supported restriction from S, then if a player believes that the others play inside B, then all his best responses are in B. Adopting standard terminology, we call sets that satisfy this property closed under rational behavior.

Definition: set A is *closed under rational behavior* if it satisfies the following two properties.

$$A = \underset{i \in I}{\times} A_i \ \ for \ some \ A_i \ \in S_i \tag{1}$$

$$BR_i(f_{-i}) \subset A_i \ \forall \ f_{-i} \in \Omega_{-i}(A) \ , \ \forall \ i \in I \tag{2}$$

The first property is that A is a product set. The second property requires that if players believe that other players play inside the candidate set, then all their best responses are in the candidate set[1]. Let $\mathcal{M}$ denote the collection of sets closed under rational behavior. The fact that supported restrictions from S have to be in $\mathcal{M}$ will be used in proving Claim 2, the main theorem in this chapter.

Now we make the assumption that whenever there is a supported restriction for some group of players, they expect each other to play inside the restriction. Formally, players can only have conjectures that are concentrated on supported restrictions.

Let A^1 be the intersection of all sets that are supported restrictions from S by some coalition. Then our assumption is that the conjecture of every player is concentrated on A^1.

In what follows we iterate this requirement and assume that players restrict their conjectures further. But before we do that, an immediate question we have to address is whether supported restrictions are compatible with each other. Formally, is A^1 always nonempty? Since a player is part of many different coalitions and those coalitions might have different supported restrictions, it could happen that not playing any strategy outside supported restrictions rules out every pure strategy that the player has. Claim 1 below implies that this cannot happen: the intersection of supported restrictions from S is nonempty.

Once we require players' conjectures and play to be concentrated on A^1, it is natural to ask if there are new restrictions which become unambiguously advantagous for a group of players given this more limited set of beliefs. The game in the Section 2 demonstrates that the answer is yes. This suggests the following iterative process of belief restriction. Let the intersection of all supported restrictions from A^1 be A^2 and assume the conjecture of every player

---

[1]That the set of allowable beliefs should include all conjectures on allowable action profiles is termed coherence requirement in Gul [96].

is concentrated on A^2. Define A^k for $k = 3, 4, ...$ the same way in an iterative fashion and assume that the conjecture of every player is concentrated on A^k for any $k \geq 0$, where let $A^0 = S$. In Claim 2 below we establish that every A^k ($k = 1, 2, ...$) is nonempty and there is $K \geq 0$ such that $A^k = A^K \; \forall \, k \geq 0$.

The procedure above can be thought of as a descriptive theory of belief formation. Players, based on the strategies and payoff functions of the game, look for restrictions that are mutually advantageous for some coalition. If such restrictions are found, then they expect the players in those coalitions to play inside the restrictions (to successfully coordinate their moves to play inside the restrictions). This requirement restricts the set of possible beliefs players can have. Then players look for mutually advantageous restrictions with respect to the new, restricted set of possible beliefs. If such restrictions are found, then they expect players in the corresponding coalitions to play inside the restrictions, and so on, until the set of possible beliefs cannot be constrained any further. No other beliefs can be ruled out confidently based only on the information summarized in the payoff functions.

We emphasize that, in the above interpretation players do not explicitly make agreements with each other. They simply go through a reasoning procedure, based on the commonly known payoff structure of the game and formulate their beliefs concerning the others' play according to this procedure. Explicit agreements would require pre-play communication among players, which we do not assume.

First we extend the definition of a supported restriction to cover cases when the set of possible beliefs is already restricted to a subset of S.

Let A and B be such that $A \subset S$, $A = \underset{i \in I}{\times} A_i$ and $A \neq \emptyset$, $B \subset A$, $B = \underset{i \in I}{\times} B_i$ and $B \neq \emptyset$.

Definition: B is a *supported restriction* from A by J if

1) $B_i = A_i \; \forall \, i \notin J$, and

2) $\forall \, j \in J$, $f_{-j} \in \Omega^*_{-j}(A_j/B_j)$ it holds that

$\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j}) \; \forall \, g_{-j}$ such that $g_{-j} \in \Omega_{-j}(B)$ and $g^{-J}_{-i} = f^{-J}_{-i}$.

Let $F_J(B)$ be the set of supported restrictions from B by J and let $F(B) = \{C : C \in F_J(B)$ for some $J \subset I\}$.

Definition: let $A^0 = S$ and let $A^k = \underset{B \in F(A^{k-1})}{\cap} B$ whenever $k \geq 1$. The decreasing sequence of sets A^0, A^1, A^2, ... represents *iterated deletion of coalitionally dominated strategies*.

Below we show that $A^\infty = \underset{k=0,1,2,...}{\cap} A^k$ is well-defined and nonempty.

To see an example of how the procedure works, consider the game of Figure 4.

	C1					C2			
	B1	B2	B3	B4		B1	B2	B3	B4
A1	4,4,3	5,3,4	9,2,0	0,0,4	A1	2,2,2	2,4,1	10,1,10	0,0,4
A2	3,2,1	3,10,3	10,1,0	0,0,4	A2	3,9,2	2,2,1	9,2,9	0,0,4
A3	0,0,4	0,0,4	0,0,4	2,1,4	A3	0,0,4	0,0,4	0,0,4	1,0,5

Figure 4

This game has multiple Nash equilibria, none of which is Pareto dominant. The equilibrium most preferred by player 1 is the one in which (A1,B1,C1) is played with probability 1. The equilibrium most preferred by player 2 is the one in which (A2,B1,C2) is played with probability 1. And the equilibrium most preferred by player 3 is the one in which (A3,B4,C2) is played with probability 1. It is not immediately obvious which outcome should be expected to occur, but as we show below, using coalitional rationality provides a sharp prediction.

From A^0, $J^0 = \{1,2\}$ has a nontrivial restriction, namely $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\} \equiv B^0$. $B3$ is a best responses for player 2 only if player 1 plays $A3$ with probability 1. But in that case player 2's payoff cannot exceed 1, while any outcome in B^0 gives him at least 2. $B4$ gives at most a payoff of 1 to player 2. For player 1, $A3$ gives a payoff of at most $1 + \gamma$, where γ is the probability of player 3 playing $C1$. And for any given γ, the minimum payoff player 1 can get if play is in B^0 is $2 + \gamma$.

It is straightforward to check that there is no more supported restriction from A^0. For example consider the coalition $\{1,3\}$. Player 3's most optimistic conjecture allocates probability 1 to player 1 playing A1 and player 2 playing B3. Against this conjecture C2 is a best response strategy (in fact the only one). Therefore any supported restriction should include C2, otherwise the restriction cannot satisfy the requirement to be supported. But if C2 is part of the restriction, then player 1 might have a conjecture that player 2 plays B4 and player 3 plays C2. Against this conjecture all of his strategies are best responses. Therefore any supported restriction by $\{1,3\}$ should include all strategies of player 1. But if the restriction does not rule out any of player 1's strategies, then player 3 can have a conjecture that player 1 plays A1 with probability 1 and player 2 plays B1 with probability 1. Against this conjecture C1 is the best response. This establishes that no supported restriction by $\{1,3\}$ can rule out any of player 3's strategies either. Verifying that every supported restriction from S by $\{1,2\}$ has to include all strategies in B^0 and that the only supported restriction from S is S itself for any other coalition can be done analogously. This concludes that $A^1 = B^0$.

Similar considerations establish that $\{A1\} \times \{B1, B2\} \times \{C1\} \equiv B^1$ is a supported restriction from by $J^1 = \{1,3\}$ and that there is no other nontrivial supported restriction from A^1. Therefore $A^2 = B^1$. And from A^2, only the singleton coalition $\{2\}$ has a nontrivial supported restriction, $\{A1\} \times \{B1\} \times \{C1\} \equiv B^2$. Therefore $A^k = B^2 \ \forall \ k \geq 2$ and so $A^\infty = B^2$.

Now we establish that in any game the intersection of all supported restrictions from a set that is coherent under rational behavior is nonempty, which is the main step in proving Claim 2, the central result of this section.

Claim 1: let A be such that $A \in \mathcal{M}$. Then $\underset{B:\ B \in F(A)}{\cap} B \neq \emptyset$.

proof: in the appendix.

We are ready to state the main claim of the section. It establishes that the iterative deletion of coalitionally dominated strategies stops in finite steps, and the set it obtains is nonempty, coherent under rational behavior and has the property that there is no nontrivial supported restriction from it by any coalition.

Claim 2: A^{∞} is nonempty, $\exists\, K < \infty$ such that $A^k = A^{\infty}$ whenever $k \geq K$, $A^{\infty} \in \mathcal{M}$ and there is no nontrivial supported restriction from A^{∞} by any coalition.

proof: in the appendix.

Now that we established that A^{∞} is well-defined and nonempty, we define it as the set of coalitionally rationalizable strategies.

Definition: the *set of coalitionally rationalizable strategies* is the limit set of the iterated deletion of coalitionally dominated strategies, A^{∞}.

We remark here that nonemptyness of the set of coalitionally rationalizable strategies and that there is no nontrivial supported restriction from it generalize to games with compact strategy spaces and continuous payoff functions, although the argument we use is necessarily more complicated, because the iterative procedure does not have to stop after a finite number of steps. We discuss this extension in Section 7.

Another remark that we make here is that although the iterative procedure and the set of coalitionally rationalizable strategies are defined on pure strategies[2], they are compatible with allowing players to use mixed strategies (analogously to rationalizability). In Section 7 we show that if we define the procedure directly on mixed strategies, we get the same solution set, A^{∞}.

In the remainder of this section we prove that coalitional rationalizability is a refinement of rationalizability and that the iterative process that defines it is insensitive to the order in which restrictions are made. Furthermore we show that the solution set can be defined without referring to any iterative process. Despite the previous two results, if coalitions make supported restrictions in an iterative manner, one might be concerned whether an agreement necessarily remains advantageous for a group of players if it might lead to further restrictions. We conclude the section by addressing those concerns and argueing that in a context where players make their moves secretly and independently of each other, a supported restriction for a coalition is unambiguously in the interest of every player in the coalition.

Now we prove that A^{∞} is a refinement of rationalizability. First we state a lemma on supported restrictions made by single-player coalitions and introduce the concept of closedness under rational behavior.

---

[2]but the players' conjectures not. We do not require players conjectures to be point-conjectures, they are defined to be probability distributions on other players' strategies.

Lemma 4: let A be such that $A \in \mathcal{M}$. Let $i \in I$. B is a supported restriction by $\{i\}$ from A iff $B = B_i \times A_{-i}$, $B_i \subset A_i$ and $s_i \in B_i/A_i$ implies that there is no $f_{-i} \in \Omega_{-i}(A)$ such that $s_i \in BR_i(f_{-i})$.

proof: follows from the fact that for a single-player coalition $\{i\}$, requirement 2 in the definition of supported restriction is equivalent to requiring that there are no s_i and f_{-i} such that $s_i \in B_i/A_i$, $f_{-i} \in \Omega_{-i}(A)$ and $s_i \in BR_i(f_{-i})$. QED

Lemma 4 implies that the intersection of the sets that are supported restrictions for a single-player coalition from A^k ($k \geq 0$) is the set of strategies of that player that are not strictly dominated on A^k. This further implies that A^{k+1} does not contain the strategies that are strictly dominated on A^k.

Definition: set A is *closed under rational behavior* if it is coherent under rational behavior (satisfies (1) and (2) above) and satisfies:

$$\underset{f_{-i}\in\Omega_{-i}(A)}{\cup} BR_i(f_{-i}) = A_i \; \forall \, i \in I \tag{3}$$

Claim 2 implies that, in particular, there is no supported restriction from A^∞ by any single-player coalition. Lemma 4 then implies that A^∞ satisfies (3). Furthermore, Claim 2 establishes that $A^\infty \in \mathcal{M}$. Combining these results establishes that A^∞ is closed under rational behavior.

Let R be the set of rationalizable strategies.

Claim 3: $A^\infty \subset R$.

proof: follows from the fact that A^∞ is closed under rational behavior, since R is the largest set that is closed under rational behavior (see Bernheim[84]). QED

Claim 3 establishes that the set of coalitionally rationalizable strategies is a refinement of rationalizability. This is not a surprising result in the light of Lemma 4, which says that at any stage of the iterative procedure all strategies are deleted that are not best responses.

Going back to Lemma 3, it states that a supported restriction from a set which is coherent under rational behavior remains a supported restriction even if other supported restrictions are made in the meantime. This result, besides providing internal consistency to iterated deletion of coalitionally dominated strategies, is critical in proving the following result. Note that we defined the set of coalitionally rationalizable strategies as the set obtained by an iterative procedure that requires supported restrictions to be made in a particular order, namely first making all supported restrictions from S simultaneously, then making all supported restrictions from the resulting set simultaneously and so on. Now we prove that the particular order does not matter. Any iterative procedure that makes some nontrivial supported restriction whenever one exists (for example just making one restriction at a time, in any possible order) would yield the same solution set, A^∞.

Claim 4: let $B^0 = S$. If there is no nontrivial supported restriction from B^0, then let $B^1 = B^0$. Otherwise let Θ^0 be a nonempty collection of nontrivial

14

restrictions from B^0 and let $B^1 = \underset{B:B\in\Theta^0}{\cap} B$. In a similar fashion once B^k is defined for some $k \geq 1$, let $B^{k+1} = B^k$ if there is no nontrivial supported restriction from B^k, otherwise let Θ^k be a nonempty collection of nontrivial restrictions from B^k and let $B^{k+1} = \underset{B:B\in\Theta^k}{\cap} B$. Then there is $L \geq 0$ such that $B^k = A^\infty \ \forall \ k \geq K$.

proof: in the appendix.

This result establishes that iterated deletion of coalitionally dominated strategies is similar to iterated deletion of strictly dominated strategies in the sense that *the order of deletions does not matter.*

It is possible to define the set of coalitionally rationalizable strategies without referring to an iterative procedure, using two properties of the set proved in the next two lemmas.

Definition: a set A^* is externally coalitionally stable if $A^* \subset A$ and $A \neq A^*$ imply $A^* \subset \underset{B:\ B\in F(A)}{\cap} B$ and $A \neq \underset{B:\ B\in F(A)}{\cap} B$.

Definition: a set A^* is internally coalitionally stable if there is no nontrivial supported restriction from A^*.

Definition: a set A^* is coalitionally stable if it is both externally and internally coalitionally stable.

Intuitively, coalitional stability of A requires that whenever we start out from a set larger than A, supported restrictions restrict that set "towards A", while A itself cannot be restricted further.

Lemma 5: let A be such that $A^\infty \subset A$ and $A \neq A^\infty$. Then $A^\infty \subset \underset{B:\ B\in F(A)}{\cap} B$.

proof: in the appendix.

Lemma 6: let A be such that $A/A^\infty \neq \emptyset$ and $A\cap A^\infty \neq \emptyset$. Then $\underset{B:\ B\in F(A)}{\cap} B \neq A$.

proof: in the appendix.

Note that in particular Lemma 6 implies that if A is such that $A^\infty \subset A$ and $A \neq A^\infty$, then $A \neq \underset{B:\ B\in F(A)}{\cap} B$. This together with Lemma 5 implies that an iterative procedure of supported restrictions goes to A^∞ from any set which is larger than A^∞, not just from S.

Claim 5: A^∞ is the only coalitionally stable set in G.

proof: in the appendix.

Claim 5 provides a direct characterization of the set of coalitionally rationalizable strategies. It is the only subset of the strategy space that has both external and internal coalitional stability.

We conclude the section by analyzing the incentives players have to make supported restrictions and pointing out that coalitional rationality can make every player worse off, just like rationality can in cases like the prisoners' dilemma.

Above we established that supported restrictions and in particular playing inside the set of coalitionally rationalizable strategies (which is equivalent to making a sequnce of supported restrictions) are self-enforcing from an individual player's point of view in the sense that if a player assumes that a supported restriction is made, then all his best responses are inside the set, so he plays inside the set. Furthermore we established that a supported restriction remains supported after other restrictions occur in the game, so assuming that supported restrictions are made in a setting in which there is a sequence of restrictions is internally consistent. Finally, we established that the iterative procedure of restricting beliefs is robust in the sense that the order in which the restrictions are made does not matter.

These results may increase confidence in beleiving in the prediction that play is inside the set of coalitionally rationalizable strategies *if players think in terms of supported restrictions*, but if players are aware of the fact that making a restriction might cause other coalitions to make other restrictions, the question arises whether a supported restriction is always unambiguously in the interest of the players in the given coalition.Consider the game in Figure 5.

	B1	B2	B3		B1	B2	B3
A1	4,4,4	9,3,0	1,1,0		5,2,0	4,2,1	0,0,0
A2	1,0,0	1,2,2	2,3,0		0,0,0	1,1,0	1,1,5

Figure 5

Here $\{A1\} \times \{B1, B2\} \times \{C1, C2\}$ is a supported restriction from S by $\{1,2\}$, and then $\{A1\} \times \{B1\} \times \{C1\}$ is a supported restriction from $\{A1\} \times \{B1, B2\} \times \{C1, C2\}$ by $\{2,3\}$, which makes $\{A1\} \times \{B1\} \times \{C1\}$ the only coalitionally rationalizable outcome in the game. In the light of this the first restriction does not seem to be unambiguously beneficial for player 1, since he could foresee that once the agreement is made, another restriction follows which results in losing the outcome with the highest payoff for him, $(A1, B2, C1)$. Maybe player 1 would not like to play $A2$, but in the meantime would like to maintain uncertainty concerning his move and make player 2 believe that he plays $A2$ with some probability, so that player 2 plays $B2$. We take the position that this plan is not feasible for player 1. Restricting play to $\{A1\} \times \{B1, B2\} \times \{C1, C2\}$ gives him a higher payoff than any payoff he could get when playing $A2$, so a "threat" by player 1 of possibly playing $A2$ is not credible if he has the option of making the agreement. If other players know that player 1 considers the possibility of coalitional restrictions, player 1 cannot maintain the possibility that he plays $A2$ with positive probability. We draw a paralell with iterative deletion of strictly dominated strategies to make this point clearer. Consider the game in Figure 6.

B1   B2

	B1	B2
A1	1,1	4,0
A2	0,0	3,2

Figure 6

Here a rational player 1 never plays $A2$. If rationality is common knowledge, then player 2 knows this and plays $B1$. Note that if it was possible, player 1 would like to make player 2 believe that he plans to play $A2$, because then a rational player 2 would play $B2$, which is clearly beneficial for player 1. But common knowledge of rationality implies that it is not credible that player 1 plans to play $A2$. It is a similar reason why in our model in the game of Figure 5 player 1 cannot credibly commit to not playing $A1$.

The fact that groups of players cannot credibly commit to not going for common gains can make all of them worse off though, as the game of Figure 7 shows.

	C1 B1	C1 B2			C2 B1	C2 B2
A1	2,2,2	0,0,0		A1	0,0,0	0,0,0
A2	0,0,0	3,3,0		A2	0,0,0	1,1,1

Figure 7

In this game the Nash equilibrium outcome $(A1, B1, C1)$ is not coalitionally rationalizable because no matter what player 3 does, playing $(A2, B2)$ always gives the highest payoff for players 1 and 2. But then player 3 is better off playing $C2$ and the players end up playing $(A2, B2, C2)$, which is strictly Pareto-dominated by $(A1, B1, C1)$. One could then argue that players 1 and 2 should not make the supported restriction $\{A2\} \times \{B2\} \times \{C1, C2\}$, to make it "more likely" that player 3 plays $C1$. But even if this argument convinces player 3 to play $C1$, players 1 and 2 have the incentive to play $(A2, B2)$ if players make their moves secretly and independently of each other.

The above point is made even starker in the next example, which we call the coalitional prisoner's dilemma.

	C1 B1	C1 B2			C2 B1	C2 B2
A1	2,2,2	0,0,0		A1	0,0,0	0,3,3
A2	0,0,0	3,3,0		A2	3,0,3	1,1,1

Figure 8

In this game no single player can profitably deviate from the $(A1, B1, C1)$ outcome, but every two-player coalition can. And indeed $\{A2\} \times \{B2\} \times \{C1, C2\}$ is a supported restriction from S by $\{1,2\}$, $\{A2\} \times \{B1, B2\} \times \{C2\}$ is

a supported restriction from S by $\{1,3\}$ and $\{A1, A2\} \times \{B2\} \times \{C2\}$ is a supported restriction from S by $\{2,3\}$, making $(A2, B2, C2)$ the only coalitionally rationalizable outcome of the game.

As these examples demonstrate, coalitional rationality can destroy efficient equilibria. But in special classes of games it promotes efficiency. In section 4 we prove that in two-player games every Pareto-undominated Nash equilibrium is in the set of coalitionally rationalizable strategies, and in section 5 we prove that in games in which there is a Pareto-dominant outcome on the set of rationalizable strategies, that outcome is the only coalitionally rationalizable outcome.

4 Relating coalitional rationalizability to other solution concepts

We examine the connection between the set of coalitionally rationalizable strategies and some standard equilibrium notions in normal-form games: the sets of Nash equilibria, Pareto-undominated Nash equilibria, coalition-proof Nash equilibria and strong Nash equilibria.

Our concept is not an equilibrium concept, so it is not surprising that the set of coalitionally rationalizable strategies is not contained in the set of Nash-equilibrium outcomes (the outcomes that can be realizations of some mixed strategy Nash-equilibrium). Consider the game of Figure 9.

	B1	B2	B3
A1	-2,1	-1,0	1,-2
A2	0,-1	0,0	0,-1
A3	1,-2	-1,0	-2,1

Figure 9

The only Nash equilibrium of the game is $(A2, B2)$. Nevertheless, the set of coalitionally rationalizable outcomes is the whole game.

	B1	B2
A1	1,1	-2,-2
A2	-1,-1	2,2

Figure 10

Furthermore, the set of Nash equilibria is not contained in the set of coalitionally rationalizable strategies, as we can see in the 2-player game of Figure

10. There $(A1, B1)$ is a Nash equilibrium outcome, but not coalitionally rationalizable. The game in Figure 11 demonstrates that in games with more than two players, even a strategy on the Pareto frontier of the Nash equilibrium set (the strategy $(A2, B2, C2)$) does not have to be in the set of coalitionally rationalizable strategies.

	C1			C2	
	B1	B2		B1	B2
A1	4,4,0	0,0,1	A1	3,3,1	0,2,0
A2	3,0,1	2,2,0	A2	2,0,3	2,2,3

Figure 11

However, below we provide two positive results concerning the inclusion of Nash equilibria in the set of coalitionally rationalizable strategies. The first one establishes that there is at least one Nash equilibrium of the game which is inside the set of coalitionally rationalizable strategies.

Claim 6: every normal form game has a Nash equilibrium such that every outcome in its support belongs to the set of coalitionally rationalizable strategies.

proof: the set of coalitionally rationalizable strategies yields a finite normal form game, which has a Nash equilibrium in mixed strategies. In this profile every player plays best response from his set of coalitionally rationalizable strategies against the other players' strategy profile. Then every pure strategy played in these mixed strategies with positive probability is a best response against the conjecture which corresponds to other players' strategy profile. But since the set of coalitionally rationalizable strategies is closed under rational behavior (Claim 2), this means that every player plays a best response from his whole strategy set against the other players' strategy profile. QED

The second result is that in 2-player games all Pareto-undominated Nash equilibria are contained in the set of coalitionally rationalizable strategies.

Claim 7: for every Pareto-undominated mixed strategy equilibria $\sigma = (\sigma_1, \sigma_2)$ of every 2-player game, $\text{supp}\sigma \subset A^\infty$.

proof: in the appendix.

Next we investigate how the set of coalitionally rationalizable strategies is related to coalition-proof equilibrium. There are lot of games which do not have coalition-proof equilibria. The next example demonstrates that coalitional rationality can play a role in these games, in the sense that the set of coalitionally rationalizable strategies is strictly smaller than the set of rationalizable strategies.

We consider a dollar division game with an external award in case the players behave "nicely". Three players vote secretly and simultaneously how to divide a dollar among each other. If two or more players vote for the same allocation, the dollar is divided accordingly, otherwise the dollar is lost to the players. The additional twist is that if every player votes for allocations that would give all players in the game at least a quarter dollar, then every player gets an additional

$2 reward for the group being "generous", independently of what happens to the original dollar (so even if it is not allocated to the players, because of lack of agreement). In this game coordinating on voting for allocations that give at least 1/4 dollar to every player is unambiguously mutually advantegous for the players. It is not clear how the original dollar should be divided, or whether it is reasonable to expect two or more players to vote for the same division, and if yes then which players vote for the winning allocation. There is a conflict of interest among players how to allocate the "last quarter", but they have a strong incentive to propose at least 1/4 dollar to every player. This intuition is captured by coalitional rationalizability. Proposing only allocations (x_1, x_2, x_3) such that $x_i \geq 1/4 \; \forall \; i \in \{1, 2, 3\}$ is a supported restriction for the coalition of all players from S, and it is the set of coalitionally rationalizable strategies. The game doesnot have any coalition-proof equilibrium.

The next question we address is whether or not the set of outcomes consistent with some coalitioon-proof equilibrium is contained in the set of coalitionally rationalizable strategies. The following example demonstrates that the answer is no.

	C1 B1	C1 B2	C1 B3		C2 B1	C2 B2	C2 B3
A1	2,1,0	0,0,0	-9,-9,-9	A1	1,2,0	0,2,1	-9,-9,-9
A2	2,0,1	1,0,2	-9,-9,-9	A2	0,0,0	0,1,2	-9,-9,-9
A3	-9,-9,-9	-9,-9,-9	-9,-9,-9	A3	-9,-9,-9	-9,-9,-9	-9,-9,-9

	C3 B1	C3 B2	C3 B3
A1	-9,-9,-9	-9,-9,-9	-9,-9,-9
A2	-9,-9,-9	-9,-9,-9	-9,-9,-9
A3	-9,-9,-9	-9,-9,-9	-9,-9,-9

Figure 12

In this game $(A3, B3, C3)$ is the unique coalition-proof equilibrium (on mixed strategies). It is straightforward to establish that there is no self-enforcing profile in which no player plays his third strategy with positive probability, by showing that from every such Nash equilibrium there is a two-player coalition that can profitably deviate to another Nash equilibrium. There is no Nash equilibrium (and therefore there is no self-enforcing profile) in which at most one player plays his third strategy with probability 1, but at least one player plays his third strategy with positive probability, since then there is a player who plays his third strategy with positive probability despite the fact that other two players both play at least one of their first two strategies with positive probability ($A3$,$B3$ and $C3$ cannot be a best response for player 1, 2 and 3 respectively if both of the other players play at least one of their first two strategies with

positive probability). And it is straightforward to show that there is no self-enforcing profile in which two players play their third strategies with probability 1 and the third player does not, because then the first two players have a joint deviation from which neither of them could deviate further profitably. So the only candidate even in mixed strategies for a self-enforcing profile is when player 1 plays $A3$ with probability 1, player 2 plays $B3$ with probability 1 and player 3 plays $C3$ with probability 1. And it is a self-enforcing profile, because no single-player or two-player coalition can have a profitable deviation, and the coalition of all three players does not have a self-enforcing deviation, because there is no other self-enforcing profile in the game. Since $(A3, B3, C3)$ is the only self-enforcing profile in the game, it is the unique coalition-proof equilibrium.

Furthermore, the set of coalitionally rationalizable outcomes is $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ (this is the set we get after the first round of the iterative procedure and then there is no more nontrivial supported restriction), so $(A3, B3, C3)$ is not coalitionally rationalizable. In fact the set of coalitionally rationalizable strategies and the set of coalition-proof equilibria are disjunct sets in this game.

Note that $(A3, B3, C3)$ is a coalition-proof equilibrium only because the subgame $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ does not have a coalition-proof equilibrium. But note that all players would strictly prefer to switch play to the subgame, no matter what happens over there (they cannot agree upon a concrete profile, but they would all agree to restricting their moves to subgame $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$). We think that the prediction that players will play something inside $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ is a more reasonable than predicting that players will play the outcome $(A3, B3, C3)$. So this example suggests that the coalition-proof equilibrium concept can support non-intuitive outcomes, driven by the fact that the original game might have a restriction which does not have a coalition-proof equilibrium.

The above game could be made generic by some small perturbation of the payoffs, in a way that the set of coalitionally rationalizable strategies and the set of coalition-proof equilibria remain the same. Thus, even in generic games a coalition-proof equilibrium might not be coalitionally rationalizable.

The following claim establishes that every *strong Nash equilibrium* of the original game must be contained in the set of coalitionally rationalizable strategies.

Claim 8: let $\sigma = (\sigma_1, ..., \sigma_I)$ be a strong Nash equilibrium profile. Then supp$\sigma \subset A^\infty$.

proof: in the appendix.

5 Coalitional rationalizability in special classes of games

In generic pure coordination games (games of pure common interest), like the game of Figure 10 in the previous section, the only coalitionally rationalizable outcome is the Pareto-dominant outcome, since it is a supported restriction from S by I, the coalition of all players.

This observation can be generalized to a larger class of games. Let R be the set of rationalizable strategies.

Claim 9: if the set of rationalizable strategies in G has a unique Pareto-dominant outcome S, then that outcome is the only coalitionally rationalizable outcome in the game.

proof: in the appendix.

In particular coalitional rationalizability predicts the Pareto-dominant outcome in the game in Figure 13.

$$\begin{array}{cc} & \text{B1} \quad \text{B2} \end{array}$$

	B1	B2
A1	3,3	0,0
A2	2,0	1,1

Figure 13

In this game it is always in player 1's interest that player 2 plays $B1$, even if he plays $A2$. So if there is a round of cheap talk before the game and player 1 proposes to play $(A1, B1)$, this proposal might not be credible, which might make player 1 to play $A2$. Similarly, without cheap talk, if players contemplate playing $(A1, B1)$, these expectations might break down if players go through the above thought procedure. This comment, originally made by Robert Aumann, appears in Farrell (88). Coalitional rationalizability implies that players do not get engaged into self-fulfilling pessimistic expectations like the one above. They expect each other to look for mutually advantageous restrictions, and if they find one, like $(A1, B1)$ in this game, then they expect each other to play inside that restriction.

In games in which players have opposite interest, not surprisingly only singular coalitions have an effective role and the set of coalitionally rationalizable strategies is the same as the set of rationalizable strategies.

Claim 10: If G is a 2-person 0-sum game, then $A^\infty = R$.

proof: let $\widehat{s}_1$, $\widehat{s}_2$ be minmax strategies for players 1 and 2 respectively. Note that $R \subset A^0$. Suppose $R \subset A^k$ for some $k \geq 0$. Since strategies in R are best

responses against some conjecture concentrated on R, they are best responses against a conjecture concentrated on A^k. Therefore no single-player coalition has a restriction B from A^k such that $B \not\subseteq R$. Now assume there is B such that B is a supported restriction from A^k by {1,2} and there are i and a_i such that $i \in \{1,2\}$, $a_i \in R$ and $a_i \notin B_i$. Since $a_i \in R$, there is f_{-i} such that $f_{-i} \in \Omega_{-i}(A^k)$ and $a_i \in BR_i(f_{-i})$. Since $u_i(\widehat{s}_1, \widehat{s}_2)$ is the minmax value for i, $u_i(a_i, f_{-i}) \geq u_i(\widehat{s}_1, \widehat{s}_2)$. But then the fact that B is a supported restriction from A^k by {1,2} implies $u_i(b_i, g_{-i}) > u_i(\widehat{s}_1, \widehat{s}_2) \; \forall \; b_i, \; g_{-i}$ such that $g_{-i} \in \Omega_{-i}(A^k)$ and $b_i \in BR_i(g_{-i})$. Note that the game that has players I, strategy sets B_i $\forall \; i \in I$ and payoff functions $\widehat{u}_i$ such that $\widehat{u}_i(s) = u_i(s) \; \forall \; s \in B$ has a Nash equilibrium on mixed strategies. Denote this profile by σ. Since $B \in \mathcal{M}$, σ is a Nash equilibrium in G, too. By the above inequality, $u_i(\sigma) > u_i(\widehat{s}_1, \widehat{s}_2)$. Since G is 0-sum, this implies $u_{3-i}(\sigma) < u_{3-i}(\widehat{s}_1, \widehat{s}_2)$. But that contradicts that B is a supported restriction from A^k by {1,2}, since $\widehat{s}_{3-i}$ is a best response against $\widehat{s}_i$. This establishes that $R \subset A^{k+1}$. By induction $R \subset A^k \; \forall \; k \geq 0$, so $R \subset A^\infty$. Since A^∞ is closed under rational behavior, $A^\infty \subset R$, so $A^\infty = R$. QED

6   Perfect coalitional rationalizability

Coalitional rationalizability requires a certain amount of confidence on the players' side in that other players use the logic behind coalitional rationality to rule out conjectures. In a lot of settings assuming that players have complete certainty in each others' reasoning procedure is too restrictive. Even if players are sure of each others' intentions, they might think that there is a small probability that the other players make mistakes when choosing their actions. This motivates us to analyze how the prediction of our model changes if we impose a certain amount of cautiousness on players when they choose actions.

We capture this motive a standard way, by requiring players' beliefs to have full support. Besides cautiousness, potential gains might motivate coalitions to make agreements based on the assumption that anything can happen at least with a small probability. Consider the game of Figure 14.

	B1	B2	B3
A1	4,4	2,2	2,2
A2	2,2	4,4	2,2
A3	2,2	2,2	2,2

Figure 14

In this game $\{A1, A2\} \times \{B1, B2\}$ is not a superior response for {1,2}. On the other hand, when players can only have beliefs with full support, they always expect a payoff more than 2 when they play according to the agreement, while $A3$ and $B3$ can only give a payoff of 2 to player 1 and 2 respectively.

The way we modify our concept to incorporate the motive of caution is by requiring players to have beliefs with full support and by requiring them to

allocate high probability in their conjectures to agreements which are mutually advantageous for a coalition whenever every member of the coalition beleives in the agreement with high enough probability.

Let $\overline{\Omega}_{-i}$ be the set of probability distributions over S_{-i} with full support, corresponding to the set of conjectures for player i that allocate positive probability to every profile of strategies played by the other players. For $p \in (0,1)$ and $A = \underset{i \in I}{\times} A_i$ let $\Omega^p_{-i}(A)$ be the set of probability distributions in $\overline{\Omega}_{-i}$ which allocate at least probability p to outcomes in A_{-i} : $\Omega^p_{-i}(A) = \{f_{-i} \mid f_{-i} \in \overline{\Omega}_{-i}, \sum_{s_{-i} \in A_{-i}} f_{-i}(s_{-i}) \geq p\}$.

For $p \in (0,1)$, A is coherent under level-p rational behavior if it satisfies (1) and the following modification of (2):

$$BR_i(f_{-i}) \subset A_i \; \forall \; f_{-i} \in \Omega^p_{-i}(A) \; , \; \forall \; i \in I \tag{4}$$

Let $\mathcal{M}'(p)$ denote the sets coherent under level-p rational behavior.

Let $A \subset S$, $A = \underset{i \in I}{\times} A_i$ and $A \neq \emptyset$.

Definition: B is a *level–p supported restriction* from A by J if

1) $B_i = A_i \; \forall \; i \notin J$, and

2) $a_j \in BR_j(f_{-j})$ implies

$u_j(a_j, f_{-j}) < \underset{g_{-j}:g_{-j} \in \Omega^p_{-j}(B) \text{ and } \widehat{g}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J}) \; \forall \; s_{-J} \in S_{-J}}{\min} \; \underset{b_j \in S_j}{\max} u_j(b_j, g_{-j}) \; \forall \; j,$

a_j and f_{-j} such that $j \in J$, $a_j \in S_j/B_j$ and $f_{-j} \in \Omega^p_{-j}(A)$.

The difference between this concept and supported restriction is that here the starting set of conjectures is not the conjectures concentrated on some set A, but conjectures which have full support and allocate at least probability p to outcomes in A. Similarly, the restrictions players consider are not restricting conjectures to ones having support inside set B, but to ones that have full support and allocate at least probability p to outcomes in B.

Let $F^p_J(A)$ be the set of level-p supported restrictions from A by J and let $F^p(A) = \{B : B \in F^p_J(A) \text{ for some } J \subset I\}$.

Now we can define an iterative procedure which is analogous to iterative deletion of coalitionally dominated strategies.

For any $p \in (0,1)$ let $A^0(p) = S$ and let $A^k(p) = \underset{B \in F^p(A^{k-1})}{\cap} B$ whenever $k \geq 1$. Let $A^\infty(p) = \underset{k \to \infty}{\lim} A^k(p)$.

Claim 11: for every $p \in (0,1)$, $A^\infty(p)$ is nonempty and closed, there is no nontrivial level-p supported restriction from $A^\infty(p)$ and $\exists \; K < \infty$ such that $A^k(p) = A^\infty(p)$ whenever $k \geq K$.

proof: similar to Claim 2 in section 3, therefore omitted.

Now we claim that the set of level-p coalitionally rationalizable strategies is decreasing in p and we use that to prove that there is a set of strategies which is level-p coalitionally rationalizable for every $p \in (0,1)$ or alternatively, for p

24

arbitrarily close to 1, which is the case we are really interested in. We call this set the set of perfectly coalitionally rationalizable strategies.

Claim 12: $A^\infty(p)$ is decreasing in p and $\underset{p:p\in(0,1)}{\cap} A^\infty(p) \neq \emptyset$.

proof: in the appendix.

Definition: the set of perfectly coalitionally rationalizable strategies is $\underset{p\to1}{\lim} A^\infty(p) = \underset{p:p\in(0,1)}{\cap} A^\infty(p)$.

The set of perfectly coalitionally rationalizable strategies consists of all that are level-p coalitionally rationalizable for p arbitrarily close to 1.

The above definition of perfect coalitional rationality is not directly useful. In finite games we could obtain it by iterated deletion of strategies that are not level-p coalitional best responses for p close enough to 1, but we do not know what is "close enough to 1". Fortunately we can provide an iterative procedure which gets around this problem and does not use the concept of level-p coalitional rationality. The procedure is similar to the Dekel-Fudenberg iterative procedure (see Dekel and Fudenberg (90), and also Borgers (93), Herings and Vannetelbosch (00)). It requires one round of deletion of strategies that are never best responses against a belief with full support (deletion of weakly dominated strategies) and then iterated deletion of strategies that are not coalitional best responses, as defined in the previous section.

Now consider the following iterative procedure. Let P^0 be the set of strategies that are not weakly dominated: $P^0 \equiv \underset{i=1}{\times} P_i^0$, where $P_i^0 = \{s_i \mid s_i \in S_i$ and $s_i \in BR_i(\Omega_{-i})$ for some $\theta_{-i} \in \overline{\Omega}_{-i}\}$. For $k \geq 1$ let $P^k = \underset{B\in F(P^{k-1})}{\cap} B$. Finally let $P^\infty = \underset{k\to\infty}{\lim} P^k$.

We claim that P^∞ is the set of perfectly coalitionally rationalizable strategies.

Lemma 9: B is a level-p supported restriction from A by J for some $p \in (0,1)$ iff B is a supported restriction from $C \equiv B \cup (A \cap P^0)$ by J.

proof: in the appendix.

Claim 13: $s \in P^\infty$ iff s is perfectly coalitionally rationalizable.
proof: in the appendix.

Here we note that allowing for correlated beliefs is important for Claim 13. In Figure 15 there is an example showing that for uncorrelated beliefs the iterative procedure defined above yields a different set than the set of perfectly coalitionally rationalizable strategies.

	C1			C2	
	B1	B2		B1	B2
A1	1,0,0	0,0,0	A1	-1,0,-1	0,0,-1
A2	0,0,0	0,0,0	A2	0,0,-1	0,0,-1

Figure 15

In this game $P^0 = \{A1, A2\} \times \{B1, B2\} \times \{C1\}$. In particular A2 is not weakly dominated in S. There is no nontrivial supported restriction from P^0,

so $P^\infty = P^0$. But we claim that $A2$ is not perfectly coalitionally rationalizable for player 1 if beliefs are constrained to be uncorrelated. It is because p-level coalitional rationalizability imples that player 1 should believe it with at least probability p that player 3 plays C1 and then uncorrelated beliefs imply that for high enough p $A1$ yields positive payoff for him, while $A2$ yields 0 (note that correlated beliefs do not imply that $A1$ yields positive payoff).

One could propose other refinements of coalitional rationalizability which incorporate the motive of caution. In particular one could define perfect coalitional rationalizability along the lines of perfect rationalizability, as proposed in Bernheim (84). This would involve obtaining perfectly coalitionally rationalizable strategies, defined as strategies belonging to the limit set of a sequence of sets of coalitionally rationalizable strategies of perturbed games, with the perturbation going to zero along the sequence. This construction implicitly assumes that the perturbations (or probabilities of players making mistakes) are commonly known, which we find less plausible than the assumptions required for our concept.

Another possibility would be to construct cautious coalitional rationalizability along the lines of cautious rationalizability, as proposed in Pearce (84). This would impose the condition that the players' conjectures give positive weight to every coalitionally rationalizable profile of the others and zero weight to profiles which are not coalitionally rationalizable. It would lead to a procedure starting with the iterated deletion of strategies which are not coalitional best responses, then a round of elimination of the weakly dominated strategies in the remaining game, then again a round of iterated deletion of strategies that are not coalitional best responses and so on. Again, as far as the stories we have in mind (like players making mistakes with small probability), we find the assumptions behind this concept less plausible than our assumptions. If we introduce the motive of caution, then we should require players to allocate positive weights to every possibility in their conjectures, not just to the possibilities consistent with the solution concept.

We saw in section 3 that coalitional rationalizability is a refinement of rationalizability. In finite games perfect coalitional rationalizability is a refinement of τ-perfect rationalizability, as defined in Gul (96), which is also called weak perfect rationalizability in the literature (Herings and Vannetelbosch (99) and (00)). The easiest way to see this is to note that the iterative procedure in the definition of the set of perfectly coalitionally rationalizable strategies deletes all the strategies deleted in the Dekel-Fudenberg iterative procedure, and the strategies that are τ-perfect rationalizable are exactly the ones that survive the Dekel-Fudenberg procedure.

Perfect coalitional rationalizability can be incorporated into the framework called τ-theories proposed by Gul. In this framework there are two types of players, rational and irrational and it is common knowledge that at least $1 - \varepsilon$ fraction of the players are rational for some $\varepsilon > 0$. Furthermore it is common belief that rational players play strategies from Ψ, a subset of the set of mixed

strategies and that irrational players play strategies from another subset Ψ'. Informally, a set-valued solution concept is a τ-theory if for every game there are Ψ, Ψ' and ε^* such that the solution set is the set of strategies that can be played by rational players if the above assumptions hold and $\varepsilon < \varepsilon^*$. A τ-theory is called perfect if Ψ' is included in the interior of the set of mixed strategies (irrational players are required to play every pure strategy with some positive probability). For the formal definition see Gul(96).

Coalitional rationality is a solution concept defined on pure strategies, so it cannot be directly fit into the framework of τ-theories. But there is a natural extension of coalitional rationalizability to mixed strategies, namely the set of mixed strategies that can be best responses against conjectures concentrated on the set of coalitionally rationalizable strategies. For more on coalitional rationality and mixed strategies see the discussion in section 11.

Claim 14: the set of mixed strategies consistent with perfect coalitional rationalizability is a perfect τ-theory.

proof: in the appendix.

This gives an indirect proof that perfect coalitional rationalizability is a refinement of τ-perfect rationalizability, since τ-perfect rationalizability in finite games is the weakest perfect τ-theory.

Relating perfect coalitional rationalizability to the solution concepts that we considered in section 4 gives the same conclusions that we obtained there when comparing coalitional rationalizability to those concepts. The notable exception is that not every strong Nash equilibrium is included in the set of perfect coalitionally rationalizable strategies, due to the fact that a weakly dominated strategy can be part of a strong Nash equilibrium. The game in Figure 16 demonstrates this. $(A1, B1)$ constitutes a strong Nash equilibrium, but $A1$ is weakly dominated for player 1, therefore it is not perfect coalitionally rationalizable.

	B1	B2
A1	2,2	0,0
A2	2,0	2,2

Figure 16

7   Games with continuous strategy spaces

We extend coalitional rationalizability to games with compact strategy spaces and continuous payoff functions although it does not provide any new conceptual insight, because it is a relatively easy excercise and because several economic applications in which we think our concept might be relevant (including the application in section 7) are analytically easier in continuous strategy spaces.

Let $G = (I, S, u)$ be such that I is finite, S_i compact $\forall\, i \in I$ and u_i continuous $\forall\, i \in I$. For every $i \in I$ let Ω_{-i} be the set of Borel probability measures on S_{-i}. For every $i \in I$ and $A \subset S$ let $\Omega_{-i}(A)$ be the set of Borel probability measures on S_{-i} which allocate measure 1 to A.

Let A be such that $A \subset S$, $A = \underset{i \in I}{\times} A_i$, A is closed and nonempty.

Definition: B is a *supported restriction* from A by J if

1) $B_i = A_i \,\forall\, i \notin J$,

2) B_i is closed $\forall\, i \notin J$

3) $a_j \in BR_j(f_{-j})$ implies

$u_j(a_j, f_{-j}) < \underset{g_{-j}:\, g_{-j} \in \Omega_{-j}(B) \text{ and } \widehat{g}_{-J} = \widehat{f}_{-J}}{\min}\ \underset{b_j \in S_j}{\max} u_j(b_j, g_{-j}) \ \forall\, j,\ a_j$ and f_{-j} such that $j \in J$, $a_j \in A_j / B_j$ and $f_{-j} \in \Omega_{-j}(A)$.

The only new element in the definition is that we require the restrictions to be closed. This is a technical requirement which is needed to guarantee nonemptyness of the set of coalitionally rationalizable strategies.

To show that the definition is well-defined, we have to show that the min-max on the right-hand side of requirement 3 is attained. $u_j(b_j, g_{-j})$ is continuous in b_j and S_j is compact, so $\underset{b_j \in S_j}{\max} u_j(b_j, g_{-j})$ is attained. Furthermore, since $u_j(b_j, g_{-j})$ is continuous in g_{-j} with respect to the weak topology, $\underset{b_j \in S_j}{\max} u_j(b_j, g_{-j})$ is continuous in g_{-j} with respect to the weak topology (by the theorem of the maximum). Then since $\Omega_{-j}(B)$ is compact if B is compact, $\underset{g_{-j}:\, g_{-j} \in \Omega_{-j}(B) \text{ and } \widehat{g}_{-J} = \widehat{f}_{-J}}{\min}\ \underset{b_j \in S_j}{\max} u_j(b_j, g_{-j})$ is attained.

Now we can define iterative deletion of coalitionally dominated strategies the same way we did in section 3 to obtain A^∞, the set of coalitionally rationalizable strategies. It is easy to establish that Claim 1 hold in this extended setting too. The following claim is the extension of Claim 2 to the new setting.

Claim 15: A^∞ is nonempty and closed.

proof: in the appendix.

One can establish that there is no nontrivial supported restriction from A^∞ in this new setting either. The sketch of the argument is that if there was a non-trivial supported restriction from A^∞, then that agreement would be supported from A^k for high enough k, contradicting that $A^\infty = \underset{k \to \infty}{\lim} A^k$, because of continuity of expected payoffs in the conjecture concerning other players' strategies, and the fact that we have a strict inequality in the definition of a supported restriction.

8 An application

In a lot of economic and political situations participants are likely to use coalitional reasoning when they form their conjectures on other participants' actions

and choose their own actions. Examples include voting in committees, trade negotiations among countries, consumers choosing platforms/facilities in the presence of network externalities, and oligopolistic competition. In all these examples there might be subgroups of players having an incentive to coordinate their action choices. Nash equilibrium doesnot capture coalitional considerations, while coalition-proof equilibrium does not always exist or can give non-intuitive predictions, as the game of Figure 12 in Section 4 demonstrates. But coalitional rationalizability can be used to make sensible predictions in these games. Another useful aspect of applying coalitional rationalizability is that the iterative procedure that defines the solution set gives a sequence of implicit agreements through which the set of coalitionally rationalizable strategies is reached. In some settings these implicit agreements can be interpreted as explicit self-enforcing agreements. And the type of agreements considered in this paper, that is agreements in which participants rule out certain actions commonly arise in business and political settings. For instance, companies agree not to enter to each others' markets and nations sign agreements banning certain weapons or prohibiting the torture of prisoners of war. In fact, in a world of incomplete contracts, all agreements restrict participants' actions in certain dimensions, while giving freedom in others.

Below we investigate a model of an arms race among three countries. The countries engage in a costly arms race because country A tries to increase its political influence in one of two possible geographic regions, while countries B and C are interested in reducing the expansion of country A's political influence (they do not want country A to get too strong). If country A chooses to expand its influence in region 1, it faces country B, while in region 2 it faces country C. Political influence in a region is determined by the amount of arms the two opposing countries have. Country A chooses the region to expand where he faces the country with less arms. Building up arms yields an increasing and convex cost function. The utility coming from country A's political influence is an increasing concave function for country A and a decreasing concave function for countries B and C.

For analytical convenience we chose a piecewise linear utility function on political influence and a simple quadratic function for the cost function of building up arms.

Let $I = \{A, B, C\}$. Let $S_A = S_B = S_C = [0, 1]$. Let the payoff functions be the following:

$$u_A(s_A, s_B, s_C) = \begin{cases} s_A - \min(s_B, s_C) - s_A^2 & \text{if } s_A < \min(s_B, s_C) \\ \frac{1}{2}[s_A - \min(s_B, s_C)] - s_A^2 & \text{if } s_A \geq \min(s_B, s_C) \end{cases}$$

$$u_B(s_A, s_B, s_C) = \begin{cases} \min(s_B, s_C) - s_A - s_B^2 & \text{if } s_A \geq \min(s_B, s_C) \\ \frac{1}{2}[\min(s_B, s_C) - s_A] - s_B^2 & \text{if } s_A < \min(s_B, s_C) \end{cases}$$

$$u_C(s_A, s_B, s_C) = \begin{cases} \min(s_B, s_C) - s_A - s_C^2 & \text{if } s_A \geq \min(s_B, s_C) \\ \frac{1}{2}[\min(s_B, s_C) - s_A] - s_C^2 & \text{if } s_A < \min(s_B, s_C) \end{cases}$$

We claim that $(1/4, 1/4, 1/4)$ is the only coalitionally rationalizable outcome of the game.

First we claim that $M \equiv ([1/4, 1/2], [0, 1], [0, 1])$ is a supported restriction

from S by $\{A\}$. Note that for country A, increasing the level of arms by $\delta > 0$ from an initial level x implies a cost increase of $(x + \delta)^2 - x^2 = \delta^2 + 2x\delta$ and increases the expected utility coming from the expected political influance by something between $\delta/2$ and δ, depending on the expectations concerning the other countries arms level choices. Now if $x < 1/4$ and δ is small, then $\delta^2 + 2x\delta < \delta/2$, so x cannot be a best response to any conjecture (a small increase in arms level is always beneficial). If $x > 1/2$ and δ is small, then $\delta^2 + 2x\delta > \delta$, so x cannot be a best response to any conjecture (a small decrease in arms level is always beneficial). This establishes that M is a supported restriction by $\{A\}$.

Next we claim that $N \equiv ([0,1],[1/4,1/2],[1/4,1/2])$ is a supported restriction from S by $\{B,C\}$. Fix any conjecture concerning country A's move and let G be the distribution function belonging to this conjecture. When country B chooses an arms level of $x \in [0, 1/4)$, his expected payoff is maximized if country C chooses an arms level of at least x with probability 1, in which case country B's expected payoff is

$$\int_{s_A \leq x} (x - s_A)/2 \; dG(s_A) + \int_{s_A > x} (x - s_A) \; dG(s_A) - x^2 \tag{5}$$

Now suppose $s_C \in [1/4, 1/2]$. Then when he plays $1/4$, country B's expected payoff is

$$\int_{s_A \leq 1/4} (1/4 - s_A)/2 \; dG(s_A) + \int_{s_A > 1/4} (1/4 - s_A) \; dG(s_A) - 1/16 \tag{6}$$

The difference between 6 and 5 is:
$\int_{s_A \leq x} (1/4 - x)/2 \; dG(s_A) + \int_{x \leq s_A \leq 1/4} (1/4 - s_A)/2 \; dG(s_A) + \int_{1/4 \leq s_A \leq 1/2} (1/4 - x) \; dG(s_A) + \int_{x \leq s_A \leq 1/4} (s_A - x) \; dG(s_A) + x^2 - 1/16 = \int_{s_A \leq x} (1/4 - x)/2 \; dG(s_A) + \int_{1/4 \leq s_A \leq 1/2} (1/4 - x) \; dG(s_A) + \int_{x \leq s_A \leq 1/4} (1/4 + s_A/2 - x) \; dG(s_A) \geq \int_{0 \leq s_A \leq 1/2} (1/4 - x)/2 \; dG(s_A)$

The last expression is positive for $x \in [0, 1/4)$. This implies that country B's expected payoff when he chooses any $x \in [0, 1/4)$ as a best response is smaller than his expected payoff when he expects country C to play inside $[1/4, 1/2]$ and best responds, fixing the conjecture concerning country A's action. Showing that country C's expected payoff when he chooses any $x \in [0, 1/4)$ as a best response is smaller than his expected payoff when he expects country B to play inside $[1/4, 1/2]$ and best responds, fixing the conjecture concerning country A's action is completely symmetric to the above. This concludes that N is a supported restriction from S by $\{B,C\}$.

Now we claim that $(1/4, 1/4, 1/4)$ is a supported restriction from $P \equiv ([1/4, 1/2], [1/4, 1/2], [1/4, 1/2])$ by $\{A, B, C\}$. To show this, we have to show that whenever play is in P and a player's best response is different from $1/4$, he expects a lower payoff than what he gets if he expects both of the other players to choose an arms quantity of $1/4$.

First note that $1/4$ is the only best response for country A against a conjecture which allocates probability 1 to both other countries choosing $1/4$. Also note that $1/4$ is the only best response for country B against a conjecture which allocates probability 1 to country A playing $1/4$ and country C playing inside $[1/4, 1/2]$. Now suppose country A has a conjecture concentrated on $([1/4, 1/2], [1/4, 1/2])$ against which x is a best response and $x \in (1/4, 1/2]$. It is straightforward to show that his expected payoff is at most as much as the expected payoff that playing x yields against a conjecture that assigns probability 1 to countries B and C both playing $1/4$. But that payoff is less than what the outcome $(1/4, 1/4, 1/4)$ gives to player 1, since $1/4$ is the only best response for player 1 against the other players playing $1/4$. Next suppose country B has a conjecture concentrated on $([1/4, 1/2], [1/4, 1/2])$ against which x is a best response and $x \in (1/4, 1/2]$. It is straightforward to show that his expected payoff is at most as much as the expected payoff that playing x yields against a conjecture that assigns probability 1 to country A choosing $1/4$ and country C choosing $1/2$. But that payoff is less than what the outcome $(1/4, 1/4, 1/4)$ gives to country B, since $1/4$ is the only best response for country B against the above conjecture and $(1/4, 1/4, 1/4)$ gives the same payoff to country B as $(1/4, 1/4, 1/2)$ does. Showing that the expected payoff that country C gets when he plays $x \in (1/4, 1/2]$ is less than what the outcome $(1/4, 1/4, 1/4)$ gives to him is completely similar to the above. This concludes that $(1/4, 1/4, 1/4)$ is a supported restriction from P.

Since M and N are both supported restrictions from S and $P = A \cap B$, it has to be that $A^1 \subset P$. Then since $(1/4, 1/4, 1/4)$ is a supported restriction from P, by Lemma 3 $(1/4, 1/4, 1/4)$ is a supported restriction from A^1. Then since the set of coalitionally rationalizable strategies is nonempty by Claim 2, $(1/4, 1/4, 1/4)$ is the only coalitionally rationalizable outcome in the game.

To summarize, coalitional rationality in this model predicts that first countries B and C make an agreement not to choose too low a level of arms (lower than $1/4$) which would make them vulnerable against country A, then the three countries together make an agreement not to chose too high an arms level (higher than $1/4$), or not to engage in too costly an arms race. Coalitional rationalizability is particularly interesting in this context because it not only gives a predicted outcome in the game, but also predicts a sequence of agreements through which the outcome is reached. Furthermore, the type of agreements it predicts are the type observed in actual political situations, for example, in agreeing on an upper limit of nuclear weapons build-up.

It is straightforward to show that the set of Nash equilibria in the above game is $\{(x, x, x) \mid x \in [1/4, 1/2]\} \cup \{(1/4, y, y) \mid y \in [0, 1/4]\}$. $(1/4, 1/4, 1/4)$ is not Pareto-dominant inside the equilibrium set, but it is the only coalition-proof Nash equilibrium. The fact that it is the only coalitionally rationalizable out-

come means that we do not have to assume equilibrium behavior exogeneously to justify $(1/4, 1/4, 1/4)$ as the only "reasonable" outcome in the game.

9   Pre-play communication

In earlier sections we mentioned that in our model we do not require pre-play communication, but we also do not preclude it. The reasoning procedure and the agreements we propose do not require pre-play communication: everything is based on public information, namely the strategy sets and payoff functions. The question arises whether our results apply to situations when there is pre-play communication before the game, so players do not have to rely only on public information when formulating their beliefs. We claim that the answer is yes, if the communication is cheap talk (does not effect the players' payoffs directly), in the sense that independently of the nature of the pre-play communication and the messages sent there, a mutually advantageous agreement for some coalition remains beneficial for everyone in the coalition, in terms of expected payoffs. Because talk is cheap and players ultimately make their moves secretly, coalitions have the incentive to make mutually advantageous agreements, independently of what threats and promises are made in the communication phase. Consider again Figure 7 from Section 3. Even if players 2 and 3 promise to play $B1$ and $C1$ in the communication phase, they are better off playing $(B2, C2)$, for any conjecture that they might have about player1's action. In general, the only reason members of some coalition would not want to play inside some mutually advantageous agreement is if they think that doing so influences the play of players outside the coalition. But that consideration is irrelevant if players make their moves secretly. So we claim that with or without pre-play communication, play has to be inside the set of coalitionally rationalizable strategies, if players use coalitional reasoning in formulating their beliefs. But that does not mean that players cannot make further agreements, even in the absence of pre-play communication. And pre-play communication definitely makes it more likely that further agreements are made, besides the ones proposed in the iterative procedure we provided in section 2. We give two examples to demonstrate this. Cheap talk can resolve symmetric coordination problems. Consider the game in Figure 17.

	B1	B2
A1	1,1	0,0
A2	0,0	1,1

Figure 17

In this game there is no nontrivial supported restriction, and any outcome is consistent with coalitional rationality, although the players' interests perfectly coincide and they would like to coordinate on either $(A1, B1)$ or $(A2, B2)$. The problem is that there is not enough public information to tell them which of these outcomes to play, so they might miscoordinate and play $(A1, B2)$ or $(B1, A2)$. But if players can communicate before the game, it is reasonable to assume that they play either $(A1, B1)$ or $(A2, B2)$ and neither of the two remaining outcomes can be played. This example shows that a coalitional-rationalizability-type concept which assumes pre-play communication would not yield solution sets having the product structure, even if players moved secretly and independently of each other.

The second example shows that in some cases players can credibly transmit information concerning their beliefs, which can help them make agreements which would not be mutually advantageous if players in the coalition did not know each others' beliefs. Consider the game of Figure 18.

	C1				C2		
	B1	B2	B3		B1	B2	B3
A1	5,5,0	3,3,0	0,0,0		3,4,0	3,3,0	0,0,0
A2	3,3,0	3,3,0	0,0,0		5,3,0	5,3,0	0,0,0
A3	0,0,0	0,0,0	4,2,0		0,0,0	0,0,0	0,0,0

Figure 18

In this game there is no mutually advantageous agreement. In particular $\{A1, A2\} \times \{B1, B2\}$ is not mutually advantageous for players 1 and 2, since player 1 might believe that player 3 plays $C1$ with probability 1, but also believe that player 2, believing that player 1 believes that player 3 plays $C2$ with probability 1, will play $B2$. In this case playing inside $\{A1, A2\}$ yields a payoff of 3 to player 1. And $A3$ can be a best response and yield a payoff of 4 for him. But one can prove that if player 1 can make his conjecture known to player 2, then they are always able to make a mutually advantageous agreement inside $\{A1, A2\} \times \{B1, B2\}$. For example if player 1 expects player 3 to play $C1$ with probability 1, $\{A1\} \times \{B1\}$ is a mutually advantageous restriction. If player 1 expects player 3 to play $C2$ with probability 1, $\{A2\} \times \{B1, B2\}$ is a mutually advantageous restriction, and so on for any possible belief concerning player 3's action. Therefore in case of pre-play communication we expect play to be inside $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$.

We leave it to a future project to work out a fully satisfactory solution concept for the case of unlimited pre-play communication, which is consistent with coalitional rationality. The two examples above suggest that the key question in building up a solution concept along these lines is what information the players can credibly communicate to each other. That question has been analyzed extensively in the literature on cheap talk. For references, see Myerson[89], Rabin[90], Farrell[93], Rabin and Farrell[96] and Zapater[97], besides the papers mentioned in the introduction.

We note that pre-play communication can have a role in determining whether or not players use coalitional reasoning. This is because coalitional rationality requires a certain confidence in other players reasoning the same way. If players are not sure of each others' ways of reasoning and belief formation, then pre-play communication can help establish the trust needed to believe in mutually advantageous agreements. Experimental game theory provides some support for this claim. In certain coordination games pre-play communication increases the propensity of the players to play Pareto optimal outcomes, and multi-sided pre-play communication may increase cooperation more than one-sided communication (see Cooper et al.[92] and Charness[00]). An interesting question is that of how much trust is needed in different games to establish coalitional rationalizability, and what happens when players are not completely sure of each others' intentions. One interpretation of perfect coalitional rationalizability, which assumes that players allocate high probability to agreements which are mutually advantageous with high probability, to be a step to this direction.

10 Related literature

In section 4 we related coalitional rationalizability to some refinements of Nash equilibrium: Pareto-undominated Nash equilibrium, coalition-proof Nash equilibrium and strong Nash equilibrium. Other concepts in the literature also incorporate coalitional reasoning into the play of normal-form games. Below we discuss the similarities and differences between the assumptions of these concepts and the underlying assumptions behind coalitional rationalizability.

Chwe[94], Mariotti[97], Xue[98] and Xue[00]) assume that players engage in a possibly infinite negotiation procedure before playing a normal-form game. At any stage of the negotiation there is a status quo outcome, but players are farsighted and only care about the final outcome of the negotiation procedure. Chwe's largest consistent set (see Chwe[94]) and Xue's concept of perfect foresight agreements (see Xue[98]) captures this in cooperative game theory spirit, while Mariotti [97] and Xue[00] examine equilibria of the noncooperative game obtained when the normal-form game is preceded by the negotiation game. These papers are similar to our approach in assuming that coalitions can freely form and make agreements, and that there are no binding agreements. The main difference, aside from the fact that these papers consider agreements which can only specify a unique strategy profile to be played, is that the above papers assume that coalitions act publicly such that negotiation is publicly observed. Even though the status quo can be changed at any stage of the negotiation, it is known by everyone if it is changed. This is not compatible with our assumption that players make their moves secretly, which implies that players cannot make their moves contingent on the other players' intended moves. Applying the above concepts to situations in which players make their moves secretly is inappropriate in the sense that it would contradict individual rationality. Xue[00] assumes

individual rationality exogeneously by assuming that the final outcome reached by the negotiation procedure has to be a Nash equilibrium, thereby accepting that players can secretly change their minds and play something other than the final stage status quo. However, this approach does not solve the problem of secret coalitional deviations, which is what our paper concentrates on. We claim that if players make their moves secretly and use coalitional reasoning, then even if there is a negotiation procedure before playing a normal-form game, independently of the status quo reached, they have an incentive to jointly deviate if the deviation is a mutually advantageous agreement. This problem cannot be satisfactorily overcome in an equilibrium framework. There are games in which, no matter what profile is played, if it is expected to be played by every player, then some coalition has a mutually advantageous agreement to play something else.

Another line of related literature is noncooperative coalitional bargaining. Noncooperative coalitional bargaining considers extensive form noncooperative games to model n-player coalitional bargaining situations based on characteristic function games (or generalizations of those, such as partition-function games). However, these characteristic forms can be derived from normal-form games, as done in Ray and Vohra[97] and Ray and Vohra[99]. In this manner the framework can be used to examine questions such as which coalitions form and which agreements players make before playing a normal-form game. The main difference between that approach and ours is that these papers maintain an underlying assumption that members inside some coalition can make binding agreements and only play among coalitions in a noncooperative fashion. This paper, more in the tradition of noncooperative game theory, maintains the assumption that players cannot make binding agreements.

Rabin's concept (see 94/I) of Consistent Behavioral Theories can be used to incorporate coalitional reasoning into playing a normal-form game, and the paper proposes one such theory, Pareto-focal rationalizability. Using the terminology of our paper, it is the smallest set which is closed under rational behavior and contains all outcomes that can be played in some Pareto-undominated Nash equilibrium. Rabin defines the concept only for 2-player games. In games with more than two players the concept would not take the possibility of subgroups of players coordinating their moves into account. In 2-player games it is easy to establish that the set of Pareto-focal rationalizable outcomes is included in the set of coalitionally rationalizable strategies. The game of Figure 9 in Section 4 provides an example that the inclusion might be strict. There (A2,B2) is the only Pareto-focal rationalizable outcome, while every outcome in the game is coalitionally rationalizable. A conceptual difference between our concept and the way Rabin constructs consistent behavioral theories is that we start out from the whole strategy set and discard strategies in an iterative manner based on coalitional considerations, while Rabin starts out with a set of focal outcomes and expands the set until it is closed under rational behavior. Our procedure can endogenously explain why certain outcomes are focal in a game, while Rabin's approach starts with some exogenously given set of focal points.

Finally, there are papers investigating the role of pre-play communication (cheap talk) before playing a normal-form game, examining whether communication leads to the type of belief restrictions we consider in this paper. To achieve a sharper prediction than rationalizability, these papers assume that messages in the cheap talk phase are interpreted according to their literal meanings whenever they are credible, where credibility is defined formally.

Farrell[88] assumes that a round of cheap talk precedes playing a normal-form game, in which one player sends a suggestion to the others. A suggestion specifies a subset of the strategy set in which players should play. Suggestions have to be consistent in the sense that every strategy of every player in the set has to be a best response to some conjecture consistent with the other players playing inside the set. The proposing player has a conjecture conditional on every consistent proposal and chooses the proposal with the highest expected payoff. This construction differs from ours in several respects. The above consistency requirement is weaker than the one specified by our solution set. In Farrell's model there is only one round of agreement-making, while our concept allows that once one agreement is made, further agreements might become desirable for coalitions. There is no highlighted player in our model who chooses between possible solution sets. Furthermore, we claim that if players use coalitional reasoning, then not every consistent proposal of Farrell is followed. Consider the game of Figure 19.

	C1			C2	
	B1	B2		B1	B2
A1	3,1,1	0,0,0	A1	0,0,0	1,5,5
A2	0,0,0	0,0,0	A2	0,0,0	1,5,5

Figure 19

Here player 1 would propose to play $\{A1\} \times \{B1\} \times \{C1\}$, but we take the position that independently of player 1's message, players 2 and 3 should play $\{B2\} \times \{C2\}$, which yields the highest payoff in the game for both of them.

Watson[91] analyzes 2-player games and introduces two rounds of pre-play communication, one arbitrary and one in which player 1 sends a message in which he suggests that both players play inside some subset of the strategy space. The message is believed iff the best response against any belief which assigns probability 1 on the other playing inside the set yields a higher expected payoff than that the best response gives against any belief which puts probability 1 on the other player playing outside the set. This is similar to the way we define an agreement to be mutually advantageous in that it compares two sets of expectations of rational players and checks whether one set is Pareto-dominated to the other. However, Watson's comparison leaves out conjectures that allocate positive probability to strategies both outside and inside the set under consideration, as the game in Figure 20 demonstrates.

36

	B1	B2	B3
A1	7,7	2,2	2,2
A2	2,2	7,7	2,6
A3	2,2	6,2	4,2

Figure 20

In this game $\{A1, A2\} \times \{B1, B2\}$ is a believable message, because the best response against any belief concentrated on it gives a payoff of at least 4.5 for both players, while the best response against $A3$ and $B3$ respectively gives a payoff of 4 to player 1 and 2 to player 2. It is not a supported restriction though, according to our definition. Player 1 can have the conjecture that player 2 plays B2 and B3 with probability 1/2-1/2. Against this belief, $A3$ is a best response and yields 5, which is larger than the minimum payoff player 1 can expect if play is inside $\{A1, A2\} \times \{B1, B2\}$. Hence, it is not always advantageous for player 1 to restrict play to $\{A1, A2\} \times \{B1, B2\}$.

Watson's concept is not iterative (there is only one agreement) and is defined only for 2-player games, in which the issues of coalitional agreements are less complex. Pre-play communication is assumed, unlike in our paper, although its role is not specified clearly, given that the proposer and the other player are treated symmetrically. Since the agreement is only made if it is beneficial for both of them, it is unclear why the agreement would not be made without pre-play communication, and if it indeed would not, then why pre-play communication helps.

Rabin's model of pre-game communication (Rabin[94/II]) deals with 2-player games and assumes a special form of pre-game communication in which the two players negotiate over Nash equilibria of a game. If no agreement is reached, then non-equilibrium outcomes may be played. The paper examines both the equilibria and the rationalizable outcomes of the negotiation game and asks how much cooperation the two players can achieve. It shows that if players can negotiate for a long time, then in any equilibrium of the extended game the players get at last as much as their minimal payoffs in equilibrium in the original game. The same result holds for expected payoffs for rationalizable outcomes of the extended game. These results do not hold for our model, as the game in Figure 21 demonstrates.

	B1	B2
A1	3,-3	-3,3
A2	0,3	-2,-2

Figure 21

Here, in the only Nash equilibrium of the game, player 1 gets -3/2. In contrast, coalitional rationalizability does not restrict the set of possible beliefs, so player 1 can expect player 2 to play $B2$ with probability 1, in which case he expects -2. Intuitively, if players' interests conflict sharply, then coalitional rationalizability does not help players to coordinate on avoiding bad outcomes, while in games like those which assume equilibrium (that expectations are correct), our reasoning procedure might put a lower bound on expected payoffs. Rabin's communication procedure puts different restrictions on beliefs than coalitional rationalizability and in some games those belief restrictions yield a higher minimum expected payoff than the minimum payoffs that players can expect if we assume coalitional rationalizability. In other games, coalitional rationalizability is more efficient in achieving coordination between players.

11   Modifications of the concept

Coalitional rationalizability is a concept defined on pure strategies (with respect to the players' actions, a player's conjecture can be any probability distribution on the other players' strategy set). The question we ask is what pure strategies can players choose if they are rational and they have "reasonable" beliefs (beliefs that players make supported restrictions), but the construction is valid if we allow players to play mixed strategies. Let $\Sigma = \underset{i \in I}{\times} \Sigma_i$ be the set of mixed strategies and for every A such that $A = \underset{i \in I}{\times} A_i$ and $A \subset S$ let $\Sigma(A)$ be the set of mixed strategies with support inside A. For any i and any $f_{-i} \in \Omega_{-i}$ let $BR_i^\Sigma(f_{-i})$ be the set of mixed strategy best responses against f_{-i}. For any $i \in I$, $\sigma_i \in \Sigma_i$ and $f_{-i} \in \Omega_{-i}$ let $u_i(\sigma_i, f_{-i})$ be the expected payoff of player i if he plays mixed strategy σ_i and has conjecture $, f_{-i}$. For any $\Phi \subset \Sigma$, let $\Omega_{-i}(\Phi)$ be the set of conjectures player i can have which can be obtained as convex combinations of mixed strategy profiles from Φ.

Let Φ and Ψ be such that $\Phi \subset \Sigma$, $\Psi \subset \Phi$, $\Phi = \underset{i \in I}{\times} \Phi_i$ and $\Psi = \underset{i \in I}{\times} \Psi_i$.

Definition: Φ is a *supported restriction on mixed strategies* from Ψ by J if

1) $\Phi_i = \Psi_i \ \forall \ i \notin J$, and
2) $\sigma_j \in BR_j^\Sigma(f_{-j})$ implies
$u_j(\sigma_j, f_{-j}) < \underset{g_{-j}: g_{-j} \in \Omega_{-j}(\Phi) \text{ and } \widehat{g}_{-J} = \widehat{f}_{-J}(s_{-J})}{\min} \underset{\tau_j \in \Sigma_j}{\max} u_j(\tau_j, g_{-j}) \ \forall \ j, \ \sigma_j \text{ and } f_{-j}$
such that $j \in J$, $\sigma_j \in \Psi_j / B_j$ and $f_{-j} \in \Omega_{-j}(\Psi)$.

This is essentially the same concept on mixed strategies than supported restriction on pure strategies.

Then it is possible to show that Φ is a mixed strategy supported restriction from $\Sigma(A)$ by J iff there is B such that B is a supported restriction from A and $\Omega_{-i}(\Phi) = \Sigma_i(B) \ \forall \ i \in I$. This establishes that the iterative procedure that we defined in section 3, applied to mixed strategies using the above definition

of supported restriction on mixed strategies, would predict conjectures concentrated on A^∞ to be the ones consistent with coalitional rationality. Then the set of mixed strategies that are consistent with coalitional rationality are the ones that can be best responses against conjectures concentrated on A^∞.

The definition of coalitional rationality can be modified in several ways. One possibility is modifying the definition of supported restriction such that the expected payoff resulting from conjectures that are compatible with playing outside the restriction are not compared to the expected payoffs of all conjectures that are compatible with the restriction, but only to those which are "minimal changes" of the original conjecture (leaving the part of the conjecture which is consistent with the restriction unchanged). This leads to a refinement of the set of coalitionally rationalizable strategies, since more restrictions become supported, and in certain applications gives much sharper predictions. We are currently investigating the properties of this and related refinements and leave the formal exposure of them to another paper.

Our concept is a non-equilibrium concept, but Claim 9 provides a way to define a refinement of Nash equilibrium which is consistent with the type of reasoning our players use. We could define coalitionally rational Nash equilibria to be the Nash equilibria of the game which are inside the set of coalitionally rationalizable strategies. According to Claim 9 every finite game has an equilibrium like that and it is easy to extend the result for games with compact strategy spaces and continuous payoff functions. An interpretation of this equilibrium concept is that players use coalitional reasoning to rule out certain strategies, but after that, in the remaining game they behave in an individualistic way and their conjectures on each others' play are correct.

12   Appendix

Lemma 1: let A, B be such that $A \in \mathcal{M}$ and B is a supported restriction from A by some $J \subset I$. Then $B \in \mathcal{M}$.

proof: suppose not. Then $\exists j$, a_j and f_{-j} such that $j \in J$, $a_j \in A_j/B_j$, $f_{-j} \in \Omega_{-j}(B)$ and $a_j \in BR_j(f_{-j})$, which contradicts that B is a supported restriction from A by J. QED

Lemma 2: Let B be a supported restriction from A by J and C be such that $C \subset A$, $C = \underset{i \in I}{\times} C_i$ and $C \cap B \neq 0$. Then $C \cap B$ is a supported restriction from C by J.

proof: let j and c_j be such that $j \in J$ and $c_j \in C_j/B_j$. Because B is a supported restriction from A by J, $u_j(c_j, f_{-j}) < u_j(b_j, f_{-j})$ $\forall$ b_j, f_{-j}, g_{-j} such that $f_{-j} \in \Omega_{-j}(A)$, $c_j \in BR_j(f_{-j})$, $g_{-j} \in \Omega_{-j}(B)$, $b_j \in BR_j(g_{-j})$ and $\widehat{g}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$. But since $C \cap B \subset C$, this implies $u_j(c_j, f_{-j}) < u_j(b_j, g_{-j})$ $\forall$ b_j, f_{-j}, g_{-j} such that $f_{-j} \in \Omega_{-j}(A)$, $c_j \in$

$BR_j(f_{-j})$, $g_{-j} \in \Omega_{-j}(C \cap B)$, $b_j \in BR_j(f_{-j})$ and $\widehat{g}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$. Since it holds for every j and c_j such that $j \in J$ and $c_j \in C_j/B_j$, $B \cap C$ is a supported restriction from C by J. QED

proof of Claim 1: let a be such that $u_j(a) = \max\limits_{s:\, s \in A} u_j(s)$. Then by the definition of a supported restriction, $a_j \in B_j$ $\forall$ $B \in F(A)$, because a_j is a best response against a_{-j} (it yields the maximum payoff in A and $A \in \mathcal{M}$). Therefore $\bigcap\limits_{B:\, B \in F(A)} B_j \neq \emptyset$. This establishes the claim since j was arbitrary and $\bigcap\limits_{B:\, B \in F(A)} B$ is a product set. QED

Lemma 3: let A be such that $A \in \mathcal{M}$. Let B be a supported restriction from A by J^B. Let $C^0, ..., C^k$ ($k \geq 1$) be such that $C^0 = A$ and C^i is a supported restriction from C^{i-1} by J^i $\forall$ $i = 1, ..., k$. Then $B \cap C^k$ is a supported restriction from C^k by J^B.

proof: by Claim 1, $B \cap C^1 \neq \emptyset$. Then by Lemma 2, $B \cap C^1$ is a supported restriction from C^1. Now suppose $B \cap C^n$ is a supported restriction from C^n for some $1 \leq n \leq k-1$. By Lemma 1, $C^n \in \mathcal{M}$. Then by Claim 1, $B \cap C^{n+1} \neq \emptyset$ and by Lemma 2, $B \cap C^{n+1}$ is a supported restriction from C^{n+1}. QED

proof of Claim 2: since S is finite and $A^{k-1} \supset A^k$ $\forall$ $k \geq 1$, the second part of the claim is immediate. Note that $A^0 = S \in \mathcal{M}$. Now assume $A^k \in \mathcal{M}$ for some $k \geq 0$. By Claim 1, $A^{k+1} \neq \emptyset$. By Lemma 2, A^{k+1} can be reached from A^k by a sequence of restrictions and then by Lemma 3, $A^{k+1} \in \mathcal{M}$. By induction, $A^k \neq \emptyset$ and $A^k \in \mathcal{M}$ $\forall$ $k \geq 0$. Since $A^\infty = A^k$ whenever $k \geq K$, this implies $A^\infty \neq \emptyset$ and $A^\infty \in \mathcal{M}$. Now suppose that there exists a supported restriction from A^∞. Since $A^\infty = A^K$, this implies that there is a supported restriction from A^K, which contradicts that $A^{K+1} = A^K$. QED

proof of Claim 4: since the sequence of sets $(B^k)_{k=0}^\infty$ is nested and S is finite, there is $L \geq 0$ such that $B^k = B^L$ $\forall$ $k \geq L$. Since $B^0 = S$, $B^0 \in \mathcal{M}$. Now assume $B^k \in \mathcal{M}$ for some $k \geq 0$. By Claim 1, $\bigcap\limits_{B:\, B \in F(B^k)} B \neq \emptyset$, so $B^{k+1} = \bigcap\limits_{B:\, B \in \Theta^k} B \neq \emptyset$. By Lemma 1 and 2, $B^{k+1} \in \mathcal{M}$. Since $B^k = B^L$ $\forall$ $k \geq L$, by definition of the sequence $(B^k)_{k=0}^\infty$ there is no nontrivial supported restriction from B^L. By definition $B^L \subset A^0$. Then if B is a supported restriction from A^0, then $B \cap B^L$ is a supported restriction from B^L. Then since there is no nontrivial supported restriction from B^L, $B^L \subset A^1$. An inductive argument shows that $B^L \subset A^k$ $\forall$ $k \geq 0$, therefore $B^L \subset A^\infty$. By definition $A^\infty \subset B^0$. Then if B is a supported restriction from B^0, then $B \cap A^\infty$ is a supported restriction from A^∞. Then since there is no nontrivial supported restriction from A^∞, $A^\infty \subset B^1$. An inductive argument shows that $A^\infty \subset B^k$ $\forall$ $k \geq 0$, therefore $A^\infty \subset B^L$. QED

proof of Lemma 5: suppose not. Then there are A and B such that $A^\infty \subset A$, $A \neq A^\infty$, B is a supported restriction from A by J and $A^\infty \not\subseteq A$. First consider $B \cap A^\infty \neq \emptyset$. Then by Lemma 2 $B \cap A^\infty$ is a supported restriction

from A^∞, contradicting that there is no nontrivial supported restriction from A^∞.

Now consider $B \cap A^\infty = \emptyset$. Then there is $k \geq 0$ such that $B \cap A^k \neq \emptyset$ and $B \cap A^{k+1} = \emptyset$. As established above, $A^k \in \mathcal{M}$. Together with $A \in \mathcal{M}$ this implies that $A \cap A^k \in \mathcal{M}$, because for any $i \in I$ and any $f_{-i} \in \Omega_{-i}(A \cap A^k)$, $BR_i(f_{-i}) \in A$ since $A \in \mathcal{M}$ and $BR_i(f_{-i}) \in A^k$ since $A^k \in \mathcal{M}$, so $BR_i(f_{-i}) \in A \cap A^k$. Since B is a supported restriction from A, by Lemma 3 $B \cap A^k$ is a supported restriction from $A \cap A^k$, so $B \cap A^k \supset \underset{B:\ B \in F(A \cap A^k)}{\cap} B$. Since $\underset{B:\ B \in F(A^k)}{\cap} B = A^{k+1}$ and $A \cap A^{k+1} \neq \emptyset$ (they both contain A^∞), if C is a supported restriction from A^k, then $A \cap C$ is a supported restriction from $A \cap A^{k+1}$, by Lemma 3. Therefore $A \cap A^k \supset \underset{B:\ B \in F(A \cap A^k)}{\cap} B$. But $(B \cap A^k) \cap (A \cap A^{k+1}) = B \cap A^{k+1} = \emptyset$, contradicting Claim 1. QED

proof of Lemma 8: for any A defined above, there is $k \geq 0$ such that $A \subset A^k$ and $A \not\subseteq A^{k+1}$. Then there is B such that B is a supported restriction from A^k by some J and $A \not\subseteq B$. By Lemma 4 $A^\infty \subset B$, therefore $A \cap B \neq \emptyset$. Then by Lemma 2, B is a supported restriction from A by J, so $\underset{B:\ B \in F(A)}{\cap} B \neq A$.
QED

proof of Claim 5: lemmas 5, 6 and Claim 2 establish that A^∞ is a coalitionally stable set. By Lemma 6, if A is such that $A/A^\infty \neq \emptyset$ and $A \cap A^\infty \neq \emptyset$, then A cannot be coalitionally stable. If $A/A^\infty = \emptyset$, then $A \subset A^\infty$, in which case either $A = A^\infty$ or A cannot be coalitionally stable, since A^∞ contains it and there is no nontrivial supported restriction from A^∞. And if $A \cap A^\infty = \emptyset$, then there is $k \geq 0$ such that $A \subset A^k$ and $A \not\subseteq A^{k+1}$, therefore A is not coalitionally stable, since $A \subset A^k$ and $A \not\subseteq \underset{B:\ B \in F(A^k)}{\cap} B$. QED

proof of Claim 7: By definition supp$\sigma \subset A^0$. Now suppose supp$\sigma \subset A^k$. Then no a_i can be outside a supported restriction for $\{i\}$, because a_i is a best response against σ_{-i}. By Claim 5, there is ζ such that ζ is a Nash equilibrium profile and supp$\zeta \subset A^*$. The latter implies supp$\zeta \subset A^{k+1}$, or supp$\zeta \subset B \ \forall \ B$ such that B is a supported restriction from A^k. σ is a Pareto-undominated Nash equilibrium, so there is i such that $i \in \{1, 2\}$ and $u_i(a_i, \sigma_{-i}) \geq u_i(\zeta)$ $\forall \ a_i \in$suppσ_i. But then by the definition of a supported restriction, there cannot be a supported restriction by $\{1, 2\}$ which does not contain every $a_i \in A_i$. By Lemma 1 $B \in \mathcal{M} \ \forall \ B$ such that B is a supported restriction from A^k by $\{1, 2\}$, so the previous statement implies $A_{3-i} \subset B \ \forall \ B$ such that B is a supported restriction from A^k by $\{1,2\}$. This proves supp$\sigma \subset A^{k+1}$. Then supp$\sigma \subset A^k$ $\forall \ k \geq 0$, so supp$\sigma \subset A^\infty$. QED

proof of Claim 8: let $A_i =$supp$\sigma_i \ \forall \ i \in I$ and let $A = \underset{i \in I}{\times} A_i$. Suppose $A \not\subseteq A^*$. Then there is $k \geq 0$ such that $A \subset A^k$, but $A \not\subseteq A^{k+1}$. Then there are B and J such that B is a supported restriction from A^k by J and $A \not\subseteq B$. Let L$=\{j \mid j \in J, \exists \ s_j$ such that $s_j \in A_j$ and $s_j \notin B_j\}$. For every $l \in L$ let a_l be

such that $a_l \in A_l$ and $a_l \notin B_l$. For every $l \in L$ let f_{-l} be the conjecture of player l corresponding to the others playing the profile $\sigma_{-l} : f_{-l}(s_{-l}) = \underset{i \in I/\{l\}}{\times} \sigma_i(s_i)$.

Note that $a_l \in BR_l(f_{-l}) \; \forall \; l \in L$. Now let G_L be the truncated game in which the set of players are L, the set of strategies are B_l, $l \in L$ and the payoff functions are $\widehat{g}_l(s_L) = g_l(s_L, \sigma_{-L})$. Since its strategy sets are compact and payoff functions are continuous, G_L has a Nash equilibrium in mixed strategies. Let $\widehat{\xi}_L$ be such a profile. Since $B \in \mathcal{M}$ by Lemma 1, for every $l \in L$, $\widehat{\xi}_l$ is a best response against the profile $(\widehat{\xi}_{L/l}, \sigma_{-L})$. Then since B is a supported restriction from A^k by J, $u_l(\widehat{\xi}_L, \sigma_{-L}) > u_l(a_l, \sigma_{-l}) = g_l(\sigma)$. But that implies $\widehat{\xi}_L$ is a profitable deviation for L from σ, contradicting that σ is a strong Nash equilibrium. Therefore $\mathrm{supp}\sigma = A \subset A^\infty$. QED

proof of Claim 9: let s be the Pareto-dominant outcome in R. Assume $s \notin A^\infty$. Then there is $k \geq 0$ such that $s \in A^k$, but $s \notin A^{k+1}$. Then there are J and B such that B is a supported restriction from A^k by J and $s \notin B$. Let j be such that $j \in J$ and $s_j \notin B_j$. Let a_j and f_{-j} be such that $a_j \in A_j^\infty$, $f_{-j} \in \Omega_{-j}(A^\infty)$ and $a_j \in BR_j(f_{-j})$ (since A^∞ is closed under rational behavior, there are a_j and θ_{-j} like that). Then $a_j \in B_j$ and $f_{-j} \in \Omega_{-j}(B)$. Since $A^\infty \subset R$ and s is Pareto-dominant in R, $u_j(a_j, f_{-j}) \leq u_j(s)$. On the other hand, since s_j is a best response to s_{-j} and B is a supported restriction from A^k by J, $u_j(a_j, f_{-j}) > u_j(s)$, a contradiction. QED

proof of Claim 10: let $\widehat{s}_1, \widehat{s}_2$ be minmax strategies for players 1 and 2 respectively. Note that $R \subset A^0$. Suppose $R \subset A^k$ for some $k \geq 0$. Since strategies in R are best responses against some conjecture concentrated on R, they are best responses against a conjecture concentrated on A^k. Therefore no single-player coalition has a restriction B from A^k such that $B \not\subseteq R$. Now assume there is B such that B is a supported restriction from A^k by $\{1,2\}$ and there are i and a_i such that $i \in \{1, 2\}$, $a_i \in R$ and $a_i \notin B_i$. Since $a_i \in R$, there is f_{-i} such that $f_{-i} \in \Omega_{-i}(A^k)$ and $a_i \in BR_i(f_{-i})$. Since $u_i(\widehat{s}_1, \widehat{s}_2)$ is the minmax value for i, $u_i(a_i, f_{-i}) \geq u_i(\widehat{s}_1, \widehat{s}_2)$. But then B being a supported restriction from A^k by $\{1,2\}$ implies $u_i(b_i, g_{-i}) > u_i(\widehat{s}_1, \widehat{s}_2) \; \forall \; b_i, g_{-i}$ such that $g_{-i} \in \Omega_{-i}(A^k)$ and $b_i \in BR_i(g_{-i})$. Note that the game that has players I, strategy sets B_i $\forall \; i \in I$ and payoff functions $\widehat{u}_i$ such that $\widehat{u}_i(s) = u_i(s) \; \forall \; s \in B$ has a Nash equilibrium on mixed strategies. Denote this profile by σ. Since $B \in \mathcal{M}$, σ is a Nash equilibrium in G, too. By the above inequality, $u_i(\sigma) > u_i(\widehat{s}_1, \widehat{s}_2)$. Since G is 0-sum, this implies $u_{3-i}(\sigma) < u_{3-i}(\widehat{s}_1, \widehat{s}_2)$. But that contradicts the fact that B is a supported restriction from A^k by $\{1,2\}$, since $\widehat{s}_{3-i}$ is a best response against $\widehat{s}_i$. This establishes that $R \subset A^{k+1}$. By induction $R \subset A^k \; \forall \; k \geq 0$, so $R \subset A^\infty$. Since A^∞ is closed under rational behavior, $A^\infty \subset R$, so $A^\infty = R$. QED

Lemma 7: Let B be a level-p supported restriction from A by J and C be such that $C \subset A$, $C = \underset{i \in I}{\times} C_i$ and $C \cap B \neq 0$. Then $C \cap B$ is a level-p supported restriction from C by J.

proof: similar to the proof of Lemma 2, therefore omitted.

Lemma 8: let $p \in (0,1)$ and A be such that $A \in \mathcal{M}$. Let B be a level-p supported restriction from A by J^B. Let $C^0, ..., C^k$ $(k \geq 1)$ be such that $C^0 = A$ and C^i is a level-p supported restriction from C^{i-1} by J^i $\forall$ $i = 1, ..., k$. Then $B \cap C^k$ is a level-p supported restriction from C^k by J^B.

proof: similar to the proof of Lemma 3, therefore omitted.

proof of Claim 12: by the definition of a level-p supported restriction, if B is a level-p supported restriction from A by J for some $p \in (0,1)$, then B is a level-q supported restriction from A by J for any $q \in (p,1)$. Therefore $A^1(q) \subset A^1(p)$ for every $p \in (0,1)$ and $q \in (p,1)$. Now suppose for some $k \geq 0$ $A^k(q) \subset A^k(p)$ for every $p \in (0,1)$ and $q \in (p,1)$. Then using again that if B is a level-p supported restriction from A by J for some $p \in (0,1)$, then B is a level-q supported restriction from A by J for any $q \in (p,1)$, and using Lemma 7, $A^{k+1}(q) \subset A^{k+1}(p)$ for every $p \in (0,1)$. By induction, $A^k(q) \subset A^k(p)$ $\forall$ $k \geq 0$, $p \in (0,1)$ and $q \in (p,1)$. This establishes that $A^\infty(q) \subset A^\infty(p)$ $\forall$ $p \in (0,1)$ and $q \in (p,1)$, so $A^\infty(p)$ is decreasing in p. Since S is finite, this implies there is $p \in (0,1)$ such that $A^\infty(q) = A^\infty(p)$ $\forall$ $q \in (p,1)$. Then $\bigcap_{p:p\in(0,1)} A^\infty(p) = A^\infty(q) \neq \emptyset$. QED

proof of Lemma 9: first suppose that B is not a supported restriction from C by J. Then there is j, b_j, a_j, f_{-j} and g_{-j} such that $j \in J$, $c_j \in C_j/B_j$, $f_{-j} \in \Omega_{-j}(C)$, $g_{-j} \in \Theta_{-j}(B)$ $c_j \in BR_j(f_{-j})$, $b_j \in BR_j(g_{-j})$, $\widehat{g}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$ and

$$u_j(c_j, f_{-j}) \geq u_j(b_j, g_{-j}) \tag{7}$$

Since $c_j \in P_j^0$ (by construction none of the elements of C_j/B_j are weakly dominated), there exists $h_{-j} \in \overline{\Omega}_{-j}$ such that

$$u_j(c_j, h_{-j}) \geq u_j(s_j, h_{-j}) \ \forall \ s_j \in S_j \tag{8}$$

Now for $p \in (0,1)$ construct the following correlated beliefs. Let $d_{-j} = p \cdot f_{-j} + (1-p) \cdot h_{-j}$ and let $m_{-j} = p \cdot g_{-j} + (1-p) \cdot h_{-j}$. Note that $d_{-j} \in \Omega_{-j}^p(C)$, $m_{-j} \in \Omega_{-j}^p(B)$, $\widehat{d}_{-J}(s_{-J}) = \widehat{m}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$ and $c_j \in BR_j(d_{-j})$. Combining (**??**) and (8) yields $u_j(c_j, d_{-j}) \geq u_j(s_j, m_{-j})$ $\forall$ $s_j \in S_j$, since the best response against m_{-j} yields a lower expected payoff than $p \cdot u_j(b_j, d_{-j}) + (1-p) \cdot u_j(\widetilde{s}_j, h_{-j})$ where $\widetilde{s}_j$ is a best response to h_{-j}. This means B cannot be a level-p supported restriction from A by J, since $c_j \in A_j/B_j$.

Now suppose that B is a supported restriction from C by J. For every $c_j \in C_j/B_j$, let $U_{-j}(c_j) = \{u \mid u = u_j(b_j, h_{-j}) - u_j(c_j, f_{-j})$ for some f_{-j}, h_{-j} and b_j such that $f_{-j} \in \Omega_{-j}(C)$, $h_{-j} \in \Omega_{-j}(B)$, $c_j \in BR_j(f_{-j})$, $b_j \in BR_j(h_{-j})$ and $\widehat{h}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}\}$. It is easy to establish that $U_{-j}(c_j)$

is closed, and because B is a supported restriction from C by J, $u > 0$ $\forall$ $u \in U_{-j}(c_j)$. Therefore $\underline{u}(c_j) \equiv \min_{u \in U_{-j}(c_j)} u$ is well-defined and positive, and so is $\delta = \min_{c_j : c_J \in C_J / B_J} \underline{u}(c_j)$ since that is a minimum of a finite set of positive numbers. Then $u_j(b_j, h_{-j}) - \delta/2 > u_j(c_j, f_{-j})$ $\forall$ j, b_j, c_j, f_{-j}, h_{-j} such that $j \in J$, $f_{-j} \in \Omega_{-j}(C)$, $h_{-j} \in \Omega_{-j}(B)$, $c_j \in BR_j(f_{-j})$, $b_j \in BR_j(h_{-j})$ and $\widehat{h}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$. Now let $\Delta = \max_{j, s_j, t_j : j \in J, \ s_j, t_j \in S_j} [u_j(s_j) - u_j(t_j)]$ (the maximum payoff difference in the game) and $\varepsilon = \frac{\delta}{4\Delta}$.

Now let $\widehat{h}_{-j} \in \Theta_{-j}^{1-\varepsilon}(B)$, $\widehat{f}_{-j} \in \Theta_{-j}^{1-\varepsilon}(C)$, $\widehat{c}_j \in C_j / B_j$, $\widehat{c}_j \in BR_j(\widehat{f}_{-j})$, $\widehat{b}_j \in BR_j(\widehat{h}_{-j})$ and $\widehat{h}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$. These conditions imply that we can decompose $\widehat{h}_{-j}$ and $\widehat{f}_{-j}$ the following way: $\widehat{h}_{-j} = (1 - \varepsilon) \cdot h_{-j} + \varepsilon \cdot \gamma g_{-j}$ and $\widehat{f}_{-j} = (1 - \varepsilon) \cdot f_{-j} + \varepsilon \cdot \zeta d_{-j}$, where $f_{-j} \in \Omega_{-j}(C)$, $h_{-j} \in \Omega_{-j}(B)$, $g_{-j}, d_{-j} \in \overline{\Omega}_{-j}$ and $\widehat{h}_{-J}(s_{-J}) = \widehat{f}_{-J}(s_{-J})$ $\forall$ $s_{-J} \in S_{-J}$. Now, by construction $u_j(\widehat{b}_j, \widehat{h}_{-j}) - u_j(b_j, h_{-j}) \leq \delta/4$ $\forall$ $b_j \in BR_j(h_{-j})$ and $g_j(\widehat{c}_j, \widehat{\theta}_{-j}) - g_j(c_j, \theta_{-j}) \leq \delta/4$ $\forall$ $c_j \in BR_j(\theta_{-j})$. Combining these with $g_j(b_j, \tau_{-j}) - \delta/2 > g_j(c_j, \theta_{-j})$ from above implies $u_j(\widehat{b}_j, \widehat{h}_{-j}) - u_j(\widehat{c}_j, \widehat{f}_{-j}) > 0$, which establishes that B is an level-$(1 - \varepsilon)$ supported restriction from C by J. QED

proof of Claim 13: suppose first that $s \in P^\infty$. Assume s is not perfectly coalitionally rationalizable. Then there is p such that $s \notin A^\infty(p)$, which implies that there is $k \geq 1$, B and J such that $s \notin B$ and B is a level-p suported restriction from $A^k(p)$. By Lemma 8 this means that there are $C_1, ..., C_k$ and $J_1, ..., J_k$ such that C_1 is a supported restriction from P^0 by J_1, C_n is a supported restriction from C_{n-1} by J_n for n=2,...,k and B is a supported restriction from C_k by J. But then $P^{k+1} \subset B$, which establishes $s \notin P^\infty$, a contradiction.

Suppose now that $s \notin P^\infty$. Then either $s \notin P^0$, in which case s is trivially not perfectly coalitionally rationalizable, or there are $C_1, ..., C_k$ and $J_1, ..., J_k$ such that C_1 is a supported restriction from P^0 by J_1, C_n is a supported restriction from C_{n-1} by J_n for $n = 2, ..., k$ and $s \notin C_n$. By Lemma 8 that means there are $p_1, ..., p_n$ such that C_1 is a level-p_1 supported restriction from S by J_1 and C_n is a level-p_n supported restriction from C_{n-1} by J_n for $n = 2, ..., k$. But then for $p = \min_{n=1,...,k} p_n$, C_1 is a level-p supported restriction from S by J_1 and C_n is a level-p supported restriction from C_{n-1} by J_n for $n = 2, ..., k$, so $s \notin A^k(p)$. This implies that s is not perfectly coalitionally rationalizable. QED

proof of Claim 14: note that if $f_{-i} \in \Omega_{-i}^p(A)$, then $f_{-i} \in \Omega_{-i}^q(A)$ $\forall$ $q \in (0, p)$, so if a strategy is a best response against some conjecture from $\Omega_{-i}^p(A)$, then it is a best response against some conjecture from $\Omega_{-i}^q(A)$, $\forall$ $q \in (0, p)$. Then $A^\infty(p) \in \mathcal{M}'(p)$ $\forall$ $p \in (0, 1)$ and $P^\infty = \cap_{p \in (0,1)} A^\infty(p)$ imply that $\Sigma_i^{BR}(P^\infty)$ is exactly the set of strategies that are best responses against some $f_{-i} \in \Theta_{-i}^p(P^\infty)$ for every $p \in (0, 1)$. Now observe that for every $p \in (0, 1)$, $\Theta_{-i}^p(P^\infty)$ is exactly the set of conjectures that can be written as $p \cdot g_{-i} + (1 - p) \cdot h_{-i}$, where

$g_{-i} \in \Theta_{-i}(P^\infty)$ and $h_{-i} \in Int(\Omega_{-i})$. But then the above claims mean that $(\Sigma^{BR}(P^\infty), Int(\Sigma))$ is a τ-theory, and since irrational players are required to play every pure strategy with some positive probability, it is a perfect τ-theory. QED

proof of Claim 15: Note that $A^0 = S \in \mathcal{M}$. Now assume $A^k \in \mathcal{M}$ for some $k \geq 0$. By Claim 1, $A^{k+1} \neq \emptyset$. By Lemma 2, A^{k+1} can be reached from A^k by sequence of restrictions and then by Lemma 3, $A^{k+1} \in \mathcal{M}$. By induction, $A^k \neq \emptyset \ \forall \ k \geq 0$. By definition, all B that is a supported restriction from A^k are closed for any $k \geq 0$, so the intersection of them, A^{k+1} is closed too. Because S is compact, this implies that A^k is a nested sequence of nonempty compact sets, so its limit, A^∞ is nonempty and compact. Now assume $A^\infty \notin \mathcal{M}$. Then there are i, a_i and f_{-i} such that $i \in I$, $f_{-i} \in \Omega_{-i}(A^\infty)$, $a_i \notin A_i^\infty$ and $a_i \in BR_i(f_{-i})$. But then for high enough k, a_i and f_{-i} satisfy $f_{-i} \in \Omega_{-i}(A^k)$, $a_i \notin A_i^k$ and $a_i \in BR_i(f_{-i})$, contradicting $A^k \in \mathcal{M}$. QED

13 References

- Aumann, R. J. : Acceptable points in general cooperative n-person games; in "Contribuitions to the theory of games IV", Princeton University Press, Princeton, NJ, 1959

- Aumann, R. J. : Nash equilibria are not self-enforcing; in Economic Decision-making: "Games, Econometrics and Optimisation", Elsevier Science Publishers B.V., 1990
- Basu, K. and Weibull, J. W. : Strategy subsets closed under rational behavior, 36, Economics Letters, 1991, p141-146

- Bernheim, B. D. : Rationalizable strategic behavior, Econometrica, 52, 1984, p1007-1028
- Bernheim, B. D., Peleg B. and Whinston, M. D. : Coalition-proof Nash equilibria I. Concepts; Journal of Economic Theory, 42, 1987, p1-12
- Bernheim, B. D. and Whinston, M. D. : Coalition-proof Nash equilibria II. Applications; Journal of Economic Theory, 42, 1987, p13-29

- Bernheim, B. D. and Whinston, M. D. : Menu auctions, resource allocation, and economic influence; Quarterly Journal of Economics, 101, 1986, p1-31
- Borgers, T. : Weak dominance and approximate common knowledge; Journal of Economic Theory, 64, 1994, p256-276
- Charness, G. : Self-serving cheap talk: a test of Aumann's conjecture; Games and Economic Behavior, 33, 2000, p177-194

- Chwe, M. S. : Farsighted coalitional stability; Journal of Economic Theory, 63, 1994, p299-325

- Cooper, R., DeJong, D. V., Forsythe, R., Ross T. W. : Communication in coordination games; Quarterly Journal of Economics, 1992, p739-771

- Dekel, E. and Fudenberg, D. : Rational behavior with payoff uncertainty; Journal of Economic Theory, 52, 1990, p243-267

- Farrell, J. : Communication, coordination and Nash equilibrium; Economics Letters 27, 1988, p209-214

- Farrell, J. : Meaning and credibility in cheap talk games; Games and Economic Behavior 5, 1993, p514-531

- Farrell, J. and Rabin, M. : Cheap talk; Journal of Economic Theory 10, 1996, p103-118

- Gul, F. : Rationality and coherent theories of rational behavior; Journal of Economic Theory, 70, 1996, p1-31

- Herings, P. J. and Vannetelbosch, V. J. : The equivalence of the Dekel-Fudenberg iterative procedure and weakly perfect rationalizability; Economic Theory 15, 2000, p677-687

- Herings, P. J. and Vannetelbosch, V. J. : Refinements of rationalizability for normal-form games, International Journal of Game Theory, 28, 1999, p53-68

- Mariotti, M. : A model of agreements in strategic form games; Journal of Economic Theory, 74, 1997, p196-217

- Myerson, R. B. : Credible negotiation statements and coherent plans; Journal of Economic Theory 48, 1989, p264-303

- Pearce, D. G. : Rationalizable strategic behavior and the problem of perfection, Econometrica 52, 1984, p1029-1050

- Rabin, M. : A model of pre-game communication; Journal of Economic Theory 63, 1994, p370-391

- Rabin, M. : Incorporating behavioral assumptions into game theory; IN: Problems of coordination in economic activity, ed. by James Friedman, Boston and Dordrecht: Kluwer Academic, 1994, p 69-87

- Rabin, M. : Communication between rational agents; Journal of Economic Theory 51, 1990, p144-170

- Ray, D. and Vohra, R. : A theory of endogeneous coalition structures; Games and Economic Behavior 26, 1999, p286-336

- Ray, D. and Vohra, R. : Equilibrium binding agreements; Journal of Economic Theory 73, 1997, p30-78

- Rubinstein, A. and Wolinsky, A. : Rationalizable conjectural equilibria: between Nash and rationalizability; Games and Economic Behavior, 6, 1994, p299-311

- Tan, T. and Werlang, S. : The Bayesian foundation of solution concepts of games; Journal of Economic Theory, 45, 1988, p370-391

- Watson, J. : Communication and superior cooperation in two-player normal form games; Economics Letters 35, 1991, p267-271

- Xue, L. : Coalitional stability under perfect foresight; Economic Theory, 11, 1998, p603-627

- Xue, L. : Negotiation-proof Nash equilibrium; International Journal of Game Theory, 29, 2000, p339-357

- Zapater, I. : Credible proposals in communication games; Journal of Economic Theory 72, 1997, p173-197