Coalitional Rationalizability*

Attila Ambrus

This Version: July 2005

Abstract

This paper investigates how groups or coalitions of players can act in their collective interest in non-cooperative normal form games even if equilibrium play is not 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. An iterative procedure of restrictions is used to define a non-cooperative solution concept, the set of coalitionally rationalizable strategies. The procedure is analogous to iterative deletion of never best response strategies, but operates on implicit agreements by different coalitions. The solution set is a nonempty subset of the rationalizable strategies.

I. Introduction

The main solution concept in noncooperative game theory, Nash equilibrium, requires stability only with respect to individual deviations by players. It does not take into account the possibility that groups of players might coordinate their moves, in order to achieve an outcome that is better for all of them. There have been several attempts to incorporate the latter consideration into the theory of noncooperative games. The most widely-known solution concepts proposed along these lines are strong Nash equilibrium (Aumann [1959]) and coalition-proof Nash equilibrium (Bernheim et al. [1987]).

However, all of the concepts introduced so far are burdened with the problem of not being able to guarantee the existence of a solution in a natural class of games. This casts doubt on the validity of the solution they provide even in games in which a solution exists. Nonexistence is especially severe in the case of strong Nash equilibrium. Coalition-proof Nash equilibrium exists in a larger set of games, but at the cost of imposing debatable restrictions on which coalitions of players can make agreements with each other. And even so it cannot get around the nonexistence issue. Despite the conceptual problems associated with the equilibrium concept, it is used to derive predictions in a wide range of economic models. Examples include menu auctions (Bernheim and Peleg [1986]), dynamic public good provision games (Chakravorti [1995]), bankruptcy rules (Dagan et al. [1997]), principle-agent games (Gupta et al. [1998]), corporate takeovers (Noe [1998]), common agency games (Konishi et al. [1999]) and oligopoly competition (Delgado and Moreno [2004]). This is a clear indication that in a wide range of settings the consideration that players with similar interest try to coordinate their play is thought to play an important role.

This paper proposes a new solution concept, coalitional rationalizability, to address the issue of coalitional agreements. We restrict attention to normal-form games, although the principles we propose can be applied to more general settings.

The main conceptual innovation is that we depart from the tradition of imposing coalitional restrictions in an equilibrium setting. We present a non-equilibrium theory, like rationalizability (Bernheim [1984] and Pearce [1984]). In particular, we do not assume that players' expectations concerning play have to be correct. The coalitional agreements players can consider in this context take the form of restrictions of the strategy space. To put it differently, in our construction players look for agreements to avoid certain strategies, without specifying play within the set of nonexcluded strategies. These agreements are therefore more flexible than the ones that uniquely pin down a strategy profile for a coalition. Furthermore, the nonequilibrium nature of the agreements implies that players in a coalition can rationalize an agreement differently (they can have different expectations concerning play after the agreement is made). Both of the above points imply that the agreements we consider are more compatible

with each other than equilibrium agreements. We show that this ultimately leads to nonemptyness of the solution concept we propose.

Potential agreements by coalitions are evaluated on the basis of whether they are of mutual interest of the players involved. We assume that players are Bayesian decision makers and compare expected payoffs associated with making the agreement versus not making it. In particular, we call a restriction supported if every group member always (for every possible expectation) expects a higher payoff if the 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 to these conjectures yield, then the above requirement means 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.

We assume that players go through the following reasoning. First every coalition of players looks for supported restrictions given the set of all strategies. Then players consider the set of strategy profiles that are consistent with all the above restrictions (their conjectures are restricted to be concentrated on these strategies) and look for further restrictions, given this smaller set of strategies. They continue this procedure and restrict the set of strategies further and further, until they reach a point at which there is no supported restriction by any coalition given the set of strategies that survived the procedure so far. This procedure is analogous to iterative deletion of never best response strategies in that the order of restrictions does not matter. Furthermore it deletes all strategies that the latter procedure eliminates. The new feature is that groups of players, as opposed to just individual players, can delete strategies.

The solution concept this paper proposes, the set of coalitionally rationalizable strategies, is the set of profiles that survive the above procedure of iterated supported restrictions by coalitions.

The interpretation we provide for this process is that it reflects a reasoning procedure by the players. The restrictions on players' conjectures come from introspection, based on the publicly known payoffs of the game. We assume that there is no pre-play communication among players, and therefore the agreements implied by supported restrictions are implicit.

We also provide a direct definition of the set of coalitionally rationalizable strategies, one that does not refer to an iterative procedure. However, the procedural definition has the advantage of being constructive. To obtain the set of coalitionally rationalizable strategies in any game, one can just use the iterative procedure in the definition. Furthermore, this procedure is computationally relatively undemanding, implying that coalitional rationalizability is easy to use in applications.

Section 2 provides three simple examples to provide intuition on the logic of coalitional restrictions. The first one is a voting game with three individuals and

3

costly voting. The second one features consumers choosing between competing platforms in a two-sided market with network externalities. The third one is a game in which players have to divide a pie, with an external reward being promised to them in case they all offer fair allocations. These examples also show that coalitional reasoning can lead to sharp predictions in certain games, even if it does not assume that players' expectations necessarily have to be correct. In the first two examples it leads to an equilibrium profile (so the correctness of expectations is established as a result, as opposed to being assumed in the first place), while in the third one it significantly reduces the set of possible outcomes.

Section 3 provides the formal construction of the theory, defining supported restrictions and the iterative procedure that obtains the solution set. We establish several properties of this procedure and show that the set of coalitionally rationalizable strategies is nonempty.

In Sections 4 and 5 we relate the solution set to other non-cooperative solution concepts and examine connections with Pareto efficiency.

II. Motivating Examples

In this Section we provide examples on how coalitional reasoning can help players narrow down the set possibilities they should consider, and to coordinate their action choices.

II.A. Voting with Costly Participation

Consider the following voting game. There are three potential voters, and three possible outcomes. Assume there is a status quo outcome S and that it can only be changed if at least two of the voters show up and vote for the same new alternative. There are two potential new alternatives to vote for, A and B. Showing up and casting a vote costs $\varepsilon \in (0,1)$. Voters rank the possible outcomes the following way. Voter 1's favorite outcome is A, then B, then S. Voter 2's favorite outcome is A, then S, then B. Finally, voter 3's favorite outcome is B, then S, then A. Assume that for every voter her favorite outcome yields a payoff of 2, her second favorite outcome yields 1 and her least favorite outcome has 0 (minus ε in each case if she showed up to vote). This game has multiple equilibria, and in fact for all three outcomes in the game there is some Nash equilibrium that yields that outcome. In one equilibrium all three voters stay at home and the status quo outcome prevails. In another equilibrium voters 1 and 3 show up and vote for outcome B, while voter 2 stays home. And in yet another equilibrium voters 1 and 2 show up and vote for outcome A, while voter 3 stays at home. The above equilibria are not Pareto-ranked (note that each voter has a distinct least favorite outcome, so in any equilibrium someone's

4

least favorite outcome is chosen). On the other hand, the following reasoning procedure selects a unique equilibrium in the game.

First, note that it is never rational (it is strictly dominated) for voter 3 to show up and vote for alternative A, since voting is costly and A is her least preferred outcome. Given that, the best possible outcome in the game for voters 1 and 2 is if they show up and vote for A. Although this point is fairly straightforward, let us make it more rigorous, to demonstrate the logic of supported restrictions that we formally define in the next section. Fix any conjecture concerning player 3's strategy that allocates zero probability to player 3 voting for A. Then the following are true. If players 1 and 2 both show up and vote for A, they expect a payoff of $2 - \varepsilon$ for sure. If player 1 shows up and votes for B, her expected payoff cannot be higher than $1 - \varepsilon$. If player 1 stays at home, her expected payoff cannot be higher than 1. The same are true for player 2, too (with the addition that voting for B is not even rational for her). Therefore for any possible conjecture concerning player 3's strategy, both player 1 and player 2 are strictly better off coordinating on playing A than not coordinating and playing some other strategy. Completing the argument, if player 1 and player 2 act accordingly, then outcome A is indeed implemented, no matter whether player 3 stays at home or shows up and votes for B. Given this, voter 3 should conclude that she is better off staying at home and not voting.

II.B. Platform Choices with Network Externalities

The second example features consumers choosing between competing platforms on a two-sided market with network externalities.[1] This means that there are two distinct set of consumers who need a common platform to interact with each other, and the payoff of a consumer depends on how many other consumers join the same platform from the other side of the market. There are many examples of markets having these characteristics, but for ease of exposition assume that the two sides are men and women looking for dates, and that the platforms are different matchmaking services. Suppose the matchmaking services use registration fees as a pricing instrument and that they can charge different prices on different sides of the market. Furthermore, suppose there is a large number of men and women on both sides of the market and that the payoff resulting from joining a matchmaking service depends linearly on the ratio of people of the other gender joining the same matchmaking service. In particular, if man i joins a network which charges registration fee p and which is joined by a fraction x of the women who are on the dating market, then his payoff is $u_i^m x - p$. Similarly, if woman i joins a network which charges registration fee p and which is joined by a fraction x of the men, then her payoff is $u_i^w x - p$. The individual-specific parameters u_i^m and u_i^w determine how much someone on the dating market cares about having a lot of choices.

---

[1] See Ambrus and Argenziano [2004] for a more general analysis of applying coalitional rationalizability in this setting.

Assume now the following concrete specification. Sixty per cent of the men care a lot about how many women are on the same network, having $u_i^m = 2$. The remaining forty per cent of men care much less about this, having $u_i^m = 0.4$. The distribution of this parameter is exactly the reverse for women. For forty per cent of them $u_i^w = 2$ and for sixty per cent of them $u_i^w = 0.4$. For ease of exposition call men and women with intensity parameter 2 "intensely searching" and those with intensity parameter 0.4 "moderately searching". Assume that there are two matchmaking services in the market, A and B, and that A charges a fee of 0.5 on men and 0.2 on women, while B charges a fee of 0.2 on men and 0.5 on women. Consumers decide simultaneously which one of them to join. At first, even if we assume that consumers try to coordinate along the lines of common interest, it is not clear which network they should coordinate on. Service A is more expensive for men but cheaper for women and service B is the other way around. There is a conflict of interest between the two sides. And indeed there are starkly different Nash equilibria of this game, corresponding to consumers coordinating on either network A or network B or not being able to coordinate and staying at home. However, the logic of coalitional restrictions selects a unique outcome in this game, as in the previous example. First note that for moderately searching women it is never a best response to join network B, since even if all men join this network, their net utilities are negative (similarly, for moderately searching men it is never a best response to join network A). Given this, intensely searching men and moderately searching women are strictly better off if they coordinate on joining network A. To see this, note that an implicit agreement along these lines guarantees an expected payoff of $1.2 - 0.5 = 0.7$ to intensely searching men, while joining network B can at most yield an expected payoff of $0.8 - .02 = 0.6$ for them, and not joining any network yields 0 for sure. As far as moderately searching women are concerned, the implicit agreement guarantees an expected payoff of $0.24 - 0.2 = 0.04$ for them, which is strictly higher than the payoff of not joining any service. Concluding the argument, given that the above group of consumers all join network A, the unique best response for intensely searching women is also joining network A, while the unique best response for moderately searching men is not joining any network.

II.C. Dollar Division Game with External Reward

The third example we provide demonstrates that even if coalitional reasoning does not lead to a unique prediction in a game, it can considerably narrow down the set of possible outcomes. Consider a classic dollar division game with the additional twist that players receive an external reward in the event that players behave "nicely". Concretely, three players vote secretly and simultaneously on how to divide a dollar. If two or more players vote for the same allocation, the dollar is divided accordingly, otherwise the dollar is lost to the players. The added element 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 $100 reward for the group being "generous", independently of what happens to

the original dollar (in particular, even in the event that it is not allocated to the players because of lack of agreement).[2] In this game coordinating on voting for allocations that give at least $1/4$ dollar to every player is unambiguously mutually advantageous 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 regarding how to allocate the "last quarter", but they have a strong incentive to propose at least $1/4$ dollar to every player, since the external reward is much bigger than the stake that is to be divided. Players who coordinate along the line of common interest therefore should expect each other to propose allocations (x_1, x_2, x_3) such that $x_i \geq 1/4$, $\forall\ i \in \{1, 2, 3\}$, even if they are uncertain about exactly what allocations the others propose. For more on this game, see Section 4.

III. Construction and the Main Results

III.A. Notation and Basic Definitions

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 (by a mixed strategy),[3] 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 Δ_{-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 $f_{-i} \in \Delta_{-i}$ let f_{-i}^{-J} be the marginal distribution of f_{-i} over S_{-J}.

---

[2] This game can be interpreted as a simple model of the following situation. Several political parties in a post-war country trying to form a coalitional government, and an international organization makes a credible promise to provide a large amount of financial aid to the country if during the negotiations parties do not try to squeeze out any of the participants from power.

[3] For the proof of this well-known result see for instance Fudenberg and Tirole [1992], pp. 52-53.

We will compare expectations of players under different conjectures. For every $f_{-i} \in \Delta_{-i}$ and $s_i \in S_i$ let $u_i(s_i, f_{-i}) = \sum\limits_{t_{-i}:t_{-i} \in S_{-i}} u_i(s_i, t_{-i}) \cdot f_{-i}(t_{-i})$ 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 $f_{-i} \in \Delta_{-i}$ 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 $B \subset S$ such that $B \neq \emptyset$ and $B = \underset{i \in I}{\times} B_i$, let $\Omega^*_{-i}(B_i) = \{f_{-i} \mid f_{-i} \in \Delta_{-i}, \exists b_i \in B_i \text{ such that } b_i \in BR_i(f_{-i})\}$. 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. We require these restrictions to be product sets. For any A such that $A \subset S$ and $A = \underset{i \in I}{\times} A_i$, let $\Delta_{-i}(A) = \{f_{-i} \mid \text{supp} f_{-i} \subset A_{-i}\}$, the set of conjectures player i can have that are concentrated on A_{-i} (the set of conjectures that are consistent with player i believing that other players play inside A).

Certain product subsets of the strategy space play an important role in the construction below. These are sets that satisfy that whenever a player believes that others play inside the set then all her best responses are inside the set. We call these sets closed under rational behavior.[4]

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

(*) $A = \underset{i \in I}{\times} A_i$ *and* $A_i \subset S_i, \forall i \in I$

(**) $BR_i(f_{-i}) \subset A_i, \forall f_{-i} \in \Delta_{-i}(A), \forall i \in I.$

Let $\mathcal{M}$ denote the collection of sets closed under rational behavior.

We assume that players coordinate on restricting their play to a subset of the strategy space whenever by doing so each of them is guaranteed to get an expected payoff that is strictly higher than any expected payoff he could get if the restriction was not made and he played a strategy outside the restriction.

---

[4]The terminology was introduced by Basu & Weibull [1991].

These restrictions, which we call supported restrictions, constitute the main building block of the construction that follows.

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

DEFINITION. *B is a supported restriction by J given A if it satisfies the following two properties:*

(1) $B_i = A_i$, $\forall\, i \notin J$
(2) $\forall\, j \in J$, $f_{-j} \in \Omega^*_{-j}(A_j/B_j) \cap \Delta_{-j}(A)$ *it is the case that* $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$ $\forall\, g_{-j} \in \Delta_{-j}(B)$ *such that* $g^{-J}_{-j} = f^{-J}_{-j}$.

Let $A \searrow_J B$ denote that B is a supported restriction by J given A. Let $A \searrow B$ denote $A \searrow_J B$ for some $J \subset I$, $J \neq \emptyset$.

Supported restrictions are defined given a nonempty product subset of the strategy space. The motivation is that players are certain that play is inside this set.[5] Note that the set is allowed to be the set of all strategies. The first condition in the above definition requires that only the strategies of those players who are members of the given group be restricted. The second condition requires that for any player in the coalition, any belief to 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.

Intuitively, this definition considers the possible expected payoffs of a player if he chose to play a strategy to be excluded by the agreement, and compares them to payoffs he could expect if the restriction is made (if all the other players in the coalition confined their play to the restriction). Arguably this is the most natural payoff comparison to use in deciding whether it is in the interest of a player to agree upon not playing the strategies to be excluded by a restriction. A restriction is then called supported if the expected payoffs that are compatible with the agreement strictly Pareto dominate those that are associated with playing strategies outside the restriction, for any fixed conjecture concerning players' strategies outside the coalition. Intuitively, a restriction is supported if every coalition member prefers the agreement to playing a strategy outside the agreement, for every possible conjecture that he can have associated with the above two scenarios. Besides its intuitive appeal, the definition above has the advantage that supported restriction is a generalization of eliminating never best response strategies by individual players (see Proposition 4 below), which ensures that the theory we propose is consistent with individual rationality.

---

[5]For a rigorous formalization of this idea see Ambrus [2005], where coalitional rationalizability is introduced in an epistemic framework.

This property does not hold for definitions that are not based on comparing expected payoffs, for example if it is only required that for every player in the coalition every payoff within the restriction is strictly higher than every payoff outside the restriction.

The reason we fix marginal conjectures concerning the play of players outside the coalition 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. The other players, after going through whatever mental procedure they use to formulate beliefs, may or may not believe that members of the coalition play according to the restriction. The point is that they do not have a chance to find this out. Note however that, as discussed above, we require the payoff comparison condition to hold for all possible conjectures concerning the play of players outside the coalition.

Besides its intuitive appeal, the definition above has the advantage that supported restrictions satisfy two desirable properties. One is that the concept is a generalization of eliminating never best response strategies by individual players (see Proposition 4 below), which ensures that the theory we propose is consistent with individual rationality. This property does not hold for definitions that are not based on comparing expected payoffs, for example if it is only required that for every player in the coalition every payoff within the restriction is strictly higher than every payoff outside the restriction. The second is that supported restrictions are self-enforcing. Note that the second condition in the above definition 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 A_{-j}/B_{-j}. This implies that if A is closed under rational behavior and A is a supported restriction given B, then B is also closed under rational behavior. If a player believes that the others play according to a supported restriction, then it is in his best interest to play according to the restriction too.

Nevertheless, comparing sets of payoffs is a far from obvious exercise, therefore one might want to consider alternative definitions to supported restriction. We do not take up that task here, instead refer the interested reader to Ambrus [2005].

III.B. The Set of Coalitionally Rationalizable Strategies and its Basic Properties

In the construction that follows we refer to certain product subsets of the strategy space. Sets that are closed under rational behavior are defined in the previous subsection. A subset is coherent if it is closed under rational behavior and every strategy in the set is a best response to some conjecture concentrated on the set.[6]

---

[6]The terminology reflects that these sets satisfy the coherence requirement of Gul [1996] in that the strategies of a player that are inside the given set are implied by the restriction that

DEFINITION. *Set A is coherent if it is closed under rational behavior and satisfies:*

$$(3) \quad \bigcup_{f_{-i} \in \Delta_{-i}(A)} BR_i(f_{-i}) = A_i, \ \forall \ i \in I.$$

Now we define an iterative procedure using supported restrictions. Starting from the set of all strategies, in each step the intersection of all supported restrictions is retained. Below we show that generalized supported restrictions are compatible with each other in the sense that taking the intersection of all supported restrictions at any step of the procedure results in a nonempty set of strategies. This holds despite the fact that a player is part of many different coalitions and those coalitions may have different supported restrictions.

The procedure we define can be thought of as a descriptive theory of belief formation. Players, based on the strategies and payoff functions of the game, look for supported restrictions given the set of all strategies. This means that at the beginning of the procedure they consider any strategy profile to be possible to be played. If such restrictions are found, then they expect the players in the corresponding coalitions to play inside the restrictions (to successfully coordinate their moves to play inside the restrictions). This requirement restricts the set of possible beliefs they can have. Then they look for supported 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 allow.

For every $A \in \mathcal{M}$ let $\mathcal{F}(A)$ denote the collection of all supported restrictions given A.

DEFINITION. *Let $A^0 = S$. For every $k \geq 1$ let $A^k = \bigcap_{B \in \mathcal{F}(A^{k-1})} B$. Let $A^* = \bigcap_{k=0,1,2,\dots} A^k$.*

---

other players play inside the set and that allowable beliefs about other players should include the convex hull of action profiles in the restriction. Investigating sets of strategies satisfying this type of requirement was first undertaken by Bernheim [1984] and Pearce [1984]. Basu and Weibull [1991] call these sets strictly closed under rational behavior.

Next we show two properties of supported restrictions that are useful in establishing one of the main results in the paper, nonemptyness of A^*. Proposition 1 establishes that the intersection of all supported restrictions given a set that is closed under rational behavior is nonempty. Proposition 2 establishes that if B is a supported restriction given A and A is closed under rational behavior, then B remains a supported restriction given any set that is obtained from A by a sequence of generalized supported restrictions. This property guarantees the internal consistency of the iterative procedure of supported restrictions, the main step in establishing that the order in which restrictions are made is inconsequential in this iterative procedure (see Proposition 5 below).

PROPOSITION 1. *Let* $A \in \mathcal{M}$. *Then* $\underset{B: B \in \mathcal{F}(A)}{\cap} B \neq \emptyset$.

PROPOSITION 2. *Let* $A \in \mathcal{M}$. *Assume* $A \searrow_{J^B} B$. *Let* $C^0, ..., C^k$ $(k \geq 1)$ *be such that* $C^0 = A$ *and* $C^i \searrow_{J^i} C^{i-1}$, $\forall\, i = 1, ..., k$. *Then* $C^k \searrow_{J^B} (B \cap C^k)$.

All formal proofs are in the Appendix. The important insight in establishing Proposition 1 is that for every player the strategies that are best responses for this player for her most optimistic conjecture (the one that yields the highest expected payoff) concentrated on A have to be included in every supported restriction by every coalition that this player is a member of. This follows from the definition of a supported restriction. For proving Proposition 2 the key step is showing that given a set that is closed under rational behavior, the intersection of two supported restrictions by the same coalition is a supported restriction itself, by the same coalition (and therefore, by Proposition 1, nonempty).

Proposition 3, a central result in this section, establishes that the iterative procedure of supported restrictions stops in a finite number of steps, and the set it obtains is nonempty, closed under rational behavior and has the property that, given this set, no coalition has a nontrivial supported restriction. The key step in proving this is showing, using Propositions 1 and 2, that A^k is nonempty and closed under rational behavior for every k. The finiteness of the game then implies all the claims in the proposition.

PROPOSITION 3. A^* *satisfies the following properties:*
  (i)   nonempty
  (ii) closed under rational behavior
  (iii) $A^* \searrow B$ implies $B = A^*$
  (iv) $\exists\, K < \infty$ such that $A^k = A^*$ whenever $k \geq K$.

We note that although the iterative procedure is defined on pure strategies, allowing players to use mixed strategies would lead to the same pure strategies

and some (not necessarily all) of the mixed strategies with the same support. For more on this, see Subsection 6.3.

To obtain some other properties of A^* it is useful to identify supported restrictions by singleton coalitions. Proposition 4 establishes that these restrictions are equivalent to elimination of never best-response strategies.

PROPOSITION 4. *Let $A \in \mathcal{M}$ and $i \in I$. Then $A \searrow_{\{i\}} B$ iff $B = B_i \times A_{-i}$, $B_i \subset A_i$ and $s_i \in B_i/A_i$ implies $-\exists\, f_{-i} \in \Delta_{-i}(A)$ such that $s_i \in BR_i(f_{-i})$.*

Proposition 3 implies that, in particular, there is no supported restriction by any single-player coalition given A^*. Proposition 4 then implies that A^* satisfies (3). Furthermore, Proposition 3 establishes that $A^* \in \mathcal{M}$. Combining these results establishes that A^* is coherent. This immediately implies that A^* is contained in the set of rationalizable strategies, since the latter is the largest coherent set (see Bernheim [1984]).

We defined A^* as the set obtained by a particular iterative procedure that requires supported restrictions to be made in a particular order, namely making all possible generalized supported restrictions simultaneously at every step. The next claim establishes that the particular order of restrictions 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^*. This result is essentially a consequence of the one in Proposition 2.

PROPOSITION 5. *Let $B^0 = S$. If there is no nontrivial supported restriction given B^0, then let $B^1 = B^0$. Otherwise let Θ^0 be a nonempty collection of nontrivial supported 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 given B^k, otherwise let Θ^k be a nonempty collection of nontrivial supported restrictions given B^k and let $B^{k+1} = \underset{B:B\in\Theta^k}{\cap} B$. Then there is $L \geq 0$ such that $B^k = A^*, \forall\, k \geq L$.*

We conclude the section by showing that the set of coalitionally rationalizable strategies is stable with respect to supported restrictions given any superset and use this property to give a direct definition of the set, which does not refer to any iterative procedure.

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

DEFINITION. *A is externally coalitionally stable if for every $C \neq A$ such that $A \subset C$ it is the case that $A \subset \underset{B:\ B \in \mathcal{F}(C)}{\cap} B$ and $C \neq \underset{B:\ B \in \mathcal{F}(C)}{\cap} B$.*

DEFINITION. *A is internally coalitionally stable if $A \searrow B$ implies $B = A$.*

DEFINITION. *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.

PROPOSITION 6. *A^* is the only coalitionally stable set in G.*

The proof generalizes the arguments behind Proposition 2 and shows that (i) A^* is coalitionally stable; (ii) for any set A which is such that A/A^* is nonempty there is a superset of A given which there is a supported restriction that does not include A/A^*, implying that these sets cannot be coalitionally stable.

IV. Relating Coalitional Rationalizability to Other Solution Concepts

We examine the connections between the set of coalitionally rationalizable strategies and some standard noncooperative solution concepts. These are Nash equilibrium, and the two most commonly used noncooperative coalitional equilibrium concepts: coalition-proof Nash equilibrium and strong Nash equilibrium.

IV.A. Nash Equilibrium

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 1.

[INSERT FIGURE I HERE]

The only Nash equilibrium of the game is $(A2, B2)$. Nevertheless, the set of coalitionally rationalizable profiles is the whole game. For all other strategy profiles the sum of payoffs is negative. Still, if conjectures do not have to be correct, then other strategies can be played since players can expect positive payoffs in the negative-sum matching pennies game $\{A1, A3\} \times \{B1, B3\}$.

Furthermore, as seen in previous examples, the set of Nash equilibria is not contained in the set of coalitionally rationalizable strategies. However, it is straightforward to show that there is always *at least one Nash equilibrium* of every game that is *inside the set of coalitionally rationalizable strategies*. This is a direct implication of the result that A^* is closed under rational behavior: any Nash equilibrium of the game with restricted strategy sets A^* (which as a finite normal-form game, has at least one Nash equilibrium in mixed strategies) is also a Nash equilibrium of G.

IV.B. Coalition-proof Nash Equilibrium and Strong Nash Equilibrium

The first question we address is whether there can be nontrivial supported restrictions in games in which coalition-proof Nash equilibrium does not exist. The next example demonstrates that the answer is yes. Coalitional rationality can have bite in these games, in the sense that the set of coalitionally rationalizable strategies is strictly smaller than the set of rationalizable strategies.

The dollar division game is a classic example of a game in which coalitional equilibrium concepts do not exist. Consider the version of the game presented in Section 2. The intuition that players should expect each other to vote for generous allocations so that the group can receive the external reward 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 given the set of all strategies, and it is the set of coalitionally rationalizable strategies. The game does not have any coalition-proof Nash equilibrium.

The next question we ask is whether the set of outcomes consistent with some coalition-proof Nash equilibrium is contained in the set of coalitionally rationalizable strategies. The following example demonstrates that the answer is no.

[INSERT FIGURE II HERE]

In the game of Figure 2 $(A3, B3, C3)$ is the unique coalition-proof equilibrium (even allowing for mixed strategies). It is straightforward to establish that there is no self-enforcing profile in which players only play strategies inside $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ 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.[7] Next, observe that in any Nash equilibrium a player can only play his third strategy with positive probability if at least one of the other players plays his third strategy with probability 1 (otherwise the third strategy cannot be a best response). But then in any self-enforcing profile at least two players have to play their third strategies with probability 1. 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. Hence, 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. Furthermore, it is a self-enforcing profile since 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 strategies is $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ (this is the set we get after the first round of the iterative procedure at which point 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 disjoint sets in this game.

Note that $(A3, B3, C3)$ is a coalition-proof equilibrium only because the game with set of strategies $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ does not have a coalition-proof equilibrium. On the other hand, all players would strictly prefer to switch play to that part of the game no matter what happens over there (they cannot agree upon a concrete profile, but they would all agree to restricting their moves to $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$). We think that the prediction that players will play something inside $\{A1, A2\} \times \{B1, B2\} \times \{C1, C2\}$ is more reasonable than predicting that players will play the profile $(A3, B3, C3)$.

The above game could easily be made generic by some small perturbation of the payoffs, such 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.

We conclude this subsection by establishing that, in generic games, *only* the nonexistence of coalition-proof Nash equilibrium in some restriction of the original game can result in coalition-proof Nash equilibria not to be coalitionally rationalizable. Proposition 7 establishes that if in every restriction of the

---

[7]In a three-player game a profile is self-enforcing iff it satisfies the following two properties: (i) it is a Nash equilibrium; (ii) no coalition of two players can deviate to a profile that is a Nash equilibrium in the component game induced on them by the third player's strategy in such a way that both of them are strictly better off. For the general definition of self-enforcing see Bernheim et al. [1987].

original game there exists a coalition-proof Nash equilibrium, then all the pure strategy coalition-proof Nash equilibria of the original game are coalitionally rationalizable. The main argument in the proof is as follows. If a profile is not included in the set of coalitionally rationalizable strategies then there is k and a coalition J such that the profile is included in A^k, but there is a supported restriction B by J given A such that the profile is not included in B. Then it can be shown that coalition J could deviate in a credible and profitable manner from the original profile to any coalition-proof Nash equilibrium of the game restricted to B, contradicting that the profile is a coalition-proof Nash equilibrium itself. The result can be extended to mixed strategy coalition-proof Nash equilibria by requiring that every game which is obtained from the original game by fixing a mixed strategy profile for some players and restricting the strategies of the other players to a subset of their original strategy set has a coalition-proof Nash equilibrium.

For every $A \subset S$ such that $A = \underset{i \in I}{\times} A_i \neq \emptyset$ let $G[A] = (I, A, u_A)$ denote the normal form game for which $u_A(s) = u(s), \forall s \in A$. It is the game obtained from G by restricting the set of strategies to A.

PROPOSITION 7. *If for every $A \subset S$ such that $A = \underset{i \in I}{\times} A_i \neq \emptyset$ it holds that $G[A]$ has a coalition-proof Nash equilibrium then every pure strategy coalition-proof Nash equilibrium of G is contained in A^*.*

Proposition 7 and the example of Figure 2 illustrate that the issue of nonexistence confounds the iteratively defined solution concept of coalition-proof Nash equilibrium. This is not surprising given the iterative definition of the concept. The set of coalition-proof Nash equilibria of a game directly depends on coalition-proof Nash equilibria of restrictions of the original game, therefore coalition-proof Nash equilibrium can only be trusted to give a reasonable prediction if it gives a reasonable prediction in every restriction of the game (namely if it exists in every restriction). It is unclear to us whether in the latter class of games coalition-proof Nash equilibria give more reasonable prediction than other Nash equilibria that lie within the set of coalitionally rationalizable strategies. This is an issue we hope to return to in future work.

Finally we investigate relations to strong Nash equilibria: profiles that satisfy that no coalition of players have a profitable joint deviation. The latter is a very strong stability requirement, resulting in nonexistence of strong Nash equilibrium in many games. The next proposition establishes that, unlike coalition-proof Nash equilibria, every strong Nash equilibrium of every game must be contained in the set of coalitionally rationalizable strategies. This means that there is no inherent contradiction between supported restrictions in a non-equilibrium setting and group deviations in an equilibrium setting. Inconsistencies only arise if the set of allowable coalitional deviations are restricted according to the definition of coalition-proof Nash equilibrium.

PROPOSITION 8. *Let $\sigma = (\sigma_1, ..., \sigma_I)$ be a strong Nash equilibrium profile. Then supp$\sigma \subset A^*$.*

V. Pareto Efficiency

Coalitional rationalizability takes the interest of coalitions other than the coalition of all players into account, therefore in general it does not guarantee Pareto efficiency. There is no containment relationship between the set of coalitionally rationalizable strategies and the set of Pareto undominated profiles, or rationalizable profiles that are Pareto undominated by other rationalizable profiles, or Nash equilibria that are Pareto undominated by other Nash equilibria.

The fact that groups of players expect each other to pursue common gains can make all of them worse off, as the game of Figure 3 shows.

[INSERT FIGURE III HERE]

In this game the (strict) Nash equilibrium profile $(A1, B1, C1)$ 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 $C2$, making $(A2, B2, C2)$ the only coalitionally rationalizable profile, which is strictly Pareto-dominated by $(A1, B1, C1)$.

This example might seem puzzling, given that coalitional rationalizability is built on the assumption that players try to attain common gains. However, if not only the coalition of all players, but also subcoalitions of players act along these lines, then nothing guarantees Pareto efficiency of the resulting outcome. The fact that coalitions cannot commit not to go for a common gain can make them worse off. We view this as a coalitional version of the insight that is obtained from the prisoner's dilemma, where the fact that players cannot commit not to play individually rational strategies makes both of them worse off. Although in the above example Pareto inefficiency results from the fact that there is a highlighted coalition that cannot commit not to go for coalitional gains, it can be shown that inefficiency can arise in perfectly symmetric games as well.[8]

Pareto efficiency can be guaranteed only in special classes of games. For example it is easy to establish that in games that have a Pareto dominant profile among rationalizable strategies, that profile is the unique element in the set of coalitionally rationalizable strategies. Also, in 2-player games the support of every Pareto-undominated Nash equilibrium is contained in the set of coalitionally rationalizable strategies. The proofs of these claims are straightforward and therefore omitted.

---

[8] See the previous version of the paper for an example.

18

VI. Discussion

In this section we comment on various aspects of the construction.

VI.A. Epistemic Definition

The definition of coalitional rationalizability provided in this paper is along the lines of the original definition of rationalizability (see Pearce [1984] and Bernheim [1984]). It is characterized by an iterative procedure which has an intuitive interpretation: players expect all supported restrictions to be made given the set of all strategies, and they expect all supported restrictions to be made given the set of strategies that are consistent with the above requirement, and so on. The main advantage of this definition, besides that it refers to an intuitive reasoning procedure, is that it is constructive, making the set of rationalizable strategies relatively easy to compute in examples and applications. However, it does not directly reveal what primitive assumptions on players' beliefs and action choices imply that they play coalitionally rationalizable strategies. The set of rationalizable strategies was subsequently shown to be equivalent to the set of strategies compatible with rationality and common certainty of rationality (Tan and Werlang [1988], Brandenburger and Dekel [1993]), using the formalism of interactive epistemology. This raises the question whether the set of coalitionally rationalizable strategies has a similar interpretation. The direct characterization of Proposition 6 is a step in this direction. We do not pursue the issue further in this paper, instead referring the interested reader to Ambrus [2005]. That paper provides alternative, epistemic definitions of coalitional rationalizability.

VI.B. Mixed Strategies

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). However, the construction is also valid if we allow players to play mixed strategies. It is straightforward to extend the concept of supported restriction and then coalitional rationalizability to the space of mixed strategies. One can then show that coalitionally rationalizable mixed strategies are a subset of mixed strategies whose support is inside the set of coalitionally rationalizable pure strategies.[9] Furthermore, they include all coalitionally rationalizable pure strategies.

VI.C. Pre-play Communication

A natural issue to investigate is the interaction of pre-play communication and coalitional rationality. The particular questions that arise include whether all supported restrictions remain credible in a context with communication among

---

[9] The inclusion can be strict. This is analogous to the relationship between rationalizable pure and mixed strategies.

players, whether there are new restrictions that become credible and whether the set of solutions compatible with the theory still has product structure. Since these issues are complicated and possibly depend on the exact specification of the rules of communication, we leave the task of formally addressing the problem of coalitional agreements in a framework with communication to a future project.

We note though that there are reasons to believe that pre-play communication can have a role in determining whether or not players use coalitional reasoning. Coalitional rationalizability requires confidence that other players reason a particular way. Pre-play communication can help establish the necessary amount of trust for the restrictions involved. Experimental game theory provides some support for this claim. In certain coordination games, pre-play communication increases players' propensity to play Pareto optimal outcomes, and multi-sided pre-play communication may increase cooperation more than one-sided communication (see Cooper et al. [1992] and Charness [2000]).

VII. Related Literature

In Section 4 we related coalitional rationalizability to various equilibrium concepts. Besides the general solution concepts mentioned, there are several contributions in the literature that incorporate coalitional reasoning into the play of normal-form games in more specific settings.

Chwe [1994], Mariotti [1997], Xue [1998] and Xue [2000] assume that players engage in a possibly infinite negotiation procedure before playing a normal-form game. We share with these papers the assumption that coalitions can freely form and that binding agreements are not available. The main difference, aside from the fact that these papers consider agreements that can only specify a unique strategy profile to be played, is that the above papers assume that coalitions act publicly, therefore the negotiation is publicly observed.

Noncooperative coalitional bargaining considers extensive form noncooperative games to model n-player coalitional bargaining situations based on characteristic function games (or generalizations of those). These characteristic forms can be derived from normal-form games, as done in Ray and Vohra [1997] and Ray and Vohra [1999]. The main difference between this approach and ours is that these papers maintain an underlying assumption that members inside a coalition can make binding agreements and only the play among coalitions is in a noncooperative fashion. The current paper conforms with the tradition of noncooperative game theory and retains the assumption that players cannot make any binding agreements.

Rabin's concept of Consistent Behavioral Theories (see Rabin [1994b]) can be used to incorporate coalitional reasoning into normal-form games, and the paper proposes one such theory, Pareto-focal rationalizability. However, while

Rabin's approach starts with some exogenously given set of focal points, our procedure can endogenously explain why certain outcomes are focal in a game.

There are papers investigating the role of pre-play communication before playing a normal-form game, examining whether it leads to the type of belief restrictions we consider in this paper. Farrell [1988] assumes that before playing a normal-form game one player can send a suggestion to the others. This suggestion is allowed to be a set of strategies, not just a single profile. He points out that there are games in which players clearly do not want to make any single-profile agreement, although his considerations are different than ours. Rabin [1990] considers similar assumptions to Farrell in a rationalizability setting. Watson [1991] introduces a model in which one player can suggest playing inside some subset of the strategy space. The definition of when this message is credible is somewhat similar to our definition of supported restriction, although it is not a generalization of best response strategies. Furthermore, Watson's concept is defined only for 2-player games, in which the issues of coalitional agreements are fairly simple.

VIII. Conclusion

In a lot of economic and political situations subgroups of the participants have an incentive to coordinate their action choices. Various coalitional equilibrium concepts were proposed in the literature that assumed that players with similar interest could coordinate their play. However, the concepts proposed so far cannot guarantee existence in a natural class of games. The way we interpret this is that allowing coalitions to be able to make only point restrictions (agreements) leads to logical inconsistencies. As this paper shows, these inconsistencies do not have to arise if one allows for set-valued restrictions, suggesting that the latter is the appropriate framework to incorporate coalitional reasoning into noncooperative game theory.

IX. Appendix

Proof of Proposition 1. Let a be such that $u_j(a) = \max_{s:\, s \in A} u_j(s)$. But then by the definition of a supported restriction, $a_j \in B_j \; \forall \, B$ such that $A \searrow B$ because a_j is a best response against a_{-j} (it yields the maximum payoff in A and $A \in \mathcal{M}$). Therefore $\bigcap_{B:\, B \in F(A)} B_j \neq \emptyset$. This establishes the claim since j was arbitrary and $\bigcap_{B:\, B \in F(A)} B$ is a product set. QED

LEMMA 1. *Let $A \in \mathcal{M}$ and $A \searrow_J B$. Then $B \in \mathcal{M}$.*

Proof of Lemma 1. Suppose $B \notin \mathcal{M}$. Then $\exists\, j$, a_j and f_{-j} such that $j \in J$, $a_j \in A_j/B_j$, $f_{-j} \in \Delta_{-j}(B)$ and $a_j \in BR_j(f_{-j})$, which contradicts $A \searrow_J B$. QED

LEMMA 2. *Let $A \searrow_J B$ and let C be such that $C \in \mathcal{M}$, $C \subset A$ and $C \cap B \neq 0$. Then $C \searrow_J (C \cap B)$.*

Proof of Lemma 2. Lemma 1 and $C \in \mathcal{M}$ imply $B \in \mathcal{M}$. Then $C \cap B \in \mathcal{M}$ by the definition of a set closed under rational behavior. Let j and c_j be such that $j \in J$ and $c_j \in C_j/B_j$. Note that $A \searrow_J B$ implies $u_j(c_j, f_{-j}) < u_j(b_j, g_{-j})$, $\forall\, b_j$, f_{-j}, g_{-j} such that $f_{-j} \in \Omega^*_{-j}(A/B) \cap \Delta_{-j}(A)$, $c_j \in BR_j(f_{-j})$, $g_{-j} \in \Delta_{-j}(B)$, $b_j \in BR_j(g_{-j})$ and $g_{-j}^{-J} = f_{-j}^{-J}$, $\forall\, s_{-J} \in S_{-J}$. But since $C \subset A$ and $C \cap B \subset B$, this implies $u_j(c_j, f_{-j}) < u_j(b_j, g_{-j})$, $\forall\, b_j$, f_{-j}, g_{-j} such that $f_{-j} \in \Omega^*_{-j}((C \cap B)) \cap \Delta_{-j}(C)$, $c_j \in BR_j(f_{-j})$, $g_{-j} \in \Delta_{-j}(C \cap B)$, $b_j \in BR_j(g_{-j})$ and $g_{-j}^{-J} = f_{-j}^{-J}$, $\forall\, s_{-J} \in S_{-J}$. Since this holds for every j and c_j such that $j \in J$ and $c_j \in C_j/B_j$, $(C) \searrow_J (C \cap B)$. QED

Proof of Proposition 2. By Lemma 1, $B \in \mathcal{M}$ and $C^i \in \mathcal{M}$ $\forall\, i = 1, ..., k$. By Proposition 1, $B \cap C^1 \neq \emptyset$. Then by Lemma 2, $C^1 \searrow_{J^B} (B \cap C^1)$. Now suppose $C^n \searrow_{J^B} (B \cap C^n)$ for some $1 \leq n \leq k - 1$. Then by definition $B \cap C^n \neq \emptyset$. Then Proposition 1, $C^n \searrow_{J^B} (B \cap C^n)$ and $C^n \searrow_{J^n} C^{n-1}$ imply $B \cap C^{n+1} \neq \emptyset$. Then Lemma 2 implies that $C^{n+1} \searrow_{J^B} (B \cap C^{n+1})$. The claim follows by induction. QED

Proof of Proposition 3. 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 Proposition 1, $A^{k+1} \neq \emptyset$. By Lemma 2, A^{k+1} can be reached from A^k by a sequence of supported restrictions and then by Proposition 2, $A^{k+1} \in \mathcal{M}$. By induction, $A^k \neq \emptyset$ and $A^k \in \mathcal{M}$, $\forall\, k \geq 0$. Since $A^* = A^k$ whenever $k \geq K$, this implies $A^* \neq \emptyset$ and $A^* \in \mathcal{M}$. Now suppose that there exists a nontrivial supported restriction given A^*. Since $A^* = A^K$, this implies that there is a nontrivial supported restriction given A^K, which contradicts that $A^{K+1} = A^K$. QED

Proof of Proposition 4. 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 $s_i \in B_i/A_i$, $f_{-i} \in \Delta_{-i}(A)$ and $s_i \in BR_i(f_{-i})$. QED

Proof of Proposition 5. 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 Proposition 1, $\bigcap_{B:\, B \in \mathcal{F}(B^k)} B \neq \emptyset$, so $B^{k+1} = \bigcap_{B:\, B \in \Theta^k} B \neq \emptyset$. By Lemmas 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 given B^L.

By definition $B^L \subset A^0$. Note that by Lemma 2 and Proposition 2 B^L can be reached from A^0 by a sequence of supported restrictions. Then by Lemma 1 $B^L \in \mathcal{M}$. Therefore by Proposition 2 $A^0 \searrow B$ implies $B^L \searrow B^L \cap B$. Then since there is no nontrivial supported restriction given B^L, $B^L \subset A^1$. Then by Lemma 2 B^L can be reached from A^1 by a sequence of supported restrictions. An inductive argument shows that $B^L \subset A^k, \forall\, k \geq 0$, and therefore $B^L \subset A^*$. A symmetric argument establishes that $A^* \subset B^k, \forall\, k \geq 0$, and therefore $A^* \subset B^L$. QED

LEMMA 3. *Let A be such that $A^* \subset A$ and $A \neq A^*$. Then $A^* \subset \underset{B:\ B \in \mathcal{F}(A)}{\cap} B$.*

Proof of Lemma 3. Suppose not. Then there are $B \subset A$ and $J \subset I$ such that $A \searrow_J B$ and $A^* \nsubseteq B$. First consider $B \cap A^* \neq \emptyset$. Then by Lemma 2 $A^* \searrow_J (B \cap A^*)$, contradicting that there is no nontrivial supported restriction given A^*.

Now consider $B \cap A^* = \emptyset$. 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 $f_{-i} \in \Delta_{-i}(A \cap A^k)$, $BR_i(f_{-i}) \in A$ 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 $A \searrow_J B$, by Proposition 2 $(A \cap A^k) \searrow_J (B \cap A^k)$. Furthermore, since $A \cap A^{k+1} \neq \emptyset$ (they both contain A^*), $(A \cap A^k) \searrow_J (A \cap C)$ $\forall\, C \in \mathcal{F}(A^k)$. The above implies $\underset{C:\ C \in \mathcal{F}(A \cap A^k)}{\cap} \subset (A \cap A^{k+1}) \cap (B \cap A^k)$. But $(B \cap A^k) \cap (A \cap A^{k+1}) = B \cap A^{k+1} = \emptyset$, contradicting Proposition 1. QED

LEMMA 4. *Let $A \in \mathcal{M}$ be such that $A/A^* \neq \emptyset$ and $A \cap A^* \neq \emptyset$. Then $\underset{B:\ B \in \mathcal{F}(A)}{\cap} B \neq A$.*

Proof of Lemma 4. There exists $k \geq 0$ such that $A \subset A^k$ and $A \nsubseteq A^{k+1}$. Then there are $B \subset A$ and $J \subset I$ such that $A^k \searrow_J B$ and $A \cap B \neq A$. Note that $A^k \searrow_J B$ implies $A^{k+1} \subset B$ which in turn implies $A^* \subset B$. Therefore $A \cap A^* \neq \emptyset$ implies $A \cap B \neq \emptyset$. Furthermore, $A^k \searrow_J B$ implies $A \in \mathcal{M}$, which together with $A \in \mathcal{M}$ and $A \cap B \neq \emptyset$ implies $A \cap B \in \mathcal{M}$. Then by Lemma 2 $A \searrow_J (A \cap B)$, so $\underset{B:\ B \in \mathcal{F}(A)}{\cap} B \neq A$. QED

Proof of Proposition 6. Lemmas 3 and 4 establish that A^* is externally coalitionally stable. Proposition 3 establishes that A^* is internally coalitionally stable. By Lemma 4, if A is such that $A/A^* \neq \emptyset$ and $A \cap A^* \neq \emptyset$, then A cannot be coalitionally stable. If $A/A^* = \emptyset$, then $A \subset A^*$, in which case either $A = A^*$ or A cannot be coalitionally stable, since A^* contains it and there is no nontrivial supported restriction given A^*. And if $A \cap A^* = \emptyset$, then there is $k \geq 0$ such that $A \subset A^k$ and $A \nsubseteq A^{k+1}$, therefore A is not coalitionally stable, since $A \subset A^k$ and $A \nsubseteq \underset{B:\ B \in \mathcal{F}(A^k)}{\cap} B$. QED

Proof of Proposition 7. suppose s^* is a pure strategy coalition-proof Nash equilibrium and $s^* \notin A^*$. Then there is $k \geq 0$ such that $s^* \in A^k$ but $s^* \notin A^{k+1}$. That implies that there is $J \subset I$ and $B \subset A$ such that $A^k \searrow_J B$ and $s^* \notin B$. Let s' be a coalition-proof Nash equilibrium of $G[B_{J'} \times s^*_{I/J'}]$. The starting assumption guarantees the existence of such profile s'. Note that for every $j \in J$ it holds that s'_j is a best response against a conjecture that allocates probability 1 to s'_{-j}, and that s^*_j is a best response against a conjecture that allocates probability 1 to s^*_{-j}. This implies that $u_j(s') > u_j(s^*)$, $\forall \, j \in J'$ because B is a supported restriction by J given A^k. Therefore s' is a profitable coalitional deviation from s^* by J'. Suppose now that there is a credible profitable coalitional deviation s'' from s' by $J'' \subset J'$. The credibility of this deviation implies that for every $j \in J''$, s''_j is a best response against the belief that allocates probability 1 to s''_{-j}. Consider first the case that $s''_j \in A^k_j$, $\forall \, j \in J''$. Since s' is a coalition-proof Nash equilibrium of $G[B_{J'} \times s^*_{I/J'}]$, it can only be that $s''_i \notin B_i$ for some $i \in J''$. But then the profitability of this deviation implies that $u_i(s'') > u_i(s')$, contradicting $A^k \searrow_J B$, since $i \in J'' \subset J$. Next consider the case that $s''_j \notin A^k_j$ for some $j \in J''$. Let $m \geq 0$ be such that $s''_j \in A^m_j$, $\forall \, j \in J''$, but there is $i \in J''$ such that $s''_i \notin A^{m+1}_i$. Note that $m < k$ since $s''_j \notin A^k_j$. Then there is some $L \subset I$ and $C \subset A^m$ such that $A^m \searrow_L C$ and $s''_i \notin C_i$. But the profitability of the above deviation implies $u_i(s'') > u_i(s')$, contradicting $A^m \searrow_L C$. This establishes that s' is a credible profitable deviation from s^* by J' since there is no credible profitable deviation from it by a subcoalition of J'. This contradicts that s^* is a pure strategy coalition-proof Nash equilibrium. QED

Proof of Proposition 8. Let $A_i = \mathrm{supp}\sigma_i$, $\forall \, i \in I$ and let $A = \underset{i \in I}{\times} A_i$. Suppose $A \nsubseteq A^*$. Then there is $k \geq 0$ such that $A \subset A^k$, but $A \nsubseteq A^{k+1}$. This implies that there are $B \subset A^k$ and $J \subset I$ such that B is a supported restriction by J given A^k, and $A \nsubseteq B$. 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 σ_{-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 by J given C, $u_l(\widehat{\xi}_L, \sigma_{-L}) > u_l(a_l, \sigma_{-l}) = u_l(\sigma)$. But that implies $\widehat{\xi}_L$ is a profitable deviation for L from σ, contradicting that σ is a strong Nash equilibrium. QED

PRINCETON UNIVERSITY AND HARVARD UNIVERSITY

X. References

Ambrus, Attila, "Theories of coalitional rationality," mimeo Harvard University, 2005.

Ambrus, Attila and Rossella Argenziano, "Network markets and consumer coordination," mimeo Harvard University and Yale University, 2004.

Aumann, Robert J., "Acceptable points in general cooperative n-person games," in Albert W. Tucker and R. Duncan Luce ed.s, *Contributions to the theory of games IV* (Princeton, NJ: Princeton University Press, 1959), pp. 287-324.

Aumann, Robert J., "Nash equilibria are not self-enforcing," in Jean J. Gabszewicz, Jean-Francois Richard and Laurence A. Wolsey ed.s, *Economic Decision-making* (Amsterdam, North-Holland: Elsevier Science Publishers B.V., 1990), pp. 201-206.

Basu, Kaushik and Jorgen W. Weibull, "Strategy subsets closed under rational behavior," *Economics Letters*, XXXVI (1991), 141-146.

Bernheim, B. Douglas, "Rationalizable strategic behavior," *Econometrica*, LII (1984), 1007-1028.

Bernheim, B. Douglas, Bezalel Peleg and Michael D. Whinston, "Coalition-proof Nash equilibria I. Concepts," *Journal of Economic Theory*, XVII (1987), 1-12.

Bernheim, B. Douglas and Michael D. Whinston, "Menu auctions, resource allocation, and economic influence," *Quarterly Journal of Economics*, CI (1986), 1-31.

Bernheim, B. Douglas and Michael D. Whinston, "Coalition-proof Nash equilibria II. Applications," *Journal of Economic Theory*, XVII (1987), 13-29.

Brandenburger, Adam and Eddie Dekel, "Hierarchies of beliefs and common knowledge," *Journal of Economic Theory*, LIX (1993), 189-198.

Charness, Gary, "Self-serving cheap talk: a test of Aumann's conjecture," *Games and Economic Behavior*, XXXIII (2000), 177-194.

Chwe, Michael Suk-Young, "Farsighted coalitional stability," *Journal of Economic Theory*, LXIII (1994), 299-325.

Cooper, Russell, Douglas V. De Jong, Robert Forsythe and Thomas W. Ross, "Communication in coordination games," *Quarterly Journal of Economics*, CVII (1992), 739-771.

Dagan, Nir, Roberto Serrano and Oscar Volij, "A Noncooperative View of Consistent Bankruptcy Rules," *Games and Economic Behavior*, XVIII (1997), 55-72.

Delgado, Juan and Diego Moreno, "Coalition-Proof Supply Function Equilibria in Oligopoly," *Journal of Economic Theory*, CXIV (2004), 231-254.

Farrell, Joseph, "Communication, coordination and Nash equilibrium," *Economics Letters*, XXVII (1988), 209-214.

Farrell, Joseph, "Meaning and credibility in cheap talk games," *Games and Economic Behavior*, V (1993), 514-531.

Fudenberg, Drew and Jean Tirole, *Game theory, 2nd edition* (Cambridge, MA: MIT Press, 1992).

Gul, Faruk, "Rationality and coherent theories of rational behavior," *Journal of Economic Theory*, LXX (1996), 1-31.

Gupta, Srabana and Richard E. Romano, "Monitoring the principle with multiple agents," *Rand Journal of Economics*, XXIX (1998), 427-442.

Herings, P. Jean-Jacques and Vincent J. Vannetelbosch, "Refinements of rationalizability for normal-form games," *International Journal of Game Theory*, XXVIII (1999), 53-68.

Konishi, Hideo, Michel LeBreton and Shlomo Weber, "On coalition-proof Nash equilibria in common agency games," *Journal of Economic Theory*, LXXXV (1999), 122-139.

Mariotti, Marco, "A model of agreements in strategic form games," *Journal of Economic Theory*, LXXIV (1997), 196-217.

Noe, Thomas H., "Rationalizable and coalition proof shareholder tendering strategies in corporate takeovers," *Review of Quantitative Finance & Accounting*, XI (1998), 269-291.

Pearce, David G., "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, LII (1984), 1029-1050.

Rabin, Matthew, "Communication between rational agents," *Journal of Economic Theory*, LI (1990), 144-170.

Rabin, Matthew, "A model of pre-game communication," *Journal of Economic Theory*, LXIII (1994a), 370-391.

Rabin, Matthew, (1994b): "Incorporating behavioral assumptions into game theory," in James Friedman ed., *Problems of coordination in economic activity* (Boston and Dordrecht: Kluwer Academic, 1994b), pp. 69-87.

Ray, Debraj and Rajiv Vohra, "Equilibrium binding agreements," *Journal of Economic Theory*, LXXIII (1997), 30-78.

Ray, Debraj and Rajiv Vohra, "A theory of endogenous coalition structures," *Games and Economic Behavior*, XXVI (1999), 286-336.

Rubinstein, Ariel and Asher Wolinsky, "Rationalizable conjectural equilibria: between Nash and rationalizability," *Games and Economic Behavior*, VI (1994), 299-311.

Tan, Tommy Chin-Chiu and Sergio Ribeiro da Costa Werlang, "The Bayesian foundations of solution concepts of games," *Journal of Economic Theory*, XVLV (1988), 370-391.

Watson, Joel, "Communication and superior cooperation in two-player normal form games," *Economics Letters*, XXXV (1991), 267-271.

Xue, Licun, "Coalitional stability under perfect foresight," *Economic Theory*, XI (1998), 603-627.

Xue, Licun, "Negotiation-proof Nash equilibrium," *International Journal of Game Theory*, XXIX (2000), 339-357.

Zapater, Inigo, "Credible proposals in communication games," *Journal of Economic Theory*, LXXII (1997), 173-197.

	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 I

C1

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

C2

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

C3

	B1	B2	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 II

C1

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

C2

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

Figure III