Theories of coalitional rationality*

Attila Ambrus†

Work in progress
First Version: July 2004
This Version: June 2005

Abstract

This paper generalizes the concept of best response to coalitions of players and offers epistemic definitions of coalitional rationalizability in normal form games. The best response of a coalition is defined to be a correspondence from sets of conjectures to sets of strategies. From every best response correspondence it is possible to obtain a definition of the event that a coalition is rational. It requires that if it is common certainty among players in the coalition that play is in some subset of the strategy space then they confine their play to the best response set to those conjectures. A strategy is epistemic coalitionally rationalizable if it is consistent with rationality and common certainty that every coalition is rational. A characterization of this set of strategies is provided for best response correspondences that satisfy two consistency properties and a weak requirement along the line of Pareto dominance for members of the coalition. Special attention is devoted to two correspondences from this class. One leads to a solution concept that is generically equivalent to the set of coalitionally rationalizable strategies as defined in Ambrus [04], while the other one leads to a solution concept exactly equivalent to it.

*I thank Satoru Takahashi for a thorough proofreading of the paper and Drew Fudenberg for useful comments.

†Department of Economics, Harvard University, Cambridge, MA 02138, ambrus@fas.harvard.edu

1 Introduction

Since Aumann's paper nearly fifty years ago (Aumann [59]) there have been numerous attempts to incorporate coalitional reasoning into the theory of non-cooperative games, but the issue is still unresolved. Part of the problem seems to be that the concept of coalitional reasoning itself is not formally defined. At an intuitive level it means that players with similar interest (a coalition) coordinate their play to achieve a common gain (to increase every player's payoff in the coalition). This intuitive definition can be formalized in a straightforward way if there is a focal strategy profile that all players expect to be played. With respect to this profile, a profitable coalitional deviation is a joint deviation by players in a coalition that makes all of them better off, supposing that all other players keep their play unchanged. This definition is a generalization of a profitable unilateral deviation, therefore concepts that require stability with respect to coalitional deviations are refinements of Nash equilibrium. The two most well-known equilibrium concepts along this line are strong Nash equilibrium (Aumann [59]) and coalition-proof Nash equilibrium (Bernheim et al [87]). However, as opposed to Nash equilibrium, these solution concepts cannot guarantee existence in a natural class of games. This casts doubt on whether these theories give a satisfactory prediction even in games in which the given equilibria do exist.

Outside the equilibrium framework Ambrus [04] proposes the concept of coalitional rationalizability, using an iterative procedure. The construction is similar to the original definition of rationalizability, provided by Bernheim [84] and Pearce [84]. The new aspect is that not only never best-response strategies of individual players are deleted by the procedure, but strategies of groups of players simultaneously too, if it is in their mutual interest to confine their play to the remaining set of strategies. These are called supported restrictions by different coalitions. The set of coalitionally rationalizable strategies is the set of strategies that survive the iterative procedure of supported restrictions. The paper also provides a direct characterization of this set. But even this characterization (stability with respect to supported restrictions given any superset) is not based on primitive assumptions about players' beliefs and behavior. Since such characterizations were provided for rationalizability by Tan and Werlang [88] and Brandenburger and Dekel [93], using the framework of interactive epistemology, the question arises whether similar epistemic foundations can be worked out for coalitional rationalizability as well.

This paper investigates a range of possible definitions of coalitional rationalizability in an epistemic framework. These theories differ in how the event that a coalition is rational is defined. We only consider definitions that are generalizations of the standard definition of individual rationality, namely that players are subjective expected utility maximizers: every player forms a conjecture on other players' choices and plays a best response to it.[1] We define the best re-

[1] This is the starting point for rationalizability as well, although Epstein [97] considers

sponse of a coalition to be a correspondence that allocates a set of strategies to certain sets of conjectures. The assumption that the correspondence is defined on sets of conjectures corresponds to the idea that in a non-equilibrium framework players in a coalition might not have the same conjecture, but it can be common certainty among them that play is within a certain subset of the strategy space. Intuitively then the best response of the coalition to this set of conjectures is a set of strategies that players in the coalition would agree upon confining their play to, given the above set of possible conjectures. Since players' interests usually do not coincide perfectly, there are various ways to formalize this intuition. Because of this we consider a wide range of coalitional best response correspondences.

Each best response correspondence can be used to obtain a definition of coalitional rationality the following way. A coalition is rational if (i) every player in the coalition is rational, and (ii) whenever it is common certainty among coalition members that play is within a certain set of strategies, then members of the coalition play within the best response set of the coalition to the set of conjectures concentrated on that set of strategies. Once the event that a coalition is rational is well-defined, the events that every coalition is rational, that a player is certain that every coalition is rational, and that it is common certainty among players that every coalition is rational can be defined the usual manner. Then a definition of coalitional rationalizability can be provided as the set of strategies that are consistent with the assumptions that every player is rational and that it is common certainty that every coalition is rational. We refer to coalitional rationalizability corresponding to best response correspondence γ as coalitional γ-rationalizability.

We then investigate the class of best response correspondences that satisfy three properties. The first is a consistency requirement imposing that the best response to a set that is closed under rational behavior is itself closed under rational behavior. The second one imposes a form of monotonicity on the correspondence that reflects the idea that if restricting play in a certain way is mutually advantageous for members of a coalition for a set of possible beliefs, then the same restriction should still be advantageous for a smaller set of possible beliefs. The third property requires that the best response of a coalition retains the strategies of players in the coalition that can be best responses to their most optimistic conjectures. This is a very weak requirement along the lines of Pareto optimality of the best response correspondence for coalition members, but turns out to be enough to establish our results. We call the above best response correspondences sensible. We show that there is a smallest and a largest sensible best response correspondence.

It is shown that for every sensible best response correspondence γ the resulting set of coalitionally γ-rationalizable strategies is nonempty and coherent,

building the concept on alternative definitions of rationality.

and it can be characterized by an iterative procedure that is defined from the corresponding best response correspondence. In generic games this procedure is fairly simple. Starting from the set of all strategies, in each step it involves taking the intersection of best responses of all coalitions, given the set of strategies that survive the previous step. In a nongeneric class of games the procedure involves checking best responses of coalitions given certain subsets (not only the entire set) of the set of strategies reached in the previous round, and it characterizes a subset of the set of strategies reached by the simpler iterative procedure.

The best response correspondence that we pay special attention to uses the concept of supported restriction as defined in Ambrus [04]. It specifies the best response of a coalition to the set of conjectures concentrated on some set of strategies to be the smallest supported restriction by the coalition given that set. The resulting definition of epistemic coalitional rationalizability requires that whenever it is common certainty among members of a coalition that play is in A, and B is a supported restriction by the coalition given A, then players in the coalition choose strategies in B. Our results then imply that the set of epistemic coalitionally rationalizable strategies defined this way is generically equivalent to the iteratively defined set of coalitionally rationalizable strategies of Ambrus [04]. In a nongeneric class of games the former can be a strict subset of the latter, providing a (slightly) stronger refinement of rationalizability. Furthermore, we show that there exists another sensible best response correspondence that leads to a set of epistemic coalitionally rationalizable strategies that is exactly equivalent to the iteratively defined set of coalitionally rationalizable strategies.

2 The model

2.1 Basic notation.

Let $G = (I, S, u)$ be a normal form game, where I is a finite set of players, $S = \underset{i \in I}{\times} S_i$, is the set of strategies, and $u = (u_i)_{i \in I}$, $u_i : S \to R$, $\forall\, i \in I$ 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 $S_{-J} = \underset{j \in I/J}{\times} S_j$, $\forall\, J \subset I$. Let $\mathcal{C} = \{J \mid J \subset I,\ J \neq \emptyset\}$. We will refer to elements of $\mathcal{C}$ as *coalitions.*

Let Δ_{-i} be the set of probability distributions over S_{-i}, representing the set of conjectures (including correlated ones) player i can have concerning other players' moves. For every $J \in C$, $i \in J$ and $f_{-i} \in \Delta_{-i}$ let f_{-i}^{-J} be the marginal distribution of f_{-i} over S_{-J}. For every $f_{-i} \in \Delta_{-i}$ and $s_i \in S_i$ let

$u_i(s_i, f_{-i}) = \sum_{t_{-i} \in S_{-i}} u_i(s_i, t_{-i}) \cdot f_{-i}(t_{-i})$ denote the expected payoff of player i if he has conjecture f_{-i} and plays pure strategy s_i. 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}.

2.2 Type spaces

Definition: a type space T for G is a tuple $T = (I, (T_i, \Phi_i, g_i)_{i \in I})$ where T_i is a compact topological space, Φ_i is a compact subset of $S_i \times T_i$ such that $\text{proj}_{S_i} \Phi_i = S_i$, and $g_i : T_i \to \triangle(\Phi_{-i})$ (where $\triangle(\Phi_{-i})$ is the set of Borel probability measures on Φ_{-i}) is a continuous mapping (with respect to the topology on T_i and the weak topology on $\triangle(\Phi_{-i})$).

T_i represents the set of epistemic types of player i. Φ is the set of states of the world. Every state of the world consists of a strategy profile (the external state) and a profile of epistemic types. A player's epistemic type determines her probabilistic belief (conjecture) about other players' strategies and epistemic types. Player i's belief as a function of her type is denoted by g_i.[2]

For every $i \in I$ and $\phi_i \in \Phi_i$ let $\phi_i = (s_i(\phi_i), t_i(\phi_i))$.

Definition: i is *certain of* $\Psi_{-i} \subset \Phi_{-i}$ at $\phi \in \Phi$ if $g_i(t_i(\phi_i))(\Psi_{-i}) = 1$.[3]

In the formulation we use a player does not have beliefs concerning her own strategy. Nevertheless, for the construction below it is convenient to extend the definition of certainty to particular events of the entire state space.

Definition: i is certain of $\Psi = \Psi_i \times \Psi_{-i} \subset \Phi$ at $\phi \in \Phi$ if i is certain of Ψ_{-i} at ϕ.

Let $\Psi = \Psi_i \times \Psi_{-i}$ and let $\mathcal{C}_i(\Psi_{-i}) \equiv \{\phi \in \Phi : g_i(t_i(\phi_i))(\Psi_{-i}) = 1\}$. $\mathcal{C}_i(\Psi)$ is the event in the state space that i is certain of Ψ.

Let $\Psi = \underset{i \in I}{\times} \Psi_i$ where $\Psi_i \subset \Phi_i$ (a product event).

Definition: *Mutual certainty* of Ψ holds at $\phi \in \Phi$ if $\phi \in \underset{i \in N}{\cap} \mathcal{C}_i(\Psi)$. Mutual certainty of $\Psi \subset \Phi$ among J holds at $\phi \in \Phi$ if $\phi \in \underset{i \in I}{\cap} \mathcal{C}_i(\Psi)$.

Let $\mathcal{MC}_J(\Psi)$ denote mutual certainty of Ψ among J.

Definition: Let $\mathcal{MC}_J^1(\Psi) \equiv \mathcal{MC}_J(\Psi)$. Let $\mathcal{MC}_J^k(\Psi) = \mathcal{MC}_J(\mathcal{MC}_J^{k-1}(\Psi))$ for $k \geq 2$. *Common certainty* of Ψ among J holds at $\phi \in \Phi$ if $\phi \in \underset{k=1,2,\ldots}{\cap} \mathcal{MC}_J^k(\Psi)$.

Let $\mathcal{CC}_J(\Psi)$ denote common certainty of Ψ among J.

[2] For more on type spaces see for example Battigalli and Bonanno[99].
[3] The terminology "i believes Ψ_{-i}" is also common in the literature.

3 Best response correspondences for coalitions and definitions of coalitional rationalizability

In this section we define the event that a coalition is rational. We start out by generalizing the concept of best response for coalitions. Since there is no one clear way of doing this, we consider all possible best response correspondences that satisfy certain desirable properties. Each of these correspondences can then be used to define rationality of a coalition.

The set of best responses of player i to a conjecture $f_{-i} \in \Delta_{-i}$ consists of the strategies of i that maximize her expected payoff given f_{-i}. When trying to generalize this definition to coalitions of multiple players, two conceptual difficulties arise. One is that in a nonequilibrium framework different players in the coalition might have different conjectures on other players' strategy choices. Second, even if they share the same conjecture, typically players' interests do not align perfectly - different strategy profiles maximize the payoffs of different coalition members to the conjecture. However, these inconsistencies can be resolved if the best response correspondence is defined such that it allocates a set of strategies to a set of conjectures.

In particular, consider the case that it is common certainty among players in the coalition that the conjecture of each of them is concentrated on a product subset of strategies $A \subset S$.[4] Then even if they are uncertain that exactly what conjectures others in the coalition have from the above set of possible conjectures, they might all implicitly agree to confine their play to a set $B \subset A$. Therefore any theory that specifies what set of strategies a coalition would implicitly agree upon confining its play to a given set of conjectures can be interpreted as a best response correspondence. The problem is that there is no one obvious definition of a restriction being of mutual interest of a coalition, since evaluating a restriction involves a comparison of two sets of expected payoffs (expected payoffs in case the restriction is made and in case the restriction is not made) for every player. One formal definition can be obtained from the concept of supported restriction of Ambrus [04].

Let $\mathcal{X}$ denote product subsets of the strategy space: $\mathcal{X} = \{A \mid A = \underset{i \in I}{\times} A_i$ st $A_i \subset S_i \ \forall\, i \in I\}$.

For any $A \in \mathcal{X}$ let $\Delta_{-i}(A) = \{f_{-i} \mid \text{supp} f_{-i} \subset A_{-i}\}$. We will refer to $\Delta_{-i}(A)$ as the set of conjectures concentrated on A.

For any $B_i \subset S_i$ let $\Delta^*_{-i}(B_i) = \{f_{-i} \mid f_{-i} \in \Delta_{-i}, \exists\, b_i \in B_i$ such that $b_i \in BR_i(f_{-i})\}$. In words, $\Delta^*_{-i}(B_i)$ is the set of conjectures to which player i has a best response strategy in B_i.

[4] For a discussion on why we only consider product sets see Section 6.

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.

Let $A, B \in \mathcal{X}$ and $\emptyset \neq B \subset A$.

Definition: B is a *supported restriction* by J given A if
1) $B_i = A_i$, $\forall\, i \notin J$, and
2) $\forall\, j \in J$, $f_{-j} \in \Delta^*_{-j}(A_j/B_j) \cap \Delta_{-j}(A)$ it is the case that $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$ $\forall\, g_{-j} \in \Delta_{-j}(B)$ such that $g_{-j}^{-J} = f_{-j}^{-J}$.

Restricting play to B given that conjectures are concentrated on A is supported by J if for any fixed conjecture concerning players outside the coalition, every player in the coalition expects a strictly higher expected payoff in case his conjecture is concentrated on B than if his conjecture is such that she has a best response strategy to it which is outside B. In short, for every fixed conjecture concerning outsiders, every coalition member is always strictly better off if the restriction is made than if the restriction is not made and she wants to play a strategy outside the restriction.

Let $\mathcal{F}_J(A)$ be the set of supported restrictions by J given A. It is possible to establish (see Ambrus [04]) that $\bigcap\limits_{B \in \mathcal{F}_J(A)} B$ is either empty or itself a member of $\mathcal{F}_J(A)$. This motivates a best response correspondence that allocates $\bigcap\limits_{B \in \mathcal{F}_J(A)} B$ to be the best response of J to the set of conjectures concentrated on A.

Definition: Let $\gamma^* : \mathcal{X} \times \mathcal{C} \to \mathcal{X}$ be such that for every $J \in \mathcal{C}$ $\gamma^*(A, J) = \bigcap\limits_{B \in \mathcal{F}_J(A)} B$ $\forall\, \emptyset \neq A \in \mathcal{X}$ and $\gamma^*(\emptyset, J) = \emptyset$.

However, supported restriction is just one possible way of formalizing the idea that a restriction is unambiguously in the interest of every player in the coalition. There are other intuitively appealing definitions. A stronger requirement (leading to larger best response sets) is that restriction B is supported by J given A iff $s \in B$, $t \in A/B$ and $s_{-J} = t_{-J}$ imply that $u_j(s) > u_j(t)$ $\forall$ $j \in J$ (fixing the strategies of players outside the coalition, the restriction payoff-dominates all other outcomes). A weaker requirement (leading to smaller best response sets) can be obtained from the following modification of supported restrictions. Note that if $B \subset A$ then every $f_{-i} \in \Delta_{-i}(A)$ can be decomposed as a convex combination of a conjecture in $\Delta_{-i}(B)$ and a conjecture in $\Delta_{-i}(A/B)$: $f_{-i} = \alpha^{f_{-i}} f_{-i}^B + (1 - \alpha^{f_{-i}}) f_{-i}^{A/B}$, where $\alpha^{f_{-i}}$ is uniquely determined. Then $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$ in the definition of supported restriction above can be required to hold only if $g_{-j} = \alpha^{f_{-i}} f_{-i}^B + (1 - \alpha^{f_{-i}}) g'_{-j}$ for some $g'_{-j} \in \Delta_{-j}(B)$. Intuitively, this corresponds to assuming that when players compare expected payoffs between the case the restriction is made and the case that it is not made, they leave the part of the conjecture that is consistent with the restriction unchanged.

Instead of selecting a particular best response correspondence, we proceed by considering a wide range of possible ones. The rest of this section defines rationalizability based on any coalitional best response correspondence. The next section considers best response correspondences that satisfy certain criteria (a set of correspondences that include γ^*) and provides a characterization result for the set of rationalizable strategies that are derived from them.

Definition: $\gamma : \mathcal{X} \times \mathcal{C} \to \mathcal{X}$ is a *coalitional best response correspondence* if $\gamma(A, J) \subset A$ and $\gamma(A, J) \neq \emptyset$ implies $(\gamma(A, J))_{-J} = A_{-J}$.

In words, coalitional best response correspondences are restrictions on the set of strategies such that only strategies of players in the corresponding coalitions are restricted. Let Γ be the set of coalitional best response correspondences.

Next we define the concept of rationality of a coalition. The definition refers to subsets of the strategy space that are called closed under rational behavior in the literature.

Definition: set $A \in \mathcal{X}/\emptyset$ is closed under rational behavior if $BR_i(f_{-i}) \subset A_i$, $\forall\, f_{-i} \in \Delta_{-i}(A)$, $\forall\, i \in I$.

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

.

For any $\gamma \in \Gamma$ we define a coalition to be γ-rational at some state of the world if the strategy profile that is played at that state is within the γ-best response of the coalition to any set which satisfies that it is common certainty among the coalition members that play is within this set.

For any $\emptyset \neq A \subset S$ let $\Psi^A = \{\phi \in \Phi \mid s(\phi) \in A\}$. Then $\mathcal{CC}_J(\Psi^A)$ is the event that there is common certainty among J that play is in A.

Definition: coalition J is γ-rational at $\phi \in \Phi$ if $\phi \in \mathcal{CC}_J(\Psi^A)$ implies $s_i(\phi_i) \in \Psi_i^{\gamma(A,J)}$ $\forall\, i \in J$, $A \in \mathcal{M}$.

In particular coalition J is γ^*-rational at $\phi \in \Phi$ if $\phi \in \mathcal{CC}_J(\Psi^A)$ and $B \in \mathcal{F}_J(A)$ together imply that $s_i(\phi_i) \in B_i$ $\forall\, i \in J$ and $A \in \mathcal{M}$.

Let R_J^γ denote the event that coalition J is γ-rational. Furthermore, let $CR^\gamma = \bigcap_{J \in \mathcal{C}, J \neq \emptyset} R_J^\gamma$, the event that every coalition is γ-rational.

Let $\overline{g}_{-i}(\phi_i)$ denote the marginal distribution of $g_i(t_i(\phi_i))$ over S_{-i}. It is the conjecture of type ϕ_i of player i regarding what strategies other players play. Following standard terminology, we call player i to be individually rational at ϕ

if $s_i(\phi_i) \in BR_i(\overline{g}_{-i}(\phi_i))$. Let R_i denote the event that player i is rational and let $R = \underset{i \in N}{\cap} R_i$ (the event that every player is rational).

A strategy profile is coalitionally γ-rationalizable if there exists a type space and a state in which the above strategy profile is played and both rationality and common certainty of coalitional γ-rationality hold.[5]

Definition: $t \in S$ is *coalitionally γ-rationalizable* if $\exists$ type space T and $\phi \in \Phi$ such that $\phi \in R \cap \mathcal{CC}_I(CR^{\gamma})$ and $s(\phi) = t$.

In particular coalitional γ^*-rationalizability implies common certainty of the event that whenever it is common certainty among players in a coalition that play is in $A \in \mathcal{M}$ and B is a supported restriction given A, then players in this coalition play strategies in B.

4 Sensible best response correspondences

This section focuses on coalitional best response correspondences that satisfy four basic requirements and investigates the resulting coalitional rationalizability concepts.

Definition: $\gamma \in \Gamma$ is a *sensible* coalitional best response correspondence if it satisfies the following properties:

(i) if $A \in \mathcal{M}$ then $\gamma(A, J) \in \mathcal{M}\ \forall\ J \in \mathcal{C}$

(ii) for every $A \in \mathcal{M}$, $i \in N$ and $a_i \in A_i$ it holds that if $-\exists\ f_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f_{-i})$ then $a_i \notin (\gamma(A, J))_i$ for every $J \ni i$

(iii) if $B \subset A$ and $\gamma(A, J) \cap B \neq \varnothing$ then $\gamma(B, J) \subset \gamma(A, J)\ \forall\ A, B \in \mathcal{M}$

(iv) $a \in \underset{s \in A}{\arg\max} u_j(s)$ implies $a_j \in (\gamma(A, J))_j\ \forall\ j \in J\ \forall\ J \in \mathcal{C}$ and $A \in \mathcal{M}$

Properties (i) and (ii) impose consistency of the coalitional best response correspondence with individual best response correspondence. Property (i) requires that the best response of any coalition to a set that is closed under rational behavior is closed under rational behavior. This corresponds to the requirement that a coalition member's individual best response strategies to any conjecture that is consistent with the coalition's best response should be included in the coalition's best response. Property (ii) requires that strategies that are never individual best responses for a player cannot be part of those coalitions' best responses that contain the player. Note that (i) and (ii) imply that the best

[5] Since Mertens and Zamir [85] and Brandenburger and Dekel [93] establish the existence of a universal type space that contains every possible type, an alternative definition for a strategy profile to be epistemic coalitionally rationalizable is that there is a state of the world in the universal type space in which rationality and common certainty of coalitional rationality hold and in which the given strategy profile is played.

response of a single-player coalition to a set $A \in \mathcal{M}$ is exactly the set of strategies that can be best responses to a conjecture in $\Delta_{-i}(A)$: (i) implies that all these strategies have to be included in the best response and (ii) implies that all other strategies are excluded from the best response.

Property (iii) is a monotonicity condition. Informally it corresponds to the idea that outcomes in $\gamma(A, J)$ in some sense (the exact meaning depends on γ) should be preferred to outcomes in in $A/\gamma(A, J)$ by players in J, which then should imply that outcomes in $B \cap \gamma(A, J)$ are preferred to outcomes in $B \cap (A/\gamma(A, J))$. Another way to say the same motivating argument is that if coalition J's best response involves not playing strategies in $A/\gamma(A, J)$ for a set of contingencies (namely when play is concentrated on A), then their best response should also involve not playing the above strategies for a smaller set of contingencies (when play is concentrated on $B \subset A$).

Property (iv) is a weak requirement along the lines of Pareto optimality for coalition members. It requires that for any coalition member the best response of a coalition to set A should include the strategies that are (individual) best responses to her most optimistic conjecture on A. Otherwise the best response of a coalition would not include strategies that could yield the highest payoff that the corresponding player could hope for, given that conjectures are concentrated on A. We consider this property as a minimal requirement for coalitional rationality. The reason that we do not impose a stronger requirement is primarily that even this weak requirement is enough to establish the main results of the section.

All four properties are only required for best responses to sets that are closed under rational behavior in the definition. They could be easily extended to all product subsets of the strategy space, but this strengthening of the requirements turns out to be immaterial, because best responses to sets that are not closed under rational behavior do not play any role in the construction below.

Let Γ^* denote the set of sensible coalitional best response correspondences. One example of a sensible coalitional best response correspondence is γ^*, the correspondence obtained from supported restrictions.

The proofs of all propositions are in the Appendix.

Proposition 1: $\gamma^* \in \Gamma^*$.

That γ^* satisfies (i)-(iv) follows from the definition of a supported restriction. In particular if a maximizes u_j on $A \in \mathcal{M}$ then a_j is a best response to the belief that allocates probability 1 to a_{-j}. Then the definition of supported restriction implies $a_j \in B$ whenever B is a supported restriction given A, by any coalition J that involves j. Note that this also holds for all coalitions not involving j, since then $B_j = A_j$.

We note that there are various ways of changing the definition of the supported restriction that lead to coalitional best response correspondences different than γ^*, but also sensible. One is when conjectures concerning players outside the coalition are not required to be fixed in expected payoff comparisons between conjectures consistent with a restriction and conjectures to which there is a best response outside the restriction (when $g_{-j}^{-J} = f_{-j}^{-J}$ is no longer required in requirement (2) of the definition of supported restriction).

It is straightforward to establish that there exists a smallest and a largest element of Γ^*. The largest is the one that only excludes (individual) never best-response strategies for coalition members.[6] The smallest one can be defined in an iterative manner. It involves starting out from the correspondence that for every $A \in \mathcal{M}$ allocates the smallest set in $\mathcal{M}$ that is consistent with property (iv) of a sensible best response correspondence and then iteratively enlarging the values of the correspondence until it satisfies property (i).

Proposition 2: There exist $\gamma^M \in \Gamma^*$ and $\gamma^m \in \Gamma^*$ such that $\gamma^M(A, J) \supset \gamma(A, J) \supset \gamma^m(A, J)$ $\forall$ $\gamma \in \Gamma^*$ and $A \in \mathcal{X}$.

It turns out that to establish important properties of the set of coalitionally γ-rationalizable strategies for a sensible coalitional best response correspondence γ it is convenient to provide a characterization of it as the set of profiles obtained by an iterative procedure.

Definition: Let $A \in \mathcal{M}$. $B \in \mathcal{X}$ is *self-supporting for J with respect to A* if $B \subset A$ and for every $j \in J$ and $b_j \in B_j$ it holds that $b_j \in BR_j(\theta_{-j})$ and $\theta_{-j} \in \Delta_{-j}(A)$ imply $\theta_{-j} \in \Delta_{-j}(B)$.

Let $\mathcal{N}_J(A)$ denote the collection of self-supporting sets for J with respect to A.

Definition: Let $A \in \mathcal{X}$ and let $J \in \mathcal{C}$. For every $j \in J$ let $A_j^-(J) = \{s_j \in A_j \mid \exists\ B \in \mathcal{N}_J(A),\ C \in \mathcal{M} \cup \{A\}$ st $B \subset C \subset A,\ s_j \in B_j$ and $s_j \notin (\gamma(C, J))_j\}$. The *generalized γ-best response* by J given A is $G^\gamma(A, J) = \underset{j \in J}{\times} (A_j / A_j^-) \underset{i \in I/J}{\times} A_i$.

The generalized γ-best response of a coalition to $A \in \mathcal{X}$ is a restriction that besides excluding strategies that are not in the best response of the coalition to A also excludes certain strategies that are not in the best response of the coalition to certain subsets of A. The latter are the sets that are closed under rational behavior and contain a set that is self-supporting for the coalition. To get an intuition why this concept is useful in characterizing γ-coalitionally rationalizable strategies, in particular why self-supporting sets play a role in the process, see the example of Figure 1 below. Note that the generalized γ-best

[6] For sets in $\mathcal{M}$. For other sets it is equal to the identity correspondence, since the definition of sensibility does not restrict.the correspondence in any way for the latter sets.

response of a coalition to a set of strategies is by definition a subset of the γ-best response of the coalition to the same set of strategies (since $A \in \mathcal{N}_J(A)\ \forall\ A \in \mathcal{X}$ and $J \in \mathcal{C}$). Moreover, Proposition 3 states that in a generic class of games the two correspondences are equivalent, therefore in this class of games the above complicated definition can be greatly simplified.

Proposition 3: Suppose that for every $A \in \mathcal{M}$ it holds that there is no $i \in I$ and $a_i \in A_i$ for which it holds that $\exists\ f_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f_{-i})$ but $-\exists\ f'_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f'_{-i})$ and $f'_{-i}(s_{-i}) > 0\ \forall\ s_{-i} \in A_{-i}$. Then $G^{\gamma}(A, J) = \gamma(A, J)\ \forall\ \gamma \in \Gamma^*$, $A \in \mathcal{X}$ and $J \in \mathcal{C}$.

The condition in the proposition, namely that no set that is closed under rational behavior has a strategy that is weakly but not strictly dominated within that set, ensures that if $B \in \mathcal{N}_J(A)$ for some $J \in \mathcal{C}$ then either $B = A$ or B consists of only never best response strategies for some player. The latter strategies cannot be in $\gamma(A, J)$ by property (ii) of a sensible best response correspondence, implying that $G^{\gamma}(A, J) = \gamma(A, J)$. It is straightforward to establish that the property in the proposition is generic.

Consider now the following procedure. Let $E^0(\gamma) = S$. For every $k \geq 1$ let $E^k(\gamma) = \underset{J \in \mathcal{C}}{\cap} G^{\gamma}(E^{k-1}(\gamma), J)$.

Definition: Let $E^*(\gamma) = \underset{k=0,1,2,\ldots}{\cap} E^k(\gamma)$.

Note that, by Proposition 3, in a generic class of games $E^*(\gamma)$ can be obtained simply by taking the intersection of γ-best responses of all possible coalitions in an iterative manner, starting from the set of all strategies. Furthermore, it is straightforward to show that the latter set contains $E^*(\gamma)$ in every game if $\gamma \in \Gamma^*$, using properties (i) and (iii) of a sensible best response correspondence. In particular the above imply that $E^*(\gamma^*)$ is always included in and generically equivalent to the set of coalitionally rationalizable strategies as defined in Ambrus [04].

The next proposition establishes nonemptyness and other basic properties of $E^*(\gamma)$ for sensible coalitional best response correspondences.

Proposition 4: For every $\gamma \in \Gamma^*$ $E^*(\gamma)$ is nonempty, $\exists\ K < \infty$ such that $E^k(\gamma) = E^*(\gamma)$ whenever $k \geq K$, $E^*(\gamma) \in \mathcal{M}$ and $G^{\gamma}(E^*(\gamma), J) = E^*(\gamma)\ \forall$ $J \in \mathcal{C}$.

The outline of the proof is the following. Condition (iii) in the definition of a sensible best response correspondence implies that $E^k(\gamma)$ is nonempty for every k, and condition (i) in the definition implies that $E^k(\gamma)$ is closed under rational behavior for every k. By construction $E^k(\gamma)$ is decreasing in k, which together with the finiteness of S implies that $E^k(\gamma) = E^*(\gamma)$ for large enough k. The rest of the claim follows straightforwardly from these results.

Definition: set A is coherent if it is closed under rational behavior and satisfies:

$$\bigcup_{f_{-i} \in \Delta_{-i}(A)} BR_i(f_{-i}) = A_i, \ \forall\, i \in I \tag{1}$$

Note that for any $\gamma \in \Gamma^*$ the result that $\gamma(E^*(\gamma), J) = E^*(\gamma)$ (which follows from $G^\gamma(E^*(\gamma), J) = E^*(\gamma)) \; \forall\, J \in \mathcal{C}$ implies that for every $i \in I$ and $s_i \in E_i^*(\gamma)$ there exists $f_{-i} \in \Delta_{-i}(E^*(\gamma))$ such that $s_i \in BR_i(f_{-i})$. This and $E^*(\gamma) \in \mathcal{M}$ together imply that $E^*(\gamma)$ is a coherent set for $\gamma \in \Gamma^*$. Next we establish the equivalence of $E^*(\gamma)$ and the set of coalitionally γ-rationalizable strategies.

Proposition 5: For every, $\gamma \in \Gamma^*$, type space T and state $\phi \in \Phi$ it holds that $\phi \in R \cap \mathcal{CC}_I(CR^\gamma)$ implies $s(\phi) \in E^*(\gamma)$. Conversely, for every $s \in E^*(\gamma)$ $\exists$ type space T and $\phi \in \Phi$ such that $s(\phi) = s$, and $\phi \in R \cap \mathcal{CC}_I(CR^\gamma)$.

The first part of the proposition can be established the following way. It is common certainty among players of any coalition that play is in S. Therefore the assumption that every coalition is γ-rational implies that players of any coalition play inside the γ-best response of the coalition to S. Moreover, if a strategy of a player is included in a self-enforcing set (implying that the given strategy can only be played if it is common certainty that play is in this set), and the γ-best response of some coalition does not include this strategy, then γ-rationality of this coalition implies that the above strategy cannot be played.[7] This establishes that play has to be within any coalition's generalized γ-best response to S, therefore it has to be included in $E^1(\gamma)$. Common certainty of coalitional γ-rationalizability then implies that it is common certainty that play is in $E^1(\gamma)$. Applying the same arguments iteratively then establishes that common certainty of every coalition being coalitionally γ-rational implies that it is common certainty that play is in $E^*(\gamma)$. Then rationality of players, together with the result that $E^*(\gamma)$ is closed under rational behavior implies that play is included in $E^*(\gamma)$. The other part of the statement can be shown by creating a particular type space. In this type space every player has a type belonging to every coalitionally γ-rationalizable strategy in the sense that he plays the given strategy and has a conjecture to which this strategy is a best response and which conjecture is concentrated on $E^*(\gamma)$. Such a conjecture exists because $E^*(\gamma)$ is coherent. Furthermore, there exists a conjecture like that with a maximal support. Then property (ii) of a sensible best response correspondence can be used to show that both rationality of every player and common certainty of every coalition being rational are satisfied in every state of the world of this model. Since by construction every coalitionally γ-rationalizable strategy is played in some state of the world, this establishes the claim.

[7] The example of Figure 1 in the next section provides more intuition on this point.

Propositions 3 and 4 imply that the set of coalitionally γ-rationalizable strategies is a nonempty and coherent set for every sensible coalitional best response correspondence γ.

5 Coalitional rationalizability and epistemic coalitional rationalizability

Ambrus [04] introduces the concept of coalitionally rationalizable strategies as follows. Let $A^0 = S$ and define A^k $k \geq 1$ iteratively such that $A^k = \bigcap_{B \in \mathcal{F}_J(A^{k-1})} B$. The set of coalitionally rationalizable strategies, A^*, is defined to be $\bigcap_{k \geq 0} A^k$ (equivalently, the limit of A^k as $k \to 0$). The propositions in the previous section imply that the set of coalitionally γ^*-rationalizable strategies is a subset of A^* and in a generic class of games the two solution concepts are equivalent.

Furthermore, in every game both solution concepts yield a nonempty, coherent set of strategies. This also implies that there is always at least one Nash equilibrium of every finite game that is fully contained in the set of coalitionally γ^*-rationalizable strategies. There are two further results on the set of coalitionally rationalizable strategies that can be extended to the epistemic solution concept. The first is that it is possible to provide a direct characterization of the solution set. It is the unique set A which satisfies that (i) $G^{\gamma^*}(A, J) = A$ $\forall\ J \in \mathcal{C}$, and (ii) $A \subset G^{\gamma^*}(B, J) \subset B$ $\forall\ B \supset A,\ J \in \mathcal{C}$. The second is that every strong Nash equilibrium (see Aumann [59]) is fully included in the set of coalitionally γ^*-rationalizable strategies. The proofs of these claims are similar to the corresponding claims in Ambrus [04] and therefore omitted.[8]

Figure 1 below provides an example that the set of coalitionally γ^*- rationalizable strategies can be a strict subset of the set the set of coalitionally rationalizable strategies.

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

Figure 1

In the above example there is no nontrivial supported restriction given S. In particular $A1$ and $B1$ are coalitionally rationalizable strategies. However, note

[8]See Propositions 6 and 7 in the cited paper.

that P1 only plays $A1$ if she is certain that P2 plays $B1$. Similarly P2 only plays $B1$ if she is certain that P1 plays $A1$. This implies that $A1$ or $B1$ are only played if P1 or P2 is certain that the other player is certain that it is common certainty that $(A1, B1)$ is played. But then P1 or P2 is also certain that it is common certainty that play is inside $\{A1, A2\} \times \{B1, B2\} \in \mathcal{M}$. And note that $\{A2\} \times \{B2\}$ is a supported restriction by {P1,P2} given $\{A1, A2\} \times \{B1, B2\}$. This concludes that $A1$ and $B1$ are not coalitionally γ^*-rationalizable. The set of coalitionally γ^*-rationalizable strategies is $\{A2, A3\} \times \{B2, B3\}$.

The example shows that (in non-generic games) there can be strict subsets of the strategy space that have the property that if rationality and common certainty of rationality hold, then it has to be common certainty that play is within the set whenever a given strategy is played. These are exactly the self-supporting sets. Since the definition of coalitional γ^*-rationalizability (and γ-rationalizability in general) refers to sets for which it is common certainty that play is within the set, supported restrictions given self-supporting sets might play a role in determining whether some strategies are coalitionally γ^*-rationalizable or not.

We conclude this section by showing that there exists a sensible best response correspondence γ' such that the resulting coalitionally γ'-rationalizable strategies is exactly equivalent to the set of coalitionally rationalizable strategies defined in Ambrus [04]. Denote the latter set of strategies by A^*.

For any $J \in \mathcal{C}$ and $A \in \mathcal{X}$ let B be a cautious supported restriction by J given A if it is a supported restriction by J given A and $B_i \supset A_i \cap A_i^* \ \forall\, i \in I$. Let $\mathcal{F}'_J(A)$ denote the set of cautious supported restrictions by J given A. Then define γ' such that $\gamma'(A, J) = \bigcap\limits_{B \in \mathcal{F}'_J(A)} B$.

Implicit in γ' is the assumption that coalitions only look for supported restrictions outside A^*, but not within. The definition of γ' is less appealing than that of γ^*, since it directly refers to the set A^*.[9] Nevertheless, as the next proposition states, γ' is a sensible best response correspondence and the set of coalitionally rationalizable strategies resulting from it is exactly equivalent to A^*.

Proposition 6: $\gamma' \in \Gamma^*$ and the set of γ'-rationalizable strategies is A^*.

Since the underlying best response correspondence can be defined in a more natural way, the set of γ^*-rationalizable strategies is an epistemically more well-founded concept than A^*. On the other hand, A^* can be defined by a simple iterative procedure and hence more easily computable in applications. Furthermore, in all games it contains all γ^*-rationalizable strategies, therefore any

[9] Note that the set A^* is defined independently of the epistemic part, by the iterative definition, therefore the definition of γ' is not self-referential.

statement that holds in a game (or in any class of games) for every strategy in A^* also holds for every γ^*-rationalizable strategy. The only drawback of using A^* as opposed to γ^*-rationalizable strategies is that one uses a slightly weaker solution concept than what could be justified. Finally, as shown above, the two concepts are generically equivalent.

6 A remark on the product structure of restrictions

The construction only considers restrictions that are product subsets of the strategy space. This has a natural interpretation if players' conjectures are required to be independent. If correlated conjectures are allowed then focusing on product subsets might seem to result in loss of generality. However, this is not the case: extending the construction to non-product sets leads to the same (product) subset of the strategy space.

The set of conjectures concentrated on a set, $\Delta_{-i}(A)$, can be extended to non-product sets. Then the definition of sensible coalitional best response can be applied to non-product sets, too. Closedness under rational behavior can be similarly extended. The event that a coalition is rational then can be defined as before, a requirement that if it is common certainty among players in the coalition that play is inside a set that is closed under rational behavior (but now not necessarily product), then they play within the best response to the set. Coalitional γ-rationalizability can then be defined as before. It is possible to show that the strategy profiles that are consistent with coalitional γ-rationalizability in this context are exactly the same as in the original construction, for the projection of γ to product sets. The intuition is similar to the reason that the set of rationalizable strategies is a product set even if one allows for correlated conjectures, namely that the non-equilibrium context imposes a product structure on solution sets. Different players can have completely different conjectures and therefore strategies they play can have completely different justifications.

7 Conclusion

There is a wide variety of solution concepts in noncooperative game theory that make an implicit assumption that groups of players can coordinate their play if it is in their common interest. Strong Nash equilibrium and coalition-proof Nash equilibrium - both of which are defined both in static and in certain dynamic games - are examples of concepts that allow this type of coordination for subgroups of players. Different versions of renegotiation-proof Nash equilibrium are concepts which assume that only the coalition of all players can coordinate their play at different stages of a dynamic game.[10] A common feature of these concepts is that assumptions concerning when coordination is feasible or credible are made either on intuitive grounds or referring to an unmodeled negotiation procedure. There is also a line of literature that explicitly models (pre-play or during the game) negotiations among players.[11] One problem associated with these models is that their predictions are sensitive to the exact specification of the negotiations, and typically there is no obvious way to specify the rules of negotiations. Second, typically further assumptions are required to ensure that players can send meaningful and credible messages to each other. And these assumptions are once again made on intuitive grounds, referring to unmodeled features of the interaction, which brings up similar concerns as in the case when negotiations are not explicitly modeled.

This paper is the first attempt to impose assumptions on players' beliefs in an epistemic context to obtain formal foundations for assuming that players with similar interest recognize their common interest and play in a way that is mutually advantageous for them. It is far from clear how to formalize the latter intuitive assumption in noncooperative games, which lead to the emergence of competing solution concepts (for example renegotiation-proof Nash equilibrium has various definitions in the context of infinitely repeated games). Therefore continuing this line of work and making explicit the underlying assumptions that these concepts impose on the knowledge, beliefs and behavior of players seems to be of highlighted importance.

8 Appendix

Proof of Proposition 1: By construction $\gamma^* \in \Gamma$.

$A \in \mathcal{M}$ implies that if $a \in \arg\max_{s \in A} u_j(s)$ then $u_j(a) = \max_{f_{-j} \in \Delta_{-j}(A)} \widehat{u}_j(f_{-j})$ and a_j is a best response strategy to the conjecture that puts probability 1 on other players playing a_{-j}. Then the definition of supported restriction implies that $a_j \in B_j\ \forall\ B \in \mathcal{F}_J(A)$, $J \in \mathcal{C}$ and $j \in J$. Then $a_j \in (\bigcap_{B \in \mathcal{F}_J(A)} B)_j = (\gamma^*(A,J))_j$ which establishes that γ^* satisfies property (iv) in the definition of a sensible supported restriction.

Suppose now that $B, B' \in \mathcal{F}_J(A)$ for some $A \in \mathcal{M}$ and $J \in \mathcal{C}$. This means that $\forall\ j \in J$, $f_{-j} \in \Delta^*_{-j}(A_j/B_j) \cap \Delta_{-j}(A)$ it is the case that $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$ $\forall\ g_{-j} \in \Delta_{-j}(B)$ such that $g_{-j}^{-J} = f_{-j}^{-J}$, and that $\forall\ j \in J$, $f_{-j} \in \Delta^*_{-j}(A_j/B'_j) \cap \Delta_{-j}(A)$ it is the case that $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})\ \forall\ g_{-j} \in \Delta_{-j}(B')$ such that $g_{-j}^{-J} = f_{-j}^{-J}$. These imply that $\forall\ j \in J$, $f_{-j} \in \Delta^*_{-j}(A_j/(B_j \cup B'_j) \cap \Delta_{-j}(A)$ it is the case that $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})\ \forall\ g_{-j} \in \Delta_{-j}(B' \cap B)$ such that $g_{-j}^{-J} = f_{-j}^{-J}$. Furthermore, as shown above, $B \cap B'$ is nonempty. By construction it is also

[10] See for example Farrell and Maskin [89], Bernheim and Ray [89], Abreu et al [93] and Benoit and Krishna [93].

[11] Without completeness some papers along this line are: Farrell [88], Myerson [89], Rabin [90] and [94], Ray and Vohra [97] and [99], Mariotti [97].

a product set which satisfies that $(B \cap B')_{-J} = A_{-J}$. This concludes that $B \cap B' \in \mathcal{F}_J(A)$. Finiteness of A then implies $\gamma^*(A, J) \in \mathcal{F}_J(A)$. It follows from the definition of a supported restriction that $A \in \mathcal{M}$ implies $B \in \mathcal{M}$ $\forall$ $B \in \mathcal{F}_J(A)$. Therefore $\gamma^*(A, J) \in \mathcal{M}$, establishing that γ^* satisfies property (i) in the definition of a sensible supported restriction.

Let $a_i \in A_i$ be such that there is no $f_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f_{-i})$. Note that $A \in \mathcal{M}$ implies that $A_i/\{a_i\} \neq \emptyset$. Then the definition of supported restriction implies that $(A_i/\{a_i\}) \times A_{-i} \in \mathcal{F}_J(A)$ and hence $\gamma^*(A, J) \subset (A_i/\{a_i\}) \times A_{-i}$ $\forall$ $J \ni i$. This establishes that γ^* satisfies property (ii) in the definition of a sensible supported restriction.

Assume again that $A \in \mathcal{M}$ and let $J \in \mathcal{C}$. Consider $C \subset A$ such that $\gamma^*(A, J) \cap C \neq \emptyset$ and $C \in \mathcal{M}$. Let $B \in \mathcal{F}_J(A)$. Note that $\gamma^*(A, J) \cap C \neq \emptyset$ implies $B \cap C \neq \emptyset$. Furthermore, $B \in \mathcal{F}_J(A)$ implies that $\forall$ $j \in J$, $f_{-j} \in \Delta^*_{-j}(A_j/B_j) \cap \Delta_{-j}(A)$ it is the case that $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$ $\forall$ $g_{-j} \in \Delta_{-j}(B)$ such that $g^{-J}_{-j} = f^{-J}_{-j}$, from which it follows that $\forall$ $j \in J$, $f_{-j} \in \Delta^*_{-j}(C_j/B_j) \cap \Delta_{-j}(C)$ it is the case that $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$ $\forall$ $g_{-j} \in \Delta_{-j}(B \cap C)$, establishing that $B \cap C \in \mathcal{F}_J(A)$. Since $B \in \mathcal{F}_J(A)$ was arbitrary, $\gamma^*(C, J) \subset \gamma^*(A, J)$. This establishes that γ^* satisfies property (iii) in the definition of a sensible supported restriction. This concludes the claim. QED

Proof of Proposition 2: Define $\gamma^M \in \Gamma$ such that $\gamma^M(A, J) = \underset{j \in J}{\times} \{a_i \in A_i \mid \exists\ f_{-i} \in \Delta_{-i}(A) \text{ st } a_i \in BR_i(f_{-i})\}$ if $A \in \mathcal{M}$ and $\gamma^M(A, J) = A$ if $A \in \mathcal{X}/\mathcal{M}$, $\forall$ $J \in \mathcal{C}$. There cannot be a larger valued coalitional best response correspondence satisfying (ii), and it trivially satisfies all the other properties in the definition of sensibility.

Correspondence γ^m can be constructed iteratively as follows:

For any $A \in \mathcal{M}$ and $J \in \mathcal{C}$ let $T^{J,0}(A)$ denote the smallest set in $\mathcal{M}$ for which $(T^{J,0}(A))_{-J} = A_{-J}$ and which contains $\{a \in A \mid \exists\ j \in J \text{ st } u_j(a) \geq u_j(s)\ \forall\ s \in A\}$. There exists a set like that since $A, A' \in \mathcal{M}$ imply $A \cap A' \in \mathcal{M}$. Moreover, $T^{J,0}(A) \subset A$ since $A \in \mathcal{M}$ and $\{a \in A \mid \exists\ j \in J \text{ st } u_j(a) \geq u_j(s) \forall s \in A\} \subset A$. Note that if a_i is such that there is no $f_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f_{-i})$ then by construction $a_i \notin T^{J,0}(A)$ (otherwise $T^{J,0}(A)$ was not the smallest set satisfying the above conditions). Properties (i) and (iv) of a sensible best response correspondence imply that $T^{J,0}(A) \subset \gamma(A, J)$ for any $\gamma \in \Gamma^*$. Suppose now that for some $k \geq 0$ we defined $T^{J,k}(A)$ for every $A \in \mathcal{M}$ and $J \in \mathcal{C}$. Assume that $T^{J,k}(A) \in \mathcal{M}$ and that $T^{J,k}(A)$ is such that if for $a_i \in A_i$ there is no $f_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f_{-i})$ then by construction $a_i \notin T^{J,k}(A)$. Furthermore, assume we established that $T^{J,k}(A) \subset \gamma(A, J)$ for any $\gamma \in \Gamma^*$. Define $\widehat{T}^{J,k}(A) = \underset{B \in \mathcal{M}:\ T^{J,k}(A) \cap B \neq \emptyset, B \subset A}{\cup} \gamma(B, J)$. Note that $T^{J,k}(A) \subset \widehat{T}^{J,k}(A)$ since $T^{J,k}(A) \in \mathcal{M}$ and $T^{J,k}(A) \cap T^{J,k}(A) \neq \emptyset$. Let $T^{J,k+1}(A)$ be the smallest set in $\mathcal{M}$ for which $(T^{J,0}(A))_{-J} = A_{-J}$ and which contains $\widehat{T}^{J,k}(A)$. Then the starting assumption that $T^{J,k}(A) \subset \gamma(A, J)$ for any $\gamma \in \Gamma^*$, and properties (i) and (iii) of a sensible best response correspondence imply that $T^{J,k}(A) \subset \gamma(A, J)$ for any $\gamma \in \Gamma^*$. Also note that by construction it holds that if for $a_i \in A_i$ there is no $f_{-i} \in \Delta_{-i}(A)$ such that $a_i \in BR_i(f_{-i})$ then $a_i \notin T^{J,k+1}(A)$. This establishes that $T^{J,0}(A), T^{J,1}(A), ...$ is an increasing sequence of sets such that $T^{J,k}(A) \in \mathcal{M}$ and $T^{J,k}(A) \subset A \ \forall \ k = 1, 2, ...$ Since S is finite, there has to be $K \geq 0$ such that $T^{J,k}(A) = T^{J,K}(A) \ \forall \ k \geq K$. Let $\gamma^m(A, J) = T^{J,K}(A) \ \forall \ A \in \mathcal{M}$ and $J \in \mathcal{C}$. The above arguments imply that $\gamma^m(A, J) \subset \gamma(A, J) \ \forall \ \gamma \in \Gamma^*$ and that properties (i) and (ii) of a sensible best response correspondence hold for γ^m. $T^{J,0}(A) \subset \gamma^m(A, J)$ implies that γ^m satisfies property (i) as well. Furthermore, $T^{J,K+1}(A) = T^{J,K+1}(A)$ implies that γ^m satisfies property (iv), establishing that it is the smallest sensible best response correspondence. QED

Proof of Proposition 3: Suppose $A \in \mathcal{M}$. Let $A' \in \mathcal{X}$ be such that $A'_i = \{s_i \in A_i \mid \exists \ f_{-i} \in \Delta_{-i}(A) \text{ st } s_i \in BR_i(f_{-i})\}$. $A \in \mathcal{M}$ implies $A' \neq \emptyset$. The starting assumption implies that if $B \in \mathcal{N}_J(A)$ for some $J \in \mathcal{C}$ then either $B = A$ or $B_j \cap A'_j = \emptyset \ \forall \ j \in J$. By property (ii) of a sensible best response correspondence $s_j \in A_j / A'_j$ for $j \in J$ implies that $s_j \notin \gamma(A, J)$. Therefore $A^-_j(J) = A_j / (\gamma(A, J))_j \ \forall \ j \in J$, which implies that $G^\gamma(A, J) = \gamma(A, J)$. QED

Lemma 1: Let $A \in \mathcal{M}$ and $\gamma \in \Gamma^*$. Then $\bigcap_{J \in \mathcal{C}} G^\gamma(A, J) \neq \emptyset$.

Proof: let a be such that $u_j(a) = \max_{s \in A} u_j(s)$. Let $A' \in \mathcal{N}(A)$ be such that $a_j \in A'_j$ and let $A'' \in \mathcal{M}$ such that $A' \subset A'' \subset A$. The assumptions $u_j(a) = \max_{s \in A} u_j(s)$ and $A' \in \mathcal{N}(A)$ together imply that $a \in A'$. Then $u_j(a) = \max_{s \in A''} u_j(s)$. But then property (iii) of a sensible best response implies that $a_j \in (\gamma(A'', J))_j$ $\forall \ J \in \mathcal{C}$. Therefore $a_j \in G^\gamma_j(A, J) \ \forall \ J \in \mathcal{C}$. This establishes the claim since j was arbitrary and $\bigcap_{J \in \mathcal{C}} G^\gamma(A, J)$ is a product set. QED

Lemma 2: Let $A \in \mathcal{M}$ and $\gamma \in \Gamma^*$. Then $G^\gamma(A, J) \in \mathcal{M}$.

Proof: Suppose not. Then:

(*) $\exists \ i \in I$, $f_{-i} \in \Delta_{-i}(G^\gamma(A, J))$ such that $a_i \in BR_i(f_{-i})$, and

(**) $\exists \ J \in \mathcal{C}$, $B \in \mathcal{N}(A)$ and $C \in \mathcal{M}$ such that $B \subset C \subset A$, $a_i \in B_i$ and $a_i \notin (\gamma(C, J))_i$.

From (*) and the assumptions that $B \in \mathcal{N}(A)$ and $a_i \in B_i$ it follows that $\text{supp} f_{-i} \subset B_{-i}$ and therefore $\text{supp} f_{-i} \subset C_{-i}$. Then $f_{-i} \in \Delta_{-i}(G^\gamma(A, J))$ implies $\text{supp} f_{-i} \subset (\gamma(C, J))_{-i}$. But then $\gamma(C, J) \in \mathcal{M}$ (which follows from $\gamma \in \Gamma^*$) implies that $a_i \in (\gamma(C, J))_i$, contradicting (**). QED

Proof of Proposition 4: since S is finite and $E^{k-1}(\gamma) \supset E^k(\gamma)$, $\forall \ k \geq 1$, the existence of $K \geq 0$ in the claim is immediate.

Next, note that $E^0(\gamma) = S \in \mathcal{M}$. Assume $E^k(\gamma) \in \mathcal{M}$ for some $k \geq 0$. By Lemma 1, $E^{k+1}(\gamma) \neq \emptyset$. By Lemma 2 $G^\gamma(E^k(\gamma), J) \in \mathcal{M} \ \forall \ J \in \mathcal{C}$ which implies $E^{k+1}(\gamma) \in \mathcal{M}$ since the intersection of sets that are closed under rational behavior is also closed under rational behavior. By induction $E^k(\gamma) \in \mathcal{M}$ and $E^k(\gamma) \neq \emptyset \ \forall \ k \geq 0$. Since $E^*(\gamma) = E^k(\gamma)$ whenever $k \geq K$, this implies $E^*(\gamma) \neq \emptyset$ and $E^*(\gamma) \in \mathcal{M}$.

Now suppose $G^\gamma(E^*(\gamma), J) \neq E^*(\gamma)$. Since $E^*(\gamma) = E^K(\gamma)$, this implies that $E^{K+1}(\gamma) \neq E^K(\gamma)$, contradicting that $E^*(\gamma) = E^k(\gamma) \ \forall \ k \geq K$. QED

Lemma 3: Let $\gamma \in \Gamma^*$. Let $C \in \mathcal{M}$, $J \in \mathcal{C}$ and $B \in \mathcal{N}_J(E^*)$ such that $B \subset A$. Then $\gamma(C, J) \supset B$.

Proof: Suppose not.

Consider first $\gamma(C, J) \cap E^* \neq \emptyset$. Note that $E^* \cap C \in \mathcal{M}$ since both $E^* \in \mathcal{M}$ and $C \in \mathcal{M}$. Therefore property (iii) of a sensible best response correspondence implies that $\gamma(E^* \cap C, J) \subset \gamma(C, J)$. But note that $B \in \mathcal{N}_J(E^*)$ and $\gamma(C, J) \not\supseteq B$, and therefore $\gamma(E^* \cap C, J) \not\supseteq B$. This implies $G^\gamma(E^*, J) \not\supseteq E^*$, contradicting Proposition 4.

Consider next $\gamma(C, J) \cap E^* = \emptyset$. Let k be such that $E^k \cap \gamma(C, J) \neq \emptyset$ but $E^{k+1} \cap \gamma(C, J) = \emptyset$. Note that $E^k \cap C \neq \emptyset$. Furthermore, $C \in \mathcal{M}$ and $E^k \in \mathcal{M}$ imply $E^k \cap C \in \mathcal{M}$. Property (iii) of the best response correspondence, together with the assumption that $E^* \cap C \neq \emptyset$ and hence $E^{k+1} \cap C \neq \emptyset$, implies that $\gamma(E^k \cap C, J') \subset \gamma(E^k, J')$ $\forall$ $J' \in \mathcal{C}$. This establishes that $\bigcap_{J' \in \mathcal{C}} \gamma(E^k \cap C, J') \subset E^{k+1}$. But note that $(E^k \cap C) \cap \gamma(C, J) = E^k \cap \gamma(C, J) \neq \emptyset$, therefore property (iii) of the best response correspondence implies $\gamma(E^k \cap C, J) \subset \gamma(C, J)$. Combining the above yields $\bigcap_{J' \in \mathcal{C}} \gamma(E^k \cap C, J') \subset E^{k+1} \cap \gamma(C, J)$. But this contradicts Lemma 1 since $E^{k+1} \cap C \neq \emptyset$. QED

Proof of Proposition 5: Suppose first that $\phi \in R \cap \mathcal{CC}_I(CR^\gamma)$. Note that $E^0(\gamma) = S$ implies $\phi \in \mathcal{CC}_I(\Psi^{E^0(\gamma)})$. Assume now that for some $k \geq 0$ it holds that $\phi \in \mathcal{CC}_I(\Psi^{E^k(\gamma)})$. Let now $J \in \mathcal{C}$ and let $B \in \mathcal{N}_J(E^k(\gamma))$ be such that $s(\phi) \in B$. Then $\phi \in \mathcal{CC}_I(\Psi^{E^k(\gamma)})$ and $\phi \in R$ together imply that $\phi \in \mathcal{CC}_J(\Psi^B)$. Therefore $\phi \in \mathcal{CC}_I(CR^\gamma)$ implies $\phi \in \mathcal{CC}_I(\{s(\phi) \in B \rightarrow s(\phi) \in B \cap \gamma(C, J)\})$. This in turn implies $\phi \in \mathcal{CC}_I(\Psi^{\gamma(E^k(\gamma), J)})$. Since $J \in \mathcal{C}$ was arbitrary, this in turn implies $\phi \in \mathcal{CC}_I(\Psi^{E^{k+1}(\gamma)})$. By induction then $\phi \in \mathcal{CC}_I(\Psi^{E^*(\gamma)})$. Then $E^*(\gamma) \in \mathcal{M}$ and $\phi \in R$ imply that $s(\phi) \in E^*(\gamma)$.

Let now $s^* \in E^*(\gamma)$. Construct the following type space. For every $i \in N$ let Φ_i be such that for every $s_i \in S_i$ there exists exactly one $\phi_i \in \Phi_i$ st $s_i(\phi_i) = s_i$. Denote it by $\phi_i^{s_i}$. For every $s_i \in E_i^*(\gamma)$ let $f_{-i}^{s_i} \in \Delta_{-i}(E^*(\gamma))$ be such that $s_i \in BR_i(f_{-i}^{s_i})$ and supp$f_{-i}^{s_i}$ $\supset$suppf_{-i} $\forall$ $f_{-i} \in \Delta_{-i}(E^*(\gamma))$ such that $s_i \in BR_i(f_{-i})$. There exists such $f_{-i}^{s_i}$ since $E^*(\gamma)$ is coherent and because $f_{-i}, f'_{-i} \in \Delta^*_{-i}(\{s_i\})$ implies $\alpha f_{-i} + (1-\alpha) f'_{-i} \in \Delta^*_{-i}(\{s_i\})$ $\forall$ $\alpha \in (0, 1)$, implying that there exists an element of $\Delta_{-i}(E^*(\gamma)) \cap \Delta^*_{-i}(\{s_i\})$ with maximal support. Now let $t_i(\phi_i^{s_i})$ be such that $t_i(\phi_i^{s_i})([\phi_j^{s_j}]_{j \in N/i}) = f_{-i}^{s_i}(s_{-i})$ $\forall$ $s_{-i} \in S_{-i}$. Consider $\phi^* \in \Phi$ such that $\phi_i^* = \phi_i^{s_i^*}$. Then by construction $s(\phi^*) = s^*$ and $\phi^* \in R$. Also by construction $\phi^* \in \mathcal{CC}_I(\Psi^{E^*(\gamma)})$. Consider now any $\phi \in \Phi$ and any $J \in \mathcal{C}$ and $A \in \mathcal{M}$ such that $\phi \in \mathcal{CC}_J(\Psi^A)$. By the construction of Φ there is $B \in \mathcal{N}_J(E^*(\gamma))$ such that $B \subset A$ and $s(\phi) \in B$. By Lemma 3 then $\gamma(A, J) \supset B$ and therefore $s_j(\phi) \in (\gamma(A, J))_j$ $\forall$ $j \in J$. This implies that $\phi \in R_J^\gamma$ $\forall$ $\phi \in \Phi$ and $J \in \mathcal{C}$. Therefore $\phi \in \mathcal{CC}_I(CR^\gamma)$ $\forall$ $\phi \in \Phi$. In particular $\phi^* \in \mathcal{CC}_I(CR^\gamma)$. QED

Proof of Proposition 6: By construction $\gamma' \in \Gamma$. Also by construction $\gamma'(A, J) \supset \gamma^*(A, J)$, therefore the fact that γ^* satisfies (iii) in the definition of a sensible best response correspondence, which follows from Proposition 1, implies that γ' satisfies the same property. Also established in the proof of Proposition 1 is that $A \in \mathcal{M}$ implies $B \in \mathcal{M}\ \forall\ B \in \mathcal{F}_J(A),\ J \in \mathcal{C}$. Then $\mathcal{F}'_J(A) \subset \mathcal{F}_J(A)$ implies $B \in \mathcal{M}\ \forall\ B \in \mathcal{F}'_J(A)$. Since $\mathcal{M}$ is closed with respect to taking intersections, this establishes that γ' satisfies (i) in the definition of a sensible best response correspondence. Suppose now that B is a supported restriction by J given $A \in \mathcal{M}$ and $B_i \supset A_i \cap A_i^*\ \forall\ i \in I$. Let $A' \in \mathcal{M}$ be such that $A' \subset A$ and $B \cap A' \neq \varnothing$. Then the definition of a supported restriction implies that $B \cap A'$ is a supported restriction by J given A'. Furthermore, by construction $(B \cap A')_i \cap A_i^* = A'_i \cap A_i^*\ \forall\ i \in I$, so $B \cap A'$ is a cautious supported restriction by J given A'. Since B was an arbitrary cautious supported restriction by J given A, this implies $\gamma'(A, J) \supset \gamma'(A', J)\ \forall\ A, A' \in \mathcal{M}$ such that $A \supset A'$ and $\gamma'(A, J) \cap A' \neq \varnothing$. Therefore γ' satisfies (ii) in the definition of a sensible best response correspondence, which concludes that $\gamma' \in \Gamma^*$.

Then by Proposition 5 the set of γ'-rationalizable strategies is $E^*(\gamma')$. By construction $E^k(\gamma') \supset A^*\ \forall\ k \geq 0$, therefore $E^*(\gamma') \supset A^*$. Next note that for every $J \in \mathcal{C}$ and every $B \in \mathcal{F}_J(A)$ it holds that $B \in \mathcal{F}'_J(A)$. Therefore $E^1(\gamma') \subset A^1$. Then by Lemma 2 of Ambrus [04] for every $J \in \mathcal{C}$ and for every $B \in \mathcal{F}_J(A^1)$ it holds that $B \cap E^1(\gamma') \in \mathcal{F}_J(E^1(\gamma'))$ (note that for any such B it holds that $B \supset A^*$ therefore $B \cap E^1(\gamma') \neq \varnothing$ and so the conditions for the above lemma hold). Since for any such B it holds that $B \cap E^1(\gamma') \supset A^*$, it also holds that $B \cap E^1(\gamma') \in \mathcal{F}'_J(E^1(\gamma'))$. This implies that $E^2(\gamma') \subset A^2$. Iterative application of the previous argument implies $E^k(\gamma') \subset A^k$, which in turn implies $E^*(\gamma') \subset A^*$. Combining the above findings yields $E^*(\gamma') = A^*$. QED

9 References

ABREU, D., D. PEARCE AND E. STACCHETTI (1993): "Renegotiation and Symmetry in Repeated Games," *Journal of Economic Theory*, 60, 217-40

AMBRUS, A. (2004): "Coalitional rationalizability," mimeo Princeton University and Harvard University

AUMANN, R. J. (1959): "Acceptable points in general cooperative n-person games," in *Contributions to the theory of games IV*, Princeton University Press, Princeton, NJ

——— (1990): "Nash equilibria are not self-enforcing," in *Economic Decision-making*, Elsevier Science Publishers B.V.

BASU, K. AND J. W. WEIBULL (1991): "Strategy subsets closed under rational behavior," *Economics Letters*, 36, 141-146

BATTIGALLI, P. AND G. BONANNO (1999): "Recent results on belief, knowledge and the epistemic foundations of game theory," *Research In Economics*, 53, 149-225

BENOIT, J. and V. KRISHNA [1993]: "Renegotiation in Finitely Repeated Games," *Econometrica*, 61, 303-23

BERNHEIM, B. D. (1984): "Rationalizable strategic behavior," *Econometrica*, 52, 1007-1028

BERNHEIM, B. D., B. PELEG AND M. D. WHINSTON (1987): "Coalition-proof Nash equilibria I. Concepts," *Journal of Economic Theory*, 42, 1-12

BERNHEIM, D. B. AND D. RAY (1989): "Collective dynamic consistency in repeated games," *Games and Economic Behavior,* 1, 327-360

BRANDENBURGER, A. AND E. DEKEL (1993): "Hierarchies of beliefs and common knowledge," *Journal of Economic Theory*, 59, 189-198

DEKEL, E. AND F. GUL (1997): "Rationality and Knowledge in Game Theory," in *Advances in economics and econometrics: Theory and applications, Seventh World Congress, Volume 1, Econometric Society Monographs, no. 26*, D. M. Kreps and K. F. Wallis, (eds.), Cambridge University Press, Cambridge, New York and Melbourne, 87-172

EPSTEIN, L. (1997): "Preference, rationalizability and equilibrium," *Journal of Economic Theory*, 73, 1-29

FARRELL, J. (1988): "Communication, coordination and Nash equilibrium," *Economics Letters*, 27, 209-214

FARRELL, J. and E. MASKIN [1989]: "Renegotiation in Repeated Games," *Games and Economic Behavior,* 1, 327-360

GUL, F. (1996): "Rationality and coherent theories of rational behavior," *Journal of Economic Theory*, 70, 1-31

KANDORI, M., G. J. MAILATH AND R. ROB (1993): "Learning, mutation, and long run equilibria in games," *Econometrica*, 61, 29-56

MARIOTTI, M. (1997): "A model of agreements in strategic form games," *Journal of Economic Theory*, 74, 196-217

MERTENS, J.-F. AND S. ZAMIR (1985): "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory,* 14, 1-29

MYERSON, R. B. (1989): "Credible negotiation statements and coherent plans," *Journal of Economic Theory*, 48, 264-303

PEARCE, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52, 1029-1050

RABIN, M. (1990): "Communication between rational agents," *Journal of Economic Theory*, 51, 144-170

——— (1994): "A model of pre-game communication," *Journal of Economic Theory*, 63, 370-391

RAY, D. AND R. VOHRA (1997): "Equilibrium binding agreements," *Journal of Economic Theory*, 73, 30-78

——— (1999): "A theory of endogenous coalition structures," *Games and Economic Behavior*, 26, 286-336

ROBSON, A. AND F. VEGA-REDONDO (1996): "Efficient equilibrium selection in evolutionary games with random matching," *Journal of Economic Theory*, 70, 65-92

TAN, T. and S. WERLANG (1988): "The Bayesian foundation of solution concepts of games," *Journal of Economic Theory*, 45, 370-391