



# DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

## Theories of Coalitional Rationality

The Harvard community has made this article openly available.  
[Please share](#) how this access benefits you. Your story matters.

<b>Citation</b>	Ambrus, Attila. 2009. Theories of coalitional rationality. <i>Journal of Economic Theory</i> 144(2): 676-695.
<b>Published Version</b>	<a href="https://doi.org/10.1016/j.jet.2007.03.010">doi:10.1016/j.jet.2007.03.010</a>
<b>Accessed</b>	February 18, 2015 12:17:52 AM EST
<b>Citable Link</b>	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:3204917">http://nrs.harvard.edu/urn-3:HUL.InstRepos:3204917</a>
<b>Terms of Use</b>	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

*(Article begins on next page)*

# Theories of coalitional rationality\*

Attila Ambrus<sup>†</sup>

## 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 an operator from sets of conjectures to sets of strategies. A strategy is epistemic coalitionally rationalizable if it is consistent with rationality and common certainty that every coalition is rational. A characterization of this solution set is provided for operators satisfying four basic properties. Special attention is devoted to an operator that leads to a solution concept that is generically equivalent to the iteratively defined concept of coalitional rationalizability.

Keywords: normal form games, coalitional agreements, coordination, epistemic foundations of solution concepts

JEL classification: C72

---

\*I thank Satoru Takahashi for careful proofreading as well as for useful comments, Drew Fudenberg and three anonymous referees for useful comments, and Eddie Dekel and Jeffrey Ely for asking me questions concerning my previous research that inspired me to write this paper.

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

# 1 Introduction

Since Aumann's paper nearly fifty years ago ([3]) there have been numerous attempts to incorporate coalitional reasoning into the theory of noncooperative 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 best-known equilibrium concepts along this line are strong Nash equilibrium ([3],[4]) and coalition-proof Nash equilibrium ([9]). 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 [2] proposes the concept of coalitional rationalizability, using an iterative procedure. The construction is similar to the original definition of rationalizability, provided by [8] and [25]. The new aspect is that not only never best-response strategies of individual players are deleted by the procedure, but also strategies of groups of players simultaneously, if it is in their mutual interest to confine their play to the remaining set of strategies. These are called supported restrictions by 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 [11] and Brandenburger and [31], 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.<sup>1</sup> We define the response of a coalition to be an operator that allocates a set of strategies to certain sets of conjectures. The assumption that the operator 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, nevertheless it can be common certainty among them that play is within a certain subset of the strategy space. Intuitively then the 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 response operators.

Each response operator can be used to obtain a definition of coalitional rationality the following way. A coalition is rational if for every subset of strategies for which it is common certainty among coalition members that play is within this set, members of the coalition play within the response set to the set of conjectures concentrated on this 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 in 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 operator  $\gamma$  as coalitional  $\gamma$ -rationalizability.

We then investigate the class of response operators that satisfy four properties. Two of these serve the purpose of establishing consistency with individual best response correspondences. The third one imposes a form of monotonicity on the operator, reflecting the idea that if restricting play in a

---

<sup>1</sup>This is the starting point for rationalizability as well, although [15] considers building the concept on alternative definitions of rationality.

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. Finally, the fourth property requires that the 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 weak requirement along the lines of Pareto optimality of the best response operator for coalition members, but it turns out to be enough to establish our main results. We call the above response operators sensible best responses. We show that there is a smallest and a largest sensible best response operator.

It is shown that for every sensible best response operator  $\gamma$  the resulting set of coalitionally  $\gamma$ -rationalizable strategies is nonempty, and it can be characterized by an iterative procedure that is defined from the corresponding best response operator. 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 yields a subset of the set of strategies reached by the simpler iterative procedure.

The best response operator that we pay special attention to uses the concept of supported restriction as defined in [2]. 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 [2]. In a nongeneric class of games the former can be a strict subset of the latter, providing a (slightly) stronger refinement of rationalizability. Finally, we show that there exists another sensible best response operator that leads to a set of epistemic coalitionally rationalizable strategies that is exactly equivalent to the iteratively defined set of coalitionally rationalizable strategies.

The relationships between the solution concepts derived in this paper and standard noncooperative solution concepts, including strong Nash equilibrium and coalition-proof Nash equilibrium, are analogous to the relationship between the set of coalitionally rationalizable strategies in [2] and the same noncooperative solution concepts, therefore we do not discuss the issue here. We refer the interested reader to the same paper for an extended discussion of how our approach is related to other approaches of incorporating coalitional reasoning into noncooperative game theory. Here we only provide a brief summary of the related literature. One set of papers assumes an explicit and public negotiation procedure among players ([13],[18],[21],[32]). Another line of literature investigates games in which players can sign binding agreements with each other (see for example [12],[26],[29],[30]). All the above papers assume that secret negotiations among subgroups of players are not possible.

## 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,  $S = \times_{i \in I} S_i$ , is the set of strategies, and  $u = (u_i)_{i \in I}$ , and  $u_i : S \rightarrow \mathbb{R}$ , for every  $i \in I$ , are the payoff functions. We assume that  $S_i$  is finite for every  $i \in I$ . Let  $S_{-i} = \times_{j \in I \setminus \{i\}} S_j$ ,  $\forall i \in I$  and let  $S_{-J} = \times_{j \in I \setminus J} S_j$ ,  $\forall J \subset I$ . Let  $\mathcal{C} = \{J \mid J \subset I, J \neq \emptyset\}$ . We will refer to elements of  $\mathcal{C}$  as *coalitions*.

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

Let  $\Delta_{-i}$  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. Let  $f_{-i}^{-J}$  be the marginal distribution of  $f_{-i}$  over  $S_{-J}$ , for every  $J \in \mathcal{C}$ ,  $i \in J$  and  $f_{-i} \in \Delta_{-i}$ . 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  $f_{-i} \in \Delta_{-i}$  and  $s_i \in S_i$ . Let  $BR_i(f_{-i}) = \{s_i \in S_i \mid \forall t_i \in S_i, u_i(s_i, f_{-i}) \geq u_i(t_i, f_{-i})\}$ , the set of pure strategy best responses of player  $i$  to conjecture  $f_{-i}$ , for every  $f_{-i} \in \Delta_{-i}$ .

## 2.2 Type spaces

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 measurable space,  $\Phi_i$  is a subset of  $S_i \times T_i$  such that  $\text{proj}_{S_i} \Phi_i = S_i$ , and  $g_i : T_i \rightarrow \Delta(\Phi_{-i})$  (where  $\Delta(\Phi_{-i})$  is the set of probability measures on  $\Phi_{-i}$ ) is a measurable function with respect to the  $\sigma$ -algebra generated by

$$\{\{\mu \mid \mu(A) \geq p\} \mid p \in [0, 1] \text{ and } A \subset \Phi_i \text{ measurable}\}$$

on  $\Delta(\Phi_{-i})$ .<sup>2</sup>

$T_i$  represents the set of epistemic types of player  $i$ .  $\Phi$  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  $g_i$ .<sup>3</sup>

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

We say that  $i$  is *certain of*  $\Psi_{-i} \subset \Phi_{-i}$  at  $\phi \in \Phi$  if  $g_i(t_i(\phi_i))(\Psi_{-i}) = 1$ .<sup>4</sup>

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.

We say that  $i$  is certain of  $\Psi = \Psi_i \times \Psi_{-i} \subset \Phi$  at  $\phi \in \Phi$  if  $i$  is certain of  $\Psi_{-i}$  at  $\phi$ .

Let  $C_i(\Psi) \equiv \{\phi \in \Phi : g_i(t_i(\phi_i))(\Psi_{-i}) = 1\}$ , for any  $\Psi = \Psi_i \times \Psi_{-i}$ .  $C_i(\Psi)$  is the event in the state space that  $i$  is certain of  $\Psi$ .

---

<sup>2</sup>All finite sets are endowed with the obvious sigma algebra, and all product spaces are endowed with the product sigma algebra. Each subset  $X$  of a measurable space  $Y$  is endowed with the sigma algebra induced by the inclusion map from  $X$  to  $Y$ . The construction is the same as in [20], in particular we assume the same sigma algebra on  $\Delta(\Phi_{-i})$ .

<sup>3</sup>For more on type spaces see for example [6] and [14].

<sup>4</sup>The terminology “ $i$  believes  $\Psi_{-i}$ ” is also common in the literature.

Let  $\Psi = \times_{i \in I} \Psi_i$  where  $\Psi_i \subset \Phi_i$  (a product event). *Mutual certainty* of  $\Psi$  holds at  $\phi \in \Phi$  if  $\phi \in \bigcap_{i \in I} C_i(\Psi)$ . Mutual certainty of  $\Psi \subset \Phi$  among  $J$  holds at  $\phi \in \Phi$  if  $\phi \in \bigcap_{i \in J} C_i(\Psi)$ .

Let  $MC_J(\Psi)$  denote mutual certainty of  $\Psi$  among  $J$ .

Let  $MC_J^1(\Psi) \equiv MC_J(\Psi)$ . Let  $MC_J^k(\Psi) = MC_J(MC_J^{k-1}(\Psi))$  for  $k \geq 2$ . *Common certainty* of  $\Psi$  among  $J$  holds at  $\phi \in \Phi$  if  $\phi \in \bigcap_{k=1,2,\dots} MC_J^k(\Psi)$ .

Let  $CC_J(\Psi)$  denote common certainty of  $\Psi$  among  $J$ .

### 3 Response operators 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.

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 extend 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 operator 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 a coalition that the conjecture of each of them is concentrated on a product subset of strategies  $A \subset S$ .<sup>5</sup> Then even if they are uncertain 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

---

<sup>5</sup>For a discussion on why we only consider product sets see Section 7.



would implicitly agree upon confining its play to, given a set of conjectures, can be interpreted as a best response operator. The problem is that there is no single obvious definition of a restriction being in the 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. For this reason, we proceed by considering a wide range of possible coalitional response operators, which lead to different definitions of coalitional rationalizability.

**Definition 1:**  $\gamma : \mathcal{X} \times \mathcal{C} \rightarrow \mathcal{X}$  is a *coalitional response operator* if

- (i)  $\gamma(A, J) \subset A$
- (ii)  $\gamma(A, J) \neq \emptyset$  implies  $(\gamma(A, J))_{-J} = A_{-J}$ .

In words, coalitional response operators are restrictions on the set of strategies such that only strategies of players in the corresponding coalitions are restricted. Let  $\Gamma$  be the set of response operators.<sup>6</sup> Although coalitional response operators are defined as operators from sets of strategies (like operators in [23]), one should interpret them as operators from sets of belief profiles. In particular, the interpretation of  $\gamma(A, J)$  for some  $A \in \mathcal{X}$  and  $J \in \mathcal{C}$  is that it is the set of strategies that players in  $J$  want to restrict play to if it is common certainty among  $J$  that play is inside  $A$ . Therefore, set of strategies  $A$  in the above definition should be thought of as a shortcut for the set of belief profiles for which all players in  $J$  think that it is common certainty among  $J$  that play is in  $A$ .

One example of a coalitional response operator, which formalizes the idea that restricting play within the response set of the coalition should be strictly in the interest of coalition members, can be obtained from the concept of supported restriction of [2].

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

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

---

<sup>6</sup>Note that the definition of a coalitional response operator does not impose any restriction along the lines of optimality for coalition members. This is why we do not call all operators in  $\Gamma$  coalitional *best* response operators.

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 2:**  $B$  is a *supported restriction* by  $J$  given  $A$  if

- 1)  $B_i = A_i, \forall i \notin J$ , and
- 2)  $\widehat{u}_j(f_{-j}) < \widehat{u}_j(g_{-j})$  for every  $j \in J, f_{-j} \in \Delta_{-j}^*(A_j \setminus B_j) \cap \Delta_{-j}(A)$ , and  $g_{-j} \in \Delta_{-j}(B)$  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 her conjecture is concentrated on  $B$  than if her 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 can be established (see [2]) that  $\bigcap_{B \in \mathcal{F}_J(A)} B$  is either empty or itself a member of  $\mathcal{F}_J(A)$ . This motivates a response operator that specifies  $\bigcap_{B \in \mathcal{F}_J(A)} B$  to be the response of  $J$  to the set of conjectures concentrated on  $A$ . Note that  $\mathcal{F}_J(A) \subset A$  since  $B \subset A$  for every  $A \in \mathcal{X}$  and  $J \in \mathcal{C}$ .

**Definition 3:** Let  $\gamma^* : \mathcal{X} \times \mathcal{C} \rightarrow \mathcal{X}$  be the operator defined by  $\gamma^*(\emptyset, J) = \emptyset$  and  $\gamma^*(A, J) = \bigcap_{B \in \mathcal{F}_J(A)} B$  for every  $A \in \mathcal{X} \setminus \{\emptyset\}$ , for all  $J \in \mathcal{C}$ .

Operator  $\gamma^*$  will be our leading example throughout the paper. However, the above is just one reasonable 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  $s \in B, t \in A \setminus B$  and  $s_{-J} = t_{-J}$  imply that  $u_j(s) > u_j(t)$  for every  $j \in J$  (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  $\Delta_{-i}(A \setminus B)$ :  $f_{-i} = \alpha^{f_{-i}} f_{-i}^B + (1 - \alpha^{f_{-i}}) f_{-i}^{A/B}$ , 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. The next section will analyze a class of response operators that satisfy certain intuitive properties. This class contains some of the examples given above, including  $\gamma^*$ .

Next we define the concept of rationality of a coalition, given some  $\gamma \in \Gamma$ . The definition refers to subsets of the strategy space that are called closed under rational behavior.<sup>7</sup> Recall that a set  $A \in \mathcal{X} \setminus \{\emptyset\}$  is closed under rational behavior if  $BR_i(f_{-i}) \subset A_i$ , for every  $i \in I$  and  $f_{-i} \in \Delta_{-i}(A)$ .

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

For any  $\gamma \in \Gamma$  we define a coalition to be  $\gamma$ -rational at some state of the world if the strategy profile that is played at that state is within the  $\gamma$ -response of the coalition to any closed under rational behavior set which satisfies that it is common certainty among the coalition members that play is within this set. We only make this restriction with respect to closed under rational behavior sets because our intention is building a coalitional rationality concept that is consistent with individual rationality and the above sets are exactly the ones that are compatible with individual rationality of coalition members and the assumption that it is common certainty among them that play is within the set.

Let  $\Psi^A = \{\phi \in \Phi \mid s(\phi) \in A\}$  for every  $\emptyset \neq A \subset S$ . Then  $CC_J(\Psi^A)$  is the event that there is common certainty among  $J$  that play is in  $A$ .

**Definition 4:** coalition  $J$  is  $\gamma$ -rational at  $\phi \in \Phi$  if  $\phi \in CC_J(\Psi^A)$  implies  $s_i(\phi_i) \in \Psi_i^{\gamma(A, J)}$  for every  $i \in J$  and  $A \in \mathcal{M}$ .

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

---

<sup>7</sup>[5] introduced the terminology and analyzed the properties of these sets.

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

Let  $\bar{g}_{-i}(\phi_i)$  denote the marginal distribution of  $g_i(t_i(\phi_i))$  over  $S_{-i}$ . This is the conjecture of type  $\phi_i$  of player  $i$  regarding what strategies other players play. Following standard terminology, we call player  $i$  individually rational at  $\phi$  if  $s_i(\phi_i) \in BR_i(\bar{g}_{-i}(\phi_i))$ . Let  $R_i$  denote the event that player  $i$  is rational and let  $R = \bigcap_{i \in N} R_i$  (the event that every player is rational).

A strategy profile is coalitionally  $\gamma$ -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  $\gamma$ -rationality hold.<sup>8</sup>

**Definition 5:**  $s \in S$  is *coalitionally  $\gamma$ -rationalizable* if  $\exists$  type space  $T$  and  $\phi \in \Phi$  such that  $\phi \in R \cap CC_I(CR^\gamma)$  and  $s(\phi) = s$ .

In particular coalitional  $\gamma^*$ -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 operators

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

**Definition 6:**  $\gamma \in \Gamma$  is a *sensible coalitional best response operator* if it satisfies the following properties:

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

---

<sup>8</sup>Since [11] and [22] 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.

- (ii) if  $\nexists f_{-i} \in \Delta_{-i}(A)$  such that  $a_i \in BR_i(f_{-i})$  then  $a_i \notin (\gamma(A, J))_i$ , for every  $A \in \mathcal{M}$ ,  $i \in N$ ,  $a_i \in A_i$  and  $J \ni i$
- (iii) if  $B \subset A$  and  $\gamma(A, J) \cap B \neq \emptyset$  then  $\gamma(B, J) \subset \gamma(A, J)$ , for every  $A, B \in \mathcal{M}$
- (iv) if  $a \in \arg \max_{s \in A} u_j(s)$  then  $a_j \in (\gamma(A, J))_j$ , for every  $J \in \mathcal{C}$ ,  $j \in J$ , and  $A \in \mathcal{M}$ .

Properties (i) and (ii) impose consistency of the coalitional response operator with individual best response operator. Property (i) requires that the 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 responses of coalitions containing the player. Note that (i) and (ii) imply that the 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 response, and (ii) implies that all other strategies are excluded from the response.

Property (iii) is a monotonicity condition. It corresponds to the idea that if outcomes in  $\gamma(A, J)$  in some sense (the exact meaning depends on  $\gamma$ ) are preferred to outcomes in  $A \setminus \gamma(A, J)$  by players in  $J$ , then outcomes in  $B \cap \gamma(A, J)$  are preferred to outcomes in  $B \cap (A \setminus \gamma(A, J))$ . Equivalently, if coalition  $J$ 's response involves not playing strategies in  $A \setminus \gamma(A, J)$  for a set of contingencies (namely when play is concentrated on  $A$ ), then their 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 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 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 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.

Let  $\Gamma^*$  denote the set of sensible coalitional best response operators. One example of an operator in  $\Gamma^*$  is the operator obtained from supported restrictions in the previous section. This will be used in the next section, where we compare the epistemically defined concept of  $\gamma^*$ -rationalizability to the iteratively defined concept of coalitional rationalizability from [2].

The proofs of all propositions are in the Appendix.

**Proposition 1:**  $\gamma^*$  is a sensible best response operator.

We note that there are various ways of changing the definition of the supported restriction that lead to coalitional best response operators different from  $\gamma^*$ , 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  $\Gamma^*$ . The largest is the one that only excludes (individual) never best-response strategies for coalition members.<sup>9</sup> The smallest one can be defined in an iterative manner. It involves starting out from the operator 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 operator and then iteratively enlarging the values of the operator until it satisfies property (iii).<sup>10</sup>

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

---

<sup>9</sup>For sets in  $\mathcal{M}$ . For other sets it is equal to the identity operator. The definition of sensibleness does not restrict the operator in any way for the latter sets.

<sup>10</sup>Formally, let  $\gamma^0$  denote the operator 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. For  $k \geq 1$ , define  $\gamma^k$  iteratively such that  $\gamma^k(A, J)$  is the smallest set in  $\mathcal{M}$  containing  $\gamma^{k-1}(B, J)$  for all  $B \in \mathcal{M}$  such that  $\gamma(A, J) \cap B \neq \emptyset$ , for every  $A \in \mathcal{M}$  and  $J \in \mathcal{C}$ .

In the remainder of this section we derive a result which characterizes the set of coalitionally  $\gamma$ -rationalizable strategies for any sensible coalitional best response operator  $\gamma$ .

In the construction that we provide below, particular sets of strategies, which we call self-supporting for some coalition, play a highlighted role. These are sets with the property that for any member  $j$  of  $J$  choosing a strategy from the set can only be a best response if  $j$  thinks that it is common certainty among  $J$  that play is in the set. To put it simply, play can only be inside these sets if the latter is common certainty. The central role of these sets in coalitional  $\gamma$ -rationalizability is due to the fact that the definition of coalition  $J$  being  $\gamma$ -rational requires that players in  $J$  play inside  $\gamma(A, J)$  for every  $A \in \mathcal{M}$  for which it is common certainty among  $J$  that play is in  $A$ .

**Definition 7:** Let  $A \in \mathcal{M}$ . A subset  $B \in \mathcal{X}$  of  $A$  is *self-supporting* for  $J$  with respect to  $A$  if  $BR_j^{-1}(b_j) \cap \Delta_{-j}(A) \subset \Delta_{-j}(B)$  for every  $j \in J$  and  $b_j \in B_j$ .

	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

For an example of a self-supporting set, consider the game of Figure 1. Here the singleton set  $\{A1\} \times \{B1\}$  is self-supporting for the coalition of the two players for  $S$ , since  $A1$  can only be a best response for player 1 if she believes that player 2 chooses  $B1$  with probability 1, and  $B1$  can only be a best response for player 2 if he believes that player 1 chooses  $A1$  with probability 1. Therefore rational players can only play  $A1$  and  $B1$  if it is common certainty among them that play is in  $C$ , for any  $C \in \mathcal{M}$  which satisfies that  $\{A1\} \times \{B1\} \subset C$ .

Let  $\mathcal{N}_J(A)$  denote the collection of self-supporting sets for  $J$  with respect to  $A$ . Note that  $A \in \mathcal{N}_J(A)$ , for every  $A \in \mathcal{M}$  and  $J \subset I$ . Let  $\overline{\mathcal{M}}^A(B) =$

$\{C \in \mathcal{M} \mid B \subset C \subset A\}$ , for every  $B \subset A$ . In words, members of  $\overline{\mathcal{M}}^A(B)$  are sets in  $A$  that are closed under rational behavior and contain  $B$ .

Below we will show that  $\gamma$ -rationalizable strategies for any  $\gamma \in \Gamma^*$  can be obtained by an iterative procedure that is associated with a decreasing sequence of sets  $E^0, E^1, E^2, \dots$ . The procedure starts from the set of all strategies ( $E^0 = S$ ) and at step  $k$  it eliminates any strategy  $a_i$  of any  $i \in I$  for which the following holds: there exists  $J \ni i$  and a set  $B$  self-supporting for  $J$  with respect to  $E^{k-1}$  such that for some  $C \in \overline{\mathcal{M}}^{E^{k-1}}(B)$  it holds that  $a_i \notin (\gamma(C, J))_i$ . In particular, since  $A \in \mathcal{N}_J(A)$  for every  $A \in \mathcal{M}$ , at each step  $k$  the procedure eliminates all strategies that are not in the  $\gamma$ -response of  $E^{k-1}$  for some coalition. However, if  $\mathcal{N}_J(E^{k-1})$  contains some proper subsets of  $E^{k-1}$ , then further strategies can be eliminated in this step.

Formally, consider the following procedure. Let  $E^0(\gamma) = S$ . For every  $k \geq 1$ , let  $E^k(\gamma) = \times_{i \in I} E_i^k(\gamma)$ , where  $E_i^k(\gamma) \equiv \{s_i \in E_i^{k-1}(\gamma) \mid \forall J \in \mathcal{C}, B \in \mathcal{N}_J(E^{k-1}) \text{ and } C \in \overline{\mathcal{M}}^{E^{k-1}}(B) \text{ s.t. } s_i \in C_i \text{ it holds that } s_i \in (\gamma(C, J))_i\}$ .

**Definition 8:** Let  $E^*(\gamma) = \bigcap_{k=0,1,2,\dots} E^k(\gamma)$ .

In order to prove our main theorem, it is useful to establish some basic properties of  $E^*(\gamma)$  for sensible coalitional best response operators.

**Proposition 3:** For every  $\gamma \in \Gamma^*$ , the following hold:

- (i)  $E^*(\gamma)$  is nonempty
- (ii)  $\exists K < \infty$  such that  $E^k(\gamma) = E^*(\gamma)$  for every  $k \geq K$
- (iii)  $E^*(\gamma) \in \mathcal{M}$
- (iv)  $E_i^*(\gamma) \equiv \{s_i \in E_i^*(\gamma) \mid s_i \in (\gamma(C, J))_j \text{ for every } B \in \mathcal{N}_J(E^*) \text{ s.t. } s_j \in B_j, \text{ and } C \in \overline{\mathcal{M}}^{E^*}(B)\}$ .

The outline of the proof is the following. Condition (iii) in the definition of a sensible best response operator 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.



Note that result (iv) in Proposition 3 implies  $\gamma(E^*(\gamma), J) = E^*(\gamma)$  for every  $J \in \mathcal{C}$  and  $\gamma \in \Gamma^*$ . Hence, there exists  $f_{-i} \in \Delta_{-i}(E^*(\gamma))$  such that  $s_i \in BR_i(f_{-i})$ , for every  $i \in I$  and  $s_i \in E_i^*(\gamma)$ . We refer to sets that are closed under rational behavior and satisfy the above property as coherent.

We are ready to establish our main result, the equivalence of  $E^*(\gamma)$  and the set of coalitionally  $\gamma$ -rationalizable strategies.

**Proposition 4:** If  $\phi \in R \cap CC_I(CR^\gamma)$  then  $s(\phi) \in E^*(\gamma)$ , for every  $\gamma \in \Gamma^*$ , type space  $T$  and state  $\phi \in \Phi$ .

Conversely, there exists type space  $T$  and  $\phi \in \Phi$  such that  $s(\phi) = s$  and  $\phi \in R \cap CC_I(CR^\gamma)$ , for every  $s \in E^*(\gamma)$ .

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  $\gamma$ -rational implies that players of any coalition play inside the  $\gamma$ -response of the coalition to  $S$ . Moreover, if a strategy of a player is included in a self-supporting set (implying that the given strategy can only be played if it is common certainty that play is in this set), and the  $\gamma$ -response of some coalition does not include this strategy, then  $\gamma$ -rationality of this coalition implies that the above strategy cannot be played. To summarize,  $\gamma$ -rationality of all coalitions implies that play is inside  $E^1(\gamma)$ . Common certainty of coalitional  $\gamma$ -rationality then implies that it is common certainty that play is in  $E^1(\gamma)$ . Applying the same arguments iteratively yields that common certainty of coalitional  $\gamma$ -rationality implies 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  $\gamma$ -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 operator 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  $\gamma$ -rationalizable strategy is played in some state of the world, this establishes the claim.

Combining Propositions 3 and 4 establishes that the set of coalitionally  $\gamma$ -rationalizable strategies is a nonempty and coherent set for every sensible coalitional best response operator  $\gamma$ .

The iterative procedure introduced above is quite complicated, because in each step it refers to the  $\gamma$ -response of various subsets of the set of strategies reached in the previous step. We conclude this section by showing that in a generic class of games the procedure is equivalent to a simple and intuitive procedure.

Let  $\Delta_{-i}^o(A) = \{f_{-i} \in \Delta_{-i}(A) \mid \forall a_{-i} \in A_{-i}, f_{-i}(a_{-i}) > 0\}$  for every  $A \in \mathcal{M}$ . Let  $\mathcal{G}^o$  denote the class of games for which  $\{a_i \in A_i \mid \exists f_{-i} \in \Delta_{-i}(A) \text{ s.t. } a_i \in BR_i(f_{-i})\} = \{a_i \in A_i \mid \exists f_{-i} \in \Delta_{-i}^o(A) \text{ s.t. } a_i \in BR_i(f_{-i})\}$ . In words,  $\mathcal{G}^o$  is the set of games in which no set that is closed under rational behavior has a strategy that is weakly but not strictly dominated within that set. It is straightforward to show that the latter property is generic.<sup>11</sup> In these games the only self-supporting set for  $J$  with respect to  $A$  is  $A$  itself, for any  $A \in \mathcal{M}$  and  $J \in \mathcal{C}$ . This considerably simplifies the iterative procedure defining  $E^*$ .

**Proposition 5:** In any game in  $\mathcal{G}^o$  it holds that  $E^k(\gamma) = \bigcap_{J \in \mathcal{C}} \gamma(E^{k-1}(\gamma), J)$  for every  $\gamma \in \Gamma^*$  and  $k \geq 1$ .

By propositions 4 and 5, in a generic class of games the set of  $\gamma$ -rationalizable strategies can be obtained simply by taking the intersection of  $\gamma$ -responses of all possible coalitions in an iterative manner, starting from the set of all strategies. Furthermore, it is straightforward to show that this simple iterative procedure leads to a set which contains the set of  $\gamma$ -rationalizable strategies in every finite game, for all  $\gamma \in \Gamma^*$ .

---

<sup>11</sup>Formally, the following statement is true: fix any finite game form (i.e. games with a given number of players and strategies) and associate the set of games with points of  $\mathbb{R}^{|I| \times |S_1| \times \dots \times |S_I|}$  according to the payoffs they allocate for different strategy profiles; then the set of points in  $\mathbb{R}^{|I| \times |S_1| \times \dots \times |S_I|}$  that are associated with games in which the stated property does not hold is negligible (it is measure zero according to the Lebesgue measure).

## 5 Coalitional rationalizability and epistemic coalitional rationalizability

The concept of coalitionally rationalizable strategies is introduced in [2] as follows. Let  $A^0 = S$  and define  $A^k$  for  $k \geq 1$  iteratively such that  $A^k = \bigcap_{J \in \mathcal{C}} \bigcap_{B \in \mathcal{F}_J(A^{k-1})} B$ . The set of coalitionally rationalizable strategies,  $A^*$ , is defined to be  $\bigcap_{k \geq 0} A^k$  (equivalently, the limit of  $A^k$  as  $k \rightarrow \infty$ ). The propositions in the previous section imply that the set of coalitionally  $\gamma^*$ -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  $\gamma^*$ -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. The second is that every strong Nash equilibrium (see [3]) is fully included in the set of coalitionally  $\gamma^*$ -rationalizable strategies. The proofs of these claims are similar to the corresponding claims in [2] and therefore omitted.<sup>12</sup>

The game of Figure 1 in the previous section provides an example that the set of coalitionally  $\gamma^*$ -rationalizable strategies can be a strict subset of the set of coalitionally rationalizable strategies. In this game there is no nontrivial supported restriction given  $S$ . In particular  $A1$  and  $B1$  are coalitionally rationalizable strategies. However, note 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 and P2 are also certain that it is common certainty that play is inside  $\{A1, A2\} \times \{B1, B2\} \in \mathcal{M}$ . Furthermore, note that  $\{A2\} \times \{B2\}$  is a supported restriction by  $\{P1, P2\}$  given  $\{A1, A2\} \times \{B1, B2\}$ . This implies that  $A1$  and  $B1$  are not coalitionally  $\gamma^*$ -rationalizable. The set of coalitionally  $\gamma^*$ -rationalizable strategies is  $\{A2, A3\} \times \{B2, B3\}$ .

---

<sup>12</sup>See Propositions 6 and 7 in the cited paper.

The key feature of the above example is that although there is no nontrivial supported restriction given  $S$ , there is a set closed under rational behavior ( $\{A1, A2\} \times \{B1, B2\}$ ), which contains a set self-supporting with respect to  $S$  ( $\{A1\} \times \{B1\}$ ), and which is such that there is a supported restriction given this set for the coalition of both players that does not contain strategies from the self-supporting set. Note that the game in Figure 1 is not in  $\mathcal{G}^o$ , since it has a self-supporting set with respect to  $S$  which is a strict subset of  $S$ .

We conclude this section by showing that there exists a sensible best response operator  $\gamma'$  such that the resulting set of coalitionally  $\gamma'$ -rationalizable strategies is exactly equivalent to the set of coalitionally rationalizable strategies defined in [2]. Denote the latter set of strategies by  $A^*$ .

For any  $J \in \mathcal{C}$  and  $A \in \mathcal{X}$ , let  $B$  be a *conservative supported restriction* by  $J$  given  $A$  if it is a supported restriction by  $J$  given  $A$  and satisfies the following requirement: if  $a_i \in A_i \cap A_i^*$  is such that  $\exists f_{-i} \in \Delta_{-i}(A)$  for which  $a_i \in BR_i(f_{-i})$  then  $a_i \in B_i$ . Let  $\mathcal{F}'_J(A)$  denote the set of conservative supported restrictions by  $J$  given  $A$ . Define  $\gamma'$  such that  $\gamma'(A, J) = \bigcap_{B \in \mathcal{F}'_J(A)} B$ .

Intuitively, the definition of  $\gamma'$  requires that coalitions only look for supported restrictions outside  $A^*$ , but not within. The definition of  $\gamma'$  is less appealing than that of  $\gamma^*$ , since it directly refers to the set  $A^*$ .<sup>13</sup> Nevertheless, as the next proposition states,  $\gamma'$  is a sensible best response operator and the set of coalitionally rationalizable strategies resulting from it is exactly equivalent to  $A^*$ .

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

Since the underlying best response operator can be defined in a more natural way, the set of  $\gamma^*$ -rationalizable strategies as a solution concept is built on more solid foundations than  $A^*$ . On the other hand,  $A^*$  can be defined by a simple iterative procedure and hence is easier to use in applications. Furthermore, in all games it contains all  $\gamma^*$ -rationalizable strategies, therefore any statement that holds in a game (or in any class of games) for every strategy in  $A^*$  also holds for every  $\gamma^*$ -rationalizable strategy. Finally, as shown above, the two concepts are generically equivalent.

---

<sup>13</sup>Note that the set  $A^*$  is defined independently of the epistemic part, by the iterative definition, therefore the definition of  $\gamma'$  is not self-referential.

## 6 Cautious coalitional rationalizability concepts

Coalitional  $\gamma^*$ -rationalizability requires that players are confident that their co-players understand the implicit agreements implied by supported restrictions and comply with it. As the game of Figure 2 illustrates, this might involve taking a lot of risk.

	B1	B2
A1	99,99	99,0
A2	0,99	100,100

Figure 2

In this game  $\{A2\} \times \{B2\}$  is a supported restriction by the coalition of both players, since it yields a payoff of 100, while strategies  $A1$  and  $B1$  can yield at most 99 for players 1 and 2. On the other hand, strategies  $A1$  and  $B1$  yield a payoff of 99 for sure, which implies that player 1's conjecture should allocate at least 0.99 probability to player 2 playing  $B2$  for  $A2$  to be a best response to the conjecture. Similarly, player 2's conjecture should allocate at least 0.99 probability to player 1 playing  $A2$  for  $B2$  to be a best response to the conjecture.

In many contexts this amount of confidence in other players complying to supported restrictions is not realistic, implying that coalitional  $\gamma^*$ -rationalizability is not the appropriate solution concept to apply. However, the consideration that players should be cautious in complying to implicit coalitional agreements is compatible with many other coalitional  $\gamma$ -responses.

Consider the following modification of supported restrictions:<sup>14</sup>

Let  $\Delta_{-i}^\varepsilon(A) = \{f_{-i} \in \Delta_{-i} \mid f_{-i}(A_{-i}) \geq 1 - \varepsilon\}$  for every  $i \in I$ ,  $A \in \mathcal{X}$ , and  $\varepsilon > 0$ .

---

<sup>14</sup>The idea in this construction is similar to the one behind cautious rationalizability ([25]).

**Definition 9:**  $B \subset A$  is an  $\varepsilon$ -cautious supported restriction by  $J \in \mathcal{C}$  given  $A \in \mathcal{X}$  if

- 1)  $B_i = A_i$  for every  $i \notin J$ , and
- 2)  $\hat{u}_j(f_{-j}) < \hat{u}_j(g_{-j})$  for every  $j \in J$ ,  $f_{-j} \in \Delta_{-j}^*(A_j \setminus B_j) \cap \Delta_{-j}^\varepsilon(A)$ , and  $g_{-j} \in \Delta_{-j}^\varepsilon(B)$  such that  $g_{-j}^{-J} = f_{-j}^{-J}$ .

Intuitively, a supported restriction is an  $\varepsilon$ -cautious supported restriction if it remains supported even if any player can think that with at most probability  $\varepsilon$  others in the coalition play outside the restriction. Using  $\varepsilon$ -cautious supported restrictions, we define coalitional response operator  $\gamma^\varepsilon$  the same way that  $\gamma^*$  is defined based on supported restrictions. Let  $\mathcal{F}_J^\varepsilon(A)$  be the set of  $\varepsilon$ -cautious supported restrictions by  $J$  given  $A$ . Let  $\gamma^\varepsilon : \mathcal{X} \times \mathcal{C} \rightarrow \mathcal{X}$  be such that  $\gamma^\varepsilon(\emptyset, J) = \emptyset$  and  $\gamma^\varepsilon(A, J) = \bigcap_{B \in \mathcal{F}_J^\varepsilon(A)} B$  for every  $A \in \mathcal{X} \setminus \{\emptyset\}$ , for every  $J \in \mathcal{C}$ .

It is possible to show that  $\gamma^\varepsilon$  is a sensible coalitional best response operator, for every  $\varepsilon > 0$ .<sup>15</sup> Moreover, the resulting set of coalitionally  $\gamma^\varepsilon$ -rationalizable strategies conforms with the idea that players should be cautious when complying to supported restrictions. In particular, in the game of Figure 2 all strategies are coalitionally  $\gamma^\varepsilon$ -rationalizable for  $\varepsilon > 0.01$ .

## 7 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. This is not the case, though: one can extend the domain of coalitional response operators to non-product sets (the set of conjectures concentrated on a non-product set  $\Delta_{-i}(A)$  is well-defined) without changing the range of resulting coalitional rationalizability concepts. The key point is that if players cannot actually correlate their play, then since in a non-equilibrium framework different players can have completely different conjectures (even if the latter allow for correlation), the strategies actually played do not need to be connected

---

<sup>15</sup>The proof is analogous to the proof of Proposition 1, therefore omitted.

in any way through beliefs. Therefore if the response of  $J$  to  $A$  includes both  $a \in A$  and  $a' \in A$  then it should also include  $a_1, \dots, a_{i-1}, a'_i, a_{i+1}, \dots, a_I$  for every  $i \in I$ . Hence the extended response operators should be mappings to  $\mathcal{X}$ , just like the ones defined in Section 3. The above implies that the coalitional rationalizability concepts that can be defined based on these extended response operators lead to solution sets that have the product structure. This is analogous to why the set of correlated rationalizable profiles is a product set for every game.

## 8 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 in static and as well as 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.<sup>16</sup> 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.<sup>17</sup> These models often impose assumptions that ensure that players can send meaningful and credible messages to each other. These assumptions are once again made on intuitive grounds, referring to unmodeled features of the interaction, which brings up concerns similar to those arising 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

---

<sup>16</sup>See for example [1],[7],[10] and [17].

<sup>17</sup>Without completeness some papers along this line are: [16],[24],[27] and [28].

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). For this reason, 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.

## 9 Appendix

**Lemma 1:** If  $A \in \mathcal{M}$  and  $B \in \mathcal{F}_J(A)$  for some  $J \in \mathcal{C}$  then  $B \in \mathcal{M}$ .

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

**Lemma 2:** If  $A, B \in \mathcal{M}$  and  $A \cap B \neq \emptyset$  then  $A \cap B \in \mathcal{M}$ .

**Proof of Lemma 2:** First note that  $A, B \in \mathcal{X}$  and  $A \cap B \neq \emptyset$  imply that  $A \cap B \in \mathcal{X}$ . Let  $i \in I$  and  $f_{-i} \in \Delta_{-i}(A \cap B)$ . From  $f_{-i} \in \Delta_{-i}(A)$  and  $A \in \mathcal{M}$  it follows that  $BR_i(f_{-i}) \subset A_i$ . From  $f_{-i} \in \Delta_{-i}(B)$  and  $B \in \mathcal{M}$  it follows that  $BR_i(f_{-i}) \subset B_i$ . Hence  $BR_i(f_{-i}) \subset (A \cap B)_i$ . The fact that this holds for every  $i \in I$  implies the claim. QED

**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  $u_j(a) = \max_{f_{-j} \in \Delta_{-j}(A)} \hat{u}_j(f_{-j})$  and  $a_j$  is a best response strategy to the conjecture that puts probability 1 on other players playing  $a_{-j}$ . Denote the latter expectation by  $\hat{f}_{-j}$ . Then  $\hat{u}_j(\hat{f}_{-j}) \geq \hat{u}_j(f_{-j})$  for every  $f_{-j} \in \Delta_{-j}(B)$  such that  $B \in \mathcal{M}$  and  $B \subset A$ . Then the definition of supported restriction implies that  $a_j \in B_j$  for every  $B \in \mathcal{F}_J(A), J \in \mathcal{C}$  and  $j \in J$ . From this it follows that  $a_j \in (\bigcap_{B \in \mathcal{F}_J(A)} B)_j = (\gamma^*(A, J))_j$ , which establishes that  $\gamma^*$  satisfies property (iv) in the definition of a sensible supported restriction.

Note that for any  $A \in \mathcal{M}$  and  $J \in \mathcal{C}$  it holds that  $A \in \mathcal{F}_J(A)$ , since  $\Delta_{-j}^*(A_j \setminus B_j) \cap \Delta_{-j}(A) = \emptyset$  implies that requirement (2) in the definition of a supported restriction trivially holds, and requirement (1) trivially holds, too. Suppose now that  $B, B' \in \mathcal{F}_J(A)$  for some  $A \in \mathcal{M}$  and  $J \in \mathcal{C}$ . This means that  $\hat{u}_j(f_{-j}) < \hat{u}_j(g_{-j})$  for every  $j \in J, f_{-j} \in \Delta_{-j}^*(A_j \setminus B_j) \cap \Delta_{-j}(A)$  and  $g_{-j} \in \Delta_{-j}(B)$  such that  $g_{-j}^{-J} = f_{-j}^{-J}$ , and that  $\hat{u}_j(f_{-j}) < \hat{u}_j(g_{-j})$  for



every  $j \in J$ ,  $f_{-j} \in \Delta_{-j}^*(A_j \setminus B'_j) \cap \Delta_{-j}(A)$  and  $g_{-j} \in \Delta_{-j}(B')$  such that  $g_{-j}^{-J} = f_{-j}^{-J}$ . These imply that  $\hat{u}_j(f_{-j}) < \hat{u}_j(g_{-j})$  for every  $j \in J$ ,  $f_{-j} \in \Delta_{-j}^*(A_j \setminus (B_j \cup B'_j)) \cap \Delta_{-j}(A)$  and  $g_{-j} \in \Delta_{-j}(B' \cap B)$  such that  $g_{-j}^{-J} = f_{-j}^{-J}$ . Furthermore, as shown above,  $B \cap B'$  is nonempty. By construction it is also a product set which satisfies that  $(B \cap B')_{-J} \subset A_{-J}$ . Therefore  $B \cap B' \in \mathcal{F}_J(A)$ . Lemma 1 implies  $B, B' \in \mathcal{M}$ , and then Lemma 2 implies  $B \cap B' \in \mathcal{M}$ . Then from the finiteness of  $A$  it follows that  $\gamma^*(A, J) \in \mathcal{F}_J(A)$  and  $\gamma^*(A, J) \in \mathcal{M}$ . The latter establishes that  $\gamma^*$  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 the latter can be rewritten as  $\Delta_{-i}^*(A_i \setminus B_i) \cap \Delta_{-i}(A) = \emptyset$ . Furthermore, note that  $A \in \mathcal{M}$  implies that  $A_i \setminus \{a_i\} \neq \emptyset$  and therefore  $(A_i \setminus \{a_i\}) \times A_{-i} \in \mathcal{X}$ . Then from  $\Delta_{-i}^*(A_i \setminus B_i) \cap \Delta_{-i}(A) = \emptyset$  it follows that  $(A_i \setminus \{a_i\}) \times A_{-i}$  trivially satisfies both requirements in the definition of supported restriction, hence  $(A_i \setminus \{a_i\}) \times A_{-i} \in \mathcal{F}_J(A)$ . Therefore  $\gamma^*(A, J) \subset (A_i \setminus \{a_i\}) \times A_{-i} \forall J \ni i$ . This establishes that  $\gamma^*$  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  $B \subset A$  such that  $\gamma^*(A, J) \cap B \neq \emptyset$  and  $B \in \mathcal{M}$ . Let  $C \in \mathcal{F}_J(A)$ . Note that  $\gamma^*(A, J) \cap C \neq \emptyset$  implies  $B \cap C \neq \emptyset$ . Furthermore,  $C \in \mathcal{F}_J(A)$  implies that  $\hat{u}_j(f_{-j}) < \hat{u}_j(g_{-j})$  for every  $j \in J$ ,  $f_{-j} \in \Delta_{-j}^*(A_j \setminus C_j) \cap \Delta_{-j}(A)$ , and  $g_{-j} \in \Delta_{-j}(C)$  such that  $g_{-j}^{-J} = f_{-j}^{-J}$ , establishing that  $B \cap C \in \mathcal{F}_J(B)$ . This implies  $\hat{u}_j(f_{-j}) < \hat{u}_j(g_{-j})$  for every  $j \in J$ ,  $f_{-j} \in \Delta_{-j}^*(A_j \setminus C_j) \cap \Delta_{-j}(A)$ , and  $g_{-j} \in \Delta_{-j}(B \cap C)$  such that  $g_{-j}^{-J} = f_{-j}^{-J}$ , establishing that  $B \cap C \in \mathcal{F}_J(B)$ . Since this holds for every  $C \in \mathcal{F}_J(A)$ , we have  $\bigcap_{D \in \mathcal{F}_J(B)} D \equiv \gamma^*(B, J) \subset \gamma^*(A, J) \equiv \bigcap_{D \in \mathcal{F}_J(A)} D$ .

This establishes that  $\gamma^*$  satisfies property (iii) in the definition of a sensible supported restriction. QED

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

Operator  $\gamma^m$  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  $\{a \in A \mid \exists j \in J \text{ s.t. } u_j(a) \geq u_j(s) \forall s \in A\}$ . There exists such a set since if  $A, A' \in \mathcal{M}$  and

both  $A$  and  $A'$  contain  $\{a \in A \mid \exists j \in J \text{ s.t. } \forall s \in A, u_j(a) \geq u_j(s)\}$ , then  $A \cap A'$  also contains  $\{a \in A \mid \exists j \in J \text{ s.t. } \forall s \in A, u_j(a) \geq u_j(s)\}$  and by Lemma 2  $A \cap A' \in \mathcal{M}$ . Moreover,  $T^{J,0}(A) \subset A$  since  $A \in \mathcal{M}$  and  $\{a \in A \mid \exists j \in J \text{ s.t. } \forall s \in A, u_j(a) \geq u_j(s)\} \subset A$ . Note that if  $a_i$  is such that there is no  $f_{-i} \in \Delta_{-i}(A)$  for which  $a_i \in BR_i(f_{-i})$ , 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 operator imply that  $T^{J,0}(A) \subset \gamma(A, J)$  for every  $\gamma \in \Gamma^*$ . 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  $a_i \notin T^{J,k}(A)$ . Suppose now that we established that  $T^{J,k}(A) \subset \gamma(A, J)$  for every  $\gamma \in \Gamma^*$ . Define  $\widehat{T}^{J,k}(A) = \bigcup_{B \in \mathcal{M}: T^{J,k}(A) \cap B \neq \emptyset, B \subset A} \gamma(B, J)$ . 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 every  $\gamma \in \Gamma^*$ , and properties (i) and (iii) of a sensible best response operator imply that  $T^{J,k+1}(A) \subset \gamma(A, J)$  for every  $\gamma \in \Gamma^*$ . Also note that by construction, 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  $T^{J,0}(A), T^{J,1}(A), \dots$  is an increasing sequence of sets such that  $T^{J,k}(A) \in \mathcal{M}$  and  $T^{J,k}(A) \subset A$  for  $k = 1, 2, \dots$ . Since  $S$  is finite, there has to be  $K \geq 0$  such that  $T^{J,k}(A) = T^{J,K}(A)$  for every  $k \geq K$ . Let  $\gamma^m(A, J) = T^{J,K}(A)$  for every  $A \in \mathcal{M}$  and  $J \in \mathcal{C}$ . The above arguments imply that  $\gamma^m(A, J) \subset \gamma(A, J)$  for every  $\gamma \in \Gamma^*$ , and that properties (i) and (ii) of a sensible best response operator hold for  $\gamma^m$ . Furthermore,  $T^{J,0}(A) \subset \gamma^m(A, J)$  implies that  $\gamma^m$  satisfies property (i), and  $T^{J,K+1}(A) = T^{J,K+1}(A)$  implies that  $\gamma^m$  satisfies property (iv). This establishes that  $\gamma^m$  is the smallest sensible best response operator. QED

Let  $\eta^\gamma(A, J) = \prod_{i \in J} (A_i \setminus \bigcup_{B \in \overline{\mathcal{M}}^A(C) : C \in \overline{\mathcal{N}}_J(A)} [B_i \setminus (\gamma(A, J))_i]) \times A_{-J}$ . Note that  $E^k(\gamma) = \bigcap_{J \in \mathcal{C}} E^k(\gamma)$  for every  $k \geq 1$ .

**Lemma 3:** Let  $A \in \mathcal{M}$  and  $\gamma \in \Gamma^*$ . Then  $\bigcap_{J \in \mathcal{C}} \eta^\gamma(A, J) \neq \emptyset$ .

**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  $a_j \in (\gamma(A'', J))_j$  for every  $J \in \mathcal{C}$ . Therefore  $a_j \in \eta_j^\gamma(A, J)$  for every  $J \in \mathcal{C}$ . This establishes the claim since  $j$  was arbitrary and  $\bigcap_{J \in \mathcal{C}} \eta^\gamma(A, J)$  is a product set. QED

**Lemma 4:** Let  $A \in \mathcal{M}$  and  $\gamma \in \Gamma^*$ . Then  $\eta^\gamma(A, J) \in \mathcal{M}$ .

**Proof:** Suppose not. Then:

(\*)  $\exists i \in I, f_{-i} \in \Delta_{-i}(\eta^\gamma(A, J))$  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  $f_{-i} \in \Delta_{-i}(\eta^\gamma(A, J))$  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 3:** Since  $S$  is finite and  $E^{k-1}(\gamma) \supset E^k(\gamma)$  for every  $k \geq 1$ , 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 3,  $E^{k+1}(\gamma) \neq \emptyset$ . By Lemma 4  $\eta^\gamma(E^k(\gamma), J) \in \mathcal{M}$  for every  $J \in \mathcal{C}$ , 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  $E^k(\gamma) \neq \emptyset$  for every  $k \geq 0$ . Since  $E^*(\gamma) = E^k(\gamma)$  whenever  $k \geq K$ , this implies  $E^*(\gamma) \neq \emptyset$  and  $E^*(\gamma) \in \mathcal{M}$ .

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

**Lemma 5:** 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 operator 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  $\eta^\gamma(E^*, J) \not\supseteq E^*$ , contradicting Proposition 3.

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 operator,

together with the assumption that  $E^* \cap C \neq \emptyset$  and hence  $E^{k+1} \cap C \neq \emptyset$ , implies that  $\gamma(E^k \cap C, J') \subset \gamma(E^k, J')$  for every  $J' \in \mathcal{C}$ . 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 operator 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 3 since  $E^{k+1} \cap C \neq \emptyset$ . QED

**Proof of Proposition 4:** Suppose first that  $\phi \in R \cap CC_I(CR^\gamma)$ . Note that  $E^0(\gamma) = S$  implies  $\phi \in CC_I(\Psi^{E^0(\gamma)})$ . Assume now that for some  $k \geq 0$  it holds that  $\phi \in CC_I(\Psi^{E^k(\gamma)})$ . Let  $J \in \mathcal{C}$  and let  $B \in \mathcal{N}_J(E^k(\gamma))$  be such that  $s(\phi) \in B$ . Then  $\phi \in CC_I(\Psi^{E^k(\gamma)})$  and  $\phi \in R$  together imply that  $\phi \in CC_J(\Psi^B)$ . Therefore  $\phi \in CC_I(CR^\gamma)$  implies  $\phi \in CC_I(\{s(\phi) \in B \rightarrow s(\phi) \in B \cap \gamma(C, J)\})$ . This in turn implies  $\phi \in CC_I(\Psi^{\gamma(E^k(\gamma), J)})$ . Since  $J \in \mathcal{C}$  was arbitrary, this in turn implies  $\phi \in CC_I(\Psi^{E^{k+1}(\gamma)})$ . By induction then  $\phi \in CC_I(\Psi^{E^*(\gamma)})$ . 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  $\Phi_i$  be such that for every  $s_i \in S_i$ , there exists exactly one  $\phi_i \in \Phi_i$  such that  $s_i(\phi_i) = s_i$ . 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  $\text{supp} f_{-i}^{s_i} \supset \text{supp} f_{-i} \forall f_{-i} \in \Delta_{-i}(E^*(\gamma))$  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  $\alpha f_{-i} + (1 - \alpha) f'_{-i} \in \Delta_{-i}^*(\{s_i\})$  for every  $\alpha \in (0, 1)$ , further 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  $t_i(\phi_i^{s_i})([\phi_j^{s_j}]_{j \in N \setminus i}) = f_{-i}^{s_i}(s_{-i})$  for every  $s_{-i} \in S_{-i}$ . 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  $\phi^* \in CC_I(\Psi^{E^*(\gamma)})$ . Consider now any  $\phi \in \Phi$  and any  $J \in \mathcal{C}$  and  $A \in \mathcal{M}$  such that  $\phi \in CC_J(\Psi^A)$ . By the construction of  $\Phi$  there is  $B \in \mathcal{N}_J(E^*(\gamma))$  such that  $B \subset A$  and  $s(\phi) \in B$ . By Lemma 5 then  $\gamma(A, J) \supset B$  and therefore  $s_j(\phi) \in (\gamma(A, J))_j$  for every  $j \in J$ . This implies that  $\phi \in R_J^\gamma$  for every  $\phi \in \Phi$  and  $J \in \mathcal{C}$ . Therefore  $\phi \in CC_I(CR^\gamma)$  for every  $\phi \in \Phi$ . In particular  $\phi^* \in CC_I(CR^\gamma)$ . QED

**Proof of Proposition 5:** Suppose  $A \in \mathcal{M}$ . Let  $A' \in \mathcal{X}$  be such that  $A'_i = \{s_i \in A_i \mid \exists f_{-i} \in \Delta_{-i}(A) \text{ s.t. } s_i \in BR_i(f_{-i})\}$ .  $A \in \mathcal{M}$  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  $B_j \cap A'_j = \emptyset$  for every  $j \in J$ . By property (ii) of a sensible best response operator  $s_j \in A_j \setminus A'_j$  for  $j \in J$  implies that  $s_j \notin \gamma(A, J)$ .

Therefore  $A_j^-(J) = A_j \setminus (\gamma(A, J))_j$  for every  $j \in J$ , which implies that  $\eta^\gamma(A, J) = \gamma(A, J)$ . QED

**Proof of Proposition 6:** By construction  $\gamma' \in \Gamma$ . Also by construction  $\gamma'(A, J) \supset \gamma^*(A, J)$ . Since by Proposition 1  $\gamma^*$  satisfies (iv) in the definition of a sensible best response operator, the previous relationship implies that  $\gamma'$  satisfies property (iv), too.

Since  $\mathcal{F}'_J(A) \subset \mathcal{F}_J(A)$ ,  $\gamma'(A)$  is defined as an intersection of supported restrictions given  $A$ . By Lemma 1,  $B \in \mathcal{M}$  for every  $B \in \mathcal{F}'_J(A)$ . Then Lemma 2 implies that  $\gamma'$  satisfies (i) in the definition of a sensible best response operator.

Let  $A \in \mathcal{M}$ ,  $i \in I$  and  $a_i \in A_i$  is such that there is no  $f_{-i} \in \Delta_{-i}(A)$  for which  $a_i \in BR_i(f_{-i})$ . Note that  $A \in \mathcal{M}$  implies that  $(A_i \setminus \{a_i\}) \times A_{-i} \neq \emptyset$ . Therefore  $a_i \notin BR_i(f_{-i})$  for every  $f_{-i} \in \Delta_{-i}(A)$  implies that  $(A_i \setminus \{a_i\}) \times A_{-i} \neq \emptyset$  is a supported restriction by any  $J \ni i$  given  $A$ . Moreover, by definition it is a conservative supported restriction. Therefore  $a_i \notin (\gamma'(A, J))_i$  for every  $J \ni i$ , which implies that  $\gamma'$  satisfies (ii) in the definition of a sensible best response operator.

Suppose now that  $B \in \mathcal{F}'_J(A)$  for some  $A \in \mathcal{M}$ . Let  $A' \in \mathcal{M}$  be such that  $A' \subset A$  and  $B \cap A' \neq \emptyset$ . Then the definition of a supported restriction implies that  $B \cap A'$  is a supported restriction by  $J$  given  $A'$  (note that  $\Delta_{-j}(B \cap A') \subset \Delta_{-j}(B \cap A)$ , and  $\Delta_{-j}^*(A'_j \setminus B_j) \cap \Delta_{-j}(A') \subset \Delta_{-j}^*(A_j \setminus B_j) \cap \Delta_{-j}(A)$ ). Furthermore, by construction  $(B \cap A')_i \cap A_i^* = A'_i \cap A_i^*$  for every  $i \in I$ , so  $B \cap A' \in \mathcal{F}'_J(A')$ . Since  $B$  was an arbitrary conservative supported restriction by  $J$  given  $A$ , this implies  $\gamma'(A, J) \supset \gamma'(A', J)$  for every  $A, A' \in \mathcal{M}$  such that  $A \supset A'$  and  $\gamma'(A, J) \cap A' \neq \emptyset$ . Therefore  $\gamma'$  satisfies (iii) in the definition of a sensible best response operator, which concludes that  $\gamma' \in \Gamma^*$ .

Then by Proposition 4 the set of  $\gamma'$ -rationalizable strategies is  $E^*(\gamma')$ . By construction  $E^k(\gamma') \supset A^*$  for every  $k \geq 0$ , therefore  $E^*(\gamma') \supset A^*$ . Next notice that  $A^* \subset A^1$  implies that  $\mathcal{F}_J(S) = \mathcal{F}'_J(S)$  for every  $J \in \mathcal{C}$ , therefore  $E^1(\gamma') \subset A^1$ . Since  $\mathcal{M}$  is closed with respect to taking intersections,  $E^1(\gamma') \in \mathcal{M}$ .

Suppose now that  $E^k(\gamma') \subset A^k$  and  $E^k(\gamma') \in \mathcal{M}$  for some  $k \geq 0$ . As established above,  $\gamma' \in \Gamma^*$ . Also note that  $E^k(\gamma') \cap A^{k+1} \supset E^k(\gamma') \cap A^* \neq \emptyset$ . Then by property (iii) of a sensible best response operator,  $E^{k+1}(\gamma') = \gamma'(E^k(\gamma')) \subset \gamma'(A^k)$  and by property (i)  $\gamma'(E^k(\gamma')) \in \mathcal{M}$ . Also since  $A^{k+1} \supset A^*$ ,  $B \in \mathcal{F}_J(A^k)$  implies  $B \in \mathcal{F}'_J(A^k)$ , therefore  $\gamma'(A^k) \subset A^{k+1}$ . Combining the previous findings establishes  $E^{k+1}(\gamma') \subset A^{k+1}$ . An iterative argument

then establishes  $E^*(\gamma') \subset A^*$ . Thus  $E^*(\gamma') = A^*$ , which in turn establishes the claim. QED

## 10 References

- [1] D. Abreu, D. Pearce, E. Stacchetti, Renegotiation and Symmetry in Repeated Games, *Journal of Economic Theory*, 60 (1993), 217-40.
- [2] A. Ambrus, Coalitional rationalizability, *Quarterly Journal of Economics*, 121 (2006), 903-929.
- [3] R. J. Aumann, Acceptable points in general cooperative n-person games, in: H. W. Kuhn, A. W. Tucker (Eds.), *Contributions to the theory of games IV*, Princeton University Press, Princeton, 1959, pp. 287-324.
- [4] R. J. Aumann, Nash equilibria are not self-enforcing, in: J. J. Gabszewicz, J. F. Richard, L. A. Wolsey (Eds.), *Economic Decision-making*, Elsevier Science Publishers B.V., Amsterdam, 1990, pp. 201-206.
- [5] K. Basu, J. W. Weibull: Strategy subsets closed under rational behavior, *Economics Letters*, 36 (1991), 141-146.
- [6] P. Battigalli, G. Bonanno, Recent results on belief, knowledge and the epistemic foundations of game theory, *Research In Economics*, 53 (1999), 149-225.
- [7] J. Benoit, V. Krishna: Renegotiation in Finitely Repeated Games, *Econometrica*, 61 (1993), 303-323.
- [8] B. D. Bernheim, Rationalizable strategic behavior, *Econometrica*, 52 (1984), 1007-1028.
- [9] B. D. Bernheim, B. Peleg, M. D. Whinston, Coalition-proof Nash equilibria I. Concepts, *Journal of Economic Theory*, 42 (1987), 1-12.
- [10] B. D. Bernheim, D. Ray, Collective dynamic consistency in repeated games, *Games and Economic Behavior*, 1 (1989), 327-360.
- [11] A. Brandenburger, E. Dekel, Hierarchies of beliefs and common knowledge, *Journal of Economic Theory*, 59 (1993), 189-198.
- [12] K. Chatterjee, K., B. Dutta, D. Ray, K. Sengupta, A noncooperative theory of coalitional bargaining, *Review of Economic Studies*, 60 (1993), 463-477.
- [13] M. S. Chwe, Farsighted coalitional stability, *Journal of Economic Theory*, 63 (1994), 299-325.
- [14] E. Dekel, F. Gul, 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, K. F. Wallis (Eds.), Cambridge University Press, Cambridge, New York and Melbourne, 1997, pp. 87-172.

- [15] L. Epstein, Preference, rationalizability and equilibrium, *Journal of Economic Theory*, 73 (1997), 1-29.
- [16] J. Farrell, Communication, coordination and Nash equilibrium, *Economics Letters*, 27 (1988), 209-214.
- [17] J. Farrell, E. Maskin, Renegotiation in Repeated Games, *Games and Economic Behavior*, 1 (1989), 327-360.
- [18] J. Greenberg, *The theory of social situations*, Cambridge University Press, Cambridge, 1990.
- [19] F. Gul, Rationality and coherent theories of rational behavior, *Journal of Economic Theory*, 70 (1996), 1-31.
- [20] A. Heifetz, D. Samet, The Bayesian formulation of incomplete information, *Journal of Economic Theory*, 82 (1998), 324-341.
- [21] M. Mariotti, A model of agreements in strategic form games, *Journal of Economic Theory*, 74 (1997), 196-217.
- [22] J.-F. Mertens, S. Zamir, Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory*, 14 (1985), 1-29.
- [23] P. Milgrom, J. Roberts, Adaptive and sophisticated learning in normal form games, *Games and Economic Behavior*, 3 (1991), 82-100.
- [24] R. B. Myerson (1989): Credible negotiation statements and coherent plans, *Journal of Economic Theory*, 48 (1989), 264-303.
- [25] D. G. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica*, 52 (1984), 1029-1050.
- [26] M. Perry, P. Reny, A noncooperative view of coalition formation and the core, *Econometrica*, 62 (1994), 795-817.
- [27] M. Rabin, Communication between rational agents, *Journal of Economic Theory*, 51 (1990), 144-170.
- [28] M. Rabin, A model of pre-game communication, *Journal of Economic Theory*, 63 (1994), 370-391.
- [29] D. Ray, R. Vohra, Equilibrium binding agreements, *Journal of Economic Theory*, 73 (1997), 30-78.
- [30] D. Ray, R. Vohra, A theory of endogenous coalition structures, *Games and Economic Behavior*, 26 (1999), 286-336.
- [31] T. Tan, S. Werlang, The Bayesian foundation of solution concepts of games, *Journal of Economic Theory*, 45 (1988), 370-391.
- [32] L. Xue, Coalitional stability under perfect foresight, *Economic Theory*, 11 (1998), 603-627.