Bayesian Coalitional Rationalizability*

Xiao Luo[a,†], Chih-Chun Yang[b]

[a]*Institute of Economics, Academia Sinica, Taipei 115, Taiwan, ROC*

[b]*Department of Economics, University of Rochester, Rochester, NY 14627*

January 2007

Abstract

In this paper we extend Ambrus's [*QJE* **121**(2006), 903-929] concept of "coalitional rationalizability (c-rationalizability)" to situations where, in seeking mutual beneficial interests, players in groups (i) make use of Bayes' rule in expectation calculations and (ii) contemplate various deviations, i.e., the validity of deviation is checked against any arbitrary sets of strategies, not only against restricted subsets of strategies. We offer an alternative notion of Bayesian c-rationalizability suitable for such complex social interactions. More specifically, we define Bayesian c-rationalizability by the idea of "coalitional rationalizable set (CRS)."Roughly speaking, a CRS is a product set of pure strategies from which no group of players would like to deviate. We show that Bayesian c-rationalizability possesses nice properties similar to those of conventional rationalizability. *JEL Classification: C70, C72, D81*

Keywords: Bayesian c-rationalizability, iterated c-dominance

*The paper was partially done while the first author was visiting Harvard University. We especially thank Attila Ambrus for his encouragement and helpful comments. We also thank Yi-Chun Chen, Steve Chiu, Kim-Sau Chung, Martin Dufwenberg, Biung-Ghi Ju, Atsushi Kajii, Man-Chung Ng and Satoru Takahashi for helpful comments. Financial support from the National Science Council of Taiwan is gratefully acknowledged. The usual disclaimer applies.

†Corresponding author. Fax:+886-2-2785 3946. E-mail: xluo@econ.sinica.edu.tw.

1 Introduction

In their seminal paper Bernheim [6] and Pearce [23] proposed the solution concept of rationalizability as the logical implication of common knowledge of Bayesian rationality; see also Tan and Werlang [29]. In (finite) strategic games, the concept of rationalizability can be defined in terms of a best-response (product) set of strategies, which reflects the idea that "rational" behavior should be justified by "rational" beliefs and conversely, "rational" beliefs should be based on "rational" behavior. The set of rationalizable strategies can be derived from iterative deletion of never best response strategies (see Bernheim's [6] Proposition 3.1 and Pearce's [23] Proposition 2) and, if moreover correlated beliefs are permitted, it is equivalent to iterated (strict) dominance. The concept of rationalizability aims to be weak; it determines not what actions should actually be taken, but what actions could be ruled out with confidence.

However, the concept of rationalizability does not take into account the intriguing and important possibility that groups or coalitions of players would be willing to coordinate their moves, in order to achieve mutual beneficial outcomes even if no binding agreements are made. To capture these important aspects of collective behavior, Ambrus [2] first offered a well-defined solution concept of "coalitional rationalizability" (henceforth, c-rationalizability) by using an iterative procedure of restrictions, in which members of a coalition will confine play to a subset of their strategies if it is in their mutual interest to do so.

The main purpose of this paper is to further study the logical implication of common knowledge of the fact that every player is Bayesian rational and also aware of mutual beneficial arrangements of strategies by conceivable coalitions. Following Ambrus [2], our approach is based on the idea that the players go through an "internal reasoning procedure."

> The coalitional agreements players can consider in this context take the form of restrictions of the strategy space. This means that players look for agreements to avoid certain strategies, without specifying play within the set of nonexcluded strategies. ... A restriction is supported if every group member always (for every possible expectation) expects a higher payoff if the agreement is made than if he instead chooses to play a strategy outside the agreement. (Ambrus [2, p.904])

In this paper we extend Ambrus's notion of c-rationalizability to situations where, in pursuit of mutual beneficial interests in the process of this kind of coalitional agreements, players in groups (i) evaluate their payoff expectations if an agreement is made by Bayesian updating, instead of holding fixed the marginal expectations concerning nonmembers as in Ambrus [2], and (ii) contemplate various deviations, i.e., the validity of deviation is checked not only against restricted subsets of strategies, *but also against arbitrary sets of strategies.*

This paper is motivated mainly by the following two considerations. First, it is natural and straightforward to consider the situations in which players use Bayesian updating to compute expected payoffs when evaluating coalitional agreements in the context of coalitional rationalizability. A rational player who conforms to Savage's [27] axioms should update his/her prior beliefs according to Bayes' rule.[1] Accordingly, when players consider implicit agreements to confine their collective actions within a set of strategies, each player should make Bayesian updating of the initial priors and calculate his/her resulting expected payoffs if a new agreement is made. We would like to emphasize that Ambrus [1, p.7] also suggested a similar idea that can be used for the study of the consistency in conjecture changes. In this paper, we offer an alternative definition of c-rationalizability to accommodate the effect of Bayesian updating when players contemplate to make mutual beneficial arrangements of strategies.

Second, perhaps more importantly, Ambrus's [2] analysis does not go far enough to account for all the aspects of collective and coalitional "stability." Ambrus's notion of c-rationalizability requires the validity of deviation to be checked *only* against restricted subsets of strategies when players in groups contemplate deviations, i.e., a valid deviation is a deviation from a product set of strategies to its subset. However, there seems no reason to suppose that coalitional deviations are restricted only to subsets of strategies. In fact, members of deviating coalitions may in general confine, enlarge, or even revise and rearrange play to any arbitrary set of strategies if doing so is in their mutual interest, and the validity of deviation should go further to be checked against completely free and unrestrained any sets of strategies.[2] The alternative notion of c-rationalizability is proposed here to accommodate the presence of very universal coalition deviations.

[1] The decision-theoretic foundation of Bayes' rule is laid by the axiomatization provided by Savage [27]; cf., e.g., Myerson [20] and Kreps [17, Chapter 10]. See also Epstein and LeBreton [11] for more extensive discussions and Ghirardato [12] and Karni [16] for recent works on Bayesian updating.

[2] Based on a similar consideration, Kahn and Mookherjee [15] proposed 'universal coalition-proof equilibrium' to accommodate more general coalition formation in Bernheim et al.'s [7] 'coalition-proof equilibrium'. Unfortunately, the notion of 'universal coalition-proof equilibrium' may fail to exist in a natural class of games.

In order to easily understand our analysis in this paper, we closely follow Bernheim's [6] and Pearce's [23] approach to conventional rationalizability and carry out a *ceteris paribus* study of the coalitional version of rationalizability. All the major features and nice properties of the conventional rationalizability, as a special case of c-rationalizability with the restriction to singleton coalitions only, are essentially preserved. More specifically, we define the notion of Bayesian c-rationalizability by using the "c-rationalizable set (CRS)," which can be viewed as the counterpart of the "best-response set" in the notion of conventional rationalizability (see Definition 1). The central result of this paper is to show that there is a nonempty largest CRS, which fully characterizes the set of all Bayesian c-rationalizable strategies (Theorem 1). Moreover, the set of Bayesian c-rationalizable strategies can be derived from an iterative procedure of restrictions to c-best response strategies (Proposition 1). In particular, any Pareto-dominant or strong pure Nash equilibrium is c-rationalizable (Proposition 2). We next formulate the coalitional version of iterated strict dominance (Definition 2), and show it is equivalent to the notion of Bayesian c-rationalizability (Theorem 2).

The rest of this paper is organized as follows. Section 2 defines the notion of Bayesian c-rationalizability. Subsection 2.1 investigates its existence and properties. Section 3 formulates the notion of "c-dominance" and shows an equivalence theorem between Bayesian c-rationalizability and c-dominance. Section 4 offers some concluding remarks. To facilitate reading, all the proofs are relegated to the Appendix.

2 Bayesian Coalitional rationalizability

Consider a finite game: $G \equiv (I, \{S_i\}_{i\in I}, \{u_i\}_{i\in I})$, where I is a nonempty finite set of players, S_i is a nonempty finite set of i's strategies, and $u_i : S \equiv \times_{i\in I} S_i \to \Re$ is i's payoff function. For (nonempty) coalition $J \subseteq I$, let $S_J \equiv \times_{j\in J} S_j$, $S_{-J} \equiv \times_{i\notin J} S_i$, and $S_{-j} \equiv \times_{i\neq j} S_i$. Throughout this paper, we restrict our attention to pure strategies, and players are allowed to hold correlated beliefs about the strategies of their opponents.

Let Δ denote the space of the probability distributions on finite state space S_{-j} (faced by player j). We denote the *space of probability distributions conditional on an event* $A_{-j} \subseteq S_{-j}$ by

$$\Delta|_{A_{-j}} \equiv \{\mu \in \Delta |\ \mu(A_{-j}) = 1\}.$$

For $\mu \in \Delta$ let $\mu|_{A_{-j}} \in \Delta|_{A_{-j}}$ denote a probability distribution conditional on A_{-j} satisfying Bayes' rule: $\mu(A_{-j} \cap B_{-j}) = \mu(A_{-j})\mu|_{A_{-j}}(B_{-j})$ for all events $B_{-j} \subseteq S_{-j}$. (We decree that $\mu|_{\emptyset} \equiv \mu$.) Note that the conditional probabilities are well defined even for zero probability events.[3]

[3]We may view $\mu|$ as a 'conditional probability system (CPS)'; see, e.g., Myerson [21, Section 1.6].

Let A and B be subsets of S in product-form. We say a coalition $\mathcal{J}_{AB}$ is a "feasible coalition from A to B" if $B = B_{\mathcal{J}_{AB}} \times A_{-\mathcal{J}_{AB}}$ – i.e. $\mathcal{J}_{AB}$ can be interpreted as a coalition by which set A can be rearranged to set B. (Note that B is not necessarily a subset of A and $\mathcal{J}_{AB}$ must include a player j if $A_j \neq B_j$.)

Definition 1. A nonempty product subset $\mathcal{R} \subseteq S$ is a *coalitional rationalizable set (CRS)* if $\mathcal{R} \Rrightarrow \mathcal{R}'$ only for $\mathcal{R}' = \mathcal{R}$, where for $\mathcal{R} \neq \emptyset$ we define $\mathcal{R} \Rrightarrow \mathcal{R}'$ as: $\exists \mathcal{J}_{\mathcal{R}\mathcal{R}'}$, $\forall j \in \mathcal{J}_{\mathcal{R}\mathcal{R}'}$

(1.1) **[Profitability]** $\forall r_j \in \mathcal{R}_j$, $\forall \mu \in \Delta|_{\mathcal{R}_{-j}}$, $u_j(r_j, \mu) < \ u_j(s_j, \mu|_{\mathcal{R}'_{-j}})$ for some $s_j \in S_j$, provided that $\mu|_{\mathcal{R}'_{-j}} \neq \mu$ if $r_j \in \mathcal{R}'_j$, and

(1.2) **[Credibility]** $\forall r'_j \in \mathcal{R}'_j \backslash \mathcal{R}_j$, $\exists \mu \in \Delta|_{\mathcal{R}_{-j}}$ such that for some $\mu|_{\mathcal{R}'_{-j}} \in \Delta|_{\mathcal{R}'_{-j}}$, $u_j(r'_j, \mu|_{\mathcal{R}'_{-j}}) \geq \ u_j(s_j, \mu|_{\mathcal{R}'_{-j}})$ $\forall s_j \in S_j$.

In particular, $r_j \in \mathcal{R}_j$ is said to be a *Bayesian c-rationalizable strategy.*

That is, a CRS is a (nonempty) product set of pure strategies from which no group of players would like to deviate. In Definition 1 $u_j(r_j, \mu)$ is j's expected payoff from using r_j given his prior belief $\mu \in \Delta|_{\mathcal{R}_{-j}}$, and $u_j(s_j, \mu|_{\mathcal{R}'_{-j}})$ is j's expected payoff from using s_j under his posterior belief $\mu|_{\mathcal{R}'_{-j}} \in \Delta|_{\mathcal{R}'_{-j}}$ after moving from $\mathcal{R}$ to $\mathcal{R}'$ via coalition $\mathcal{J}_{\mathcal{R}\mathcal{R}'}$. The "profitability" condition of Definition 1(1.1) is in the spirit of Ambrus's [2] original concept of c-rationalizability. This condition says that: every coalition member always (for every possible expectation) expects a higher payoff if the new agreement is made. More specifically, every player j in coalition $\mathcal{J}_{\mathcal{R}\mathcal{R}'}$, whatever prior belief j may hold, can always obtain, by using Bayes' rule, a strictly higher expected payoff $u_j(s_j, \mu|_{\mathcal{R}'_{-j}})$ if the move from $\mathcal{R}$ to $\mathcal{R}'$ is made. Definition 1(1.2) states a moderate "credibility" requirement that, whenever a move is made, each incremental change of strategies should be justifiable by individual rationality.

It is easy to see that a CRS $\mathcal{R}$ necessarily satisfies both the "best response" and "closed under rational behavior" properties (see Basu and Weibull [5]), i.e. $\mathcal{R}_i = BR(\mathcal{R}_{-i})$ $\forall i$, where

$$BR(\mathcal{R}_{-i}) \equiv \{s_i \in S_i | \ \exists \mu \in \Delta|_{\mathcal{R}_{-i}} \text{ s.t. } u_i(s_i, \mu) \geq u_i(s'_i, \mu) \forall s'_i \in S_i\}.$$

Therefore, every $\mathcal{R}$ admits a Nash equilibrium with its support in $\mathcal{R}$. Definition 1, with the restriction of $|\mathcal{J}_{\mathcal{R}\mathcal{R}'}| = 1$, is essentially the correlated version of rationalizability (cf. Luo [18, Section 4.1]). Thus, every Bayesian c-rationalizable strategy is rationalizable.

To see how Bayes' rule plays a role, consider a two-person game (where the first player picks the row and the second player picks the column):

	a	b	c
a	$3,0$	$0,3$	$0,2$
b	$0,3$	$3,0$	$0,0$
c	$2,0$	$0,0$	$1,1$

In this game, $S \Rrightarrow \{a,b\} \times \{a,b\}$ via coalition $\{1,2\}$. Intuitively, if each player assigns a prior probability of less than 0.5 to a of the opponent, then the player could achieve an expected payoff of less than 1.5 by using c and, hence, it is beneficial for the two players to move to $\{a,b\} \times \{a,b\}$ because each can guarantee a higher expected payoff of 1.5; if each player assigns a prior probability of more than or equal to 0.5 to a, then it is also beneficial for the two players to move to $\{a,b\} \times \{a,b\}$ because in this case each can achieve a higher expected payoff by using a instead of c. (Without using Bayes' rule, the player could achieve an expected payoff of 2 by using c, higher than that by using a or b after the deviation. In this game every strategy is c-rationalizable in Ambrus's [2] sense.)[4]

While the above example shows that the set of Bayesian c-rationalizable strategies can be strictly smaller than the set of Ambrus's c-rationalizable strategies, the opposite strict inclusion can happen too. Consider the following three-person game (where players 1, 2, and 3 pick the row, column, and matrix, respectively):[5]

a

	a	b
a	$8,8,0$	$0,0,0$
b	$0,0,0$	$4,4,0$

b

	a	b
a	$2,2,0$	$0,0,0$
b	$0,0,0$	$1,1,0$

From the perspective of players 1 and 2, coordinating on (a,a) seems to be a good idea by holding beliefs that player 3's behavior fixed. Ambrus's [2] notion of c-rationalizablility is developed to capture this intuition. In this game the set of Ambrus's c-rationalizable strategies is $\{a\} \times \{a\} \times \{\mathbf{a}, \mathbf{b}\}$. However, b can be Bayesian c-rationalizable; in particular, from a Bayesian viewpoint, players 1 and 2 should be reluctant to make this agreement because of worrying about the possibility of player 3's playing the strategy **b** if the agreement is made.

Another important, and somewhat subtle, difference is that in Ambrus's [2] notion of c-rationalizablility, a coalition member might suffer a payoff loss caused by other coalition members' discarding strategies when a supported restriction is made. Consider the following two-person parametric game where sufficiently small $\epsilon > 0$ and $\theta = 0, 3$:

[4] Ambrus's [2] notion of c-rationalizability requires that the marginal expectation concerning the strategies of players outside the coalition be fixed. This sort of 'updating beliefs' is in general different from the Bayesian updating rule.

[5] We thank the referee for pointing this example out to us.

	a	b	c
a	$3, \epsilon$	$\epsilon, 3$	$0, 2$
b	$0, 3$	$3, 0$	$\theta, 0$
c	$2, \theta$	$0, 0$	$1, 1$

It is easy to verify that the set of Ambrus's c-rationalizable strategies is $\{a, b\} \times \{a, b\}$, regardless of $\theta = 0, 3$. On the contrary, while the set of Bayesian c-rationalizable strategies remains $\{a, b\} \times \{a, b\}$ if $\theta = 0$, every strategy can be Bayesian c-rationalizable if $\theta = 3$. Intuitively, in the latter case players 1 and 2 would be unwilling to form a coalition and make an agreement on $\{a, b\} \times \{a, b\}$ because each player can attain the highest payoff of 3 if this agreement is not made and if the opponent happens to play the strategy c. Consequently, in general we cannot expect a certain and unambiguous relationship between Bayesian c-rationalizablity and Ambrus's c-rationalizablity.[6]

Remark 1. Definition 1 is harmonious with Ambrus's [2] definition of "supported restriction" in which no constraint is imposed on the remainder part of strategies. This is because condition (1.2) is trivially satisfied if the special case of restriction-form move from $\mathcal{R}$ to its subset $\mathcal{R}'$ is considered. Technically, this mild "credible" condition (1.2) purports to overcome the notorious problem of emptiness, typified by Condorcet's paradox, under the "core-like" blocking arrangements. For example, in the Prisoner's Dilemma game, the noncooperative Nash outcome is the unique CRS because the Pareto-optimal cooperative outcome is not "credible."[7]

Remark 2. It is worthwhile to emphasize that, if $r_j \in \mathcal{R}_j \cap \mathcal{R}'_j$, we do not require condition (1.1) for $\mu = \mu|_{\mathcal{R}'_{-j}}$. That is, condition (1.1) requires that any deviating coalition member j who uses an unaffected strategy r_j in $\mathcal{R}_j \cap \mathcal{R}'_j$ can obtain a strictly higher expected payoff after the deviation takes place, *whereas j's posterior belief* $\mu|_{\mathcal{R}'_{-j}}$ *is changed.* This condition is designed to accommodate the requirement that each coalition member takes into account payoff loss caused by other coalition members' discarding strategies when a deviating arrangement of strategies is made, as illustrated by the aforementioned parametric game.

[6] Another subtle reason for the absence of relationship between the two solution concepts is because in Ambrus's [2, p.909] notion, only some, not all, initial beliefs are taken into account when players in coalitions contemplate deviations – i.e. each player $j \in J$ considers only the beliefs in $\Omega^*_{-j}(A_j \backslash B_j)$ to which j has a best response strategy.

[7] The credibility requirement is in the same spirit as Milgrom and Roberts's 'strongly coalition proof equilibria' and Kaplan's 'semistrong equilibria' (see Milgrom and Roberts [19, p.115]). See also Roth's [26] 'protected' condition.

2.1 Existence and Properties

The central result in this paper is that there is a largest (w.r.t. set inclusion) CRS:

$$\mathcal{R}^* \equiv \bigcup_{\mathcal{R} \text{ is a CRS}} \mathcal{R}$$

that consists of the union of all CRSs. Formally, we have

Theorem 1. *$\mathcal{R}^*$ is a largest CRS.*

The proof of Theorem 1 in Appendix shows that the set of Bayesian c-rationalizable strategies can be derived from an iterative procedure of restrictions to c-best response strategies. Furthermore, every such procedure yields the same outcome. Let "$A \Rightarrow B$" denote "$A \Rrightarrow B$ with $B \subseteq A$." Formally, we have[8]

Proposition 1. *$\mathcal{R}^* = \mathcal{D}$ where $\mathcal{D} \equiv \bigcap_{k=0}^{\infty} \mathcal{D}^k$ with $\mathcal{D}^0 = S$, $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$, and $\mathcal{D} \Rightarrow \mathcal{D}'$ only for $\mathcal{D}' = \mathcal{D}$.*

As individual rationality can be regarded as a special case of c-rationality where only singleton coalitions are allowed to form, in general we would not expect Bayesian c-rationalizable outcomes to be Pareto efficient. In the Prisoner's Dilemma game, for example, the "noncooperation" action remains to be the only Bayesian c-rationalizable strategy. Given that the notion of Bayesian c-rationalizability is built on the idea that players try to attain common aspirations, we can, however, establish some relationships to Pareto efficient (Nash) outcomes. Let $G' = (I, \{S_i'\}, \{u_i\})$ denote the reduced game after iterated elimination of strictly dominated strategies. Formally, we have

Proposition 2. *Every Pareto-dominant pure Nash equilibrium is Bayesian c-rationalizable; every strong pure Nash equilibrium is Bayesian c-rationalizable.*

Proposition 3. *If $s_J^* \in S_J$ is a (strong) Pareto-best strategy profile for coalition J in the reduced game G', i.e., $\forall j \in J$, $\left(u_j\left(s_J^*, s_{-J}'\right) > u_j\left(s\right) \; \forall s, s' \in S'$ with $s_J^* \neq s_J\right)$ $u_j\left(s_J^*, s_{-J}\right) \geq u_j\left(s\right) \; \forall s \in S'$, then s_J^* is a (unique) Bayesian c-rationalizable strategy profile for J. In particular, any (strong) Pareto-best strategy profile is a (unique) Bayesian c-rationalizable strategy profile.*

[8]See Apt [3], Chen et al. [8], and Dufwenberg and Stegeman [10] for a similar formulation of iterative procedures.

Propositions 2 and 3 are simple but useful in the study of c-rationalizable behavior. Since the Nash equilibrium concept does not take into account coalitional rationality, a Nash equilibrium may fail to be Bayesian c-rationalizable. Proposition 2 asserts that Pareto-dominant and strong pure Nash equilibria must be Bayesian c-rationalizable. Ambrus's [2] motivating example of "Voting with Costly Participation" can also be easily analyzed from the perspective of Proposition 3.[9] In addition, any Pareto-best element in the set of strategies surviving iterated deletion of dominated strategies is a c-rationalizable strategy profile and, moreover, is a coalition-proof equilibrium for any admissible coalition communication structure; see Milgrom and Roberts [19, Theorem 1]. We would like to point out that the notion of c-rationalizability can provide a unique prediction in some class of games where there is a wide range of Nash equilibria and rationalizable strategies, which is particularly interesting for games with no strong Nash equilibrium. In the class of common interest games with a strong Pareto-best outcome, for example, the notion of Bayesian c-rationalizablility predicts the unique Pareto-best outcome even if there are other strict Nash equilibria.

3 Iterated c-dominance vs. c-rationalizability

The iterative procedure in Proposition 1 requires that, whatever prior beliefs are held, players in a coalition exclude "inferior" strategies from their considerations if it is in their mutual interest to do so. This procedure, however, needs complicated calculations of expected utility. To the extent that players are concerned about their decisions, it is convenient to have a similar notion of (strict) dominance under the consideration of coalitions.[10] In this section we formulate the belief-free notion of "c-dominance" and show an equivalence theorem between iterated c-dominance and Bayesian c-rationalizablity. Let $\Delta(S_j)$ denote the set of j's mixed strategies.

[9] In that example it is a strictly dominated strategy for voter 3 to show up and vote for alternative A. Thus, it is Bayesian c-rationalizable only for voters 1 and 2 to show up and vote for alternative A and, then, voter 3 would choose to stay at home. Indeed, for Ambrus' notion of c-rationalizability, the 'uniqueness' condition in Proposition 3 can be relaxed as follows: for all $j \in J$, $u_j(s_J^*, s_{-J}) > u_j(s)$ $\forall s \in S'$ with $s_J^* \neq s_J$.

[10] In the standard case of individual rationality, it is well known that a strategy is a never best response to any correlated conjecture concerning the opponents' moves if, and only if, it is strictly dominated possibly by a mixed strategy; see, e.g., Pearce [23, Proposition 2]. See also Shimoji and Watson's [28] related work on extensive-form rationalizability.

Definition 2. A subset $B \subseteq A$ is a *c-dominated restriction from* A, denoted by $A \Downarrow B$, if $\exists \mathcal{J}_{AB}$, $\forall j \in \mathcal{J}_{AB}$

(2.1) $\forall c_j \in A_j \backslash B_j$, $\exists \sigma_j \in \Delta(S_j)$ s.t. $u_j(\sigma_j, b_{-j}) > u_j(c_j, b_{-j})$ $\forall b_{-j} \in B_{-j}$, and

(2.2) $\forall a_j \in A_j$, $\exists \sigma_j \in \Delta(S_j)$ s.t. $u_j(\sigma_j, b_{-j}) > u_j(a_j, c_{-j})$ $\forall b_{-j} \in B_{-j}$ $\forall c_{-j} \in A_{-j} \backslash B_{-j}$,

where $u_j(\sigma_j, b_{-j}) \equiv \sum_{s_j \in S_j} \sigma_j(s_j)\, u_j(s_j, b_{-j})$.

In words, from coalition member j's point of view, c-dominated strategy $c_j \in A_j \backslash B_j$ is strictly dominated given set B in the usual sense. Furthermore, given any $-j$'s c-dominated profile $c_{-j} \in A_{-j} \backslash B_{-j}$, j can always do better if coalition members jointly move from A to B, regardless of whatever $b_{-j} \in B_{-j}$ will be eventually used (cf. Fig. 1).

	b_{-j}	c_{-j}
a_j	$u_j(a_j, b_{-j})$	$u_j(a_j, c_{-j})$
c_j	$u_j(c_j, b_{-j})$	$u_j(c_j, c_{-j})$

Figure 1: *The problem faced by* j.

The following Theorem 2 shows that the set of Bayesian c-rationalizable strategies can be derived from an iterative elimination of c-dominated strategies. Moreover, the order of elimination does not matter. Formally, we have

Theorem 2. $\mathcal{R}^* = \widehat{\mathcal{D}}$ *where* $\widehat{\mathcal{D}} \equiv \bigcap_{k=0}^{\infty} \widehat{\mathcal{D}}^k$ *with* $\widehat{\mathcal{D}}^0 = S$, $\widehat{\mathcal{D}}^k \Downarrow \widehat{\mathcal{D}}^{k+1}$, *and* $\widehat{\mathcal{D}} \Downarrow \widehat{\mathcal{D}}'$ *only for* $\widehat{\mathcal{D}}' = \widehat{\mathcal{D}}$.

It is usually helpful to analyze the complex games by using the notion of (iterated) c-dominance. For example, the set of Bayesian c-rationalizable strategies in Ambrus's [2] example of "Dollar Division Game with External Reward" can be easily derived by applying one round of elimination of all c-dominated strategies. Let us reconsider the first example of two-person game in Section 2. In this game, $S \Downarrow \{a, b\} \times \{a, b\}$ via coalition $\{1, 2\}$. Intuitively, under restriction $\{a, b\} \times \{a, b\}$, (i) a strictly dominates c (that is ruled out by the restriction); (ii) $0.5a + 0.5b$ can guarantee an expected payoff of 1.5 that is higher than the best payoff of 1 resulting from the opponent's using c (that is ruled out by the restriction).

4 Concluding remarks

In many real life situations, groups of individuals often have an incentive to choose, voluntarily and without binding agreement, to coordinate their action choices and make joint decisions in noncooperative environments. Ambrus [2] took the first step to offer a solution concept of c-rationalizability for situations in which coalitions can account for profitable deviations from an initial proposal to subsets of strategies. Following this line of research, we have presented in this paper an alternative notion of Bayesian c-rationalizability suitable for situations where, in seeking mutual beneficial interests, members in groups (i) make use of Bayes' rule in expectation calculations and (ii) contemplate various deviations, i.e., the validity of deviation is checked not only against restricted subsets of strategies, but also against arbitrary sets of strategies.

We have shown that the notion of Bayesian c-rationalizability is a well-defined solution concept that possesses similar nice properties of the conventional notion of rationalizability. We have shown that the set of Bayesian c-rationalizable strategies can be fully characterized by the largest CRS, which can derived from any iterative procedure of restrictions to c-best response strategies. We have formulated the coalitional version of dominance and shown that the set of Bayesian c-rationalizable strategies can be solved by performing any iterative deletion procedure of c-dominated strategies. Thus, we have also offered a coalitional analogue to the connection between conventional rationalizability and iterated strict dominance, which is valuable on both practical and conceptual levels. We would like to make some final remarks and conclusions as follows:

(1) Various coalitional equilibrium concepts in the literature, e.g. Aumann's [4] strong Nash equilibrium, Bernheim et al.'s [7] coalition-proof Nash equilibrium, and Ray and Vohra's [24, 25] equilibrium binding agreements, often fail to exist in a natural class of games. Yet, while Ambrus [2] suggested a well-defined coalitional solution concept, his solution concept fails to be immune against being blocked by coalitions using more flexible arrangements of strategies. For example, consider Ambrus's [2, Figure III] three-person game:

	a	b
a	$2, 2, 2$	$0, 0, 0$
b	$0, 0, 0$	$3, 3, 0$

a

	a	b
a	$0, 0, 0$	$0, 0, 0$
b	$0, 0, 0$	$1, 1, 1$

b

. Ambrus's notion of c-rationalizability prescribes the outcome $(1, 1, 1)$ which can, however, be improved on by the appealing Pareto-dominant Nash equilibrium outcome $(2, 2, 2)$. The notion of c-rationalizability presented in this paper is a logically consistent solution concept that accommodates such complex coalitional reasoning. In particular, any Pareto-dominant Nash equilibrium strategy profile must be Bayesian c-rationalizable (see Proposition 2).

(2) Proposition 1 shows that the set of Bayesian c-rationalizable strategies can be derived from *any* iterative procedure of restrictions to c-best response strategies. This paper thereby provides an additional rationale for defining a coalitional version of rationalizability directly by an iterative procedure of restrictions as in Ambrus [2]. Ambrus [1] observed that there are other possible definitions of "supported restriction" by a class of sensible best response correspondences. It can be shown that there is a "fast" iterative procedure of restrictions generated by a sensible correspondence that yields the set of Bayesian c-rationalizable strategies.[11]

(3) Within a non-equilibrium framework of coalitional reasoning, Greenberg [13] offered an integrative approach to social interactions, and proposed several coalitional negotiation processes where coalitions openly negotiate to make contingent threats or to make irrevocable commitments in social situations. Within a similar framework in which coalitional moves are publicly observed, Chwe [9] and Xue [30] studied the "stable" outcomes under coalitional interactions where players are farsighted; Herings et al. [14] analyzed the social environment by using Pearce's [23] extensive-form rationalizability in the associated multistage game. The main differences of our approach in this paper are: (i) coalitional moves are secretly conducted and cannot be publicly observed; and, (ii) implicit agreements made by coalitions are in general in the form of constraint sets of strategies to be confined, rather than stringent specifications of a particular course of actions.[12]

(4) Finally, we would like to point out that the exploration of epistemic foundations of Bayesian c-rationalizability remains an interesting subject for further study. Extension of this paper to more general class of games, e.g., extensive games with imperfect or incomplete information and games with general preferences, is clearly an important subject for further research. Extension of this paper to permit nonproduct-set deviations is also an intriguing topic worth further investigation.

[11] Ambrus [1] defined a sensible best response correspondence by four properties. Because the iterative procedures defined in this paper are rather flexible to allow for eliminating some, *but not all,* never c-best response strategies at each stage of an iteration, the procedures may fail to satisfy Ambrus's [1] Properties (ii) and (iii) – the "individual rationality" and "monotonicity" properties. However, we can show that a "fast" iterative procedure of restrictions, which eliminates *all* never c-best response strategies at each stage, is a sensible correspondence and yields the set of Bayesian c-rationalizable strategies. We thank the referee for pointing this out to us.

[12] See Ambrus [2, Section VII] for extensive discussions on this form of agreements. We would also like to point out that the minimal CRS in Definition 1 can be viewed as a coalitional version of Basu and Weibull's [5] minimal curb set, which may be an interesting set-valued solution for strategic games.

Appendix: Proofs

To prove Theorem 1, we need the following Lemmas 1-4.

Lemma 1. *Suppose* $A \cap B \neq A$. *For any* $\nu \in \Delta|_B$ *there is* $\mu \in \Delta|_A$ *such that* $\mu|_B = \nu$.

Proof: Let $\mu \in \Delta|_{A \setminus B}$. Then $\mu \in \Delta|_A$ and $\mu(B) = 0$. Therefore, for given $\nu \in \Delta|_B$ we may define $\mu|_B = \nu$.■

Lemma 2. *Suppose* $A \Rightarrow B$ *with* $A \in \mathcal{M}$. *Then,* $B \neq \emptyset$ *and* $B \in \mathcal{M}$, *where*

$$\mathcal{M} \equiv \{\mathcal{A} |\ BR(\mathcal{A}_{-i}) \subseteq \mathcal{A}_i\ \forall i\}.$$

Proof: Since $A \Rightarrow B$, $A \Rrightarrow B$ and hence $A \neq \emptyset$. For any fixed i, let $u_i(a^*) \equiv \max\limits_{a \in A} u_i(a)$. Since $A \in \mathcal{M}$, for all $\mu \in \Delta|_{A_{-i}}$,

$$\max_{s_i \in S_i} u_i(s_i, \mu) = \max_{a_i \in A_i} u_i(a_i, \mu) \leq \max_{\mu \in \Delta|_{A_{-i}}} \max_{a_i \in A_i} u_i(a_i, \mu) = u_i(a^*).$$

Since $a^* \in A$ and $A \Rightarrow B$, by Definition 1(1.1), $a_i^* \in B_i$. Thus, $B \neq \emptyset$.

Let $\mu \in \Delta|_{B_{-i}}$. Since $B \subseteq A$, $\mu \in \Delta|_{A_{-i}}$ and $\mu = \mu|_{B_{-i}}$. Therefore, for all $s_i \in BR(B_{-i})$, there is $\mu \in \Delta|_{A_{-i}}$ such that $u_i(s_i, \mu) \geq\ u_i(s_i', \mu|_{B_{-i}})\ \forall s_i' \in S_i$. Let $A \Rightarrow B$ via $J \equiv \mathcal{J}_{AB}$. Since $A \in \mathcal{M}$, it follows that $BR(B_{-j}) \subseteq B_j\ \forall j \in J$. For $i \notin J$, $BR(B_{-i}) \subseteq B_i$ since $B_i = A_i$.■

Lemma 3. *Suppose* $\mathcal{A} \subseteq A \Rrightarrow B$, $\mathcal{A} \in \mathcal{M}$, *and* $\mathcal{A} \cap B = \emptyset$. *Then,* $\mathcal{A}_{-j} \cap B_{-j} = \emptyset\ \forall j$.

Proof: Since $\mathcal{A} \cap B = \emptyset$, $\mathcal{A}_j \cap B_j = \emptyset$ for some j. Since $\mathcal{A} \in \mathcal{M}$, $A_j \supseteq \mathcal{A}_j \supseteq BR(s_{-j})$ for all $s_{-j} \in \mathcal{A}_{-j} \cap B_{-j}$. But, since $A \Rrightarrow B$, by Definition 1(1.1), $B_j \supseteq BR(s_{-j})$ for all $s_{-j} \in \mathcal{A}_{-j} \cap B_{-j}$. Therefore, $\mathcal{A}_{-j} \cap B_{-j} = \emptyset$. That is, $\mathcal{A}_{j'} \cap B_{j'} = \emptyset$ for some $j' \neq j$ and, hence, $\mathcal{A}_{-j} \cap B_{-j} = \emptyset\ \forall j$.■

Lemma 4. *Suppose* $\mathcal{A} \subseteq A \Rrightarrow B$. *Then, (4.1)* $\mathcal{A} \Rightarrow \mathcal{A} \cap B$ *if* $\mathcal{A} \cap B \neq \emptyset$*;(4.2)* $\mathcal{A} \Rrightarrow \mathcal{B} \subseteq B$ *if* $\mathcal{A} \cap B = \emptyset$ *and* $\mathcal{A}, B \in \mathcal{M}$.

Proof: Let $A \Rrightarrow B$ via $J \equiv \mathcal{J}_{AB}$. (4.1) For any j consider $\mu \in \Delta|_{\mathcal{A}_{-j}}$. Note that $\mu \in \Delta|_{A_{-j}}$ and $\mu|_{(\mathcal{A} \cap B)_{-j}} \in \Delta|_{B_{-j}}$. Since $\mathcal{A} \subseteq A \Rrightarrow B$, it follows that $\forall j \in J$, $\forall a_j \in \mathcal{A}_j$, $u_j(a_j, \mu) < u_j(s_j, \mu|_{(\mathcal{A} \cap B)_{-j}})$ for some $s_j \in S_j$, where $\mu|_{(\mathcal{A} \cap B)_{-j}} \neq \mu$ if $a_j \in \mathcal{A}_j \cap B_j$. But, since $\mathcal{A}_{-J} \subseteq A_{-J} = B_{-J}$, $\mathcal{A} \cap B = (\mathcal{A} \cap B)_J \times \mathcal{A}_{-J}$. Therefore, $\mathcal{A} \Rightarrow \mathcal{A} \cap B$ via J.

(4.2) Define $\widetilde{\mathcal{B}} \equiv B_J \times \mathcal{A}_{-J}$. Since $\mathcal{A}_{-J} \subseteq A_{-J} = B_{-J}$, $\widetilde{\mathcal{B}} \subseteq B$. Since $B \in \mathcal{M}$, $\widetilde{\mathcal{B}}_j \supseteq BR\left(\widetilde{\mathcal{B}}_{-j}\right)\ \forall j \in J$. Let $\mathcal{B}$ be the (nonempty) set of surviving iterated elimination of never-best responses for all the players in coalition J in the finite subgame restricted on $\widetilde{\mathcal{B}}$. Clearly, $\mathcal{B} = \mathcal{B}_J \times \mathcal{A}_{-J}$, $\mathcal{B} \subseteq B$, and $\mathcal{B}_j = BR(\mathcal{B}_{-j})\ \forall j \in J$. To complete the proof, it remains to verify $\mathcal{A} \Rrightarrow \mathcal{B}$ via J. Since $\mathcal{A} \in \mathcal{M}$, by Lemma 3, $\mathcal{A}_{-j} \cap B_{-j} = \emptyset$.

(1) Since $\mathcal{A}_{-j} \subseteq A_{-j}$, $\mu \in \Delta|_{A_{-j}}$ for all $\mu \in \Delta|_{\mathcal{A}_{-j}}$. Since $A \Rrightarrow B \supseteq \mathcal{B}$, by Lemma 1, it follows that $\forall j \in J = \mathcal{J}_{\mathcal{AB}}$, $\forall a_j \in \mathcal{A}_j \subseteq A_j$,

$$u_j(a_j, \mu) < \; u_j(s_j, \nu) \text{ for some } s_j \in S_j,$$

for all $\mu \in \Delta|_{\mathcal{A}_{-j}}$ and $\nu \in \Delta|_{\mathcal{B}_{-j}}$. That is, Definition 1(1.1) is satisfied.

(2) Since $\mathcal{B}_j = BR(\mathcal{B}_{-j})$, by Lemma 1, $\forall j \in J = \mathcal{J}_{\mathcal{AB}}$, $\forall b_j \in \mathcal{B}_j \backslash \mathcal{A}_j \subseteq \mathcal{B}_j$, for any $\mu \in \Delta|_{\mathcal{A}_{-j}}$

$$u_j(b_j, \mu|_{\mathcal{B}_{-j}}) \geq \; u_j(s_j, \mu|_{\mathcal{B}_{-j}}) \; \forall s_j \in S_j,$$

for some $\mu|_{\mathcal{B}_{-j}}$. Thus, Definition 1(1.2) is satisfied.■

Proof of Theorem 1: The proof is split into two parts. Let "$A \Rightarrow B$" denote "$A \Rrightarrow B$ with $B \subseteq A$." Define an iterative procedure of restrictions as follows:

$$\mathcal{D} \equiv \bigcap_{k=0}^{\infty} \mathcal{D}^k$$

where $\mathcal{D}^0 = S$, $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$, and $\mathcal{D} \Rightarrow \mathcal{D}'$ only for $\mathcal{D}' = \mathcal{D}$. By Lemma 2, $\mathcal{D} \neq \emptyset$.
Part I: $\mathcal{D}$ is a CRS. Assume, in negation, that $\mathcal{D}$ is not a CRS, i.e., $\mathcal{D} \Rrightarrow \mathcal{D}' \neq \mathcal{D}$. Clearly, $\mathcal{D}' \neq \emptyset$ since $\mathcal{D} \Rrightarrow \emptyset$ implies $\mathcal{D} \Rightarrow \emptyset$. We distinguish two cases.

Case 1.1. $\mathcal{D}' \cap \mathcal{D} \neq \emptyset$. By Lemma 2, $\mathcal{D} \in \mathcal{M}$. By Lemma 4(4.1), $\mathcal{D} \Rightarrow \mathcal{D}' \cap \mathcal{D}$ and hence $\mathcal{D}' \supseteq \mathcal{D}$. Since $\mathcal{D} \in \mathcal{M}$, for every $d'_j \in \mathcal{D}'_j \backslash \mathcal{D}_j$, $\forall \mu \in \Delta|_{\mathcal{D}_{-j}}$, $u_j(d'_j, \mu|_{\mathcal{D}'_{-j}}) < u_j(s_j, \mu|_{\mathcal{D}'_{-j}})$ for some $s_j \in S_j$. By Definition 1(1.2), it is impossible that $\mathcal{D} \Rrightarrow \mathcal{D}' \neq \mathcal{D}$.

Case 1.2. $\mathcal{D}' \cap \mathcal{D} = \emptyset$. Let $\mathcal{D}^k \supseteq \mathcal{D}'$ and $\mathcal{D}^{k+1} \not\supseteq \mathcal{D}'$ (cf. Fig. 3). Thus, $\exists d'_{j^0} \in \mathcal{D}'_{j^0} \backslash \mathcal{D}^{k+1}_{j^0}$ and, hence, $d'_{j^0} \in \mathcal{D}'_{j^0} \backslash \mathcal{D}_{j^0}$. Clearly, $j^0 \in \mathcal{J}_{\mathcal{D}\mathcal{D}'} \cap \mathcal{J}_{\mathcal{D}^k \mathcal{D}^{k+1}}$. Since by Lemma 2, $\mathcal{D}^{k+1} \in \mathcal{M}$, $BR(\mathcal{D}^{k+1}_{-j^0}) \subseteq \mathcal{D}^{k+1}_{j^0}$. Since $\mathcal{D} \Rrightarrow \mathcal{D}'$, by Definition 1(1.2), $d'_{j^0} \in BR(\mathcal{D}'_{-j^0})$. Therefore, $\mathcal{D}'_{-j^0} \nsubseteq \mathcal{D}^{k+1}_{-j^0}$, i.e., $\exists d'_{-j^0} \in \mathcal{D}'_{-j^0} \backslash \mathcal{D}^{k+1}_{-j^0}$. We proceed in two steps.

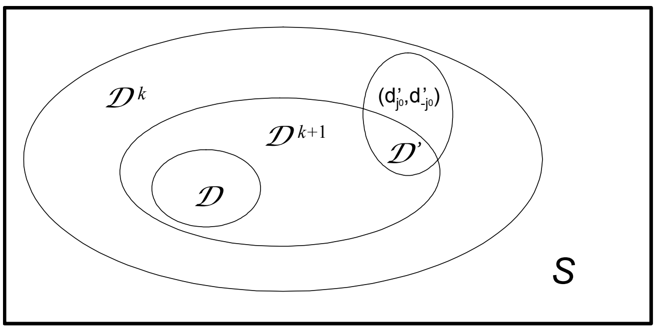

Figure 3.

Step 1. Let $d_{-j^0} \in \mathcal{D}_{-j^0}$. By Lemma 2, $\mathcal{D} \in \mathcal{M}$. Since $\mathcal{D} \Rrightarrow \mathcal{D}'$, by Lemma 3, $d_{-j^0} \notin \mathcal{D}'_{-j^0}$. Consider $\mu \in \Delta|_{\mathcal{D}_{-j^0}}$ with $\mu(d_{-j^0}) = 1$. Since $\mathcal{D} \Rrightarrow \mathcal{D}'$, by Lemma 1,

$$u_{j^0}(d_{j^0}, d_{-j^0}) = u_{j^0}(d_{j^0}, \mu) < u_{j^0}(s_{j^0}, d'_{-j^0}) \text{ for some } s_{j^0} \in S_{j^0},$$

for all $d_{j^0} \in \mathcal{D}_{j^0}$ and $d'_{-j^0} \in \mathcal{D}'_{-j^0}$. Since $\mathcal{D}' \subseteq \mathcal{D}^k \in \mathcal{M}$,

$$u_{j^0}(d_{j^0}, d_{-j^0}) < u_{j^0}(d'_{j^0}, d'_{-j^0}) \text{ for some } d'_{j^0} \in \mathcal{D}^k_{j^0},$$

for all $d_{j^0} \in \mathcal{D}_{j^0}$, $d_{-j^0} \in \mathcal{D}_{-j^0}$ and $d'_{-j^0} \in \mathcal{D}'_{-j^0}$.

Step 2. Let $d'_{-j^0} \in \mathcal{D}'_{-j^0} \backslash \mathcal{D}^{k+1}_{-j^0}$ (since $\mathcal{D}'_{-j^0} \backslash \mathcal{D}^{k+1}_{-j^0} \neq \emptyset$). Consider $\nu \in \Delta|_{\mathcal{D}^k_{-j^0}}$ with $\nu\left(d'_{-j^0}\right) = 1$. Since $\mathcal{D}' \subseteq \mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$, by Lemma 1,

$$u_{j^0}(d'_{j^0}, d'_{-j^0}) = u_{j^0}(d'_{j^0}, \nu) < u_{j^0}(s_{j^0}, d_{-j^0}) \text{ for some } s_{j^0} \in S_{j^0},$$

for all $d_{-j^0} \in \mathcal{D}_{-j^0}$ and $d'_{j^0} \in \mathcal{D}^k_{j^0}$. Since $\mathcal{D} \in \mathcal{M}$, there is $d'_{-j^0} \in \mathcal{D}'_{-j^0}$ such that

$$u_{j^0}(d'_{j^0}, d'_{-j^0}) < u_{j^0}(d_{j^0}, d_{-j^0}), \text{ for some } d_{j^0} \in \mathcal{D}_{j^0},$$

for all $d_{-j^0} \in \mathcal{D}_{-j^0}$ and $d'_{j^0} \in \mathcal{D}^k_{j^0}$, contradicting Step 1.

Part II: $\mathcal{R}^* = \mathcal{D}$. Assume, in negation, that there is a CRS $\mathcal{R} \nsubseteq \mathcal{D}$. Then, there exists $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$ via $J = \mathcal{J}_{\mathcal{D}^k\mathcal{D}^{k+1}}$ such that $\mathcal{R} \subseteq \mathcal{D}^k$ and $\mathcal{R} \nsubseteq \mathcal{D}^{k+1}$. If $\mathcal{R} \cap \mathcal{D}^{k+1} \neq \emptyset$, by Lemma 4(4.1), $\mathcal{R} \Rrightarrow \mathcal{R} \cap \mathcal{D}^{k+1} \neq \mathcal{R}$, contradicting that $\mathcal{R}$ is a CRS. If $\mathcal{R} \cap \mathcal{D}^{k+1} = \emptyset$, by Lemma 4(4.2), $\mathcal{R} \Rrightarrow \mathcal{R}'$ and $\mathcal{R} \cap \mathcal{R}' = \emptyset$, contradicting that $\mathcal{R}$ is a CRS.■

Proof of Proposition 2: Let $s^* \in S$ be a Nash equilibrium that Pareto-dominates all other (mixed) Nash equilibria, i.e., $\forall i$, $u_i(s^*) \geq u_i(s^{**})$ for all (mixed) equilibrium s^{**}. Assume, in negation, that s^* is not c-rationalizable. By Proposition 1, $s^* \in \mathcal{D}^k \backslash \mathcal{D}^{k+1}$ for some $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$. Since each player i's Nash equilibrium strategy s^*_i is a best response to s^*_{-i}, by Definition 1(1.1), $s^*_j \in \mathcal{D}^k_j \backslash \mathcal{D}^{k+1}_j$ and $s^*_{-j} \notin \mathcal{D}^{k+1}_{-j}$ for some $j \in \mathcal{J}_{\mathcal{D}^k\mathcal{D}^{k+1}}$. Note that there is a (mixed) Nash equilibrium s^{**} with the support in $\mathcal{D}^{k+1}$. Since $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$, by Lemma 1, $u_j(s^*) < u_j(s^{**})$, contradicting that s^* is a Pareto-dominant Nash equilibrium. Hence, s^* is c-rationalizable.

Now, let $s^* \in S$ be a strong Nash equilibrium, i.e., for each (mixed) strategy profile $s \neq s^*$, there is a player j such that $s_j \neq s^*_j$ and $u_j(s) \leq u_j(s^*)$ (see Aumann [4]). Assume, in negation, that s^* is not c-rationalizable. By Proposition 1, $s^* \in \mathcal{D}^k \backslash \mathcal{D}^{k+1}$ for some $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$. Let $J \equiv \left\{j | \; s^*_j \in \mathcal{D}^k_j \backslash \mathcal{D}^{k+1}_j\right\}$. Consider a subgame restricted on $\mathcal{D}_J^{k+1} \times \{s_{-J}^*\}$. There is a Nash equilibrium s^{**} in the subgame. Since by Lemma 2, $\mathcal{D}^{k+1} \in \mathcal{M}$, s_j^{**} is a best response to s_{-j}^{**} for $j \in J$. Similar to the proof for the Pareto-dominant Nash equilibrium case, we have that $\forall j \in J$, $u_j(s^*) < u_j(s^{**})$ for some $s_J^{**} \in \mathcal{D}_J^{k+1}$, contradicting that s^* is a strong Nash equilibrium. Hence, s^* is c-rationalizable.■

Proof of Proposition 3: Let s_J^* be a Pareto-best strategy profile for J in the reduced game $G' = (I, \{S_i'\}, \{u_i\})$, i.e., for all $j \in J$, $u_j(s_J^*, s_{-J}) \geq u_j(s)$ $\forall s \in S'$. Assume, in negation, that s_J^* is not c-rationalizable. By Proposition 1, $s_J^* \in \mathcal{D}_J^k \backslash \mathcal{D}_J^{k+1}$ for some $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$. Consider $s \in \mathcal{D}$. Clearly, $s \in S'$ and $(s_J^*, s_{-J}) \in \mathcal{D}^k$. As any dominated strategy is dominated by an undominated strategy in any finite game (see, e.g., Milgrom and Roberts [19, Lemma 1]), it follows that s_j^* is a best response to $\left(s_{J\backslash j}^*, s_{-J}\right)$ for $j \in J$. By Definition 1(1.1), it follows that $s_{j^0}^* \in \mathcal{D}_{j^0}^k \backslash \mathcal{D}_{j^0}^{k+1}$ and $\left(s_{J\backslash j^0}^*, s_{-J}\right) \notin \mathcal{D}_{-j^0}^{k+1}$ for some $j^0 \in J$. By $\mathcal{D}^k \Rightarrow \mathcal{D}^{k+1}$ and $\mathcal{D} \in \mathcal{M}$, we have that $u_{j^0}(s_J^*, s_{-J}) < u_{j^0}(s_{j^0}, s_{-j^0})$ for some $s_{j^0} \in \mathcal{D}_{j^0} \subseteq S_{j^0}'$, contradicting that s_J^* is Pareto-best for J. Thus, s_j^* is c-rationalizable.

Now, assume that for all $j \in J$, $u_j\left(s_J^*, s_{-J}'\right) > u_j(s)$ $\forall s, s' \in S'$ with $s_J^* \neq s_J$. Thus, J would be willing to confine their play to s_J^* that is a unique c-rationalizable strategy profile for J. By setting $J = I$, we have that a (strong) Pareto-best strategy profile is a (unique) c-rationalizable strategy profile.■

Proof of Theorem 2: The proof follows immediately from the following.
Lemma 5: *$A \Downarrow B$ iff $A \Rightarrow B$.*
Proof: The result is clearly true if $A = B$, so we assume $B \subsetneq A$.

"only if part": Suppose $A \Downarrow B$ via $J = \mathcal{J}_{AB}$. Let $j \in J$ and $\mu \in \Delta|_{A_{-j}}$. We distinguish three cases.

Case 1.1 $\mu(B_{-j}) = 1$. By Definition 2(2.1), $c_j \in A_j \backslash B_j$ is strictly dominated in the subgame restricted on $S_j \times B_{-j}$. Thus, $c_j \in A_j \backslash B_j$ is a never-best response in the subgame restricted on $S_j \times B_{-j}$. Since $\mu(B_{-j}) = 1$, $\mu = \mu|_{B_{-j}}$. Therefore, $u_j(c_j, \mu) < u_j(s_j, \mu|_{B_{-j}})$ for some $s_j \in S_j$.

Case 1.2 $\mu(B_{-j}) = 0$. By Definition 2(2.2), for any $a_j \in A_j$ there is $\sigma_j \in \Delta(S_j)$ such that

$$u_j(a_j, \mu) \leq \max_{c_{-j} \in A_{-j} \backslash B_{-j}} u_j(a_j, c_{-j}) < \min_{b_{-j} \in B_{-j}} u_j(\sigma_j, b_{-j}) \leq u_j(\sigma_j, \nu),$$

where $\nu \in \Delta|_{B_{-j}}$. Therefore, $u_j(a_j, \mu) < u_j(s_j, \mu|_{B_{-j}})$ for some $s_j \in S_j$.

Case 1.3 $\mu(B_{-j}) \neq 0$ *or* 1. Then, $\mu = q\mu' + (1-q)\,\mu''$ where $q \equiv \mu(B_{-j})$, $\mu' \in \Delta|_{B_{-j}}$, and $\mu'' \in \Delta|_{A_{-j}\backslash B_{-j}}$. Since $\mu'(B_{-j}) = 1$, for any given $a_j \in A_j$,

$$u_j(a_j, \mu') \leq u_j(s'_j, \mu') = u_j(s'_j, \mu'|_{B_{-j}}) \text{ for some } s'_j \in S_j.$$

Since $\mu''(B_{-j}) = 0$, by Case 1.2,

$$u_j(a_j, \mu'') < u_j(s''_j, \mu'|_{B_{-j}}) \text{ for some } s''_j \in S_j.$$

Therefore,

$$\begin{aligned} u_j(a_j, \mu) &= q u_j(a_j, \mu') + (1-q)\, u_j(a_j, \mu'') \\ &< q u_j(s'_j, \mu'|_{B_{-j}}) + (1-q)\, u_j(s''_j, \mu'|_{B_{-j}}) \\ &= u_j(q s'_j + (1-q)\, s''_j, \mu'|_{B_{-j}}). \end{aligned}$$

But, since $\mu''(B_{-j}) = 0$, $\mu|_{B_{-j}} = \mu'|_{B_{-j}}$. Thus, $u_j(a_j, \mu) < u_j(q s'_j + (1-q)\, s''_j, \mu|_{B_{-j}})$. Hence, $u_j(a_j, \mu) < u_j(s_j, \mu|_{B_{-j}})$ for some $s_j \in S_j$.

"if part": Suppose $A \Rightarrow B$ via $J = \mathcal{J}_{AB}$. Let $j \in J$ and $\mu \in \Delta|_{A_{-j}}$. We distinguish two cases.

Case 2.1 $\mu(B_{-j}) = 1$. Since $\mu(B_{-j}) = 1$, $\mu|_{B_{-j}} = \mu$. By Definition 1(1.1), for every $c_j \in A_j \backslash B_j$, $u_j(c_j, \mu) < u_j(s_j, \mu)$ for some $s_j \in S_j$. That is, $c_j \in A_j \backslash B_j$ is a never-best response in the subgame restricted on $S_j \times B_{-j}$. Therefore, $c_j \in A_j \backslash B_j$ is strictly dominated in the subgame restricted on $S_j \times B_{-j}$ (see, e.g., Osborne and Rubinstein's [22] Lemma 60.1). Thus, Definition 2(2.1) holds.

Case 2.2 $\mu(c_{-j}) = 1$ *for some* $c_{-j} \in A_{-j} \backslash B_{-j}$. Since $\mu(B_{-j}) = 0$, by Definition 1(1.1) and Lemma 1, $u_j(a_j, \mu) < \max_{\sigma_j \in \Delta(S_j)} u_j(\sigma_j, \sigma_{-j})$ for all $\sigma_{-j} \in \Delta\,(B_{-j})$. Since $\Delta\,(B_{-j})$ is compact, by the well-known Maximum Theorem,

$$u_j(a_j, \mu) < \min_{\sigma_{-j} \in \Delta(B_{-j})} \max_{\sigma_j \in \Delta(S_j)} u_j(\sigma_j, \sigma_{-j}).$$

By the Minmax Theorem,

$$\begin{aligned} u_j(a_j, c_{-j}) &= u_j(a_j, \mu) \\ &< \min_{\sigma_{-j} \in \Delta(B_{-j})} \max_{\sigma_j \in \Delta(S_j)} u_j(\sigma_j, \sigma_{-j}) \\ &= \max_{\sigma_j \in \Delta(S_j)} \min_{\sigma_{-j} \in \Delta(B_{-j})} u_j(\sigma_j, \sigma_{-j}) \\ &= \min_{\sigma_{-j} \in \Delta(B_{-j})} u_j(\overline{\sigma}_j, \sigma_{-j}) \\ &\leq u_j(\overline{\sigma}_j, b_{-j}) \ \forall b_{-j} \in B_{-j}, \end{aligned}$$

where $\overline{\sigma}_j \in \arg\max_{\sigma_j \in \Delta(S_j)} \min_{\sigma_{-j} \in \Delta(B_{-j})} u_j(\sigma_j, \sigma_{-j})$. Thus, Definition 2(2.2) holds.■

References

[1] A. Ambrus, Theories of coalitional rationality, mimeo, Harvard University, 2005.

[2] A. Ambrus, Coalitional rationalizability, Quart. J. Econ. 121(2006), 903-929.

[3] K.R. Apt, Order independence and rationalizability, mimeo, National University of Singapore, 2005.

[4] R.J. Aumann, Acceptable points in general cooperative n-person games, in: Contributions to the Theory of Games IV, Princeton University Press, 1959.

[5] K. Basu and J.W. Weibull, Strategy subsets closed under rational behavior, Econ. Letters 36(1991), 141-146.

[6] B.D. Bernheim, Rationalizable strategic behavior, Econometrica 52(1984), 1007-1028.

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

[8] Y.C. Chen, N.Y. Long and X. Luo, Iterated strict dominance in general games, forthcoming in: Games Econ. Behav. (2007).

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

[10] M. Dufwenberg and M. Stegeman, Existence and uniqueness of maximal reductions under iterated strict dominance, Econometrica 70(2002), 2007-2023.

[11] L. Epstein and M. LeBreton, Dynamically consistent beliefs must be Bayesian, J. Econ. Theory 61(1993), 1-22.

[12] P. Ghirardato, Revisiting Savage in a conditional world, Econ. Theory 20(2002), 83-92.

[13] J. Greenberg, The Theory of Social Situations: An Alternative Game-Theoretic Approach, Cambridge University Press, Cambridge, 1990.

[14] P.J.-J. Herings, A. Mauleon, and V.J. Vannetelbosch, Rationalizability for social environments, Games Econ. Behav. 49(2004), 135-156.

[15] C.M. Kahn and D. Mookherjee, Universal Coalition-proof equilibrium: Theory and applications, mimeo, Bell Communications - Economic Research Group, 1994.

[16] E. Karni, Foundations of Bayesian theory, J. Econ. Theory 132 (2007), 167-188.

[17] D. Kreps, Notes on the Theory of Choice, Westview, Bouldler, 1988.

[18] X. Luo, General systems and φ-stable sets – a formal analysis of socioeconomic environments, J. Math. Econ. 36(2001), 95-109.

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

[20] R. Myerson, Axiomatic foundations of Bayesian decision theory, mimeo, Northwestern University, 1986.

[21] R. Myerson, Game Theory: Analysis of Conflict, Harvard University Press, 1991.

[22] M.J. Osborne and A. Rubinstein, A Course in Game Theory, The MIT Press, 1994.

[23] D. Pearce, Rationalizable strategic behavior and the problem of perfection, Econometrica 52(1984), 1029-1051.

[24] D. Ray and R. Vohra, Equilibrium binding agreements, J. Econ. Theory 73(1997), 30-78.

[25] D. Ray and R. Vohra, A theory of endogenous coalition structures, Games Econ. Behav. 26(1999), 286-336.

[26] A. Roth, Subsolutions and the supercore of cooperative games, Math. Oper. Res. 1 (1976), 43-49.

[27] L. Savage, The Foundations of Statistics, Wiley, NY, 1954.

[28] M. Shimoji and J. Watson, Conditional dominance, rationalizability, and game forms, J. Econ. Theory 83(1998), 161-195.

[29] T. Tan and S. Werlang, The Bayesian foundations of solution concepts of games, J. Econ. Theory 45(1988), 370-391.

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