Coalitions, Agreements and Efficiency*

Effrosyni Diamantoudi and Licun Xue†

Revised: February 2006

Abstract

If agents negotiate openly and form coalitions, can they reach efficient agreements? We address this issue within a class of coalition formation games with externalities where agents' preferences depend solely on the coalition structure they are associated with. We derive Ray and Vohra's (1997) notion of *equilibrium binding agreements* using von Neumann and Morgenstern abstract stable set and then extend it to allow for arbitrary coalitional deviations (as opposed to nested deviations assumed originally). We show that, while the extended notion facilitates the attainment of efficient agreements, inefficient agreements can nevertheless arise, even if utility transfers are possible.

Journal of Economic Literature Classification Number: C71 C72.

Keywords: coalition formation, externalities, efficiency, agreements

*Parts of this work were conducted while the authors were visiting the Wallis Institute of Political Economy at the University of Rochester and the Department d'Economia i d'Història at the Universitat Autònoma de Barcelona. We would like to acknowledge both institutions for their hospitality and the seminar participants for helpful comments. We would also like to thank Hubert Kempf and Per Krussel for their comments and suggestions. Thanks also go to two anonymous referees and an associate editor for their comments and suggestions. Part of this research was funded by SSHRC and FQRSC.

†Concordia University and McGill University, respectively. Corresponding author: Effrosyni Diamantoudi, Department of Economics, Concordia University, 1455 de Maisonneuve Blvd West, Montreal, Quebec, Canada H3G 1M8. E-mail: ediamant@alcor.concordia.ca

1 Introduction

If agents negotiate openly and are able to form coalitions, can they reach efficient agreements? We address this issue within a class of simple coalition formation games with externalities where each agent's preferences depend only on the coalition structure or partition of the agents. Arguably, if binding agreements can be written without any informational imperfections, then all the gains from cooperation should be extracted. The resulting agreement must be Pareto-optimal; moreover, if utility is transferable, aggregate surplus must be maximized. Such an assertion encapsulates the Coase (1960) theorem. However, in Ray and Vohra (1997), it is shown that when coalitions can form, efficiency can no longer be guaranteed. Indeed, the authors define the notion of *equilibrium binding agreements* (henceforth EBA) for strategic form games, a more general framework than ours, and construct examples where the only agreements that can be reached are inefficient, even when utility is transferrable. This negative result[1] casts a shadow on the validity of Coase theorem in environments where coalitions can form.

What is even more puzzling is that in Ray and Vohra's notion agents are sophisticated. In particular, they are assumed to be *farsighted* in that when contemplating a deviation, a coalition takes into consideration that further deviations may occur and that other deviating coalitions also apply similar reasoning. For farsighted agents, it is the *final* agreement their deviations lead to that matters. Moreover, agents examine the *credibility* of the final outcome, thus, the notion is defined consistently. One feature of EBA is the assumption of internal deviations, that is, only a subset of an existing coalition can deviate. The negotiation process underlining EBA is as follows: Starting with the grand coalition a subset of agents can break away anticipating that other coalitions (subsets) may further break apart. While this feature makes a recursive definition possible, it precludes the possibility of coalition merging and renegotiation. As Ray and Vohra (1997) wrote,

> We must state at the outset that our treatment is limited by the assumption that agreements can be written only between members of an existing coalition; once a coalition breaks away from a larger coalition it cannot forge an agreement with any member of its complement. Thus, deviations can only serve to make an existing coalition structure finer never coarser. This is also the assumption in the definition of a coalition proof Nash equilibrium. It must be emphasized that an extension of

[1] Similar negative results also appear in Bloch (1996) and Ray and Vohra (1999) among others.

these notions to the case of arbitrary blocking is far from trivial. (p.33)

It remains an interesting question whether inefficiency arises due to the restriction of internal deviations. Within our framework, we extend the definition of EBA to allow for arbitrary coalitional deviations and, hence, an open and unrestricted negotiation process. Starting with any coalition structure (including the grand coalition), any coalition, not necessarily a subset of an existing one, can deviate and bring about a new coalition structure. Put differently, even if agents temporarily belong to some coalition they are not prevented from being in contact with members of other coalitions and orchestrating a joint objection/deviation. With open negotiations it is no longer necessary to start with the grand coalition as a coalition structure is reachable from any other structure.

In order to formalize our extension we first reformulate Ray and Vohra's definition using von Neumann and Morgenstern (1944) stable set. Such a reformulation is akin to that of coalition proof Nash equilibrium[2] (Bernheim, Peleg, and Whinston, 1987) by Greenberg (1989) and it offers a simple definition of EBA in our framework. More importantly, the use of stable set enables us to deal with the circularity that results from allowing arbitrary coalitional deviations while maintaining consistency as in the original definition of EBA. Can inefficient outcomes emerge from an open, unrestricted negotiation as entailed by the new definition? We show that while the assumption of internal deviations hinders efficiency to some extent, inefficient agreements can still arise in open unrestricted negotiations. In particular, we identify a class of games where efficiency can be attained and construct a counter example where no agreement is efficient. We also analyze the consequence of transfers. Although transfers further facilitate the attainment of efficient agreements, we construct a robust example where only inefficient agreements can emerge from open negotiations. The negative results reinforce the inefficiency puzzle posed by Ray and Vohra (1997).

In the negotiation process underlining our extension of EBA, it is *feasible* for any coalition to form and object to/deviate from any coalition structure. However, being farsighted, a coalition engages in a deviation if and only if it can ultimately benefit from doing so. Therefore, a coalition structure is "stable" if no coalition wishes to deviate, anticipating the final outcome its deviation may lead to and a coalition

[2]Note that the difference between EBA and coalition proof Nash equilibrium (CPNE) lies beyond the fact that the former considers binding agreements. In the definition of EBA, agents are farsighted in that each coalition considers the final outcome its deviation leads to, while in the definition of CPNE, a deviating coalition takes its complement's choice as given.

structure is "unstable" as long as one coalition wishes to deviate, again anticipating the final outcome of its deviation. While our approach captures the foresight of selfish agents, it differs from the sequential "non-cooperative" approach of coalition formation (at least) in that we do not specify the exact order with which individual agents act. Moreover, in the latter approach additional restrictions are often placed on the coalition formation process. For instance, Bloch (1996) studies the same class of games as ours in a noncooperative framework of sequential coalition formation, while Ray and Vohra (1999)[3] study a more general framework with endogenized payoff division. A common feature of their models is that once a coalition forms, the game is only played among the remaining players and established coalitions may not seek to attract new members nor break apart. The requirement for coalitions to commit plays an important role in determining the equilibrium coalition structures[4]. However, inefficient coalition structures can arise in their equilibria as well.

The organization of the paper is as follows: After the presentation of the preliminaries in the next Section we reformulate EBA in Section 3 and extend the notion in Section 4. Section 5 presents sufficient conditions for a game to admit efficient coalition structures and a counter example where no agreement is efficient. In Section 6 we discuss the impact of transfers. Section 7 concludes the paper. Detailed analysis of the first counter example is delineated in the appendix.

2 Preliminaries

We start with some basic notations:

- Let N be a finite set of players.
- A coalition S is a non-empty subset of N.
- A partition of $S \subset N$ is $P = \{S_1, S_2, ..., S_k\}$ such that $\bigcup_{j=1}^{k} S_j = S$ and for all $i \neq j$, $S_i \bigcap S_j = \emptyset$ and $\mathcal{P}(S)$ is the set of partitions of S. A partition of N is called a coalition structure and $\mathcal{P} \equiv \mathcal{P}(N)$ is the set of all coalition structures.

In the simple coalition formation games with externalities we study here, each player's preferences depend on the *entire* coalition structure. Formally, a *simple*

[3]See also Ray and Vohra (2001).

[4]Macho-Stadler, Pérez-Castrillo and Porteiro (2002) analyze a Cournot oligopoly, a special case of the class of games we study here. They assume that coalitions form by merging bilaterally, thereby relaxing the commitment of a formed coalition. They show that if the number of firms is sufficiently large then the monopoly (grand coalition) is the equilibrium outcome.

coalition formation game with externalities G is $(N, \{\succeq_i\}_{i\in N})$ or $(N, \{u_i\}_{i\in N})$ where

- N is the finite set of players;
- for all $i \in N$, $\succeq_i$ is a complete, reflexive, and transitive binary relation on $\mathcal{P}$, the set of coalition structures, $\succ_i$ denotes the asymmetric part of $\succeq_i$ (i.e., strict preferences) and $\sim_i$ the indifference relation;
- for all $i \in N$, $u_i : \mathcal{P} \to \Re$ is i's payoff function.

Throughout this paper we write $P \succ_S Q$ for $P, Q \in \mathcal{P}$ and $S \subset N$, if $P \succ_i Q$ or $u_i(P) > u_i(Q)$ for all $i \in S$.

The simple class of games we study here extends the class of "hedonic games" [see, e.g., Banerjee, Konishi and Sönmez (2001), Bogomolnaia and Jackson (2002), Barberà and Gerber (2003), and Diamantoudi and Xue (2003)], where each agent's preferences depend only on the coalition he belongs to. The class of games studied in this paper can also be viewed as a special class (with fixed payoff division) of partition function games introduced by Thrall and Lucas (1963), which, in turn, is a special case of normal form TU games studied by Zhao (1992). Yi (1997) studies a more restricted class of games than ours; in particular, he studies symmetric games and examines three models of coalition formation: simultaneous coalition formation model of Yi and Shin (1995), sequential coalition formation model of Bloch (1996), and EBA of Ray and Vohra (1997). Moreover, Yi classifies symmetric games into two categories, one with positive externalities (e.g., Cournot oligopoly, a class of public good economies) and one with negative externalities (e.g., customs unions).

3 EBA and Inefficiency

As discussed in the introduction, Ray and Vohra's (1997) notion of EBA is defined for strategic form games, a more general framework than ours. We shall first adapt their definition to our setting. The negotiation process underlying EBA is as follows. Suppose the grand coalition N is under consideration. A coalition $S \subsetneq N$ can break away from N and in doing so, it induces the coalition structure $\{S, N \setminus S\}$. This coalition structure is likely to be a temporary one since S or $N \setminus S$ may further break apart. More generally, given a coalition structure $P \in \mathcal{P}$, any coalition $T \subsetneq S$, for some $S \in P$, can break away from S. In addition, once T breaks away from S, it cannot forge an agreement with any member of $N \setminus S$; thus, deviations can only lead to finer coalition structures. Such an "internal" or "nested" deviation can be formalized as follows.

Internal Coalitional Deviation Given a coalition structure $P \in \mathcal{P}$ and some $S \in P$, coalition $T \subsetneq S$, by breaking away from S, induces a temporary coalition structure given by $P' = P \setminus \{S\} \cup \{T, S \setminus T\}$. We write $P \overset{T}{\rightarrowtail} P'$ in this case. Call Q a *refinement* of P if Q can be reached from P through a sequence of nested coalitional deviations.

Agents are assumed to be farsighted: Each deviating coalition or *perpetrator*, following Ray and Vohra (1997), is aware that further deviations may occur; thus, in contemplating a deviation, a coalition considers the ultimate consequence of its deviation. Given the nature of the negotiation process, EBA can be defined recursively.

- Start with the finest coalition structure, $\bar{P}$, of singleton coalitions. Since no further deviations are possible, $\bar{P}$ is the finest EBA.

- Now consider a coalition structure P such that $P \overset{T}{\rightarrowtail} \bar{P}$; thus, P comprises all singleton coalitions but one coalition of size 2. Let $\{ij\} \in P$. Then P is an EBA if and only if neither i nor j has an incentive to break away from the coalition and induce $\bar{P}$.

  ...

- Consider $Q \in \mathcal{P}$. Suppose EBAs have been defined for all refinements of Q. Then Q is an EBA if and only if it is not "blocked" by another EBA Q'. Q is blocked by Q' if Q' is an EBA and it can be can reached from Q via a sequence of internal coalitional deviations (i.e., Q' is a refinement of Q) such that

  – each deviating coalition/perpetrator[5] belongs to Q',

  – one of the deviating coalitions (the *leading perpetrator*) prefers Q' to Q, and

  – any coalition structure attained by re-merging of the other perpetrators with their corresponding residual coalitions (new coalitions in Q' other than the perpetrators) is blocked by Q' with one of these perpetrators as a leading perpetrator.

Given that deviations can only make a coalition structure finer, never coarser it is necessary that the negotiation process starts from the grand coalition. The

[5]If a coalition in Q breaks into k coalitions in Q' then $k-1$ of these coalitions can be identified as deviating coalitions or perpetrators while the remaining coalition is labeled as residual coalition.

solution of a game is, therefore, defined to be the set of coarsest EBAs. Note that in the above recursive definition of EBA, the notion of "blocking" itself is also defined recursively. To formalize this notion of blocking, we first introduce the following definition.

Definition 1 *P' is RV-reachable from P via coalitions/perpetrators $T^0, T^1, \cdots, T^{k-1}$ if $T^j \in P'$ for all $j = 0, \cdots, k-1$ and $P^0 \overset{T^0}{\rightarrowtail} P^1 \overset{T^1}{\rightarrowtail} P^2 \overset{T^2}{\rightarrowtail} P^3 \overset{T^3}{\rightarrowtail} P^4 ... P^{k-1} \overset{T^{k-1}}{\rightarrowtail} P^k$ where $P^0 = P$ and $P^k = P'$. If, in addition, $P^j \prec_{T^j} P'$, for all $j = 0, \cdots k-1$, P' is said to sequentially dominate[6] P, or $P' \gg^{Seq} P$, through coalitions $T^0, T^1, \cdots, T^{k-1}$.*

Given that all perpetrators $T^0, T^1, \cdots, T^{k-1}$ are in the final coalition structure P', if P' is RV-reachable from P via $T^0, T^1, \cdots, T^{k-1}$ then P' is also RV-reachable via any re-ordering of $T^0, T^1, \cdots, T^{k-1}$. Moreover, given the set of perpetrators $T^0, T^1, \cdots, T^{k-1}$ in P' one can uniquely identify the set of residual coalitions, $P' \setminus \left(P \cup \left\{T^0, T^1, \cdots, T^{k-1}\right\}\right)$, which are new coalitions in P' but not active deviators. Therefore, any re-merging of perpetrators and their corresponding residuals generates an intermediate coalition structure that is RV-reachable from P via a subcollection of $\left\{T^0, T^1, \cdots, T^{k-1}\right\}$. Ray and Vohra's (1997) definition of EBA does not simply employ sequential dominance as defined above for the following consideration: Suppose that P' sequentially dominates P via $P^0 \overset{T^0}{\rightarrowtail} P^1 \overset{T^1}{\rightarrowtail} P^2 \overset{T^2}{\rightarrowtail} P^3 \overset{T^3}{\rightarrowtail} P^4 ... P^{k-1} \overset{T^{k-1}}{\rightarrowtail} P^k$ where $P^0 = P$ and $P^k = P'$, and P' is an EBA. To rule out P as an EBA on the basis that P' sequentially dominates P implies that T^0, in contemplating its deviation, is certain/optimistic about the exact order of other coalitional deviations. To circumvent this optimism, Ray and Vohra (1997) define the following dominance relation (RV-dominance, henceforth):

Definition 2 *P' RV-dominates P, or $P' \gg^{R\&V} P$, if there exist $T^0, T^1, \cdots, T^{k-1} \in P'$ such that*

(1) P' is RV-reachable from P via $T^0, T^1, \cdots, T^{k-1}$,

(2) $P \prec_T P'$ for some leading perpetrator $T \in \left\{T^0, T^1, \cdots, T^{k-1}\right\}$, and

(3) if $Q = \hat{P}$ or Q is RV-reachable from $\hat{P}$ via a subcollection of $\left\{T^0, \cdots, T^{k-1}\right\} \setminus \{T\}$, where $P \overset{T}{\rightarrowtail} \hat{P}$, then Q is RV-dominated by P' with some $S \in \{T^0, \cdots, T^{k-1}\} \setminus \{T\}$ as the leading perpetrator.

[6] In the orginal definition of sequential dominance in Ray and Vohra (1997), the condition that $T^j \in P'$ for all j is not imposed.

Note that the above definition is recursive. In particular, condition (3) implies that no matter in which order coalitions in $\{T^0, T^1, \ldots, T^{k-1}\} \setminus \{T\}$ deviate, every intermediate coalition structure generated as a result is RV-dominated by P'. Once RV-dominance is formulated, EBA can now be defined: The finest coalition structure, $\bar{P}$, of singleton coalitions is an EBA since no further deviations are possible. P is an EBA if and only if it is not RV-dominated by another EBA. The fact that a coalition structure is an equilibrium binding agreement if and only if it is not "defeated" by another equilibrium binding agreement signifies the consistency embedded in the definition of EBA. Such consistency is precisely what is captured by "von Neumann and Morgenstern (vN-M) stable set". A vN-M stable set is a set of agreements, called a solution set, that is free of inner contradictions and accounts for every elements it excludes; in particular, no agreement in the solution set is defeated by another agreement in the same solution set and if an agreement is excluded from the solution set, it must be defeated by an agreement in the solution set. More formally,

vN-M stable set of $(\mathcal{P}, >)$: Let $>$ be a binary relation on $\mathcal{P}$ and $\mathcal{R} \subset \mathcal{P}$. Then,

- $\mathcal{R}$ is *vN-M internally stable* for $(\mathcal{P}, >)$ if there do not exist $P, P' \in \mathcal{R}$ such that $P' > P$;
- $\mathcal{R}$ is *vN-M externally stable* for $(\mathcal{P}, >)$ if for all $P \in \mathcal{P} \setminus \mathcal{R}$, there exists $P' \in \mathcal{R}$ such that $P' > P$.
- $\mathcal{R}$ is a *vN-M stable set* for $(\mathcal{P}, >)$ if it is both internally and externally stable.

Within our framework, Ray and Vohra's (1997) EBA can be reformulated as a vN-M stable set[7].

EBA (*A Reformulation*) Let Ω be the vN-M stable set of $(\mathcal{P}, \gg^{R\&V})$. Then P is an EBA if and only if $P \in \Omega$. The coarsest coalition structures in Ω are referred to as the solution of the game[8].

[7] This reformulation of EBA using a vN-M stable set cannot be directly generalized to strategic form games.

[8] The uniqueness of the vN-M stable set for $(\mathcal{P}, \gg^{R\&V})$ follows from the acyclicity of $\gg^{R\&V}$. Moreover, the vN-M stable set of $(\mathcal{P}, \gg^{R\&V})$ is different from the core of $(\mathcal{P}, \gg^{R\&V})$ because of the presence of externalities. In the context of cooperative games, when only internal deviations are allowed, the stable set and the core are closely related [see Ray (1989) and Greenberg (1990)].

Ray and Vohra (1997) show that if $P' \gg^{R\&V} P$ then $P' \gg^{\text{Seq}} P$. The following lemma characterizes RV-dominance using sequential dominance.

Proposition 1 *P' RV-dominates P if and only if there exists $T^0, T^1, \cdots, T^{k-1} \in P'$ (with T^0 being the leading perpetrator) such that*

(a) *P' sequentially dominates P via $T^0, T^1, \cdots, T^{k-1}$* and

(b) if $Q = \hat{P}$ or Q is RV-reachable from $\hat{P}$ via a subcollection of $\left\{T^1, \cdots, T^{k-1}\right\}$, where $P \overset{T^0}{\rightarrowtail} \hat{P}$, then P' *sequentially dominates Q via a subcollection of* $\{T^1, \ldots, T^{k-1}\}$.

Proof. Only if: Suppose P' RV-dominates P via a collection of perpetrators $\left\{E^0, \cdots, E^{k-1}\right\}$. We first show part (a). Let T^0 be the leading perpetrator and Q^1 be such that $P^0 \overset{T^0}{\rightarrowtail} Q^1$. Then, by definition, $P \prec_{T^0} P'$ and Q^1 is RV-dominated via $\left\{E^0, \cdots, E^{k-1}\right\} \backslash \{T^0\}$. Let T^1 be the leading perpetrator and $Q^1 \overset{T^1}{\rightarrowtail} Q^2$. Then by definition, $Q^1 \prec_{T^1} P'$ and Q^2 is RV-dominated via $\left\{E^0, \cdots, E^{k-1}\right\} \backslash \left\{T^0, T^1\right\}$. Continuing in this fashion, we obtain $T^0, \cdots, T^{k-1}$ such that P' sequentially dominates P via $T^0, \cdots, T^{k-1}$. Part (b) can be proved by similar constructions.

If: The proof is by induction on ℓ, the number of coalitions via which P' can be reached from P. If $\ell = 1$, (3) in Definition 2 and (b) in Proposition 1 are superfluous and (a) in Proposition 1 implies (1) and (2) in Definition 2. Now, assume that (a) and (b) imply (1), (2), and (3) for all $\ell < k$. To show that (a) and (b) imply (1), (2), and (3) for $\ell = k$, note that (a) implies (1) and (2) with $T = T^0$ as the leading perpetrator. Let $P \overset{T}{\rightarrowtail} \hat{P}$. Then by the induction hypothesis, $P' \gg^{R\&V} \hat{P}$ via a subcollection of $\left\{T^1, \ldots, T^{k-1}\right\}$ with some $\hat{T} \in \left\{T^1, \ldots, T^{k-1}\right\}$ as the leading perpetrator; moreover, for all Q such that Q is RV-reachable from $\hat{P}$ via a subcollection of $\left\{T^1, \ldots, T^{k-1}\right\}$, we have $P' \gg^{R\&V} Q$ via a subcollection of $\left\{T^1, \ldots, T^{k-1}\right\}$ with some $\tilde{T} \in \left\{T^1, \ldots, T^{k-1}\right\}$ as the leading perpetrator. Therefore, $P' \gg^{R\&V} P$. ■

The following Corollary points to another alternative definition of RV-dominance that is not explicitly recursive.

Corollary 1 *P' RV-dominates P if and only if there exist perpetrators $T^0, \cdots, T^{k-1}$ such that*

(1) P' is RV-reachable from P via $T^0, \cdots, T^{k-1}$,

(2) $P \prec_{T^0} P'$ and

(3') if $Q = \hat{P}$ or Q is RV-reachable from $\hat{P}$ via a subcollection of $\{T^1, \cdots, T^{k-1}\}$, where $P \overset{T^0}{\rightarrowtail} \hat{P}$, then there exists $S \in \{T^1, T^2, \ldots, T^{k-1}\} \backslash Q$ such that $Q \prec_S P'$.

Proof. Only if: Following from Proposition 1 and the definition of sequential dominance.

If: From any Q as defined in (3') we can construct a sequence of coalitions such that P' sequentially dominates Q via this sequence of coalitions. Also, P' sequentially dominates P given that P' sequentially dominates $\hat{P}$ and $P \prec_{T^0} P'$. Thus, by Proposition 1 $P' \gg^{R\&V} P$. ■

The reformulation of EBA with vN-M stable set does not only offer a simple definition of EBA but also enables us to extend it in various directions. One of the motivations for extending it was the inefficiency puzzle presented in Ray and Vohra (1997): although binding agreements are possible, inefficient outcomes can, nevertheless, emerge. The following example illustrates the inability of players to reach efficient binding agreements[9]. In the table below are the payoff vectors associated with ten partitions and all other partitions are assumed to yield 0 payoff vectors.

a $\{123, 45\}$ $5, 5, 5, 9, 9$	b $\{234, 51\}$ $5, 5, 5, 9, 9$	c $\{345, 12\}$ $5, 5, 5, 9, 9$	d $\{451, 23\}$ $5, 5, 5, 9, 9$	e $\{512, 34\}$ $5, 5, 5, 9, 9$
f $\{12, 34, 5\}$ $4, 4, 4, 8, 8$	g $\{23, 45, 1\}$ $4, 4, 4, 8, 8$	h $\{34, 51, 2\}$ $4, 4, 4, 8, 8$	i $\{45, 12, 3\}$ $4, 4, 4, 8, 8$	j $\{51, 23, 4\}$ $4, 4, 4, 8, 8$

Table 1

It is easy to see that f, g, h, i, and j are all EBAs. So are those for which f, g, h, i, or j is not a refinement (for example, $\{135, 24\}$). On the other hand, a, b, c, d and e are not EBAs. Take a, for example. a is not an EBA because $a \overset{1}{\rightarrowtail} g$ and $a \prec_1 g$ or $a \ll^{R\&V} g$. Therefore, all the EBAs of this game are inefficient. Indeed, f, g, h, i, and j are Pareto dominated by a, b, c, d and e, respectively and so are all other EBAs. One might ask, why are matters not renegotiated at this stage to the dominating outcome? The assumption of nested deviations rules out the possibility of such renegotiation. As Ray and Vohra wrote,

[9] Ray and Vohra have a 3-player example of strategic form game where the only equilibrium binding agreements are inefficient. Because of the simpler framework we use, we need a 5-player game to illustrate the inefficiency of EBA.

> This is a serious issue that is neglected in our model, because we only permit "internal" deviations. (p.51)

This is precisely the reason why we extend the definition of EBA to allow for arbitrary coalitional deviations. Equipped with the notion of vN-M stable set, we can define a notion consistently (as the original definition) while allowing for arbitrary coalitional deviations.

4 Extended EBA

In this section we extend the notion of EBA by relaxing the assumption of internal deviations and allowing arbitrary coalitions to deviate. Moreover, a deviating coalition is not constrained to stay together; that is, we empower the deviating coalition with the ability to restructure itself. These features (of open negotiation) are introduced to facilitate the attainment of Pareto efficient coalition structures[10].

Given a coalition structure, when a coalition of players, $T \subset N$, deviates by partitioning itself in a certain way, the new coalition structure is the one consisting of the partition of T, all the unaffected coalitions, as well as all the disrupted coalitions. Formally, a coalitional deviation is defined as follows:

Coalitional Deviation Given a coalition structure $P = \{S_1, \cdots, S_k\} \in \mathcal{P}$, a coalition $T \subset N$ can reorganize itself to some partition $\{T_1, \cdots, T_\ell\} \in \mathcal{P}(T)$. The resulting coalition structure, before any further regrouping and restructuring, is $P' \in \mathcal{P}$ such that

(i) $\{T_1, \cdots, T_\ell\} \subset P'$, that is, the new partitioning of T is included in the new coalition structure.

(ii) $\forall j = 1, \cdots, k,\ S_j \cap T \neq \emptyset \implies S_j \setminus T \in P'$, that is, the residuals of all coalitions affected by the deviation of T are also included in the new coalition structure.

(iii) $\forall j = 1, \cdots, k,\ S_j \cap T = \emptyset \implies S_j \in P'$, that is, all those coalitions that were unaffected by the deviation of T remain members of the new coalition structure.[11]

[10] Indeed, if we require the deviating coalition to stay together, the negative results remain while some of our positive results (e.g., Proposition 2) can no longer hold.

[11] Note that points (ii) and (iii) can be written more concisely as follows: $\forall j = 1, ..., k,\ S_j \backslash T \neq \emptyset \implies S_j \backslash T \in P'$.

We write $P \xrightarrow{T} P'$ to denote that "T induces P' **temporarily** from P".

Once it is precisely defined what a coalition can (directly) induce, we proceed to define our dominance relation. While myopic agents look only at the next step, farsighted players consider the ultimate outcome of their actions. Thus, a coalition may choose to "deviate" to a coalition structure, which does not necessarily make its members better off, as long as its deviation leads to a final coalition structure that benefits all its members; similarly, a coalition may choose not to deviate to a coalition structure it prefers if its deviation *eventually* leads to coalition structures that make its members worse off. Similar to the sequential dominance defined in the previous section the following "indirect dominance[12]" captures foresight when arbitrary coalitions can deviate.

Indirect dominance: P' indirectly dominates P, or $P' \gg P$, if there exist a sequence of coalition structures $P^1, P^2, \cdots, P^k \in \mathcal{P}$, where $P^1 = P$ and $P^k = P'$, and a sequence of coalitions $T^1, T^2, \cdots, T^{k-1}$ such that for all $j = 1, \cdots k-1$

(i) $P^j \xrightarrow{T^j} P^{j+1}$ and

(ii) $P^j \prec_{T^j} P'$.

An extended notion of EBA with unrestricted coalitional deviations can be defined consistently by using the vN-M stable set of $(\mathcal{P}, \gg)$, where the RV-dominance, $\gg^{R\&V}$, is replaced with indirect dominance[13], $\gg$.

Extended EBA (EEBA) Let $\mathcal{Q} \subset \mathcal{P}$ be a vN-M stable set of $(\mathcal{P}, \gg)$. $P \in \mathcal{P}$ is an extended EBA or EEBA if $P \in \mathcal{Q}$.

Given a vN-M stable set $\mathcal{Q}$, if negotiation commences with some (stable) agreement in $\mathcal{Q}$, no coalition can ultimately benefit from any deviation while starting with any (unstable) agreement in $\mathcal{P} \setminus \mathcal{Q}$, at least one coalition can benefit ultimately by initiating a deviation[14]. Although the vN-M stable set of $(\mathcal{P}, \gg^{R\&V})$ has a similar interpretation, $\gg^{R\&V}$ stipulates that deviations can only be internal, and because of this the solution of a game is considered to be coarsest partitions in the stable set[15].

[12] See also Harsanyi (1974), Chwe (1994) and Xue (1998).

[13] Note that with abritrary coalitional deviations we can no longer embed the conservative aspect of RV-dominance.

[14] Thus, if all agents expect $\mathcal{Q}$, $\mathcal{Q}$ will be "self-enforcing". In fact, subgame perfect equilibrium mapping (that specifies the continuation equilibrium for each subgame) has a similar interpretation.

[15] For instance, the singleton partition is always in the stable set.

Revisiting the example presented in the previous section we can see that efficiency can be restored. We will argue that a is an EEBA. Indeed, $Q = \{a\}$ is a vN-M stable set of $(\mathcal{P}, \gg)$. The following table summarizes how all other coalition structures are indirectly dominated by a. Coalition structure p denotes any arbitrary coalition structure that is not listed in Table 1.

$a \gg b:$	$b \xrightarrow{4} j \xrightarrow{5} \{23,1,4,5\} \xrightarrow{N} a$	$a \gg g:$	$g \xrightarrow{5} \{23,1,4,5\} \xrightarrow{N} a$
$a \gg c:$	$c \xrightarrow{5} f \xrightarrow{N} a$	$a \gg h:$	$h \xrightarrow{4} \{15,2,3,4\} \xrightarrow{N} a$
$a \gg d:$	$d \xrightarrow{5} \{14,23,5\} \xrightarrow{N} a$	$a \gg i$:	$i \xrightarrow{4} \{12,3,4,5\} \xrightarrow{N} a$
$a \gg e:$	$e \xrightarrow{5} f \xrightarrow{N} a$	$a \gg j:$	$j \xrightarrow{5} \{23,1,4,5\} \xrightarrow{N} a$
$a \gg f:$	$f \xrightarrow{N} a$	$a \gg p:$	$p \xrightarrow{N} a$

Table 2

Due to the symmetry of the game, b, c, d and e are also EEBAs. Moreover, none of f, g, h, i and j is an EEBA. Assume in negation that there exists a vN-M stable set $\mathcal{Q}$ of $(\mathcal{P}, \gg)$ that supports an inefficient coalition structure. Then, given the indirect dominance relation depicted in Table 2, $\mathcal{Q}$ cannot contain any of the efficient outcomes due to internal stability. If, however, all a, b, c, d and e are excluded, $\mathcal{Q}$ must contain more than one coalition structures to account for their exclusion. For example, if $j \in \mathcal{Q}$, then j cannot dominate e. But, coalition structures f, g, h, i and j dominate each other (and all of them dominate p) hence they cannot coexist in $\mathcal{Q}$. The following table illustrates how j indirectly dominates every other inefficient outcome. Similar arguments can be developed for the rest of the outcomes due to the symmetry of the game.

$j \gg f$:	$f \xrightarrow{3} \{12,3,4,5\} \xrightarrow{N} j$
$j \gg g$:	$g \xrightarrow{4} \{23,1,4,5\} \xrightarrow{N} j$
$j \gg h$:	$h \xrightarrow{3} \{15,2,3,4\} \xrightarrow{N} j$
$j \gg i$:	$i \xrightarrow{4} \{12,3,4,5\} \xrightarrow{N} j$
$j \gg p$:	$p \xrightarrow{N} j$

Table 3

Existence of vN-M stable set of $(\mathcal{P}, \gg)$ is not guaranteed. Case in point is a version of the roommate problem where $\{ij,k\} \succ_i \{ik,j\} \succ_i \{ijk\} \succ_i \{i,j,k\} \succ_i \{i,jk\}$ and i prefers to have j as a roommate, while j prefers to have k as a roommate and lastly k prefers to have i as a roommate. It is easy to see that no stable set exists for this game since $\{ij,k\} \ll \{jk,i\} \ll \{ik,j\} \ll \{ij,k\}$. While no two coalition structures can coexist in a stable set because of internal stability, a single coalition structure is externally unstable. In contrast, all structures involving pairs and the singletons structure are equilibrium binding agreements since the cyclicality is assumed away through the assumption of internal deviations: once $\{ij\}$ are formed j is not allowed to collude with k. Fortunately, the presence of cycles is not always a problem as can be seen in the example in Table 1. In the following section we identify classes of games where both existence and efficiency problems are resolved.

Implicit in the definition of vN-M stable set is the optimism on the part of deviating coalitions: a coalition engages in a deviation as long as its members benefit form one of the final outcomes its deviation may lead to. An alternative behavioral assumption on deviating coalitions is caution: a coalition engages in a deviation if its members benefit form all the final outcomes its deviation may lead to. Under the assumption of caution, a more inclusive notion (than vN-M stable set) can be defined and existence is guaranteed in our framework. See Chwe (1994), Xue (1998) and Diamantoudi and Xue (2003), among others, for notions that are built on cautious behavior of deviating coalitions.

5 (In)Efficiency

5.1 Positive Results

To proceed with the study of efficiency we need to distinguish between several notions of efficiency.

Pareto Efficiency $P \in \mathcal{P}$ is Pareto efficient if there does not exist $P' \in \mathcal{P}$ such that $P \prec_N P'$.

Strong Efficiency $P \in \mathcal{P}$ is strongly efficient if there does not exist $P' \in \mathcal{P}$ such that $\sum_{i \in N} u_i(P') > \sum_{i \in N} u_i(P)$.

When transfer payments within each coalition are possible, we need to consider not only coalition structures but also payoff allocations. Consider a partition $P \in \mathcal{P}$. A payoff allocation $x \in \Re^{|N|}$ is *feasible* for P if for all $S \in P$, $\sum_{i \in S} x_i = \sum_{i \in S} u_i(P)$. Pareto efficiency is then defined as follows.

Pareto Efficiency (with transfers) (P, x), where $P \in \mathcal{P}$ and $x \in \Re^{|N|}$ is feasible for P, is Pareto efficient if there does not exist (P', x') such that $P' \in \mathcal{P}$, $x' \in \Re^{|N|}$ is feasible for P', and $x'_i > x_i$ for all $i \in N$.

The following simple example from Ray and Vohra (1997) illustrates the inability of players to reach strongly efficient binding agreements.

Inefficiency Puzzle – A Cournot Oligopoly

The market demand is given by $p = a - by$, where p is the market price and y is aggregate demand. We assume symmetric firms with constant marginal cost c. Consider a coalition structure $P = \{S_1, S_2, \cdots, S_m\} \in \mathcal{P}$. Firms within the same coalition cooperate by maximizing their joint profit, while coalitions behave non-cooperatively across each other. Then, the profit of each firm in S_i is

$$\pi_i = \frac{1}{s_i(m+1)^2}\frac{(a-c)^2}{b}$$

where $s_i = |S_i|$ for all $i = 1, ..., m$.

The following table displays the per firm profit for the simple case of $n = 5$ and $\frac{(a-c)^2}{b} = 1$. The last two columns indicate which coalition structures survive the notion of Ray and Vohra (1997). In symmetric games, coalitions need to be identified only by their sizes. For example, $\langle 4, 1 \rangle$ denotes coalition structures with a coalition of size 4 and a coalition of size 1.

Structure size-wise	Perpetrator size-wise	Per Firm Profit	EBA
$\langle 5 \rangle$	——	$\frac{1}{20}$	×
↓	1		
$\langle 4, 1 \rangle$	——	$\frac{1}{36}, \frac{1}{9}$	×
↓	2		
$\langle 2, 2, 1 \rangle$	——	$\frac{1}{32}, \frac{1}{32}, \frac{1}{16}$	✓
↓	1		
$\langle 2, 1, 1, 1 \rangle$	——	$\frac{1}{50}, \frac{1}{25}, \frac{1}{25}, \frac{1}{25}$	×
↓	1		
$\langle 1, 1, 1, 1, 1 \rangle$	——	$\frac{1}{36}$	✓

Table 4

$\langle 2, 2, 1 \rangle$ associated with per firm profit of $\frac{1}{32}, \frac{1}{32}$ and $\frac{1}{16}$ is the coarsest EBA[16]. $\langle 2, 1, 1, 1 \rangle$, on the other hand is not an equilibrium coalition structure because the size-2 coalition will have incentive to break apart. $\langle 2, 2, 1 \rangle$ is an EBA because if a coalition of size 2 breaks apart and induces $\langle 2, 1, 1, 1 \rangle$, the other coalition of size

[16]To be more precise, all coalition structures with 2 size-2 coalitions and 1 singleton are EBAs.

2 will also break apart as we already argued and they will end up at the finest structure $\langle 1,1,1,1,1\rangle$ which makes the members of the first doubleton worse off. Coalition structure $\langle 4,1\rangle$ is not an EBA since a coalition of size 2 will break away to induce $\langle 2,2,1\rangle$ which is an EBA and hence a credible deviation. Similarly, coalition structure $\langle 5\rangle$ is not an EBA since one member will break away to temporarily induce $\langle 4,1\rangle$ and then a coalition of size 2 will break away to induce $\langle 2,2,1\rangle$. Observe that the leading perpetrator 1 receives $\frac{1}{16}$ under $\langle 2,2,1\rangle$ and $\frac{1}{20}$ under $\langle 5\rangle$. Similarly, $\langle 3,2\rangle$ and $\langle 3,1,1\rangle$, that are omitted from the above table for simplicity, are not EBAs.

Observe that none of the EBAs of the above game is strongly efficient[17]. However, there exists a strongly efficient EEBA. Indeed, the grand coalition alone, with payoff of $\frac{1}{20}$ per firm, constitutes a vN-M stable set of $(\mathcal{P},\gg)$. Nevertheless, this example admits other EEBAs that are not strongly efficient: Each permutation of the $\langle 2,2,1\rangle$ coalition structure constitutes an EEBA as well, since $\langle 2,2,1\rangle$ is a Pareto efficient partition that satisfies both conditions (a) and (b) of Proposition 1 below.

Proposition 2 *Let $P^* \in \mathcal{P}$ be Pareto efficient. P^* is an EEBA if*

(a) $\{1,2,\cdots,n\} \prec_N P^*$ and

(b) for all $P \in \mathcal{P}$ such that $P \neq P^*$ and $P \neq \{1,2,\cdots,n\}$, there is a coalition $S \in P$ such that $|S| > 1$ and $P \prec_i P^*$ for some $i \in S$.

Proof. We shall show that $\{P^*\}$ is a vN-M stable set of $(\mathcal{P},\gg)$. Obviously, $\{P^*\}$ is internally stable. We now need to show that $P \ll P^*$ for all $P \in \mathcal{P} \setminus P^*$. First, note that $\{1,2,\cdots,n\} \xrightarrow{N} P^*$. Since $\{1,2,\cdots,n\} \prec_N P^*$, we have $\{1,2,\cdots,n\} \ll P^*$. Let $P^1 \in \mathcal{P}$ be such that $P^1 \neq P^*$ and $P^1 \neq \{1,2,\cdots,n\}$. By condition (b), there is a coalition $S_1 \in P^1$ such that $|S_1| > 1$ and $P^1 \prec_{i_1} P^*$ for some $i_1 \in S_1$. Thus, $P^1 \xrightarrow{i_1} P^2 \equiv \{i_1, S_1 \setminus \{i_1\}, P^1 \setminus S_1\}$. Again by condition (b), there is a coalition $S_2 \in P^2$ such that $|S_2| > 1$ and $P^2 \prec_{i_2} P^*$ for some $i_2 \in S_2$. Thus, $P^2 \xrightarrow{i_2} P^3 \equiv \{i_2, S_2 \setminus \{i_2\}, P^2 \setminus S_2\}$. Continuing in this fashion, we can identify a sequence of agents $i_1, i_2, \ldots, i_k$ such that $P^1 \xrightarrow{i_1} P^2 \xrightarrow{i_2} P^3 \xrightarrow{i_3} \cdots \xrightarrow{i_k} \{1,2,\cdots,n\}$ and $P^{i_\ell} \prec_{i_\ell} P^*$ for all $\ell = 1,\ldots,k$. Given that $\{1,2,\cdots,n\} \xrightarrow{N} P^*$ and $\{1,2,\cdots,n\} \prec_N P^*$, we have $P^1 \ll P^*$. ■

In a symmetric game, the grand coalition constitutes an EEBA under similar conditions.

Corollary 2 *In a symmetric game $\{N\}$ is an EEBA if*

[17] They are, however, Pareto efficient.

(i) $\{N\}$ *is Pareto efficient,*

(ii) $\langle 1, 1, \cdots, 1\rangle$ *is not Pareto efficient and*

(iii) for all $P \in \mathcal{P}$ *such that* $P \neq \{N\}$ *and* $P \neq \langle 1, 1, \cdots, 1\rangle$ *there is a coalition* $S \in P$ *such that* $|S| > 1$ *and* $P \prec_i \{N\}$ *for some* $i \in S$.

An alternative corollary applies when agents' payoffs are comparable.

Corollary 3 *Consider a symmetric game* $(N, \{u_i\}_{i\in N})$ *where agents' payoffs are comparable.* $\{N\}$ *is an EEBA if*

(i) for all $P, P' \in \mathcal{P}$, *if* $|P| > |P'|$ *then* $\sum_{i\in N} u_i(P) < \sum_{i\in N} u_i(P')$ *and*

(ii) for all $P \in \mathcal{P}$, *if* $S, T \in P$ *and* $|S| > |T|$, *then* $u_i(P) < u_j(P)$ *for* $i \in S$ *and* $j \in T$.

Condition (i) in Corollary 3 states that as coalition structures become coarser aggregate payoff increases. Condition (ii) in Corollary 3 states that in a given coalition structure smaller coalitions yield higher per member payoffs, implying thus, that coalition formation has positive externalities. Games with positive externalities were defined in Yi (1997) and one property is that when any two coalitions merge, others benefit. Note that Yi's definition of positive externalities is stronger than conditions (i) and (ii) in Corollary 3.

The above results apply to the symmetric oligopoly model studied earlier in this section as well as to a public good economy [see Ray and Vohra (1997) and Yi (1997)].

5.2 A Counter Example

In the previous section, we identified sufficient conditions for a game to admit an efficient EEBA. However, as the example in Table 5 illustrates, it is possible for a game to have only inefficient EEBAs. This result reinforces the inefficiency puzzle posed by Ray and Vohra (1997).

Observe that the first row is Pareto efficient while every other cell is inefficient. In particular, the first entry in each column Pareto dominates all other entries in the same column.

Table A in the appendix shows how A (indirectly) dominates all other outcomes except D and d. Therefore, we cannot construct a stable set containing A alone since it cannot account for the exclusion of D and d. If A is included in an stable set, according to internal stability none of the outcomes it dominates can be included in the same stable set. Moreover, D and d cannot be included in the same stable set either, since they (indirectly) dominate A as illustrated in Table A. By the symmetry of the game the same arguments extend to all the other efficient outcomes B, C, D and E. Hence no stable set can support (contain) efficient outcomes.

A $\{512, 34\}$ 19,27,32,17,22	B $\{123, 45\}$ 19,27,32,17,22	C $\{234, 51\}$ 19,27,32,17,22	D $\{345, 12\}$ 19,27,32,17,22	E $\{451, 23\}$ 19,27,32,17,22
a $\{512, 3, 4\}$ 18,26,31,16,21	b $\{123, 4, 5\}$ 18,26,31,16,21	c $\{234, 5, 1\}$ 18,26,31,16,21	d $\{345, 1, 2\}$ 18,26,31,16,21	e $\{451, 2, 3\}$ 18,26,31,16,21
f $\{5, 12, 3, 4\}$ 15,25,30,15,20	g $\{1, 23, 4, 5\}$ 15,25,30,15,20	h $\{2, 34, 5, 1\}$ 15,25,30,15,20	i $\{3, 45, 1, 2\}$ 15,25,30,15,20	j $\{4, 51, 2, 3\}$ 15,25,30,15,20
k $\{51, 23, 4\}$ 15,25,30,15,20	ℓ $\{12, 34, 5\}$ 15,25,30,15,20	m $\{23, 45, 1\}$ 15,25,30,15,20	n $\{34, 51, 2\}$ 15,25,30,15,20	o $\{45, 12, 3\}$ 15,25,30,15,20
p $\{5, 14, 2, 3\}$ 15,25,20,30,15	q $\{1, 25, 3, 4\}$ 15,25,20,30,15	r $\{2, 31, 4, 5\}$ 15,25,20,30,15	s $\{3, 42, 5, 1\}$ 15,25,20,30,15	t $\{4, 53, 1, 2\}$ 15,25,20,30,15
u $\{5, 142, 3\}$ 15,25,20,30,15	v $\{1, 253, 4\}$ 15,25,20,30,15	w $\{2, 314, 5\}$ 15,25,20,30,15	x $\{3, 425, 1\}$ 15,25,20,30,15	y $\{4, 531, 2\}$ 15,25,20,30,15
		z $\{1, 2, 3, 4, 5\}$ 15,15,15,15,15		

Table 5

Next we argue that $\mathcal{Q} = \{k, \ell, m, n, o, z\}$ is stable. For external stability we need to show that all outcomes not in $\mathcal{Q}$ are (indirectly) dominated by some outcome in $\mathcal{Q}$. Table B in the appendix lists all such paths of dominance. To prove the internal stability of $\mathcal{Q}$ we have to show that no element of $\mathcal{Q}$ indirectly dominates another element of $\mathcal{Q}$. We start with k and ℓ. First we argue that $k \not\gg \ell$. Note that $k \succ_{124} \ell$ and from ℓ coalition 124 and its subsets can induce u, f, p, h, s and z. But, u, f and p have the same payoffs for all players as k does, thus, once at u, f or p, no sequence will initiate with destination k. Note that although $k \succ_{12} h$ from h coalition 12 can only induce ℓ. The following table summarizes the above information:

$$
\begin{array}{llll}
\ell \xrightarrow{124} & u \underset{N}{\sim} k & & \\
\ell \xrightarrow{124} & f \underset{N}{\sim} k & & \\
\ell \xrightarrow{124} & p \underset{N}{\sim} k & & \\
\ell \xrightarrow{12} & h \underset{12}{\prec} k & h \xrightarrow{12} & \ell
\end{array}
$$

Table 6

We still have to show that s and z will not lead to k. Observe that $k \succ_{12} s$ and from s coalition 12 can induce f and z. While f does not lead to k, z is still to be checked. But $k \succ_{124} z$ and from z coalition 124 can induce u, f, p and s which are all examined already. The following table summarizes this information:

$$
\begin{array}{llll}
\ell \xrightarrow{124} & s \underset{12}{\prec} k & s \xrightarrow{12} & f \underset{N}{\sim} k \\
 & {}_{12} \swarrow & & \\
\ell \xrightarrow{124} & z \underset{124}{\prec} k & z \xrightarrow{124} & u \underset{N}{\sim} k \\
 & & z \xrightarrow{124} & f \underset{N}{\sim} k \\
 & & z \xrightarrow{124} & p \underset{N}{\sim} k \\
 & & z \xrightarrow{124} & s \not\ll k
\end{array}
$$

Table 7

Next we show that $\ell \not\gg k$ in Table C in the appendix. Similarly, we show that k does not dominate and is not dominated by o, m, and n, respectively, by outlining the blocked paths in Tables D, E and F in the appendix. Moreover, it is already shown, in Table 7 above, that $k \not\gg z$. It is also easy to see that z cannot indirectly dominate k, ℓ, m, n and o since no player strictly prefers z to any of them. The rest of the proof of internal stability follows from symmetry.

6 Transfers and (In)Efficiency

When intra-coalitional transfers are possible, there are two ways of specifying a game with externalities. One is to start with our simple coalition formation game with externalities $(N, \{u_i\}_{i \in N})$ and then allow for transfers, thereby treating $u_i(P)$ as the "primitive payoff" of agent i for a given coalition structure P. For example, when two firms with convex cost functions collude and maximize their joint profits, their individual outputs and profits are also determined. In this case, transfers enable them to reallocate their "primitive profit". The second way of specifying a game is to use the partition function game of Thrall and Lucas (1963) with the joint payoff of each coalition in a coalition structure as the primitive. The main example (Table 9) in this section can be specified in either way and our main conclusion remains valid. For convenience, however, we shall follow the first approach.

It is easy to construct examples where transfer payments within coalitions are necessary to achieve strongly efficient agreements. Consider a simple example with $N = \{1, 2\}$ and the following payoffs.

$\{12\}$ $10, 5$
$\{1, 2\}$ $6, 6$

Table 8

$\{1, 2\}$ is the only EBA and EEBA and it is Pareto efficient but not strongly efficient. If transfers are possible, then there exist strongly efficient EEBAs. For example, $\{12\}$ with $(8, 7)$ payoff allocation constitutes an EEBA.

When intra-coalitional transfers are possible, each agreement comprises two components: a coalition structure and a payoff allocation feasible for this coalition structure as defined in Section 5.1. Let (P, x) be an agreement (i.e., $P \in \mathcal{P}$ and x is feasible for P). If a coalition S deviates from this agreement, it can re-partition itself as in Section 4 and at the same time agree on how to share the proceeds within each of its coalitions, but it remains to be specified how the payoffs of agents in $N \setminus S$ are affected. One approach is to reset these payoffs to their primitive levels as specified by $(N, \{u_i\}_{i \in N})$. The second approach is to allow each coalition to make *conditional statements* with respect to transfers in the sense of Ray and Vohra (1999) – each coalition, upon formation, specifies transfers to be made for every coalition structure that contains it. In other words, when a coalition forms, its members not only agree on how to share their present payoff but also agree on a *contingency plan* in the event their total payoff changes due to externalities imposed by the restructuring of its complement. For simplicity, we shall follow the first approach here. Our main conclusion regarding the example in Table 9 still holds even if we take the second approach.

Transfers do facilitate the attainment of efficient agreement. For the counter example in Table 5, the following is a stable set for $(\mathcal{P}, \gg)$:

$$\{(\{512, 34\}, 30, 30, 18, 21, 18), (\{234, 51\}, 18, 30, 30, 18, 21)\}.$$

However, there is also a stable set with only inefficient elements. In fact, $\{k, l, m, n, o\}$ with the original payoffs remains a stable set with transfers. Some alternative assumptions as to what ensues a deviation also yield efficient stable sets.

It is then natural to ask the following question: Does there exist at least one efficient agreement for every game when transfers are possible? Unfortunately, as the next example illustrates, it is possible for a game to admit only inefficient agreements (i.e., no stable set contains efficient elements).

Consider a game with 5 players, $N = \{i, j, k, l, m\}$. In the table below are the payoff vectors associated with 21 partitions and all other partitions are assumed to yield 0 payoff vectors.

$\{12345\}$	$\{ijk, lm\}$	$\{ijk, l, m\}$
$10, 10, 10, 10, 10$	$15, 15, 15, 1, 1$	$15, 15, 15, 1, 1$

Table 9

It is easy to see that Pareto efficiency requires the formation of the grand coalition. However, the example admits the following stable set:

$$\{(\{ijk, lm\}, 15, 15, 15, 1, 1), (\{ijk, l, m\}, 15, 15, 15, 1, 1) \,|\text{for all distinct } i, j, k, l, m\}.$$

Note that all elements in the stable set are inefficient. Moreover, as the following Proposition establishes there does not exist a stable set that contains efficient agreements.

Proposition 3 *The above example does not admit an efficient EEBA if transfers are possible.*

Proof. The proof is conducted in several steps. Let Σ be a stable set for $(\mathcal{P}, \gg)$.

Lemma 4 *Let $(P, x) \in \Sigma$. Then there does not exist an agent $i \in N$ such that $x_i \leq 0$.*

First we show that if (P, x) is such that $x_i \leq 0$ for some $i \in N$, then there does not exist (Q, y) such that $y_k \geq 0$ for all $k \in N$ and $(Q, y) \ll (P, x)$. There are two cases. First, if P is a partition with 0 payoff vector (that is, P is not one of the partitions in Table 9), there does not exist $S \subset N$ such that $(Q, y) \prec_S (P, x)$ and therefore no coalition will initiate a sequence from (Q, y) to (P, x). If P is a partition in Table 9, i must have participated actively in the inducement of (P, x)

from (Q, y) for otherwise, i gets a payoff of at least 1. Now, let S be the first deviating coalition containing i in this process and suppose that (Q', y') was the status quo when S deviated. Then $y'_i \geq 0$ and $(Q', y') \prec_S (P, x)$, which imply that $x_i > 0$, a contradiction.

Now, we proceed to show that if (P, x) is such that $x_i \leq 0$ for some $i \in N$, then $(P, x) \notin \Sigma$. To this end, let (Q, y) be such that $(P, x) \xrightarrow{i} (Q, y)$ and assume in negation that $(P, x) \in \Sigma$. Then the payoff structure of the game implies that either $y_j = 0$ for all $j \in N$ or $y_i = 1$. If $y_j = 0$ for all $j \in N$, then we claim $(Q, y) \notin \Sigma$. Indeed, if $(Q, y) \in \Sigma$, then by internal stability, Σ does not contain any (Q', y') such that $y'_j > 0$ for all $j \in N$. Then by the result in the preceding paragraph, Σ is not externally stable. Thus, $(Q, y) \notin \Sigma$. Then there exists $(P', x') \in \Sigma$ such that $x'_j > 0$ for all $j \in N$ and $(P', x') \gg (Q, y)$. Since $(P, x) \xrightarrow{i} (Q, y)$ and $x_i \leq 0$ but $x'_i > 0$, we have $(P', x') \gg (P, x)$, a contradiction. Now, consider the case that $y_i = 1$. By internal stability, $(Q, y) \notin \Sigma$ since $(P, x) \ll (Q, y)$. This leads to a similar contradiction as the previous case.

Lemma 5 *Σ cannot contain two elements that are efficient and give strictly positive payoff to each agent.*

Let (N, x) and (N, y) be two elements in Σ. Then, there exists i such that $x_i \neq y_i$. Without loss of generality, let $y_i > x_i$. Then, $(N, y) \gg (N, x)$ by the following sequence: $(N, x) \xrightarrow{i} (P, z) \xrightarrow{N} (N, y)$ where $z_j = 0$ for $j \in N$.

Lemma 6 *A single efficient element is not externally stable.*

Assume in negation that (N, x) is the only element in Σ. By external stability, $(N, x) \gg \{ijk, l, m\}$. Then, some player among i, j and k gets more than 15 in (N, x). Without loss of generality, let $x_i > 15$. Now consider $\{jkl, m, i\}$, it must also be the case that $(N, x) \gg \{jkl, m, i\}$. Then some player among j, k and l gets more than 15 in (N, x). Without loss of generality, let $x_j > 15$. Finally consider $\{klm, i, j\}$ again it must be the case that $(N, x) \gg \{klm, i, j\}$ and $x_k > 15$. Thus, $x_l + x_m < 5$. Now consider $(\{klm, i, j\}, y)$ with $y_k > x_k, y_l > x_l$, and $y_m > x_m$. Such an allocation is feasible because $x_i > 15$ and $x_j > 15$ (so $x_k < 20$). It is easy to see that $(\{klm, i, j\}, y) \not\ll (N, x)$.

Now we conclude the proof of Proposition 3 by showing that Σ does not contain any efficient element. Assume otherwise that $(N, x) \in \Sigma$. By Lemmata 4, 5 and 6, we know that Σ contains (P, y) such that $P \neq N$ and $y_j > 0$ for all $j \in N$ and

$x_j > 0$ for all $j \in N$. Obviously, $P = \{ijk, lm\}$ or $P = \{ijk, l, m\}$. By internal stability it must be the case that $x_i = y_i, x_j = y_j$, and $x_k = y_k$. Then $x_l + x_m = 5$. Assume w.l.o.g. that $x_k = \min\{x_i, x_j, x_k\}$ and consider $(\{klm, i, j\}, z) \equiv (Q, z)$ with $z_k > x_k, z_l > x_l$, and $z_m > x_m$. Then, $(N, x) \ll (Q, z)$. By internal stability, $(Q, z) \notin \Sigma$. By external stability, there exists $(Q', z') \in \Sigma$ such that $(Q', z') \gg (Q, z)$. Consequently, some agent, say i', in $\{klm\}$ strictly prefers (Q', z') to (Q, z), i.e., $z'_{i'} > z_{i'}$. It follows that $z'_{i'} > x_{i'}$. Then $(Q', z') \gg (N, x)$ by the following sequence: $(N, x) \overset{i'}{\to} (P', x') \overset{N}{\to} (Q', z')$ where $x'_j = 0$ for all $j \in N$. ■

Lastly, the following example, a perturbed roommate problem, illustrates how transfers *across* coalitions may be necessary to achieve strong efficiency.

Structure	Per-person Payoff	EEBA
$\langle 3 \rangle$	4	×
$\langle 2, 1 \rangle$	5, 11	×
$\langle 1, 1, 1 \rangle$	6	✓

Table 10

$\langle 1, 1, 1 \rangle$ is the only EBA and EEBA; it is Pareto efficient but not strongly efficient. Inter-coalitional transfers are necessary to achieve strong efficiency in this case. A similar phenomenon was identified as beneficial altruism in Greenberg (1980) and Dréze and Greenberg (1980).

7 Conclusion

In this paper we extend Ray and Vohra's (1997) notion of EBA to allow for arbitrary coalitional deviations. We show that despite this extension, the new notion, EEBA, while facilitating the attainment of efficient agreements, does not guarantee efficiency. In particular, we identify sufficient conditions for a game to admit an efficient EEBA and construct a counter example where none of the EEBAs is efficient. Moreover, we illustrate that the introduction of transfers does not guarantee efficiency either. A second counter example is provided where there is no efficient EEBA, despite the possibility of transfers. The negative results strengthen Ray and Vohra's (1997) inefficiency puzzle. Therefore, this subject warrants further study to explore solution concepts and negotiation processes, among agents who exercise their decision power freely, that lead to efficient agreements.

8 References

1. Banerjee, S., Konishi, H. & Sonmez, T. (2001). "Core in a Simple Coalition Formation Game," *Social Choice and Welfare* **18**, 135-153.

2. Barbera, S., Gerber, A. (2003). "On Coalition Formation: Durable Coalition Structures", *Mathematical Social Sciences* **45**, 185-203.

3. Bernheim, D., Peleg, B. and Whinston, M. (1987). "Coalition-Proof Nash Equilibria" *Journal of Economic Theory* **42**, 1-12.

4. Bloch, F. (1996). "Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division," *Games and Economic Behavior* **14**, 90-123.

5. Bogomolnaia, A. & Jackson M.O. (2002). "The Stability of Hedonic Coalition Structures," *Games and Economic Behavior* **38**, 201-230.

6. Chwe, M. S.-Y. (1994). "Farsighted Coalitional Stability," *Journal of Economic Theory* **63**, 299-325.

7. Coase, R. (1960). "The Problem of Social Cost", *Journal of Law and Economics* **3,** 1-44.

8. Diamantoudi, E. & Xue, L. (2003). "Farsighted Stability in Hedonic Games," *Social Choice and Welfare* **21**, 39-61.

9. Dreze, J. & Greenberg, J. (1980). "Hedonic Coalitions: Optimality and Stability," *Econometrica* **48**, 987-1003.

10. Greenberg, J. (1980). "Beneficial Altruism," *Journal of Economic Theory* **22**, 12-22.

11. Greenberg, J. (1989). "Deriving Strong and Coalition-Proof Nash Equilibrium from an Abstract System", *Journal of Economic Theory* **49**, 195–202.

12. Greenberg, J. (1990). *The Theory of Social Situations – An Alternative Game-theoretic Approach*, Cambridge University Press.

13. Harsanyi, J.C. (1974). "Interpretation of Stable Sets and a Proposed Alternative Definition," *Management Science* **20**, 1472-1495.

14. Macho-Stadler, I., Perez-Castrillo, D. & Porteiro, N. (2002). "Sequential Formation of Coalitions through Bilateral Agreements," mimeo.

15. Miller, N. R. (1980). "A New Solution Set for Tournaments and Majority Voting: Further Graph Theoretical Approaches to the Theory of Voting" *American Journal of Political Science* **24**, 68-96.

16. Ray, D. (1989). "Credible Coalitions and the Core", *International Journal of Game Theory*, **18**, 185-187.

17. Ray, D. and Vohra, R. (1997). "Equilibrium Binding Agreements," *Journal of Economic Theory* **73**, 30-78.

18. Ray, D. and Vohra, R. (1999). "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior* **26**, 286-336.

19. Ray, D. and Vohra, R. (2001). "Coalitional Power and Public Goods", *Journal of Political Economy* **109**, 1355-1384.

20. Thrall, R. M. and Lucas, W. F. (1963). "N-person Games in Partition Function Form," *Naval Research and Logistics* **10**, 281-297.

21. von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*, Princeton University Press.

22. Xue, L. (1998). "Coalitional Stability under Perfect Foresight," *Economic Theory* **11**, 603-627.

23. Yi, S. (1997). "Stable Coalition Structures with Externalities," *Games and Economic Behavior* **20**, 201-237.

24. Yi, S. & Shin H. S. (1995). "Endogenous Formation of Coalitions in Oligopoly," mimeo.

25. Zhao, J. (1992). "The Hybrid Solutions of an N-Person Game," *Games and Economic Behavior* **4**, 145-160.

9 Appendix

Table A

$A \gg a$	$: a \xrightarrow{N} A$	$A \gg B$	$: B \xrightarrow{2} \phi_1 \xrightarrow{N} A$
$A \gg f$	$: f \xrightarrow{N} A$	$A \gg b$	$: b \xrightarrow{12} f \xrightarrow{N} A$
$A \gg k$	$: k \xrightarrow{N} A$	$A \gg g$	$: g \xrightarrow{12} f \xrightarrow{N} A$
$A \gg p$	$: p \xrightarrow{N} A$	$A \gg \ell$	$: \ell \xrightarrow{4} f \xrightarrow{N} A$
$A \gg u$	$: u \xrightarrow{N} A$	$A \gg q$	$: q \xrightarrow{12} f \xrightarrow{N} A$
$A \gg z$	$: z \xrightarrow{N} A$	$A \gg v$	$: v \xrightarrow{124} \phi_2 \xrightarrow{N} A$
$A \gg C$	$: C \xrightarrow{125} A$	$A \gg E$	$: E \xrightarrow{4} k \xrightarrow{N} A$
$A \gg c$	$: c \xrightarrow{125} A$	$A \gg e$	$: e \xrightarrow{24} \phi_3 \xrightarrow{N} A$
$A \gg h$	$: h \xrightarrow{125} A$	$A \gg j$	$: j \xrightarrow{24} \phi_3 \xrightarrow{N} A$
$A \gg m$	$: m \xrightarrow{125} a \xrightarrow{N} A$	$A \gg o$	$: o \xrightarrow{4} f \xrightarrow{N} A$
$A \gg r$	$: r \xrightarrow{125} a \xrightarrow{N} A$	$A \gg t$	$: t \xrightarrow{24} \phi_8 \xrightarrow{N} A$
$A \gg w$	$: w \xrightarrow{125} A$	$A \gg y$	$: y \xrightarrow{24} \phi_4 \xrightarrow{N} A$
$A \gg i$	$: i \xrightarrow{13} \phi_1 \xrightarrow{N} A$	$A \not\gg D$	while
$A \gg n$	$: n \xrightarrow{123} f \xrightarrow{N} A$	$D \gg A$	$: A \xrightarrow{345} D$
$A \gg s$	$: s \xrightarrow{12} f \xrightarrow{N} A$	$A \not\gg d$	while
$A \gg x$	$: x \xrightarrow{13} \phi_5 \xrightarrow{N} A$	$d \gg A$	$: A \xrightarrow{45} o \xrightarrow{2} i \xrightarrow{N} d$

Table B

$A \ll \ell$	$: A \xrightarrow{5} \ell$	$B \ll m$	$: B \xrightarrow{1} m$
$a \ll \ell$	$: a \xrightarrow{35} \phi_6 \xrightarrow{N} \ell$	$b \ll m$	$: b \xrightarrow{14} \phi_9 \xrightarrow{N} m$
$f \ll \ell$	$: f \xrightarrow{35} \phi_6 \xrightarrow{N} \ell$	$g \ll m$	$: g \xrightarrow{14} \phi_9 \xrightarrow{N} m$
$p \ll \ell$	$: p \xrightarrow{35} \phi_7 \xrightarrow{N} \ell$	$q \ll m$	$: q \xrightarrow{14} \phi_{10} \xrightarrow{N} m$
$u \ll \ell$	$: u \xrightarrow{35} \phi_2 \xrightarrow{N} \ell$	$v \ll m$	$: v \xrightarrow{14} \phi_{11} \xrightarrow{N} m$
$C \ll n$	$: C \xrightarrow{2} n$	$D \ll o$	$: D \xrightarrow{3} o$
$c \ll n$	$: c \xrightarrow{25} \phi_{12} \xrightarrow{N} n$	$d \ll o$	$: d \xrightarrow{13} \phi_1 \xrightarrow{N} o$
$h \ll n$	$: h \xrightarrow{25} \phi_{12} \xrightarrow{N} n$	$i \ll o$	$: i \xrightarrow{13} \phi_1 \xrightarrow{N} o$
$r \ll n$	$: r \xrightarrow{25} \phi_{13} \xrightarrow{N} n$	$s \ll o$	$: s \xrightarrow{13} \phi_{15} \xrightarrow{N} o$
$w \ll n$	$: w \xrightarrow{25} \phi_{14} \xrightarrow{N} n$	$x \ll o$	$: x \xrightarrow{13} \phi_5 \xrightarrow{N} o$
$E \ll k$	$: E \xrightarrow{4} k$	$t \ll k$	$: t \xrightarrow{24} \phi_8 \xrightarrow{N} k$
$e \ll k$	$: e \xrightarrow{24} \phi_3 \xrightarrow{N} k$	$y \ll k$	$: y \xrightarrow{24} \phi_4 \xrightarrow{N} k$
$j \ll k$	$: j \xrightarrow{24} \phi_3 \xrightarrow{N} k$		

Let ϕ_i, where $i = 1, ..., 15$, represent coalition structures that do not appear in Table 5 and whose payoff is 0 for all the players. Obviously all ϕ_i's are directly dominated by any element of $\mathcal{Q}$. The ϕ_i's are provided in the table below.

ϕ_1 ={13,45,2}	ϕ_4 ={135,24}	ϕ_7 ={14,35,2}	ϕ_{10} ={14,25,3}	ϕ_{13} ={13,25,4}
ϕ_2 ={124,53}	ϕ_5 ={245,13}	ϕ_8 ={24,35,1}	ϕ_{11} ={235,14}	ϕ_{14} ={134,25}
ϕ_3 ={24,51,3}	ϕ_6 ={12,35,4}	ϕ_9 ={14,23,5}	ϕ_{12} ={34,25,1}	ϕ_{15} ={13,24,5}

Table C

$\ell \not\ggg k$ while $\ell \succ_{35} k$	$k \xrightarrow{5}$	$g \underset{N}{\sim} \ell$		
	$k \xrightarrow{3}$	$j \underset{23}{\prec} \ell$	$j \xrightarrow{23}$	k
	$k \xrightarrow{35}$	$t \underset{23}{\prec} \ell$	$t \xrightarrow{23}$	$g \underset{N}{\sim} \ell$
		${}_{23}\swarrow$		
	$k \xrightarrow{35}$	$z \underset{235}{\prec} \ell$	$z \xrightarrow{235}$	$g \underset{N}{\sim} \ell$
				$q \underset{N}{\sim} \ell$
				$v \underset{N}{\sim} \ell$
				$t \not\lll \ell$

Table D

$k \not\ggg o$ while $k \succ_{24} o$	$o \xrightarrow{4}$	$f \underset{N}{\sim} k$		
	$o \xrightarrow{24}$	$z \not\lll k$	see Table 7	
	$o \xrightarrow{24}$	$s \not\lll k$	see Table 7	
	$o \xrightarrow{2}$	$i \underset{12}{\prec} k$	$i \xrightarrow{12}$	o
$o \not\ggg k$ while $o \succ_{135} k$	$k \xrightarrow{3}$	$j \underset{N}{\sim} o$		
	$k \xrightarrow{135}$	$y \underset{N}{\sim} o$		
	$k \xrightarrow{135}$	$t \underset{N}{\sim} o$		
	$k \xrightarrow{15}$	$g \underset{15}{\prec} o$	$g \xrightarrow{15}$	k
	$k \xrightarrow{135}$	$r \underset{15}{\prec} o$	$r \xrightarrow{15}$	$j \underset{N}{\sim} o$
		${}_{15}\swarrow$		
	$k \xrightarrow{135}$	$z \underset{135}{\prec} o$	$z \xrightarrow{135}$	$y \underset{N}{\sim} o$
			$z \xrightarrow{135}$	$j \underset{N}{\sim} o$
			$z \xrightarrow{135}$	$t \underset{N}{\sim} o$
			$z \xrightarrow{135}$	$r \not\lll o$

Table E

$k \not\gg m$	$m \xrightarrow{12}$	$o \underset{24}{\prec} k$	$o \not\ll k$	see Table D		
while $k \succ_{12} m$	$m \xrightarrow{12}$	$i \underset{12}{\prec} k$	$i \xrightarrow{12}$	o		
$m \not\gg k$	$k \xrightarrow{34}$	$n \underset{134}{\prec} m$	$n \xrightarrow{134}$	$w \underset{N}{\sim} m$		
while $m \succ_{34} k$			$n \xrightarrow{134}$	$r \underset{N}{\sim} m$		
			$n \xrightarrow{134}$	$h \underset{N}{\sim} m$		
			$n \xrightarrow{34}$	$j \underset{34}{\prec} m$	$j \xrightarrow{34}$	n
			$n \xrightarrow{134}$	$p \underset{34}{\prec} m$	$p \xrightarrow{34}$	$h \underset{N}{\sim} m$
				${}_{34}\swarrow$		
			$n \xrightarrow{134}$	$z \underset{134}{\prec} m$	$z \xrightarrow{134}$	$w \underset{N}{\sim} m$
						$r \underset{N}{\sim} m$
						$h \underset{N}{\sim} m$
						$p \not\ll m$
	$k \xrightarrow{34}$	$j \underset{34}{\prec} m$	$j \xrightarrow{34}$	$n \not\ll m$		

Table F

$k \not\gg n$	$n \xrightarrow{12}$	$\ell \underset{124}{\prec} k$	$\ell \not\ll k$	Tables 6 & 7		
while $k \succ_{12} n$	$n \xrightarrow{12}$	$h \underset{12}{\prec} k$	$h \xrightarrow{12}$	$\ell \not\ll k$	Tables 6 & 7	
$n \not\gg k$	$k \xrightarrow{45}$	$m \underset{25}{\prec} n$	$m \xrightarrow{25}$	$i \underset{N}{\sim} n$		
while $n \succ_{45} k$			$m \xrightarrow{25}$	$g \underset{45}{\prec} n$	$g \xrightarrow{45}$	m
			$m \xrightarrow{25}$	$q \underset{45}{\prec} n$	$q \xrightarrow{45}$	$i \underset{N}{\sim} n$
				${}_{45}\swarrow$		
			$m \xrightarrow{25}$	$z \underset{245}{\prec} n$	$z \xrightarrow{245}$	$x \underset{N}{\sim} n$
					$z \xrightarrow{245}$	$s \underset{N}{\sim} n$
					$z \xrightarrow{245}$	$i \underset{N}{\sim} n$
					$z \xrightarrow{245}$	$q \not\ll n$
	$k \xrightarrow{45}$	$g \underset{45}{\prec} n$	$g \not\ll n$			