DEPARTMENT OF ECONOMICS

Working Paper

EQUILIBRIUM BINDING AGREEMENTS
UNDER DIVERSE BEHAVIORAL ASSUMPTIONS

Effrosyni Diamantoudi

Working Paper No. 2001-2
Centre for Dynamic Modelling in Economics

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

CENTRE FOR DYNAMIC MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

WORKING PAPER

EQUILIBRIUM BINDING AGREEMENTS UNDER DIVERSE BEHAVIORAL ASSUMPTIONS

Effrosyni Diamantoudi

Working Paper No. 2001-2

DEPARTMENT OF ECONOMICS

SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Equilibrium Binding Agreements under Diverse Behavioral Assumptions

Effrosyni Diamantoudi
Department of Economics, University of Aarhus*

February 2001

Abstract

This paper extends the work of Ray and Vohra (1997). It ascertains which partitions of players will emerge and what actions will these players choose under each partition, when they can sign binding agreements and their actions have externalities. The emphasis, however, is placed on situations with multiple outcomes and how agents behave in the presence of such multiplicity. In particular, a deviating coalition considers all the likely outcomes that may prevail upon its deviation, and selects (if possible) a subset of them, avoiding thus an excessively confident behavior, which is the case under Ray and Vohra's notion. Three augmentations of their solution concept are defined, capturing three distinct behavioral assumptions. Their efficiency and relation to each other are discussed. *Keywords*: coalition structure, binding agreement, stability. *Journal of Economic Literature Classification Numbers*: C70, C71.

1 Introduction

This paper studies situations with externalities (as captured by normal form games) where agents can negotiate over their course of actions. If any group of agents (coalition) reaches an agreement over the strategies its members will adopt, the agreement is binding. Our analysis ascertains the equilibrium coalition structures and the equilibrium agreements that will be signed by each coalition in the structure. A coalition structure describes how the entire set of agents partition themselves; by studying coalition structures, therefore, we explicitly take into account every agent's behavior, which is of particular importance when externalities are present.

The basic question we address is not new. A very important and influential work on binding agreements is that of Ray and Vohra (1997) and we would refer the reader to their work for a more detailed exposition of and motivation to the general topic of *binding*

**Mailing Address*: Department of Economics, University of Aarhus, Building 350, DK-8000 Aarhus C., Denmark. *E-mail*: faye@econ.au.dk. I wish to thank Licun Xue, Joseph Greenberg and the participants of PET 2000 for very useful comments. The paper has been previously circulated under the title *Binding Agreements.*

agreements. According to their analysis a coalition structure and an agreement are in equilibrium if the structure is immune to deviations. A coalition will deviate (withdraw from the negotiating table) if one specific partition (to which it belongs) and one specific equilibrium binding agreement associated with the partition make its members better off. Although the partition the coalition considers deviating to is viable -the complementary agents do not object to it,- it is not necessarily the only viable partition to emerge upon the coalitional deviation, and even more daringly, it may not be the most likely one. And even if the partition does emerge, the contemplated binding agreement, although credible, may not be the one to prevail if the partition is associated with more than one binding agreements.

In this paper, we extend the authors' work by offering a range of solutions with diverse behavioral assumptions. We maintain the supposition that a coalition structure and an agreement are in equilibrium if the structure is immune to deviations. We allow, however, the deviating coalition to be fearful or hopeful of what may ensue from its departure, both in term of the coalition structures likely to emerge and the agreements each structure may adopt, depending on the institutional assumptions governing the deviation as well as the agents' behavioral predisposition. In fact, the motivation for this work is best captured by the words of Ray and Vohra (1997):

> "Our definition of what a coalition can induce is based on an optimistic view of what transpires after the initial deviation. A leading perpetrator need only find some equilibrium binding agreement in some coalition structure induced by the act of its deviation. ... Clearly, there are alternatives to optimism. ... Thus versions of our definition are certainly possible that incorporate increasing degrees of pessimism, culminating in the requirement that a leading perpetrator must be better off in every equilibrium binding agreement of every coalition structure induced by it. However, this pessimistic version has a serious drawback. In many interesting cases where transfers of utility are possible within a coalition, a coalition may have a choice between several equilibria such that its complement is indifferent between all of them. It would then be unreasonable to assume that members of coalition should be so pessimistic as to focus on the least desirable of these equilibria for them. ... On the other hand, a degree of optimism that ignores the possible multiplicity of responses by players external to a coalition (in the sense of simply anticipating the coalition structure that is best for the leading perpetrator) is also open to criticism. A satisfactory definition based on pessimism will, therefore have to treat these two sets of issues differently. "

2 The model

We provide three augmentations to the original solution concept by Ray and Vohra (1997) depicting three different behavioral assumptions, namely optimism, pessimism and conservatism.

Recall the negotiation process adopted by Ray and Vohra (1997) and consider a strategic form game. The grand coalition, N, contemplates the Pareto optimal[1] outcomes, and if it reaches an agreement then its members sign a contract. If not, then it is so due to some coalition believing that it can do better by deviating. Once a coalition $S \subset N$ breaks away, no more communication or cooperation takes place with the rest of the players, $N \backslash S$. The coalition contemplates which strategies its members should select and if an agreement is reached it is binding. Note that the agreement among the members of S may be reached before they actually depart from N, thus upon their departure they "know" what strategy they are going to play. Whether the agreement is reached before or after the deviation is modeled explicitly by the different schemes we propose. In any event, no mergers are allowed in the negotiation process: once a group of players forms coalition S by deviating from the grand coalition they can only split further apart into subsets of S. By restricting possible deviations to only subsets we essentially limit the applicability of the theory to models describing such predicaments. We acknowledge, along with Ray and Vohra (1997), that although it is a common situation for those that leave the negotiating table to be excluded from further contact, it is certainly not always the case. Nevertheless, in permitting only finer coalition structures to evolve out of disagreements we can easily achieve consistency, in the sense that a deviating coalition, while contemplating its deviation, counts only on credible outcomes. This is the case since we can analyze a game recursively by starting from a totally disintegrated coalition structure and then proceed to coarser and coarser coalition structures. In this manner, any deviating coalition departs by counting only on likely and not any feasible events. We return to this point in the last section of the paper and discuss how the model can be modified to accommodate mergers.

The very assumption of lack of contact across coalitions once they form suggests that any play between them is noncooperative. Thus, we follow the same approach with both Ray and Vohra (1997) and Zhao (1996), in assuming a Nash-like play between coalitions, where every member S of the partition plays a best response strategy vector -no other strategy vector can strictly improve all the coalition members- given a certain strategy vector by $N \backslash S$.

Although agreements are binding, their signing is voluntary. Thus, for an agreement to be reached by a coalition it should assign to every subcoalition at least as much as that subcoalition can attain on its own by deviating[2]. Each coalition's guarantee varies according to the behavioral assumption imposed on the players, and further specifics on the issue are stipulated in the following sections that provide the formal definitions of the

[1] An outcome is considered to be Pareto optimal if there does not exist another outcome that strictly improves every player.

[2] The difference (at least in this work) between a deviating coalition and a cheating coalition is that the latter presumes that everything else (partitions and strategies) stays put, whereas the former expects that everyone else will fend for themselves, by adjusting their behavior to some equilibrium response. Since cheating can be excluded via the binding agreements we can concentrate on what subcoalitions can guarantee given that others will optimize their choices.

different cases.

2.1 Preliminaries

Consider a game in normal form $G = (N, \{X_i\}_{i\in N}, \{u_i\}_{i\in N})$ where $N = \{1, ..., n\}$ is the finite set of players, X_i is the strategy set of player i and $u_i : \prod_{i\in N} X_i \to \mathbb{R}$ is the payoff function of player i.

- A coalition S is a nonempty subset of N. For any $S \subset N$[3] let X_S denote $\prod_{i\in S} X_i$ and $x_S \equiv \{x_i\}_{i\in S} \in X_S$ be a strategy vector for coalition S. Similarly, for any $x \in X_N$ and $S \subset N$ let $u_S(x) \equiv \{u_i(x)\}_{i\in S}$ be the payoff vector of coalition S.

- A coalition structure is a partition of N, P such that $P = \{P_1, P_2, ..., P_k\}$, $\bigcup_{j=1}^{k} P_j = N$ and for all $i \neq j$, $P_i \bigcap P_j = \emptyset$. Let $\mathcal{P}$ be the collection of all coalition structures and $\mathcal{P}_S = \{P \in \mathcal{P} \mid S \in P,\ S \subset N\}$ be the collection of all coalition structures that contain coalition S. Similarly, P_S denotes that $S \in P$.

- Given a partition P, the set $\mathcal{R}(P)$ denotes all coalition structures that are refinements of P; that is, if $P' \in \mathcal{R}(P)$ then for every coalition $T \in P'$ there exists a coalition $S \in P$ such that $T \subset S$[4]. Similarly, a coalition structure P is coarser than a coalition structure P' if $P' \in \mathcal{R}(P)$. Lastly, a coalition structure P is a coarsest coalition structure relative to some set of coalition structures $\mathcal{P}' \subset \mathcal{P}$ if $\nexists P' \in \mathcal{P}'$ such that $P \neq P'$ and $P \in \mathcal{R}(P')$.

- A strategy profile $x \in X_N$ satisfies the best response property relative to a partition P if for every coalition $S \in P$ there does not exist a strategy vector $y_S \in X_S$ such that $u_i(y_S, x_{N\setminus S}) > u_i(x)$ for every $i \in S$. Let $\beta(P)$ be the set of such strategy profiles.

- Let $P \in \mathcal{P}$, $T \in P$ and $S \subsetneq T$. If S departs from partition P, then S's deviation can only bring about $\mathcal{P}_S(P) = \mathcal{R}(P) \cap \mathcal{P}_S$ collection of partitions. If $S \in P$ then $\mathcal{P}_S(P) \equiv \emptyset$. Let $B(S; P) = \{y \in X_N \mid y \in \bigcup_{P_S \in \mathcal{P}_S(P)} \beta(P_S)\}$ be the collection of all the feasible outcomes satisfying the best response property that can arise once S forms from within a coalition structure P. If $S \in P$ then $B(S; P) = \emptyset$.

- A proposal is denoted by a pair (P, z) where $P \in \mathcal{P}$ and $z \in \beta(P)$. Let $\sigma = \{(P, z) \mid$ where (P, z) is a *credible* proposal, $P \in \mathcal{P}$, and $z \in \beta(P)\}$. Credibility will be defined in later sections according to the different behavioral assumptions, which will appear as superscripts on σ.

[3] $\subset$ denotes week inclusion.

[4] Note that $P \in \mathcal{R}(P)$.

- Let $\mathcal{P}|_\sigma \subset \mathcal{P}$ be the collection of all credible partitions, i.e., partitions associated with a credible proposal. That is,

$$\mathcal{P}|_\sigma = \{P \in \mathcal{P} \mid \exists(Q, z) \in \sigma \text{ and } Q = P\}.$$

  Let $\mathcal{P}|_\sigma^*$ be the collection of coarsest partitions in $\mathcal{P}|_\sigma$. Similarly, let $X|_\sigma \subset X_N$ be the collection of all credible strategy profiles, i.e., strategy profiles associated with a credible proposal. That is,

$$X|_\sigma = \{x \in X_N \mid \exists(Q, z) \in \sigma \text{ and } x = z\}.$$

  Let $X|_\sigma^*$ be the collection of strategy profiles associated with a credible partition in $\mathcal{P}|_\sigma^*$.

- Then $\mathcal{P}_S(P)|_\sigma = \mathcal{P}_S(P) \cap \mathcal{P}|_\sigma$ is the set of all credible partitions that can arise after S's formation from within a partition P. And $\mathcal{P}_S(P)|_\sigma^* \subset \mathcal{P}_S(P)|_\sigma$ is the set of all *plausible* (besides credible) partitions that are *likely* to arise after S's formation from within a partition P. In constraining S to consider only $\mathcal{P}_S(P)|_\sigma^*$ as the result (in terms of partitions) of its deviation from P we essentially enable it to fear (or hope for) the "next" finer credible proposal(s). Therefore, avoiding an unreasonable fear (or hope) that complementary coalitions may split up further into finer ones.

- Similarly, $B(S; P)|_\sigma = B(S; P) \cap X|_\sigma$ is the set of all credible outcomes (already satisfying the best response property) that can arise once S forms from within a partition P. And $B(S; P)|_\sigma^* \subset B(S; P)|_\sigma$ is the set of all credible outcomes associated with any partition in $\mathcal{P}_S(P)|_\sigma^*$, and therefore the set of all *plausible* (besides credible) outcomes that are *likely* to arise after S's formation from within a partition P.

- Finally, let $\Sigma = \{(P, z) \in \sigma \mid P \in \mathcal{P}|_\sigma^*\}$ be the set of *binding agreements.* The behavioral assumptions will be denoted in later sections by superscripts on Σ.

It is evident that once a coalition S deviates from partition P it is not always the case that there exists a unique plausible (coarsest among the credible) partition (to which it belongs) likely to arise; it can be the case that $|\mathcal{P}_S(P)|_\sigma^*| \neq 1$. Moreover, it is not always the case that only one strategy profile is supported by such coalition structure(s), that is, it can be the case that $|B(S; P)|_\sigma^*| \neq 1$. In the case of emptiness of the two, S's formation is not credible. The multiplicity of plausible outcomes, both in terms of partitioning and of strategy profiles warrants behavioral assumptions to illuminate and systematize players' choices.

3 Optimistic Approach

The optimistic approach assumes that the deviating coalition has agreed upon a strategy vector before its departure and can, therefore, leave the negotiating table while announcing its decision. Thus, an optimistic coalition proceeds with the deviation if the following considerations *guarantee* an improvement on the payoff of every one of its members.

The coalition considers all *plausible* proposals with coalition structures containing it. By *plausible* we refer to *credible* proposals whose partition is (i) *feasible*, i.e., conforms with the nestedness assumption and (ii) *likely* in the sense that it is the coarsest among the credible, and thus most likely to prevail, among all the credible ones. Then, it considers all strategy profiles associated with the plausible proposals restricted to the *agreed strategy vector*. Each coalition member fears that the worst (for himself alone) *plausible strategy profile* will arise and conceives such a payoff as the (individual) guarantee of the deviation under consideration. If the guarantee of every member is greater than their payoff currently on the negotiating table, then the coalition proceeds with the deviation.

The optimism is reflected in the choice and subsequent declaration of the strategy vector. Essentially, the deviating coalition believes that it can select the (entire) best response strategy profile to the extend that the complementary coalitions are not indifferent among some of their own best response strategy vectors[5].

Optimistic Credible Proposal The set of optimistic credible proposals, σ^o, is defined as follows:

- Given that a proposal (P, z) is under consideration, an *optimistic credible deviation* by a coalition $S \subset N$,where $S \subsetneq S'$ and $S' \in P$, via strategy vector x_S is denoted by the pair $(S, x_S; P)$. Moreover, $x_S \in X_S$ such that $(x_S, x_{N\backslash S}) \in B(S;P)|^*_{\sigma^o}$[6].

- Let $g^o(S, x_S; P) = \{g^o_i(S, x_S; P)\}_{i \in S}$ denote the *optimistic guarantee* of coalition S forming from within partition P and declaring x_S. In particular, for some $i \in S$,

$$g^o_i(S, x_S; P) = \min_{y \in B(S;P)|^*_{\sigma^o}} \{u_i(y) \mid y_S = x_S\}$$

- A proposal (P, z) is *optimistically credible,* i.e., $(P, z) \in \sigma^o$ if there does not exist an *optimistic credible deviation* $(S, x_S; P)$ such that $g^o_i(S, x_S; P) > u_i(z)$ for every $i \in S$[7].

The above solution concept is less optimistic then that of Ray and Vohra (1997). Indeed, the deviating coalition is behaving cautiously by fearing the *worst plausible* situation that may stabilize after its departure and its strategy declaration. This cautiousness

[5]It can be case that more than one strategy profiles in $B(S;P)|^*_{\sigma^o}$ inlvolve coalition S playing the same stategy strategy vector, and only members of $N\backslash S$ play different strategies. In such a case the deviating coalition S does not impose a strategy profile but imposes a restriction to those strategy profiles associated with S's chosen strategy vector.

[6]The fact that $(x_S, x_{N\backslash S}) \in B(S;P)|^*_{\sigma^o}$ implies that $\exists Q_S \in \mathcal{P}_S(P)|^*_{\sigma^o}$ associated with $(x_S, x_{N\backslash S})$.

[7]Note that by using S's strategy vector instead of an entire strategy profile to determine the guarantee of each member of S, we maintain an even stronger conservative attitude since not all coalition members fear the worst from the *same* strategy profile. In the event $B(S;P)|^*_{\sigma^o}$ restricted to x_S contains more than one strategy profiles, members of S do not have to coordinate their fears, instead each one can fear the worst payoff vector for him to come about.

is captured by the guarantee function. Consistency is embedded in the domain of the guarantee function namely, the (restricted to x_s) subset of $B(S;P)|^*_{\sigma^o}$ which selects the plausible from the set of feasible and credible strategy profiles. Moreover, the credibility of S itself, as well as all the partitions that may come about is screened by concentrating only on credible partitions.

Due to the nestedness assumption, the partition involving the singletons is always credible, and it supports all the Nash equilibria. Therefore, as long as a game admits a Nash equilibrium in pure strategies it also admits a binding agreement.

The following example illustrates how the optimistic approach differs from Ray and Vohra 's (1997) notion.

	L	C	R
U	$9,2$	$0,0$	$1,2$
D	$0,0$	$5,5$	$0,6$

Notice that the above game admits two Nash equilibria (in pure strategies), namely, UL and UR. Moreover, strategy profiles UL, DC and DR are the only Pareto optimal ones. It is easy to see that both UL and DC are *optimistic binding agreements* while DR is not, since $g^o_1(\{1\}, U; \{1,2\}) = 1$ therefore player 1 cannot block UL or DC but he blocks DR. Similarly, $g^o_2(\{2\}, L; \{1,2\}) = g^o_2(\{2\}, R; \{1,2\}) = 2$, thus player 2 cannot block any of the Pareto optimal outcomes. The Ray and Vohra (1997) approach rules out DC (besides DR) on the grounds that player 1 can deviate and bring about UL, effectively dictating player 2's strategy (L) and ignoring the possibility that, once apart, player 2 may actually choose R.

4 Pessimistic Approach

The antipode of the optimistic approach is introduced by the pessimistic one. In this case, the deviating coalition does not have the opportunity to agree upon a strategy vector before its departure and therefore announce it. Instead, it only departs from the negotiating table fearing the worst among the *plausible* outcomes.

Pessimistic Credible Proposal The set of pessimistic credible proposals, σ^p, is defined as follows:

- Given that a proposal (P, z) is under consideration, a *pessimistic credible deviation* by a coalition $S \subset N$ such that $S \subsetneq S'$, $S' \in P$ is denoted by $(S;P)$, where S belongs to at least one *plausible* partition Q, that is, $S \in Q$, $Q \in \mathcal{P}_S(P)|^*_{\sigma^p}$.

- Let $g^p(S;P) = \{g^p_i(S;P)\}_{i\in S}$ denote the *pessimistic guarantee* of coalition S once formed. In particular, for some $i \in S$,

$$g^p_i(S;P) = \min_{y\in B(S;P)|^*_{\sigma^p}} \{u_i(y)\}$$

- A proposal (P,z) is *pessimistically credible,* i.e., $(P,z) \in \sigma^p$ if there does not exist a *pessimistic credible deviation* $(S;P)$ such that $g_i^p(S;P) > u_i(z)$ for every $i \in S$.

In the same way as in the optimistic approach, consistency is captured by focusing only on the plausible outcomes rather than on all the feasible ones. Pessimism is assumed not only through the guarantee function but also from its domain that includes all the plausible outcomes.

The following example illustrates how different outcomes may emerge from the pessimistic and optimistic treatments. The interesting element in the following game is that optimistic players fail to reach and sign an agreement and revert, essentially, to non-cooperative play, whereas pessimistic players behave more "cooperatively" and agree to playing one of the Pareto optimal outcomes. Similarly to the previous example, there are more than one best response strategy profiles supported by every coalition structure.

	L	R
U	$5,1,1$	$0,0,0$
M	$2,2,10$	$1,5,1$
D	$0,0,0$	$0,0,0$

A

	L	R
U	$0,0,0$	$0,0,0$
M	$10,2,2$	$1,0,0$
D	$2,10,2$	$0,0,0$

B

	L	R
U	$0,0,0$	$0,0,0$
M	$0,0,0$	$0,0,0$
D	$1,1,5$	$1,0,0$

C

The game admits three Nash equilibria, ULA, MRA and DLC, where each player may get 5 or 1. With the pessimistic approach each singleton's guarantee is 1, whereas, with the optimistic approach each player can choose a strategy and therefore guarantee 5[8]. Moving up one step to the next coarser coalition structure, we examine the partitions of the form $\{\{i,j\}\{k\}\}$. The pessimistic approach does not exclude any best response strategy profiles while the optimistic one rules them all out.

Lastly, we move one more step and investigate the grand coalition. To do that we first have to reexamine the guarantees of the individuals, which under the pessimistic approach do not change. In particular, considering player 1, we observe that $(\{\{2,3\}\{1\}\}, DLC)$ is a pessimistic credible proposal and thus, $DLC \in B(\{1\};\{\{1,2,3\}\})|^*_{\sigma^p}$. Moreover, DLC assigns the worst payoff to player 1 among the pessimistic credible proposals associated with partition $\{\{2,3\}\{1\}\}$ [that is, $(\{\{2,3\}\{1\}\}, ULA)$, $(\{\{2,3\}\{1\}\}, MRA)$ and $(\{\{2,3\}\{1\}\}, MLB)$].

According to the optimistic approach, the individuals' guarantees remain the same, albeit, for very different reasons. Recall that no proposal associated with coalition structures of the form $\{\{i,j\},\{k\}\}$ is optimistically credible. Thus, the only partition that is likely to arise when a singleton deviates from the grand coalition is $\{\{1\}\{2\}\{3\}\}$[9].

[8] For example, player 1's guarantees when deviating from some pair are $g_1^p(\{1\};\{\{1,2\}\{3\}\}) = g_1^p(\{1\};\{\{1,3\}\{2\}\}) = 1$ for the pessimistic case and $g_1^o(\{1\},U;\{\{1,2\}\{3\}\}) = g_1^o(\{1\},U;\{\{1,3\}\{2\}\}) = 5$ for the optimistic case.

[9] The guarantees of player 1 when departing from the grand coalition are $g_1^p(\{1\};\{1,2,3\}) = 1$ for the pessimistic case and $g_1^o(\{1\},U;\{\{1,2,3\}\}) = 5$ for the optimistic case.

We still have to examine the guarantees of the pairs under the pessimistic approach[10] before we proceed to investigating the grand coalition. It is easy to see, for example, for player 1 that $g_1^p(\{1,2\};\{1,2,3\}) = g_1^p(\{1,3\};\{1,2,3\}) = 1$ as well, since $(\{\{1,2\}\{3\}\}, MRA)$ and $(\{\{1,3\}\{2\}\}, MRA)$ are pessimistic credible proposals (among others).

Now we can examine which (if any) of the three Pareto optimal outcomes associated with the grand coalition (MLA, MLB, DLB) are binding agreements. According to the pessimistic approach all of them are pessimistic binding agreements since single players or pairs can guarantee only 1 per person, thus no one has an incentive to break away from the grand coalition.

According to the optimistic approach no binding agreement will be signed by the grand coalition, since all Pareto optimal outcomes are blocked by singletons whose guarantees are equal to 5[11].

5 Conservative Approach

Undoubtedly, the pessimistic perspective being so antithetic to the optimistic one, ignores any power the deviating coalition may possess, in the sense that it may have the opportunity to choose among a set of alternatives, *without* dictating others choices. In this section we introduce the conservative approach in an effort to rectify exactly this aspect of the pessimistic one.

To illustrate the deviating coalition's power consider a scenario where some proposal is on the negotiating table and suppose that some coalition contemplates deviating from it. In such an event, it is possible that one of the (plausible) partitions the deviating coalition may bring about supports two (plausible) best response strategy profiles where the complementary coalitions play the exact same strategy vectors. Then, essentially, the deviating coalition has a choice over these two strategy profiles. Note that although both profiles attribute S-Pareto efficient outcomes to S, it can be the case that only one of the two strictly dominates the original proposal under consideration. According to the pessimistic approach, S will not proceed with the deviation fearing the worst, which in this predicament can *only* occur if S *induces* it.

The conservative approach is an amalgam of the previous ones, providing thus, a more tenable analysis of a coalition's conjectures. In this treatment a deviating coalition fears the worst, but not its own actions. When departing it does not announce a strategy, but prior to its departure its members can agree on contingency plans and commit to a favorable strategy vector for *every* eventuality that is likely to arise from the complemen-

[10] This game does not admit an optimistic credible deviation involving a pair.

[11] More specifically, $(\{\{1,2,3\}\}, MLA)$ and $(\{\{1,2,3\}\}, DLB)$ are not optimistic binding agreements since there exists an optimistic credible deviation $(\{1\}, U; \{\{1,2,3\}\})$ such that $g_1^o(\{1\}, U; \{\{1,2,3\}\}) = 5 > 2 = u_1(MLA) = u_1(DLB)$. Moreover, $(\{\{1,2,3\}\}, MLB)$ is not an optimistic binding agreement since there exists an optimistic credible deviation $(\{2\}, R; \{\{1,2,3\}\})$ such that $g_2^o(\{2\}, R; \{\{1,2,3\}\}) = 5 > 2 = u_2(MRA)$.

tary set of players, provided (through credibility) that no subcoalitions intend to further deviate.

Conservative Credible Proposal The set of conservative credible proposals, σ^c, is defined as follows:

- Given that a proposal (P, z) is under consideration, a *strategic plan* $C_S \subset B(S;P)|^*_{\sigma^c}$ specifies a course of action by coalition S (at least) for every plausible behavior by $N \backslash S$. That is, $\forall x \in B(S;P)|^*_{\sigma^c}$ $\exists y \in C_S$ such that $x_{N\backslash S} = y_{N\backslash S}$.
- A *conservative credible deviation* by a coalition $S \subset N$ where $S \subsetneq S'$, $S' \in P$, via strategic plan C_S is denoted by the pair $(S, C_S; P)$, where $C_S \subset B(S;P)|^*_{\sigma^c}$.
- Let $g^c(S, C_S; P) = \{g^c_i(S, C_S; P)\}_{i \in S}$ denote the *conservative guarantee* of coalition S choosing C_S. In particular, for some $i \in S$,

$$g^c_i(S, C_S; P) = \min_{y \in C_S}\{u_i(y)\}$$

- A proposal (P, z) is *conservatively credible*, i.e., $(P, z) \in \sigma^c$ if there does not exist a *conservative credible deviation* $(S, C_S; P)$ such that $g^c_i(S; C_S) > u_i(z)$ for every $i \in S$.

The fusion of the first two treatments is captured by the concept of strategic plan. A strategic plan endows a deviating coalition with the power to choose its own actions without dictating the action of its complements. More precisely, the deviating coalition has no control over which, among the plausible partitions will emerge, or which strategy vector, among the plausible ones, the complementary coalitions will select, but it can optimize its own response (strategy vector), if such a choice exists.

Recalling the situation described in the beginning of this section, S can exclude from C_S the strategy profile that involves S playing a “not so beneficial” strategy vector, compared to the proposal on the negotiating table. Notice that C_S specifies at least one response from S to every plausible occurrence by the complementary coalitions. Thus, when the guarantee of the coalition is estimated, it expects the worst (among the plausible) from everyone else.

The following example, due to Einy and Peleg (1995), illustrates how the pessimistic approach may support counter intuitive outcomes.

	L	R
U	$3,2,0$	$0,0,0$
D	$2,0,3$	$2,0,3$

A

	L	R
U	$3,2,0$	$0,3,2$
D	$0,0,0$	$0,3,2$

B

Notice that all outcomes are Pareto optimal, and the game admits three Nash equilibria, $\beta(\{\{1\}\{2\}\{3\}\}) = \{ULA, DRA, URB\}$. According to both the pessimistic and conservative treatments the singletons' guarantees are 0[12]. Consequently, when we investigate proposals of the form $(\{\{i,j\}\{k\}\}, x)$ where $x \in \beta(\{\{i,j\}\{k\}\})$ all of them are pessimistically and conservatively credible.

Finally, we move one step up to the investigation of the grand coalition and the Pareto optimal outcomes. Firstly, we have to determine the guarantees of the pairs. For example, consider the pair $\{1,2\}$ and note that $\beta(\{\{1,2\}\{3\}\}) = \{ULA, ULB, URB\}$.

According to the pessimistic approach players' guarantees, by forming coalition $\{1,2\}$, are $(0,2)$[13]. Therefore, $\{1,2\}$ cannot block anything available to the grand coalition, essentially due to 1's low guarantee, including outcomes like URA that attribute 0 to every one.

However, 1's fears are unsubstantiated and the conservative approach captures exactly that. The choice between ULB and URB depends completely upon the coalition $\{1,2\}$ itself. In other words, $\{1,2\}$ prior to their formation and departure from $\{1,2,3\}$ can sign an agreement where they commit to play UL whether player 3 plays A or B after their departure. This element of commitment is captured by the concept of strategic plans. In particular, players 1 and 2 can choose between three strategic plans, namely $C^*_{\{1,2\}} = \{ULA, ULB\}$, $C'_{\{1,2\}} = \{ULA, URB\}$ and $C''_{\{1,2\}} = \{ULA, ULB, URB\}$[14]. Thus, by choosing $C^*_{\{1,2\}}$, $\{1,2\}$ will block all Pareto optimal outcomes that assign to 1 and 2 anything less than 3 and 2 respectively. Due to the symmetry of the game the other two pairs, $\{1,3\}$ and $\{2,3\}$ can block, in a similar manner, the rest of the Pareto optimal outcomes.

Notice that, in the absence of a binding agreement, player 2 would actually have an incentive to cheat on strategic plan $C^*_{\{1,2\}}$ after the deviation, and play R in the event player 3 chooses B. However, such a mishap is averted through the binding aspect of the agreement.

In conclusion, according to the pessimistic approach all outcomes are supported by binding agreements signed by the grand coalition. Whereas, according to the conservative approach the binding agreements are:

$$\begin{array}{lll}
(\{\{1,2\}\{3\}\}, ULA) & (\{\{1,3\}\{2\}\}, ULA) & (\{\{1\}\{2,3\}\}, URB) \\
(\{\{1,2\}\{3\}\}, ULB) & (\{\{1,3\}\{2\}\}, DLA) & (\{\{1\}\{2,3\}\}, DRA) \\
(\{\{1,2\}\{3\}\}, URB) & (\{\{1,3\}\{2\}\}, DRA) & (\{\{1\}\{2,3\}\}, DRB)
\end{array}$$

It is interesting to observe that the optimistic approach supports a subset of the conservative binding agreements. For example, for player 1, $g_1^o(\{1\}, D; \{\{1,2\}\{3\}\}) = 2$,

[12] That is, $g_i^p(\{i\}; \{\{i,j\}\{k\}\}) = 0$. Similarly, $g_i^c(\{i\}, C_{\{i\}}; \{\{i,j\}\{k\}\}) = 0$, where $C_{\{i\}} = \{ULA, DRA, URB\}$.

[13] That is, $g^p(\{1,2\}; \{1,2,3\}) = (0,2)$, due to URB for player 1 and due to ULA and ULB for player 2.

[14] According to $C''_{\{1,2\}}$ players 1 and 2 essentially do not reach an agreement prior to their departure, thus they anticipate everything. Their guarantees according to $C^*_{\{1,2\}}$ are $g^c(\{1,2\}, C^*_{\{1,2\}}; \{\{1,2,3\}\}) = (3,2)$, whereas according to $C'_{\{1,2\}}$ and $C''_{\{1,2\}}$ the guarantees are $g^c(\{1,2\}, C'_{\{1,2\}}; \{\{1,2,3\}\}) = g^c(\{1,2\}, C''_{\{1,2\}}; \{\{1,2,3\}\}) = (0,2)$.

leading to a smaller (compared to the previous two approaches) set of optimistic credible proposals, where the coalition structure is of the form$\{\{i,j\}\{k\}\}$. That is, proposal $(\{\{1,2\}\{3\}\},URB)$ is not optimistically credible, since it is blocked by player 1. Accordingly, the guarantees of the pairs are maintained at high levels, i.e., $g^o(\{1,2\},UL;\{\{1,2,3\}\}) = (3,2)$, thus, blocking all the Pareto optimal outcomes available to the grand coalition. The optimistic binding agreements are:

$$\begin{array}{ccc} (\{\{1,2\}\{3\}\},ULA) & (\{\{1,3\}\{2\}\},DLA) & (\{\{1\}\{2,3\}\},URB) \\ (\{\{1,2\}\{3\}\},ULB) & (\{\{1,3\}\{2\}\},DRA) & (\{\{1\}\{2,3\}\},DRB) \end{array}.$$

6 Properties

The following example from Ray and Vohra (1997) is used by the authors to show how inefficient results can prevail. In fact, the same (inefficient) binding agreement is supported by all three of our augmentations as well, highlighting therefore, the persistence of inefficiency in the presence of binding agreements. Specifically, the coarsest equilibrium[15] binding agreement is supported by the coalition structure of the form $\{\{i,j\},\{k\}\}$, and assigns to player k payoff 3.7 and to each i and j payoff 2.7. Note that it is Pareto dominated by payoff vector $(2.9,2.9,3.9)$. Since there does not exist a Pareto optimal outcome that assigns to every player at least 3.7, every singleton can rely on partition $\{\{i,j\},\{k\}\}$ and aim for the 3.7. Observe that singleton $\{k\}$ does not have to fear, upon deviation, partition $\{\{i\},\{j\},\{k\}\}$, since pair $\{i,j\}$ will not split further apart.[16]

c_1:

	b_1	b_2	b_3
a_1	2.6, 2.6, 2.6	3.2, 2.2, 3.2	3.7, 1.7, 3.7
a_2	2.2, 3.2, 3.2	2.7, 2.7, 3.7	3.1, 2.1, 4.1
a_3	1.7, 3.7, 3.7	2.1, 3.1, 4.1	2.6, 2.6, 4.6

c_2:

	b_1	b_2	b_3
a_1	3.2, 3.2, 2.2	3.7, 2.7, 2.7	4.1, 2.1, 3.1
a_2	2.7, 3.7, 2.7	3.1, 3.1, 3.1	3.6, 2.6, 3.6
a_3	2.1, 4.1, 3.1	2.6, 3.6, 3.6	2.9, 2.9, 3.9

c_2:

	b_1	b_2	b_3
a_1	3.7, 3.7, 1.7	4.1, 3.1, 2.1	4.6, 2.6, 2.6
a_2	3.1, 4.1, 2.1	3.6, 3.6, 2.6	3.9, 2.9, 2.9
a_3	2.6, 4.6, 2.6	2.9, 3.9, 2.9	3.3, 3.3, 3.3

A conjecture, arising from the examples presented so far, is how the three solution concepts proposed in this paper relate to each other. The (perhaps intuitive) speculation that the three solutions should be nested, that is, if a proposal is optimistically credible, then it is also conservatively and pessimistically credible, and if it is conservatively credible then it is pessimistically credible as well, holds only in certain cases as asserted by Theorem 1. In general, the conjecture does not hold, and the example presented in section 4 to

[15] By equilibrium in this case we refer to the Ray and Vohra (1997) notion as well as to the three augmentations we propose.

[16] Under partition $\{\{i,j\},\{k\}\}$ players i and j get 2.7 each, whereas if they split further apart their payoff is 2.6.

illustrate the difference between the optimistic and pessimistic approach can serve as a counter example, where the optimistic binding agreement is not a pessimistic one.

Theorem 1 *Consider a normal form game G that admits a Nash equilibrium in pure strategies. An optimistic credible proposal (P, z) is pessimistically and conservatively credible if all coalition structures $Q \in R(P)$, such that $Q \neq P$ and $\beta(Q) \neq \emptyset$, are associated with some optimistic credible proposal. Similarly, a conservative credible proposal (P, z) is pessimistically credible if all coalition structures $Q \in R(P)$, such that $Q \neq P$ and $\beta(Q) \neq \emptyset$, are associated with some conservatively credible proposal.*

Proof. We will prove the theorem by induction on the size of a coalition structure.

Let P^l denote a coalition structure $P \in \mathcal{P}$ of size l, i.e., $|P^l| = l$.

Notice that for any $x^* \in \beta(P^n)$, (P^n, x^*) is an optimistic, conservative and pessimistic credible proposal. Let the next coarser partition that admits a best response strategy profile be of size $n-k$, where $1 \leq k < n$. That is, consider $P^{n-k} \in \mathcal{P}$ such that $\beta(P^{n-k}) \neq \emptyset$ while $\beta(P^l) = \emptyset$ for all l such that $n - k < l < n$, and for all $i \in N$ such that $\{i\} \notin P^{n-k}$. Then, for any such P^{n-k} and i we have, directly from the definition of the various guarantees,

$$\begin{aligned} g_i^o(\{i\}, x^*_{\{i\}}; P^{n-k}) &\geq g_i^c(\{i\}, C_{\{i\}}; P^{n-k}) \geq g_i^p(\{i\}; P^{n-k}) \\ \text{for every } x^* &\in \beta(P^n), \text{ and for every } C_{\{i\}}. \end{aligned}$$

Observe that if a singleton has more than one strategic plans they all trivially lead to the same guarantee, i.e., $g_i^c(\{i\}, C_{\{i\}}; P^{n-k}) = g_i^c \in \mathbb{R}$ for every $C_{\{i\}} \subset B(\{i\}; P^{n-k})|^*_{\sigma^c} = \beta(P^n)$.

Now consider $y^* \in \beta(P^{n-k})$. Proposal (P^{n-k}, y^*) is optimistically credible if for any $\{i\} \notin P^{n-k}$ we have $u_i(y^*) \geq g_i^o(\{i\}, x^*_{\{i\}}; P^{n-k})$ for every $x^* \in \beta(P^n)$, but then $u_i(y^*) \geq g_i^c(\{i\}, C_{\{i\}}; P^{n-k})$ for every $C_{\{i\}}$, which implies that (P^{n-1}, y^*) is conservatively credible and lastly $u_i(y^*) \geq g_i^p(\{i\}; P^{n-k})$ which implies that (P^{n-1}, y^*) is pessimistically credible. Notice that no coalition larger than the singletons could credibly deviate since the only (optimistic, conservative, and pessimistic) credible partition that can arise is P^n. It is easy to see that if (P^{n-k}, y^*) is not optimistically credible but is nevertheless conservatively credible, then, it will also be pessimistically credible.

Next consider some optimistic credible proposal (P^m, y^m) and assume that for every $Q \in R(P^m)$, $Q \neq P^m$ if $(Q, x) \in \sigma^o$ then $(Q, x) \in \sigma^c \cap \sigma^p$. Then, we have $\mathcal{P}_S(P^m)|^*_{\sigma^o} = \mathcal{P}_S(P^m)|^*_{\sigma^p} = \mathcal{P}_S(P^m)|^*_{\sigma^p} = \mathcal{P}_S(P^m)|^*_{\sigma}$ and hence, $B(S; P^m)|^*_{\sigma^o} = B(S; P^m)|^*_{\sigma^c} = B(S; P^m)|^*_{\sigma^p} = B(S; P^m)|^*_{\sigma}$ for any coalition S such that $S \subsetneq T$, $T \in P^m$. Since (P^m, y^m) is optimistically credible $\nexists$ optimistic credible deviation $(S, x_S; P^m)$ such that $g_i^o(S, x_S; P^m) > u_i(y^m)$ for every $i \in S$. That is, no S such that $S \subsetneq T$, $T \in P^m$ can find x_S such that $(x_S, x_{N \setminus S}) \in B(S; P^m)|^*_{\sigma}$, that improves all its members compared to y^m. But then, no S such that $S \subsetneq T$, $T \in P^m$can find $C_S \subset B(S; P^m)|^*_{\sigma}$ that improves all its members either, i.e., $\nexists$ conservative credible deviation $(S, C_S; P^m)$ such that $g_i^c(S, C_S; P^m) > u_i(y^m)$ for every $i \in S$. Lastly, obviously, the entire $B(S; P^m)|_{\sigma}^{*}$ cannot improve all of S's members, i.e., $\nexists$ pessimistic credible deviation $(S; P^m)$ such that $g_i^p(S; P^m) > u_i(y^m)$ for every $i \in S$. Therefore, (P^m, y^m) is conservatively and pessimistically credible as well. It is easy to see that if (P^m, y^m) is not optimistically credible but is nevertheless conservatively credible, then, it will also be pessimistically credible.

Returning to our proposition, since every $Q \in R(P), Q \neq P$, $\beta(Q) \neq \emptyset$, is associated with some optimistic credible proposal, then all these proposals are also conservatively and pessimistically credible, and thus, (P, z) is also pessimistically and conservatively credible. Similarly, if every $Q \in R(P)$ is associated with some conservative credible proposal, then all these proposals are also pessimistically credible, and thus, (P, z) is also pessimistically credible. ■

Another conjecture arising from the example presented in this section is whether the three solution concepts differ from each other when each coalition structure admits at most one best response strategy profile. Indeed, the three solutions may differ from each other even in such a case. To illustrate how this can occur consider the case where some coalition S departs from proposal (P, z). Its deviation can induce, say, two coarsest optimistic, pessimistic, and conservative credible partitions, i.e., $\mathcal{P}_S(P)|_{\sigma^o}^{*} = \mathcal{P}_S(P)|_{\sigma^p}^{*} = \mathcal{P}_S(P)|_{\sigma^c}^{*} = \{Q, V\}$, each associated with a distinct optimistic, pessimistic, and conservative credible strategy profile, i.e., $B(S; P)|_{\sigma^o}^{*} = B(S; P)|_{\sigma^p}^{*} = B(S; P)|_{\sigma^c}^{*} = \{q, v\}$ such that $q_S \neq v_S$. In the optimistic case the deviating coalition will fear the least harmful of the two, since it is essentially choosing through the announcement of q_S or v_S. Whereas in the pessimistic case, the deviating coalition fears both q and v. In the conservative case if $q_{N\backslash S} = v_{N\backslash S}$ then the deviating coalition can choose between q and v by excluding the undesired one from its strategic plan (reducing to the optimistic case), whereas if $q_{N\backslash S} \neq v_{N\backslash S}$ then S will fear both (reducing to the pessimistic case).

7 Discussion

As we briefly touched upon in earlier parts of the paper, the most serious caveat of our approach in general is that we allow the emergence of *only* finer coalition structures out of the blocking of some partition. The benefits of doing so are on a conceptual level that we easily achieve consistency, while on a technical level existence is immediate (as long as a game admits Nash Equilibria in pure strategies).

Once mergers are allowed, a deviating coalition has to consider partitions (perhaps) coarser than the one it is departing from. In such an event solving a game recursively does not attribute consistency anymore and the only answer would be the use of a solution concept in the spirit of the stable set by von Neumann and Morgenstern (1944). The stable set is defined by satisfying two fundamental properties: all outcomes included in the solution cannot contradict each other (internal stability), while every outcome excluded from the solution is done so because of (at least) one outcome in the solution dominating it (external stability). Such a treatment attributes perfect consistency, yet as originally defined, it may not always exist. More sophisticate dominance relations, however, can rectify the problem, as shown in Greenberg (1990), who extended the merits of the stable set by incorporating it in a more general and unifying framework. Greenberg identifies large sets of situations and various dominance relations (not necessarily binary) where existence is satisfied and the very appealing features brought about by internal and external stability can be enjoyed. An interesting extension of the present work as well as of that by Ray and Vohra (1997) could therefore relax the nestedness assumption while preserving consistency.

8 References

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

2. Einy, E., and Peleg, B. (1995). "Coalition-Proof Communication Equilibria" *Social Choice, Welfare, and Ethics,* (W. A. Burnett, H. Moulin, M. Sales, and N. J. Schofield, Eds.). Cambridge: Cambridge University Press.

3. Greenberg, J. (1990). *The Theory of Social Situations: An Alternative Game-Theoretic Approach.* Cambridge: Cambridge University Press.

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

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

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

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

Working Paper

2000-8	Niels Haldrup, Antonio Montanés and Andreu Sanso: Measurement Errors and Outliers in Seasonal Unit Root Testing.
2000-9	Erik Harsaae: En kritisk vurdering af den generelle ligevægtsmodel.
2000-10	Rasmus Højbjerg Jacobsen: Why the ECB Should be Ultra-Liberal.
2000-11	Michael Rosholm and Michael Svarer: Structurally Dependent Competing Risks.
2000-12	Efforsyni Diamantoudi and Licun Xue: Farsighted Stability in Hedonic Games.
2000-13	Licun Xue: Stable Agreements in Infinitely Repeated Games.
2000-14	Bo Sandemann Rasmussen: Government Debt and Capital Accumulation in the Blanchard-Cass-Yaari OLG Model.
2000-15	Nikolaj Malchow-Møller and Bo Jellesmark Thorsen: Investment under Uncertainty - the Case of Repeated Investment Options.
2000-16	Boriss Siliverstovs: The Bi-parameter Smooth Transition AutoRegressive model.
2000-17	Peter Skott: Demand Policy in the Long Run.
2000-18	Paul Auerbach and Peter Skott: Skill Assymmetries, Increasing Wage Inequality and Unemployment.
2000-19	Torben M. Andersen: Nominal Rigidities and the Optimal Rate of Inflation.
2001-1	Jakob Roland Munch and Michael Svarer: Mortality and Socio-economic Differences in a Competing Risks Model.
2000-2	Effrosyni Diamantoudi: Equilibrium Binding Agreements under Diverse Behavioral Assumptions.

CENTRE FOR DYNAMIC MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Working papers, issued by the Centre for Dynamic Modelling in Economics:

1999-21	Martin Paldam: Corruption and Religion. Adding to the Economic Model?
2000-1	Niels Haldrup and Peter Lildholdt: On the Robustness of Unit Root Tests in the Presence of Double Unit Roots.
2000-2	Niels Haldrup and Peter Lildholdt: Local Power Functions of Tests for Double Unit Roots.
2000-3	Jamsheed Shorish: Quasi-Static Macroeconomic Systems.
2000-4	Licun Xue: A Notion of Consistent Rationalizability - Between Weak and Pearce's Extensive Form Rationalizability.
2000-6	Graham Elliott and Michael Jansson: Testing for Unit Roots with Stationary Covariates.
2000-8	Niels Haldrup, Antonio Montanés and Andreu Sanso: Measurement Errors and Outliers in Seasonal Unit Root Testing.
2000-12	Effrosyni Diamantoudi and Licun Xue: Farsighted Stability in Hedonic Games.
2000-13	Licun Xue: Stable Agreements in Infinitely Repeated Games.
2000-16	Boriss Siliverstovs: The Bi-parameter Smooth Transition AutoRegressive model.
2000-17	Peter Skott: Demand Policy in the Long Run.
2000-18	Paul Auerbach and Peter Skott: Skill Asymmetries, Increasing Wage Inequality and Unemployment.
2000-19	Torben M. Andersen: Nominal Rigidities and the Optimal Rate of Inflation.
2001-1	Jakob Roland Munch and Michael Svarer: Mortality and Socio-economic Differences in a Competing Risks Model.
2001-2	Effrosyni Diamantoudi: Equilibrium Binding Agreements under Diverse Behavioral Assumptions.