JOURNAL OF ECONOMIC THEORY **73**, 30–78 (1997)
ARTICLE NO. ET962236

Equilibrium Binding Agreements*

Debraj Ray

Boston University, Boston, Massachusetts ; and
Instituto de Análisis Económico (CSIC), 08193 Bellaterra, Barcelona, Spain

and

Rajiv Vohra

Brown University, Providence, Rhode Island 02912; and
Indian Statistical Institute, New Delhi, India 110 016

Received December 16, 1994; revised May 4, 1996

We study equilibrium binding agreements, the coalition structures that form under such agreements, and the efficiency of the outcomes that result. We analyze such agreements in a context where the payoff to each player depends on the actions of all other players. Thus a game in strategic form is a natural starting point. Unlike the device of a characteristic function, explicit attention is paid to the behavior of the complementary set of players when a coalition blocks a proposed agreement. A solution concept and its applications are discussed. *Journal of Economic Literature* Classification Numbers: C70, C71. © 1997 Academic Press

1. INTRODUCTION

The aim of this paper is to study equilibrium binding agreements, the coalition structures that form under such agreements, and the efficiency of the outcomes that result. The approach that we take is in the spirit of cooperative game theory, in the sense that the concept of "blocking" by a coalition is one of the primitive features of our analysis. A companion

* We thank Francis Bloch, Tatsuro Ichiishi, Andreu Mas-Colell, Paul Milgrom, Bezalel Peleg, Robert Rosenthal, Roberto Serrano, Sang-Seung Yi, and an anonymous referee for useful comments. We gratefully acknowledge support under National Science Foundation Grants SBR-9414114 [Ray] and SBR-9414142 [Vohra]. Partial assistance under Grant PB90-0172 from the Ministerio de Educación y Ciencia, Government of Spain [Ray], and a Fulbright Research Award [Vohra] also supported this research. An earlier version of this paper was circulated as Working Paper 92–8, Department of Economics, Brown University, Providence, RI 02912.

0022-0531/97 $25.00

paper (Ray and Vohra [28]) studies an alternative theory based on bargaining.

Our work is motivated by several considerations. First, we shall argue that a satisfactory description of what constitutes free and unrestrained negotiation, unhampered by the inability to write agreements that are binding on all agents, does not appear to exist in the literature. Our paper takes a step in the direction of a precise concept.

Our second consideration is of central concern to us. It appears to be a matter of consensus among economists that if binding agreements can be written in the absence of informational imperfections, then all the gains from cooperation will indeed be exploited. The outcome *must* be Pareto-optimal. The argument goes back at least to Coase [11], and finds explicit expression in textbooks such as Milgrom and Roberts [24]. Indeed, in the presence of transferable utility, the assertion of ubiquitous efficiency trivially implies that the aggregate surplus (and the overall agreement, under mild conditions) will be independent of the assignment of rights to various parties. It is in this latter form that the Coase Theorem is well known. The theory that we develop does not support this argument.[1] Both in general situations and in natural economic environments, we shall use our concept to demonstrate the existence of robust inefficient outcomes. The possibility of such inefficiency stems both from the possible intervention of coalitions in the negotiation process, as well as from our explicit consideration of widespread externalities across players (and therefore across coalitions).

Finally, we are interested in the coalition structures that endogenously form in the process of writing agreements.

We analyze such agreements in a model where the payoff to a player depends on the *actions* of all the others. Thus, the natural, primitive framework to consider is a game in strategic or normal form. Cooperative equilibrium notions such as the core and the bargaining set study binding agreements through the characteristic function form. If the actions of the players outside a coalition do not affect the payoffs to the members of the given coalition, then the characteristic function form is appropriate.[2] The standard approach to the problem in normal form (due to Aumann [1], see also Scarf [31]) is to convert the normal form game into characteristic function form, and analyze the core of the cooperative game so induced. There are several options to choose from in making such a conversion. But, in general, the specific conversions used do not enjoy obvious consistency properties. Consider, for instance, the notion of the α-core. This notion

[1] Because of some restrictions that we impose on coalition formation, it does not demolish it entirely either.

[2] Indeed, in this special case our solution concept does coincide with the core of a coalition structure.

presumes that when a coalition deviates, it does not expect to receive more than what it does when members of the complementary coalition act to minimax this coalition's payoffs. There is no reason why the complementary coalition should behave in this bloodthirsty fashion, and there is no reason for the deviating coalition to *necessarily* expect or fear such behavior.[3]

The easiest way to see the problem is to consider the example of a Cournot duopoly. Here, the α-core is the set of *all* individually rational Pareto optimal allocations. The reason is simple: under weak assumptions, one player can always be pushed to the point where it is not possible for him to earn any profits. But it should be obvious that any agreement that yields a player less than his Cournot–Nash payoff cannot constitute a binding agreement: by breaking off negotiations, this payoff is what he can credibly expect.

Matters are, of course, far more complicated when there are more than two players and non-singleton subcoalitions might form. One ingredient of the theory to follow will be the idea of noncooperative play across coalitions in a coalition structure.

Our discussion of the Cournot duopoly hints at a second crucial feature of our equilibrium notion. This is an explicit consideration of *consistency*. When a coalition deviates it should not take as given the strategies of its complement, nor should it fear the worst. It should look ahead to a resulting "equilibrium" that its actions induce.[4] Suppose that a proposal is made for N, the grand coalition, to which a subcoalition S objects. In the light of the construction mentioned above, this translates into: in the noncooperative environment induced by the deviation of S, namely, the coalition structure $\{S, N-S\}$, S can, in "equilibrium", be better off relative to the original proposal. Two considerations are crucial in determining what the resulting "equilibrium" will be after S deviates:

(1) S may break up even further.

(2) $N-S$ may break up even further.

The former would suggest that the original objection of S was not "credible," since S is itself vulnerable to further defections. In the context of characteristic function cooperative games, this issue has received attention (see Ray [26], Dutta and Ray [13, 14], Mas-Colell [23], and Dutta, Ray, Sengupta and Vohra [15]). The noncooperative analogue of this problem has been analyzed in Bernheim, Peleg and Whinston [4].

[3] A similar conceptual criticism applies to the β-core and the Strong Nash equilibrium.

[4] The phrase "equilibrium" will, of course, be given a precise meaning in the formal analysis to follow.

A general taxonomy of consistency approaches, relating them to the definition of a vN-M stable set, has been described by Greenberg [17].

However, the latter consideration (2), namely that the complementary coalition(s) may also break up, introduces *entirely new features*, as we shall see later in this paper. For cooperative games in characteristic function form this consideration is absolutely irrelevant to what S can achieve, and so its implications are assumed away. In refinements of Nash equilibria such as Bernheim, Peleg and Whinston [4], this consideration does not arise because *the deviating coalition takes the strategy vector of the complement as given.* Indeed, as we argue below (Section 3), it is *precisely* this difference in specification that lies at the heart of the distinction between a binding agreement and a coalition-based refinement of Nash equilibrium.

For a more complete discussion of the relevant literature, see Section 3. In the remainder of this introduction, we explain the concept that we use, and summarize the main results of the paper.

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 these notions to the case of arbitrary blocking is far from trivial.[5]

With this in mind, consider the following story. Initially, all the players are gathered together in a grand negotiation room. In the course of their deliberations, subsets of these players might irrevocably leave (or threaten irrevocable departure). Each defecting coalition is now cloistered in its own negotiation room. Players in a single room may cooperatively choose (and having chosen, enforce) a strategy vector among themselves. But they must do so *independently* of what the players in the other rooms will do. Indeed, *this last postulate is taken as a defining feature of coalitional structure.*

Of course, in the case of the grand coalition (where all negotiations are presumed to "begin"), this last consideration is empty. Nevertheless, it is necessary to describe what happens in all other coalition structures, to understand what it is that the grand coalition can achieve.

Therefore, each player must look ahead and try to predict the behavior of other players, for every possible assignment of players to negotiation rooms. It is well-known that such introspection does not guarantee Nash-like

[5] For discussions in other contexts, see, for example, Chakravorti and Kahn [8], Dutta *et al.* [15], and Greenberg [17]. A theory based on noncooperative coalitional bargaining may also throw some light on the matter: see Bloch [5] and Ray and Vohra [28]. A forthcoming paper will take up these issues explicitly in the blocking context.

behavior (Bernheim [3], Pearce [25]). We abstain from these considerations as they lead us far afield of our program, and suppose that there is not only common knowledge, but common beliefs regarding outcomes. We presume, then, that these assignments will indeed lead to Nash-like play across coalitions, with players in the same room playing (vector) best responses to the (commonly known) strategies of the others.[6] With these outcomes internalized by all players, negotiations may proceed.

Suppose that a proposal (strategy vector) x is under discussion by the grand coalition, and must be collectively accepted or rejected the next day. People are free to intermingle and discuss the proposal, individually or in groups. Suppose now, that a coalition S (we shall call it the *leading perpetrator* in our formal definition) understands that it can actually do better than x, *provided that a particular coalition structure forms, and a strategy vector y is played.* The question is: what are the minimum requirements that need to be fulfilled before S can actually convince all concerned that such a structure might indeed come into existence?

First, it must be the case that under this coalition structure, y satisfies the *best response property*: no coalition can do (Pareto) better than play their piece of y, given that all other coalitions are doing the same.

Second, the strategy vector y under this new coalition structure must itself be immune to the kind of foul play that S is currently plotting against the grand coalition. No new coalition T must be able to perpetrate a further reorganization of the coalitional structure. This is the requirement of *consistency* (see (1) and (2) above).

Third, each of the other coalitions that are needed to form the new structure have the option to *not* do what S is suggesting they *will* do. In the formal definitions, these coalitions will be referred to as (secondary) perpetrators. These secondary perpetrators (if there are any)[7] must visualize, independently, the consequences of not having defected. This must have two implications. First, the "intermediate" coalition structure thereby achieved must in itself be unstable, just as the grand coalition is currently in danger of being. Second, at least one of the coalitions that are responsible for this instability must be one of the secondary perpetrators that are contemplating this counterfactual, and in its blocking it must use precisely the coalition structure suggested by S.

[6] There is only one additional mild condition that needs to be met. It is that each coalition must be aware of any defections that might occur *from that coalition*, and that this ability (to be aware) is commonly known. This is not really an additional assumption at all. After all, an agreement for each coalition *must* be signed by all its members. In particular, this is why an entire coalition plays a best response to its complement, whereas a subcoalition of a coalition in a coalition structure cannot do the same.

[7] By the way, there may be no such additional coalitions needed, in which case S's task is made far easier!

These three conditions are necessary for S to convince the others. In this paper, we take them to be sufficient as well, though it is certainly reasonable to claim that they are not sufficient. In particular, S is being granted a high degree of optimism, for even if it were able to induce the desired structure, strategies *other* than y may be possible outcomes. We discuss this issue in detail in Sections 2 and 4.

Return, now, to our original proposal x. If S can engineer a defection as just described, then x cannot be an agreement for the grand coalition. Only those proposals that are immune to these considerations qualify as binding agreements for the grand coalition.

A similar definition applies to proposals for any other coalition structure, though the term "binding agreement" is a bit of a misnomer in such cases: the agreement *is* noncooperative across players in different negotiation rooms. What we have is really a collection of such agreements, one for each coalition in the coalition structure.

It might be useful to keep this story in mind when considering the formal definition in Section 2. Section 2 also contains a detailed discussion of a number of issues and alternatives relevant to the definition. In Section 3, we relate our definition to the existing literature, and a list of references for the interested reader is to be found there.

A central point of this paper is that binding agreements, once carefully defined, *are not necessarily efficient*. In Section 4, we show that for any assignment of strategy sets to players (satisfying some mild restrictions), there are open sets of payoff functions with the property that *every* binding agreement is inefficient, provided that there are at least three players.[8] Moreover, we argue that this result is robust to substantial changes in the definition, ranging from very optimistic to very pessimistic predictions by potential perpetrators.

In Section 5, we consider the first of our two applications in detail: a public goods economy. One aim of these applications is to describe the coalition structures that form in natural economic situations, in addition to establishing efficiency or inefficiency of the final outcomes. The propositions in this section establish that efficient outcomes recur only along a subsequence (in the total number of agents); this subsequence is, in general, "sparse" in a sense made clear in that section. In the remaining inefficient cases, the grand coalition breaks up into a subcoalition, which carries out production of the public good, and a number of free riders who enjoy the good but do not contribute to its production. We reiterate that this inefficient outcome occurs *despite* the presence of complete and perfect information.

[8] Two-player games in strategic form have a natural superadditive structure, which precludes any reasonable examples of inefficiency in such games.

In Section 6, we turn our attention to symmetric *transferable utility* games. Transferable utility (TU) is a special case but an important one, we feel. For instance, the propositions of Coase are all based on the presumption that payoffs are transferable among the players. In addition, the symmetry condition allows us to obtain an enormous computational simplification: for a large subclass of these games, one can obtain the equilibrium coalition structures by simply considering allocations that involve equal divisions of utility among players in a coalition, and a rudimentary concept of blocking.

Section 7 uses the results of Section 6 to study our second application: the case of a symmetric Cournot oligopoly. Here again, our interest is not only in the issue of efficiency but also in the nature of equilibrium coalition structures. We observe that if the outcome is inefficient, equilibrium coalition structures *must* be asymmetric even though the game is symmetric, and moreover, the coarsest of these must be "sufficiently" coarse (see Proposition 7.1). The existence of coarse structures is a general insight: if coalition structures were to be too fine, then the grand coalition must be able to achieve a binding agreement. Finally, we present an algorithm for studying equilibrium coalition structures in the Cournot oligopoly. We compute such structures up to 9 agent games. We show that if the number of agents is 5, 6, 7, or 8, the outcome must be inefficient, whereas in the remaining cases there exists an efficient equilibrium.

In both the economic applications of Section 5 and Section 7, there emerges a cyclical pattern of efficiency as the numbers of players increases. This is a curious observation that might bear more general investigation. To see the intuition, at least for symmetric games, observe that if the grand coalition does not have an equilibrium binding agreement, then there must be some *intermediate sized* (and asymmetric) coalition structure which is stable, destroying the grand coalition. If no such structure were to be viable, we would have Nash equilibrium as the outcome of interaction among singletons, which as we know is generally dominated by the grand coalition. Put another way, the grand coalition survives if there exist "large" zones of instability in intermediate coalition structures. This suggests that as the number of players increases, there might be a cyclical pattern in the viability of the grand coalition. This is borne out in both the applications studied.

2. BINDING AGREEMENTS

Consider a game in normal form $\Gamma = (N, (X_i, u_i)_{i \in N})$, where N denotes a (finite) set of players, X_i the strategy set of player i and u_i: $\prod_{i \in N} X_i \mapsto \mathbb{R}$ the payoff function of player i. A *coalition* is any nonempty subset of N.

Let $\mathcal{N}$ be the collection of all coalitions. For any $S \in \mathcal{N}$ we will use X_S to denote $\prod_{j \in S} X_j$ and X_{-S} to denote $\prod_{j \in N \setminus S} X_j$. We will also use X to denote X_N. For any $x \equiv (x_i)_{i \in N} \in X$ and $S \subseteq N$ we will use x_S to denote $(x_j)_{j \in S}$ and (if $N \setminus S$ is nonempty) x_{-S} to denote $(x_j)_{j \in N \setminus S}$. Similarly, for $x \in X$ and $S \in \mathcal{N}$, denote $(u_i(x))_{i \in S}$ by $u_S(x)$. A partition of N will be called a *coalition structure*. For a coalition structure $\mathcal{P}$, let $\mathcal{R}(\mathcal{P})$ denote all coalition structures that are refinements of $\mathcal{P}$.

The primary objective of this section is to define the set of *equilibrium binding agreements* $\mathcal{B}(\mathcal{P})$ that can arise should negotiations commence from some arbitrary coalition structure $\mathcal{P}$. A typical binding agreement will be a strategy vector x, to be interpreted as "equilibrium actions" taken by each of the agents. If equilibrium binding agreements do exist for a given coalition structure $\mathcal{P}$, we shall refer to $\mathcal{P}$ as an *equilibrium coalition structure*.

A central feature of what follows is the possible formation of new coalition structures from old ones. As discussed in the Introduction, we consider only the "internal" case in this paper, where new coalition structures can form only by the disintegration of existing coalitions.

We wish to capture the idea that the interaction between coalitions is noncooperative and that within each coalition, binding agreements, while feasible, must be constrained by consistency considerations. Thus we shall model interaction between coalitions in the spirit of Nash, retaining the feature of cooperation within coalitions. But there is one crucial qualification. Every "equilibrium" of this kind is not necessarily an equilibrium binding agreement. Specifically, such "equilibria" must also be immune to the possibility of defection by a subcoalition. To be sure, the outcomes that defecting subcoalitions can achieve will also be constrained in a consistent way.

We proceed, therefore, in two steps. In the first step we formalize equilibrium noncooperative play across coalitions in a coalition structure. We will say that a strategy vector $x \in X$ *satisfies the best response property* (*relative to* $\mathcal{P}$) if for each coalition $S \in \mathcal{P}$, there is no $y_S \in X_S$ with $u_S(y_S, x_{-S}) \gg u_S(x)$. We shall denote by $\beta(\mathcal{P})$ the set of such strategy profiles. Observe that strategy vectors satisfying the best response property do not permit outcomes that require precommitment across coalitions. At the same time, by allowing each coalition to choose a (restricted) Pareto optimal outcome, they permit cooperation *within* coalitions.

Consider two coalition structures $\mathcal{P}$ and $\mathcal{P}'$, with $\mathcal{P}' \in \mathcal{R}(\mathcal{P})$. Think of having "moved" from $\mathcal{P}$ to $\mathcal{P}'$ by the formation of one or more new coalitions, each a subset of some element of $\mathcal{P}$. Some of these coalitions may be thought of as "active movers", or *perpetrators*, in the creation of $\mathcal{P}'$, and others might be residual coalitions, or simply *residuals*, of individuals left behind by the perpetrators. Observe that we cannot uniquely identify a

class of perpetrators. But we *can* say this: if a coalition in $\mathscr{P}$ breaks into n new coalitions, $n-1$ of them must be labeled perpetrators, and the remaining coalition must be taken to be a residual. A *collection of perpetrators and residuals in the move from $\mathscr{P}$ to $\mathscr{P}'$* is any labeling of the relevant elements of $\mathscr{P}'$ which satisfies the requirement in the previous sentence.

Let $\mathscr{P}$ and $\mathscr{P}'$ with $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$ be given. Fix a collection of perpetrators and residuals in the move from $\mathscr{P}$ to $\mathscr{P}'$. A *re-merging* of $\mathscr{P}'$ is a coalition structure $\hat{\mathscr{P}}$ formed by merging any collection of perpetrators with their respective residuals. Below, this will be used to capture situations in which some perpetrators contemplate not moving to $\mathscr{P}'$.

We now recursively define equilibrium binding agreements. We will denote by $\mathscr{B}(\mathscr{P})$ the set of equilibrium binding agreements for a coalition structure $\mathscr{P}$. We begin with the finest possible coalition structure, $\mathscr{P}^*$, of singleton coalitions. In this case, $\beta(\mathscr{P}^*)$ is just the set of Nash equilibria of the game and $\mathscr{B}(\mathscr{P}^*) = \beta(\mathscr{P}^*)$.

Next, consider coalition structures $\mathscr{P}$ which have $\mathscr{P}^*$ as their only refinement. Let $x \in \beta(\mathscr{P})$. Say that $(\mathscr{P}^*, x^*)$ *blocks* $(\mathscr{P}, x)$ if $x^* \in \mathscr{B}(\mathscr{P}^*)$ and there exists a perpetrator S such that $u_S(x^*) \gg u_S(x)$.

Recursively, suppose that for some $\mathscr{P}$ the set $\mathscr{B}(\mathscr{P}')$ has already been defined for all $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$. Moreover, assume that for each $x' \in \beta(\mathscr{P}')$ we have defined all $(\mathscr{P}'', x'')$ that block $(\mathscr{P}', x')$.

Let $x \in \beta(\mathscr{P})$. We will say that $(\mathscr{P}, x)$ is *blocked* by $(\mathscr{P}', x')$ if $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$, and there exists a collection of perpetrators and residuals in the move from $\mathscr{P}$ to $\mathscr{P}'$ such that

(B.1) x' is a binding agreement for $\mathscr{P}' : x' \in \mathscr{B}(\mathscr{P}')$.

(B.2) There is a *leading perpetrator* S which gains from the move: $u_S(x') \gg u_S(x)$, and

(B.3) Any re-merging of the *other* perpetrators is blocked by $(\mathscr{P}', x')$ as well, with one of these perpetrators as a leading perpetrator. Formally, let $\mathscr{T}$ be the set of all perpetrators, other than S, in the move from $\mathscr{P}$ to $\mathscr{P}'$. Let $\hat{\mathscr{P}}$ be a coalition structure formed by merging some of the elements of $\mathscr{T}$ with their respective residuals.[9] Then $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$ and there is $\hat{x} \in \beta(\hat{\mathscr{P}})$ and $S' \in \mathscr{T}$, such that $(\hat{\mathscr{P}}, \hat{x})$ is blocked by $(\mathscr{P}', x')$ with S' as the leading perpetrator.

Note that the notion of blocking itself appears in (B.3), which is why a recursive definition of blocking is needed as well.

We may now complete the recursion. A strategy profile x is an *equilibrium binding agreement for* $\mathscr{P}$ if $x \in \beta(\mathscr{P})$ and there is no $(\mathscr{P}', x')$

[9] Of course, $\hat{\mathscr{P}} \in \mathscr{R}(\mathscr{P})$.

that blocks $(\mathscr{P}, x)$. Denote by $\mathscr{B}(\mathscr{P})$ the set of all equilibrium binding agreements for $\mathscr{P}$.

Thus, objections or blocks are defined perfectly consistently. A perpetrator can only expect to induce some binding agreement in some refinement of the coalition structure $\mathscr{P}$ (and such agreements are well-defined by our recursive procedure). Moreover, if this refinement involves the defection of *other* subcoalitions, conditions must be imposed that make it worthwhile for such coalitions to have defected. (B.3) captures this. To see this, observe that a re-merging partially reverses the defection process, returning to intermediate coalition structures of the form $\hat{\mathscr{P}}$. What (B.3) states is that each such merger should lack the ability to write equilibrium binding agreements, and moreover that there is some allocation with the best response property (relative to $\hat{\mathscr{P}}$) which is blocked by the original defection(s).

Note that the re-merging always excludes the leading perpetrator, and indeed in the rest of the paper, the term "re-merging" will always be taken to exclude the leading perpetrator.

Typically, many coalition structures admit equilibrium binding agreements. Which of these should be considered as *the* set of equilibrium binding agreements for the *game*? The answer to this question depends on what we consider to be the "initial" coalition structure under which negotiations commence. In keeping with the spirit of our exercise, which is to understand the outcomes of free and unconstrained negotiation, we take it that the initial structure is the grand coalition itself. Under this supposition, it is natural to focus on the set of binding agreements for the grand coalition, or, if this set is empty, on the next level of refinement for which the set of equilibrium binding agreements is non-empty.

The following remarks highlight various aspects of the definition.

Remark 2.1. 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. Note that this optimism on the part of the leading perpetrator is consistently mirrored in the presumed pessimism of the other perpetrators (see condition (B.3)).[10]

Clearly, there are alternatives to optimism. Observe that there are two components here: a leading perpetrator feels (i) that a coalition structure will be formed (subject to the described constraints) that is best from its point of view; and (ii) that an equilibrium will be played under this structure which is also best from its point of view. Thus versions of our

[10] This is in the sense that one may view the pessimism of the other perpetrators as an optimistic conjecture by the leading perpetrator regarding their behavior.

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 a coalition should be so pessimistic as to focus on the least desirable of these equilibria for them. In this sense, (ii) interacts with (i), and a proper specification of (ii) is required so as not to eliminate "reasonable" coalition structures following a deviation. 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.[11] In general, though, we realize that there is very little to be said about choosing from among these alternatives, and so proceed with one of them (see Greenberg [17] for a discussion of these issues in a general context). We return to this issue briefly in Section 4.

Remark 2.2. Suppose $(\mathscr{P}', x')$ blocks $(\mathscr{P}, x)$ with S^1 as a leading perpetrator. Suppose there are several other perpetrators as well. Let $\mathscr{T} = \{S^2, ..., S^m\}$ be the set of other perpetrators. Condition (B.3) in our definition requires, in particular, that even if several perpetrators from $\mathscr{T}$ are simultaneously re-merged, the resulting coalition structure is blocked by $(\mathscr{P}', x')$. One can explore several interesting variations on precisely what the leading perpetrator should be allowed to assume regarding the behavior of other perpetrators. While we shall leave a more comprehensive study of this issue to another paper, it will be instructive to consider one variation in which the leading perpetrator suggests a particular sequence in which the other perpetrators move, and at each intermediate step, the final outcome $(\mathscr{P}', x')$ justifies the move to the next step. As we shall see, this form of blocking is implied by our basic definition of blocking. We begin by formally defining a sequential notion of blocking.

Let $x \in \beta(\mathscr{P})$. $(\mathscr{P}', x')$ is said to *sequentially block* $(\mathscr{P}, x)$ if there exists a sequence $\{(\mathscr{P}^0, x^0), (\mathscr{P}^1, x^1), ..., (\mathscr{P}^m, x^m)\}$ such that:

[11] For instance, as a referee (who initiated and clarified this discussion) points out, it is possible to define conjectures where each deviating coalition supposes that it can choose intra-coalitional transfers, but anticipates the worst possible (equilibrium) action from external players. This is only one of many possibilities.

(S.1) $(\mathscr{P}^0, x^0) = (\mathscr{P}, x)$, $(\mathscr{P}^m, x^m) = (\mathscr{P}', x')$ and for every $i = 1, ..., m$, there is a coalition S^i such that S^i is the only perpetrator in the move from $\mathscr{P}^{i-1}$ to $\mathscr{P}^i$. Moreover, for every i, $x^i \in \beta(\mathscr{P}^i)$.

(S.2) $x' \in \mathscr{B}(\mathscr{P}')$.

(S.3) $\mathscr{B}(\mathscr{P}^i) = \varnothing$ for all i such that $0 < i < m$.

(S.4) $u_{S^i}(x') \gg u_{S^i}(x^{i-1})$ for all $i = 1, ..., m$.

Note that this notion of blocking does not require a recursive definition. (Of course, $\mathscr{B}(\mathscr{P})$ still needs to be defined recursively.) It has an explicitly sequential account of how coalitions move, unlike our definition. Nevertheless, our notion of blocking subsumes the sequential notion:

PROPOSITION 2.1. *If* $(\mathscr{P}', x')$ *blocks* $(\mathscr{P}, x)$, *then* $(\mathscr{P}', x')$ *sequentially blocks* $(\mathscr{P}, x)$.

Proof. Suppose $(\mathscr{P}', x')$ blocks $(\mathscr{P}, x)$. Let S^1 be the leading perpetrator in this move. If there are no other perpetrators, then it is clear that $(\mathscr{P}', x')$ sequentially blocks $(\mathscr{P}, x)$ with $m = 1$. Suppose, therefore, that the set of other perpetrators is $\mathscr{T} = \{S^2, ..., S^m\}$. Define $\mathscr{P}^1$ to be the coalition structure obtained by re-merging all other perpetrators in $\mathscr{T}$. Since $(\mathscr{P}', x')$ blocks $(\mathscr{P}, x)$, by condition (B.3), $\mathscr{B}(\mathscr{P}^1) = \varnothing$, and there exists $x^1 \in \beta(\mathscr{P}^1)$ such that $(\mathscr{P}', x')$ blocks $(\mathscr{P}^1, x^1)$, with a leading perpetrator from $\mathscr{T}$. Without loss of generality let this leading perpetrator be S^2. Define $\mathscr{P}^2$ to be the coalition structure obtained by re-merging all perpetrators $S^3, ..., S^m$ with their respective residuals. By appealing to condition (B.3) we can assert that there exists $x^2 \in \beta(\mathscr{P}^2)$ such that $(\mathscr{P}^2, x^2)$ is blocked by $(\mathscr{P}', x')$ with S^3 (say) as a leading perpetrator. In this way we obtain a sequence $(\mathscr{P}, x)$, $(\mathscr{P}^1, x^1), ..., (\mathscr{P}^m, x^m)$ such that (a) $(\mathscr{P}^m, x^m) = (\mathscr{P}', x')$, (b) for every $i = 1, ..., m$, $x^i \in \beta(\mathscr{P}^i)$, (c) for every $i = 1, ..., m$, $(\mathscr{P}^{i-1}, x^{i-1})$ is blocked by $(\mathscr{P}', x')$ with S^i as the leading perpetrator. To complete the proof it suffices to show that the sequence $\{(\mathscr{P}, x), (\mathscr{P}^1, x^1), ..., (\mathscr{P}^m, x^m)\}$ satisfies conditions (S.1)–(S.4). Condition (S.1) is clearly satisfied. Condition (S.2) follows from (B.1). Conditions (S.3) and (S.4) follow from (B.3) and properties (b) and (c). ∎

Remark 2.3. Our definition of binding agreements, in effect, considers the "intersection" of the set of best response strategies for any coalition structure with the set of unblocked strategies for that structure. Alternatively, one might wish to consider the set of best response strategies *subject to the condition that they not be blocked.* Because we insist on Pareto optimality (and no more) in the definition of best responses, it is easy to see that these two approaches yield equivalent outcomes. But this equivalence fails to hold in cases where best responses are defined by additional

considerations (e.g., Nash bargaining within coalitions, or the equal split of surplus over threat points). Extensions to these cases will proceed through the introduction of a generalized game.[12]

Remark 2.4. Under reasonable assumptions, the possible emptiness of the set of binding agreements for any coalition structure will stem from the blocking of best response strategies, *but not from the more technical* (*and less interesting*) *consideration that the set of best responses* $\beta(\mathscr{P})$ *is empty*. For instance, the set $\beta(\mathscr{P})$ for the partition of singleton coalitions is simply the set of Nash equilibria of the game. For the grand coalition, it is simply the Pareto frontier of the game. For this paper, we are not interested in situations where *these*, and related sets are empty. For completeness, the following Proposition guarantees conditions under which $\beta(\mathscr{P})$ is non-empty for each structure $\mathscr{P}$.

PROPOSITION 2.2. *Suppose for all* i, X_i *is non-empty, compact, convex and* u_i *is continuous and quasi-concave. Then* $\beta(\mathscr{P}) \neq \varnothing$ *for all* $\mathscr{P} \in \Pi$.

Proof. Consider $\mathscr{P} \in \Pi$. For every $S \in \mathscr{P}$ define a preference ordering on X as follows. For a coalition S, the "better than set" relative to $x \in X$ is defined as

$$P_S(x) = \{ y_S \in X_S \mid u_S(y_S, x_{-S}) \gg u_S(x) \}.$$

By quasi-concavity of u_i it follows that $P_S(x)$ is convex for all S and x. It is easy to see that while this does not define a complete ordering, the graph of P_S is open. We can, therefore, appeal to the existence result of Shafer and Sonnenschein [32] to assert that there exists $\bar{x}$ such that $\bar{x}_S$ is a best response of S to $\bar{x}_{-S}$ for all $S \in \mathscr{P}$, i.e., $\bar{x} \in \beta(\mathscr{P})$. ∎

3. RELATIONSHIP TO THE LITERATURE

Various aspects of our equilibrium notion are related to existing literature. We discuss these connections under the following headings.

Extensions of the Characteristic Function

In response to some of the considerations that motivate this paper, characteristic function forms have been extended to partition function

[12] This will require a notion of a generalized game (along the lines of Debreu [12], Ichiishi [20] and Shafer and Sonnenschein [32]) in which each coalition has a constraint set which, depending on the prevailing coalition structure, consists of strategies that are unblocked. Of course, the definition will again proceed by recursion.

forms (Lucas [22], Thrall and Lucas [34] and Lucas and Maceli [22])[13] A partition function is based on the idea that the worth of a coalition depends on the *entire* coalition structure of which that coalition is a part. But partition functions themselves represent a reduced form, and do not involve the complex consistency relationships that are inherent in starting from the normal form.[14] In any event, the normal form represents the primitive specification of such games and agreements in cartels or partnerships are typically written over *strategies* rather than over a division of the aggregate *payoff*. Accordingly, we derive our solution concept directly from the normal form.

Coalition Formation

Consider a game given in characteristic function or partition function form. Suppose that there is a *given* rule for payoff division within a coalition. One can then ask the question: which coalition structures will form? In Hart and Kurz [19] and Aumann and Myerson [2] the division rule is based on the Shapley value. This induces a game in partition function form, and coalition formation can now be studied in this partition function.

Hart and Kurz [19] study the stability of coalition structures in a fresh normal form game derived from the above partition function form. They are explicit in their motivation for doing so:

> ... dynamic theories usually rely on additional (arbitrary) assumptions (in our case, for example, the order in which players 'talk' to one another) which significantly affect the outcome. Our stable coalition structures may be regarded as 'universal' outcomes, independent of the specification of the process.

They consider two notions of coalitional stability based on the concept of a strong equilibrium. Their notion of δ-stability corresponds to a strong equilibrium of a game in which a deviation by a coalition $T \subset S$ leaves $S \setminus T$ as a residual, and all other coalitions remain unchanged. Their notion of γ-stability is based on the idea that when a coalition $T \subset S$ deviates, the members of $S \setminus T$ break up into singletons, while all other coalitions remain the same.[15]

A key difference between the Hart–Kurz approach and ours is that they do not address the consistency issue. An advantage of their formulation is that the formation of arbitrary coalitions is permitted, not just those which are subsets of coalitions in the existing coalition structure. As in Bernheim,

[13] See also Rosenthal [29] on the related notion of effectiveness forms.

[14] Moreover, this earlier literature seeks generalizations of the vN-M solution or the Shapley value, which is not of concern in this paper. It should be said, moreover, that in introducing some of these generalizations, notions of the kind involved in constructing an α-characteristic function were reintroduced.

[15] They consider other equilibrium notions as well, based on the α-core and the β-core.

Peleg and Whinston [4], we permit only internal deviations. The problem lies not only with the conceptual extension to a more general case, but also possibly with existence: in Hart and Kurz [19], as also in Shenoy [33], a stable structure may not exist for precisely this reason.

Aumann and Myerson [2] study coalition formation through an extensive form game where a given rule of order specifies the sequence in which players are allowed to form links (coalitions). There are restrictions on the players who can move after a certain node. In particular, once a coalition has formed, it is not allowed to break up. This ensures that the extensive form game is finite and, therefore, possesses a subgame perfect equilibrium. Bloch [5] considers a similar model in partition function form, which is more explicitly one of coalition formation rather than a linking game. This approach is also explored in Ray and Vohra [28], where coalition formation *and* the payoff division between players in a coalition are determined endogenously and simultaneously, in the context of a bargaining game.

In an extensive form game, the consistency notion—looking ahead at resulting equilibria—is, of course, interpretable as subgame perfection. However, as these authors point out, the subgame perfect equilibria of their extensive form games depend on the exogenously defined rule of order. The fact that arbitrary details in the specification of the extensive form have a significant influence on the equilibria is, of course, well known. We are able to sidestep this problem since we do not model consistency through an extensive form.[16] In this respect, our treatment of consistency is analogous to that of Bernheim, Peleg and Whinston [4], Chwe [10] and Greenberg [17]. For a more detailed comparison of our present approach with one that relies on an extensive-form description, we refer the reader to Ray and Vohra [28].

Another difference between these papers and ours is that we start the analysis *explicitly* from a game in strategic form. In a sense, our approach can be divided into two parts. First, we *derive* a partition function (simply by considering for each partition the set of strategy profiles satisfying the best response property). Our second step may be regarded as using this partition function to provide an analysis of coalition structures. This second phase can be compared to the cited papers. In general, this two-step process has the drawback that it separates the question of *what* a coalition can achieve from the question of *which* structures might form. As we pointed out in Remark 2.3, when the internal behavior of coalitions is specified as a (constrained) best response, the two-step procedure is valid.[17]

[16] This also seems to have been a motivation of Hart and Kurz [19] in avoiding the specification of a "dynamic process" as discussed above.

[17] In Section 6 we introduce a class of games in which the separation of these two issues is even more stark.

Moreover, our definition can readily be modified, as indicated in Remark 2.3, to allow for other internal rules by imposing the given rule over those strategies of a coalition that are immune to blocking.

Noncooperation Across Coalitions

Ichiishi's [20] notion of noncooperative play between coalitions is similar to the one we develop in this paper. He studies the existence of an equilibrium notion that is based on the idea underlying the Strong Nash equilibrium. Zhao [37] considers a similar model and considers equilibria in which each coalition is constrained by the requirement that it choose elements of *its* α-core. There are also several papers on economic models with externalities in which the "equilibrium" outcome corresponding to a coalition structure is a best response strategy profile ($\beta(\mathscr{P})$ in the language of the present paper). For example, Dutta and Suzumura [16] study coalition formation in a model of research joint ventures. However, they assume that when a coalition deviates, the remaining coalitions do not alter in any way, as in the Hart–Kurz notion of δ-stability. Carraro and Siniscalco [7] study a similar equilibrium concept in a model with international pollution. Chander and Tulkens [9] analyze the pollution problem assuming that when a coalition deviates, the rest of the players disintegrate into singletons, as in the Hart–Kurz notion of γ-stability. Not surprisingly, the results depend on the specific assumption regarding the new coalition structure that emerges following a deviation. In this paper, in contrast, the notion of a best response strategy is just one of the building blocks in the definition of equilibrium: what a coalition can achieve is limited by consistency considerations that are fully incorporated in the equilibrium concept.

While it will be interesting to apply our notion of equilibrium binding agreements to these economic models, in the present paper we shall confine ourselves to two economic applications: public goods and a Cournot oligopoly.[18]

Consistent Self-Enforcing Agreements

This notion has been formalized in Bernheim, Peleg and Whinston [4]. Despite the common concern for consistency, there is a fundamental difference between our concept and that of *coalition-proof Nash equilibrium* (CPE), introduced by these authors. The latter applies only to games where *no* binding agreements are possible (CPE are always Nash). On the other hand, our concept applies only to models where binding agreements

[18] We note, however, that similar considerations of consistency have already been applied to a model of customs unions by Yi [36] (this study also uses our solution concept), and to research joint ventures by Bloch [6].

are possible. Our solutions are not subsets of Nash equilibria and do not involve passive behavior on the part of complementary coalitions. For example, in the Prisoner's dilemma the non-cooperative outcome is the unique CPE, while the only equilibrium binding agreement for the grand coalition is the cooperative outcome. For a three-person example in which the set of equilibrium binding agreements is disjoint from the CPE, we refer the reader to Section 3 of an earlier version of this paper, Ray and Vohra [27].[19]

The difference is formally captured by the specification of what follows a coalitional deviation. Imagine two players at a negotiating table. If no binding agreements are possible, the CPE are precisely those Nash equilibria which are Pareto-optimal in the class of all Nash equilibria. If binding agreements are possible, the word "deviation" translates as "breaking-off of negotiations." In this example, the environment then shifts to the coalition structure of the two players acting independently (two singleton coalitions). When one of the two players "deviates," there is no question of taking the other player's strategy as given, as in CPE. Therefore, it is imperative to explicitly model the "game" that results after the deviation, and use the "equilibria" of this game to determine the limits of the original negotiation process.

Social Situations

Greenberg [17] develops the theory of "social situations." This very useful classification scheme relates numerous solution concepts (such as subgame perfection, coalition-proof Nash equilibrium, the core, and others) to one another, by viewing each of these concepts as inducing a partial order on a space of possible outcomes, and studying the von-Neumann-Morgenstern stable set with respect to that order. The present concept can possibly be embedded within this general taxonomy, but this does not yield anything new, either conceptually or in the way of general results. In this context, it might be useful to summarize what a reasonable theory of binding agreements should incorporate. In our opinion, there are two features.

First, the theory should predict (under conditions of complete and perfect information), an efficient outcome for a two-person game. For instance, if binding agreements can be written, then in the Prisoner's dilemma the cooperative outcome should be the only equilibrium.

[19] There is, however, a special case in which the set of binding agreements has a nonempty intersection with CPE. Suppose there exists a strong Nash equilibrium, x, and that there is a unique best response equilibrium in every coalition structure other than the grand coalition (this must, of course, be x). Then it is easy to see that x is an equilibrium binding agreement for every coalition structure. Of course, x is also a CPE.

Second, any theory of binding agreements must have a precise description of counterfactuals: most importantly, what happens if the grand coalition disintegrates. Our theory rests on the idea that in such cases, we predict noncooperative play across coalitions, coupled with attempts to cooperate within coalitions.

These two features are absent in the theory of social situations. Indeed, it is unfair to that theory to expect them to be present, because the taxonomy is intentionally designed to be far more abstract and general (as can easily be seen from the plethora of different solution concepts that come under it).

4. INEFFICIENCY

Our definition permits the writing of *any* agreement to which players can jointly agree. Indeed, implicit in our formulation is the idea that any outcome can, in principle, be costlessly precommitted to. The main theme of our paper is that despite this ability, inefficient outcomes are possible. In later sections of this paper, we shall argue that such inefficiency crops up in natural *economic* contexts. Our goal in this section is to record the fact that such situations are robust to *arbitrary* small perturbations of the underlying game, unrestricted in any way by the underlying economic context.

In keeping with the spirit of our discussion in Section 2 (see Remark 2.1), we are unwilling to seriously consider instances of inefficiency that rely *solely* on extreme optimistic views of blocking. Such examples rely on the multiplicity of equilibria following a block, each coalition optimistically anticipating the equilibrium most beneficial to it.[20] They will not be robust to reasonable alternative definitions that rely on a lesser degree of optimism.

We therefore demand of inefficient outcomes that they be compatible with a *unique* best response strategy for coalition structures that are involved in acts of blocking (modulo possible transfers of utility among players in the same coalition that leave the complement unaffected). In this way, one rules out a notion of blocking that relies on a deviating coalition

[20] Consider, for example, a modified version of the battle of the sexes. There are two players, and two pure Nash equilibria yielding payoff vectors of (5, 1) and (1, 5). Suppose, now, that there are (non-Nash) payoffs that Pareto-dominate either of these two outcomes, but do not dominate the vector (5, 5). This game has no efficient binding agreement. But this is so solely because of the assumption of optimism, and we discard such candidates as serious examples of inefficiency.

gaining in *some* (but not every) coalition structure that it can induce. Thus, in our construction in the proof of the proposition below, each perpetrator will be able to induce only one possible coalition structure.

Fix the number of players and a (finite) space of strategies for each. Then the set of games (payoff functions) may be identified with an appropriate Euclidean space, open sets of payoff functions being identified with open sets in this space.

PROPOSITION 4.1. *Suppose that there are at least three agents with at least three strategies each. Then there exists an open set of games such that for every game in this set, there is (in terms of payoffs) at most one binding agreement under each coalition structure. Moreover, no binding agreement is efficient.*

Remark 4.1. We have already remarked on the robustness of this proposition to various degrees of optimism. Note, moreover, that in a three-player game there is no difference between sequential blocking and blocking. From our proof of Proposition 4.1 it will be clear that this result remains valid for a notion of binding agreements based on sequential blocking. We therefore also have "robustness" in this additional sense.

Remark 4.2. Suppose there is a unique Nash equilibrium or there exists a strategy profile that Pareto dominates every Nash equilibrium. Then it is easy to see that in any two-player game, *every* equilibrium binding agreement must necessarily be efficient; this follows simply from the natural superadditive structure of a two-player normal form game.[21] To demonstrate inefficiency, therefore, we need to construct a game with at least three players.

Remark 4.3. The proposition holds under the condition that there are at least three strategies for at least three of the players. This assumption cannot be dropped free of charge, but we do not know whether some weakening is possible.

Proof of Proposition 4.1. We start with the case of three players and exactly three strategies of each of the players. The remaining cases are treated by extending this argument; we outline the extension below.

Consider the following normal form. Player 1 chooses rows, player 2 chooses columns and player 3 chooses matrices.

[21] It is worth mentioning, in passing, that in general, superadditivity breaks down with more than two players, though such breakdowns are not needed for inefficiency. The Cournot example in Section 7 illustrates this breakdown.

x_{3a}

	x_{2a}	x_{2b}	x_{2c}
x_{1a}	2.6, 2.6, 2.6	3.2, 2.2, 3.2	3.7, 1.7, 3.7
x_{1b}	2.2, 3.2, 3.2	2.7, 2.7, 3.7	3.1, 2.1, 4.1
x_{1c}	1.7, 3.7, 3.7	2.1, 3.1, 4.1	2.6, 2.6, 4.6

x_{3b}

	x_{2a}	x_{2b}	x_{2c}
x_{1a}	3.2, 3.2, 2.2	3.7, 2.7, 2.7	4.1, 2.1, 3.1
x_{1b}	2.7, 3.7, 2.7	3.1, 3.1, 3.1	3.6, 2.6, 3.6
x_{1c}	2.1, 4.1, 3.1	2.6, 3.6, 3.6	2.9, 2.9, 3.9

x_{3c}

	x_{2a}	x_{2b}	x_{2c}
x_{1a}	3.7, 3.7, 1.7	4.1, 3.1, 2.1	4.6, 2.6, 2.6
x_{1b}	3.1, 4.1, 2.1	3.6, 3.6, 2.6	3.9, 2.9, 2.9
x_{1c}	2.6, 4.6, 2.6	2.9, 3.9, 2.9	3.3, 3.3, 3.3

We claim that in this example there is no efficient equilibrium binding agreement, and that the grand coalition breaks up into an intermediate coalition structure. Notice first that every player i has a dominant strategy, x_{ia}. Thus the unique Nash equilibrium, and the only equilibrium binding agreement for $\mathscr{P}^*$, is (x_{1a}, x_{2a}, x_{3a}), which is Pareto dominated by (x_{1c}, x_{2c}, x_{3c}).

Next, we examine the equilibrium binding agreements for an intermediate coalition structure $\mathscr{P} = (\{i\}, \{j, k\})$. Since the game is symmetric, there is no loss of generality in considering the coalition structure $\mathscr{P} = (\{1\}, \{2, 3\})$. Since player 1's dominant strategy is x_{1a}, any $z \in \beta(\mathscr{P})$ must be such that $z_1 = x_{1a}$. Thus we need only look at the first row of each matrix. Clearly, both (x_{2a}, x_{3a}) and (x_{2c}, x_{3c}) are dominated by (x_{2b}, x_{3b}). In fact, it is easy to see that $(x_{1a}, x_{2b}, x_{3b}) \in \beta(\mathscr{P})$. Moreover, this strategy cannot be blocked by a deviation to $\mathscr{P}^*$. It is, therefore, an equilibrium binding agreement. Indeed, this is the only one for this coalition structure. To see this, notice that in all other best response equilibria, either player 2 or player 3 receives less than 2.6, the unique Nash payoff. Since the game is symmetric, we can now claim that for every intermediate coalition structure $(\{i\}, \{j, k\})$, the only equilibrium strategy profile is (x_{ia}, x_{jb}, x_{kb}). The payoffs to i, j and k are 3.7, 2.7 and 2.7 respectively. But this outcome is not efficient. It is Pareto dominated by (x_{ib}, x_{jc}, x_{kc}).

Finally, consider the grand coalition. For any strategy profile it must be the case that there exists a player, i, who gets less than 3.7. This player can then block this proposal by deviating to $(\{i\}, \{j, k\})$ and earning 3.7. This in fact, is the *only* coalition structure that i can induce by deviating from the grand coalition. Thus, the grand coalition breaks up into some intermediate coalition structure with an inefficient equilibrium. And the only equilibrium in the finest coalition structure too is inefficient.

Since all the best response equilibria are strict, it follows that this example is robust. An open set of payoff functions that yield the same qualitative outcome can therefore be constructed.

We now extend this example to consider additional strategies and/or additional players. First, consider the case of additional strategies for any of these three players. For each player, say i, and each such strategy, say x_{id}, let the payoff to *every* player be zero whenever x_{id} is played. This specification guarantees that such additional strategies are irrelevant, and we continue to have a robust set of games satisfying the properties of the proposition.

Finally, consider additional players. Take any such player $j \geqslant 4$. Define his payoff from any strategy vector to be 1 provided (x_{1c}, x_{2c}, x_{3c}) is played by the first three players, and zero otherwise. For the first three players, treat all such additional players as dummies, holding to the payoff matrix described above regardless of the actions of the additional players. It is easy to see that, in terms of payoffs, this modification does not change the earlier conclusions. This completes the proof of the proposition. ∎

Recall that the three-person example used in the proof of this proposition did not allow for utility transfers among players within a coalition. As a referee pointed out, it is natural to ask whether this feature is critical for the result. The answer is no. It can be shown that the conclusion remains unchanged even if such transfers are permitted. In fact, in the next section we shall study in some detail an interesting economic model in which such utility transfers are permitted, and one in which it is easy to generate three-person examples with the same basic properties as the one discussed above.

It is useful to end this section with an intuitive description of the factors that drive (and limit) our inefficiency result. We have here a three-person structure where a *single* player, say i, by inducing the coalition structure consisting of just himself in one coalition, and the other two players in another, can do better than the average payoff to a player in the (efficient) grand coalition. The consequent inefficiency hinges on the fact that in such a case, the other two players are better off staying together than also breaking apart.[22] Observe that while the payoffs to the three potential singleton deviants jointly dominate the outcome that can be achieved by the grand coalition, the game is still superadditive in the sense that the grand coalition can still Pareto-dominate each inefficient outcome. The point is that the dominating outcome will need to be changed for each inefficient outcome.

Nevertheless, one might ask, why are matters not renegotiated at this stage to the dominating (yet unequal) outcome (x_{ib}, x_{jc}, x_{kc})? This is a

[22] A reading of the Cournot example in Section 7 will show that the absence of this feature is what *maintains* efficiency in the three-firm case. There, one firm can deviate profitably as well, provided the other two stay together, but this latter event will not occur. Thus inefficiency in the Cournot example does not arise until a minimum of five firms is present.

serious issue that is neglected in our model, because we only permit "internal" deviations (see the Introduction). However, one should note that matters cannot end there, as further negotiation may follow on return to this outcome (now firm j might object). At any rate, a satisfactory study of such issues requires an explicit consideration of the dynamics of negotiation.

5. PUBLIC GOODS ECONOMICS: BINDING AGREEMENTS AND FREE RIDERS

Consider a public goods economy. The free rider problem for such an economy is well known. If agents make voluntary contributions towards the production of the public good, the equilibrium outcome will not, in general, be Pareto optimal. Is it possible to sustain a Pareto optimal allocation in such an economy? Without exception, the literature addresses this question under the assumption that individual characteristics are imperfectly observed, analyzing game forms or mechanisms to implement socially desirable outcome (see, for example, Groves and Ledyard [18] and Walker [35]).

An implicit presumption underlies such an approach. *The inability to enforce Pareto optimal outcomes is taken to be a consequence of incomplete information.* In other words, if all characteristics are commonly known, the issue of attaining *some* Pareto optimal outcome can be trivially resolved. We will argue that there is a difficulty with this presumption. In the process, we obtain a natural application of our theory of binding agreements.

The aim of this section, then, is to characterize equilibrium binding agreements in a simple public goods economy. We shall demonstrate that in general, there do not exist efficient equilibrium binding agreements.

We consider a very simple economy with a single public good. There are n identical agents. Each agent owns one unit of a private good x. The good may be used to produce a pure public good y, according to some linear production function f. We will find it more convenient to represent the technology through the linear cost function cy (the inverse of f). Each person derives linear utility from x and strictly concave utility from y. We write the utility function of agent i as $u_i(x_i, y) = g(y) + x_i$, where g is an increasing, strictly concave function. Assume that g satisfies natural endpoint conditions, so that unbounded production of the public good is never optimal.

We will use lower case letters to denote the cardinality of coalitions. For example, for a grand coalition N and a coalition S, n and s refer to the number of agents in N and S respectively.

If there is a nontrivial coalition structure, then each coalition makes its own decision about how much to contribute towards the public good. It

will be useful to start with a description of what a coalition S with cardinality s would produce *in isolation*. Note that for any such coalition, an allocation is efficient if and only if production levels of the public good maximize $sg(y)-cy$. For each s, let $y(s)$ maximize $sg(y)-cy$. By our assumptions on c and g, $y(s)$ is well-defined for each s, and if $y(s)>0$, it is characterized by the familiar condition $sg'(y(s))=c$. Note that $y(s)$ is an increasing function of s. The *per capita* payoff to coalition S is then given by

$$a(s) \equiv g(y(s)) + 1 - \frac{cy(s)}{s}.$$

It is easy to see that for two coalitions of cardinalities s and t, with $s>t$, $a(s)>a(t)$ whenever $y(s)>0$. So larger coalitions have higher *per capita* payoffs, revealing an increasing returns property that models of public goods quite naturally possess.

Now we will describe best response strategy vectors and payoffs for each coalition structure $\mathscr{P}$. It is easy to see that each coalition $T \in \mathscr{P}$ will make a contribution so that, given the other contributions, the production of the public good is no less than (but as close as possible to) $y(t)$. If the complement contributes z, coalition T will contribute $\max(cy(t)-z, 0)$. If $\mathscr{P}$ has a *unique* maximal coalition of size s, given that $y(s)$ is increasing in s, it follows that the maximal coalition is the only one that contributes towards the production of the public good. Thus, there is a unique per capita payoff accruing to each coalition in $\mathscr{P}$ under any strategy vector with the best response property: the maximal coalition carries out *its* optimal production, earning a *per capita* return of $a(s)$, while all other coalitions free-ride, earning a *per-capita* payoff $a^f(s)=g(y(s))+1>a(s)$.

If $\mathscr{P}$ has more than one maximal coalition (say of size s), then there are many possible *per capita* payoff vectors. Non-maximal coalitions continue to be free riders, earning a per capita payoff $a^f(s)$, while a maximal coalition earns a *per capita* payoff anywhere in the interval $[a(s), a^f(s)]$, depending on how much of the cost of provision is borne by this coalition in equilibrium.[23]

Recall that larger coalitions are more efficient than smaller coalitions.[24] The characterization of best response equilibria in the last two paragraphs now yields the conclusion that if $\mathscr{P} \neq \{N\}$, then $\mathscr{B}(\mathscr{P})$ cannot be efficient. To establish inefficiency, therefore, it will suffice to show that the grand coalition will write no binding agreement.

[23] Of course, these latter payoffs are not "independent" over the maximal coalitions. In each equilibrium, the total cost of provision must be borne by the union of the maximal coalitions.

[24] As seen above, this is true under the mild requirement that the efficient level of provision of the public good is positive in the larger coalition.

We will show that, given the utility functions and the technology, the stability of the grand coalition depends crucially on the size of the economy. We begin by proving that the stability of a coalition structure, say $\mathscr{P}$, is closely related to the stability of a coalition structure that consists of $\mathscr{P}$ along with an arbitrary number of singleton coalitions. To do this we need some additional notation.

We begin by abusing existing notation. In this section, we take $\mathscr{B}(\mathscr{P})$ and $\beta(\mathscr{P})$ to be the set of *payoffs* corresponding to an equilibrium strategy profile of $\mathscr{P}$ (and not the strategy profiles themselves).

For any coalition structure $\mathscr{P}$ and any set of additional agents $K=\{k_1, ..., k_m\}$ who do not belong to a coalition in $\mathscr{P}$, we denote by $\mathscr{P}^K$ the coalition structure consisting of all coalitions in $\mathscr{P}$ and m singleton coalitions containing the agents in K, i.e.,

$$\mathscr{P}^K=\{\mathscr{P}, \{k_1\}, ..., \{k_m\}\}.$$

If $\mathscr{P}$ consists of a single coalition S, then use S^K to denote the coalition structure consisting of S and all members of K as singletons.

For $z\in\mathscr{B}(\mathscr{P})$ we use z^K to denote the feasible utility profile in $\mathscr{P}^K$, where $z_i^K=z_i$ for all $i\in\mathscr{P}$ and $z_j^K=a^f(s)$ for all $j\in K$, where s is the size of a maximal coalition in $\mathscr{P}$.

LEMMA 5.1. *Fix some coalition structure $\mathscr{P}$. Then*

(1) *For any positive integer K, $(\mathscr{P}', z')$ blocks $(\mathscr{P}, z)$ with S as the leading perpetrator if and only if $(\mathscr{P}'^K, z'^K)$ blocks $(\mathscr{P}^K, z^K)$ with S as the leading perpetrator.*

(2) *$z\in\mathscr{B}(\mathscr{P})$ if and only if $z^K\in\mathscr{B}(\mathscr{P}^K)$.*

(3) *$\mathscr{B}(\mathscr{P})=\varnothing$ if and only if $\mathscr{B}(\mathscr{P}^K)=\varnothing$.*

Proof. We will proceed by induction. For $\mathscr{P}=\mathscr{P}^*$, the lemma is trivially true as there are no subpartitions to consider, and because $z\in\beta(\mathscr{P})$ if and only if $z^K\in\beta(\mathscr{P}^K)$. Consider, then, a coalition structure $\mathscr{P}$ and suppose the lemma holds for all refinements of $\mathscr{P}$. We show that it holds for $\mathscr{P}$.

Pick $\mathscr{P}'$ with $\mathscr{P}'\in\mathscr{R}(\mathscr{P})$, payoff vectors z, z' with $z\in\beta(\mathscr{P})$, and any positive integer K.

Suppose, first, that $(\mathscr{P}', z')$ blocks $(\mathscr{P}, z)$. Fix a collection of perpetrators and residuals. Form the coalition structures $\mathscr{P}^K$, $\mathscr{P}'^K$ and payoff vectors z^K, z'^K. Assign the same collection of perpetrators and residuals. We will show that $(\mathscr{P}'^K, z'^K)$ blocks $(\mathscr{P}^K, z^K)$. To do so, we must check conditions (B.1)–(B.3).

That $(\mathscr{P}'^K, z'^K)$ satisfies (B.1) follows right away from the induction hypothesis applied to part [2] of the lemma, and the fact that $(\mathscr{P}', z')$ satisfies (B.1). Next, note that for the leading perpetrator S, $z'_S = z'^K_S$ and $z_S = z^K_S$, so that (B.2) is satisfied as well. To check (B.3), consider any re-merging of perpetrators in $\mathscr{P}'^K$, leading to the coalition structure $\hat{\mathscr{P}}^K$. The corresponding re-merging of perpetrators in $\mathscr{P}'$ leads to $\hat{\mathscr{P}}$, of course. Because $(\mathscr{P}', z')$ blocks $(\mathscr{P}, z)$, $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$. By induction applied to part [3] of the lemma, $\mathscr{B}(\hat{\mathscr{P}}^K) = \varnothing$ as well.

Finally, we are to show that (B.3) holds for the re-merging. Since $(\mathscr{P}', z')$ blocks $(\mathscr{P}, z)$, and $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$, it follows that there exists $\hat{z} \in \beta(\hat{\mathscr{P}})$ such that $(\mathscr{P}', z')$ blocks $(\hat{\mathscr{P}}, \hat{z})$, with one of the original perpetrators as a leading perpetrator in the move from $\hat{\mathscr{P}}$ to $\mathscr{P}'$. Then by the induction hypothesis applied to part [1] of the lemma, $(\mathscr{P}'^K, z'^K)$ blocks $(\hat{\mathscr{P}}^K, \hat{z}^K)$, so that (B.3) holds for $\hat{\mathscr{P}}^K$.

The converse argument is very similar. Suppose that $(\mathscr{P}'^K, z'^K)$ blocks $(\mathscr{P}^K, z^K)$. Fix a collection of perpetrators and residuals. Since all agents in K are singletons, none of them can be perpetrators or residuals. We can, therefore, assign the same collection of perpetrators and residuals in the corresponding move from $\mathscr{P}$ to $\mathscr{P}'$. To show that $(\mathscr{P}', z')$ blocks $(\mathscr{P}, z)$, we must check conditions (B.1)–(B.3). The arguments are very close to those just described and we omit the details. This verifies part [1] of the lemma for $\mathscr{P}$.

To verify [2], let $z \in \mathscr{B}(\mathscr{P})$. Suppose that $z^K \notin \mathscr{B}(\mathscr{P}^K)$. It is certainly the case that $z^K \in \beta(\mathscr{P}^K)$, so the previous sentence means that there exists $\mathscr{P}'^K \in \mathscr{R}(\mathscr{P}^K)$ and $\hat{z} \in \mathscr{B}(\mathscr{P}'^K)$ such that $(\mathscr{P}'^K, \hat{z})$ blocks $(\mathscr{P}^K, z^K)$. Fix a collection of perpetrators and residuals.

Let s be the size of a maximal coalition in $\mathscr{P}'^K$. If $s \geqslant 2$, it must be the case that

$$\text{for all} \quad j \in K, \qquad \hat{z}_j = a^f(s). \tag{5.1}$$

But in this case, $\hat{z}$ is of the form z'^K for some z'. By part [1] of the lemma (already proved), $(\mathscr{P}', z')$ must block $(\mathscr{P}, z)$, a contradiction.

If $s = 1$, then there must exist $i \in \mathscr{P}$ who is a residual. Further, in this finest coalition structure, there exists a best response equilibrium in which i bears the full cost of producing $y(1)$ and all others are free riders. Clearly, this must also be an equilibrium that can block $(\mathscr{P}^K, z^K)$. To see this, note that the new allocation keeps all perpetrators just as well off, and that it is a binding agreement (because $\mathscr{P}'^K = \mathscr{P}^*$). It follows trivially that all the requirements for blocking are satisfied by the new best-response equilibrium. With this adjustment made, however, $\hat{z}$ can again be taken to be of the form z'^K, which leads to a contradiction as in the case of $s \geqslant 2$.

Conversely, if $z^K \in \mathscr{B}(\mathscr{P}^K)$, then by a subset of the arguments above (for the case $s \geqslant 2$), we can show that $z \in \mathscr{B}(\mathscr{P})$ (the details are omitted). This establishes part [2] of the lemma.

To prove part [3], note first that if $\mathscr{B}(\mathscr{P}^K) = \varnothing$, then for no z is $z^K \in \mathscr{B}(\mathscr{P}^K)$. By part [2], then, $\mathscr{B}(\mathscr{P}) = \varnothing$. Conversely, suppose that $\mathscr{B}(\mathscr{P}) = \varnothing$ but that $\mathscr{B}(\mathscr{P}^K) \neq \varnothing$. Then there exists $\hat{z} \in \mathscr{B}(\mathscr{P}^K)$. Moreover, by part [2], $\hat{z}$ *cannot* be of the form z^K for some z. Because $\hat{z} \in \beta(\mathscr{P}^K)$, it follows that $\mathscr{P}^K$ and therefore $\mathscr{P}$ are coalition structures of singletons. But then, $\mathscr{B}(\mathscr{P}) = \beta(\mathscr{P}) \neq \varnothing$, a contradiction. This completes the proof of part [3] of the lemma. ∎

Given any positive integer n, define $\alpha(n)$ to be the smallest integer such that

$$\alpha(n) > n \qquad \text{and} \qquad a(\alpha(n)) \geqslant a^f(n).$$

Put another way, if $\alpha(n) > n' > n$ then in an economy with n' agents any agent who contributes at least the average cost, would prefer to be a free rider in a coalition structure in which a maximal coalition has size n. Note that $\alpha(n)$ is always well-defined (finite) for every n.[25]

PROPOSITION 5.1. *Suppose that* $\{N\}$ *is an equilibrium coalition structure. Then for every coalition structure* $\mathscr{P}$ *that has a unique maximal coalition of size* n', *with* $n < n' < \alpha(n)$, $\mathscr{B}(\mathscr{P}) = \varnothing$.

Proof. Let $\mathscr{P} = \{S_1, ..., S_m\}$, where S_1 is the unique maximal coalition, with $s_1 = n'$. Consider any $z \in \beta(\mathscr{P})$ and denote by p any player in S_1 such that $z_p \leqslant a(n')$. Note that since $\alpha(n) > n' > n$, player p would prefer to be a free rider in a coalition structure with a maximal coalition of size at least n. In other words,

$$z'_p \leqslant a(s_1) < a^f(\hat{n}) \qquad \text{for any} \quad \hat{n} \geqslant n. \tag{5.2}$$

We will now use this fact to show that $(\mathscr{P}, z)$ can be blocked by such a coalition structure.

Consider the class of coalition structures Π^* that are derived from $\mathscr{P}$ in the following way: (i) each coalition S_i in $\mathscr{P}$ is "split" into a (possibly singleton) coalition T_i and a (possibly empty) collection of singletons, (ii) there is a maximal coalition of size at least n, (iii) $\{p\}$ belongs to the coalition structure. Call p the leading perpetrator and the coalitions of the form

[25] For each n, the fact that g is increasing implies that $y(m) > y(n)$ for some integer m (if $y(n) > 0$, m can be taken to be $n+1$). It follows that for all $n' \geqslant m$ and sufficiently large, $(a(n') \geqslant g(y(m)) + 1 - (cy(m)/n') > g(y(n)) + 1$. This proves that $\alpha(n)$ is always finite for each positive integer n.

T_i residuals; this terminology will soon be reconciled with our earlier definition. Thus an element of Π^* is of the form $\{T_1^{K_1}, ..., T_m^{K_m}\}$ where $T_i \subseteq S_i$ for all i, $t_i \geqslant n$ for some i, and $p \in K_1$.

Let $\mathscr{P}' \in \Pi^*$ be a coalition structure such that $\mathscr{B}(\mathscr{P}') \neq \varnothing$ and for any re-merging, in the move from $\mathscr{P}$ to $\mathscr{P}'$, say $\hat{\mathscr{P}}$, $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$. The existence of such a $\mathscr{P}'$ follows from the fact that by Lemma 5.1 the coalition structure consisting of any n players from S_1, excepting p, and all other singletons is stable.

Let $\mathscr{P}' = \{T_1^{K_1}, ..., T_m^{K_m}\}$ and let $\bar{t} \geqslant n$ be the size of a maximal coalition in $\mathscr{P}'$. Consider any $z' \in \mathscr{B}(\mathscr{P}')$ in which every perpetrator is a free rider, i.e., $z'_i = a^f(\bar{t})$ for all $i \in \bigcup_j K_j$.[26] We claim that $(\mathscr{P}', z')$ blocks $(\mathscr{P}, z)$. To establish this claim, we must check the blocking conditions (B.1)–(B.3).

Choose the residuals T_j as described above, and let every other singleton coalition be a perpetrator. Let p be the leading perpetrator. Condition (B.1) is satisfied by construction of $\mathscr{P}'$. From (5.2) it follows that (B.2) is satisfied.

To check condition (B.3), we must re-merge other perpetrators with their corresponding residuals.[27] In doing so, we obtain a new element of Π^*—call it $\hat{\mathscr{P}}$. By the definition of $\mathscr{P}'$, it follows that $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$. To complete the verification of (B.3) we must show that there exists $\hat{z} \in \beta(\hat{\mathscr{P}})$ such that $(\hat{\mathscr{P}}, \hat{z})$ is blocked by $(\mathscr{P}', z')$ with a leading perpetrator from $\bigcup_j K_j$.

We shall prove this by induction on the number of perpetrators involved in the remerging. Consider the remerging of a single perpetrator j. Condition (B.3) will follow if we can show that:

$$\text{there exists } \hat{z} \in \beta(\hat{\mathscr{P}}) \text{ such that } \hat{z}_j < z'_j. \tag{5.3}$$

Case 1. If j is a member of some maximal coalition $\hat{S}$ in $\hat{\mathscr{P}}$ (of size $\bar{t}+1$), then we can choose $\hat{z}$ to be an allocation achieved by letting $\hat{S}$ carry out all the production, with each member of that coalition receiving equal payoffs i.e., $\hat{z}_j = a(\bar{t}+1)$. Since $a(\bar{t}+1) \leqslant a(s_1)$, (5.3) then follows from (5.2).

Case 2. If j is not a member of a maximal coalition in $\hat{\mathscr{P}}$. It is easy to see that if $z \in \beta(\hat{\mathscr{P}})$, then any $z' \in \beta(\hat{\mathscr{P}})$, where j is made to transfer his entire endowment to some other member(s) in his coalition. In particular, we can find $\hat{z} \in \beta(\hat{\mathscr{P}})$ such that $\hat{z}_j = a^f(\bar{t}+1) - 1 \leqslant a(\bar{t}+1) < a(s_1) < a^f(\bar{t})$. Since $z'_j = a^f(\bar{t})$, this establishes (5.3).

Cases 1 and 2 complete the argument for a single perpetrator involved in re-merging.

[26] If $\mathscr{P}'$ contains a nonsingleton coalition, this isn't a requirement at all: all singleton coalitions will be free riders. Otherwise, $\mathscr{P}'$ is the coalition structure of singletons and every best response allocation is a binding agreement, including the one described in the text.

[27] If there are no other perpetrators, then (B.3) is trivially satisfied.

Now suppose inductively that whenever there are no more than m perpetrators involved in the re-merging, (B.3) is satisfied. Consider the case in which there are $m+1$ perpetrators involved in the re-merging. Given the induction hypothesis and the fact that $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$ for any re-merging from $\mathscr{P}'$, in order to prove that (B.3) is satisfied for this re-merging, it will suffice to prove that condition (5.3) holds for some perpetrator j.

Again, we distinguish between two cases. If *some* perpetrator j is a member of a maximal coalition in $\hat{\mathscr{P}}$, follow the argument above in Case 1. If *no* perpetrator is a member of some maximal coalition in $\hat{\mathscr{P}}$, then fix any perpetrator j and follow the argument in Case 2. ▌

Proposition 5.1 shows that if the number of agents is taken as a parameter, then efficient outcomes are few and far between. The outcome of free, unrestrained negotiation is not necessarily a Pareto optimal binding agreement. An immediate implication of the proposition is the following: suppose that in an economy with agent set N, there exists an efficient binding agreement. Then for any economy with agent set N' satisfying the condition $\alpha(n) > n' > n$, there exists no efficient binding agreement.

The significance of this observation derives from the fact that in many cases, for every n, $\alpha(n)$ is considerably larger than $n+1$. This serves to establish our claim that inefficiency is quite pervasive. We shall now provide a simple example in which $\alpha(n)$ can be computed quite easily.

Suppose the utility functions are specified as

$$u_i(x_i, y) = x_i + \sqrt{y}$$

and

$$cy = y.$$

Then it is easy to see that $y(s) = 0.25s^2$, $a(s) = 1 + 0.25s$ and $g(y(s)) = 0.5s$. Thus, for any n, $\alpha(n)$ is the smallest integer greater than n satisfying:

$$1 + 0.25\alpha(n) \geqslant 1 + 0.5n,$$

which implies that $\alpha(n) = 2n$. Proposition 5.1 therefore allows us to assert that, in this particular example, if an efficient binding agreement exists for an n agent economy then full cooperation will not obtain in all larger economies which are less than twice as large as n. Moreover, by Proposition 5.1, for $n' < 2n$, the grand coalition could break up into one coalition with n agents bearing the full cost of producing $y(n)$ and all other agents free riding. Continuing with this example, efficient equilibrium agreements exist for 1 and 2 agent economies, while the grand coalition in a 3 agent economy breaks up into one coalition with two agents and another coalition with the third agent. It can also be shown that the grand coalition is

stable in the case of 4 agents. Again, Proposition 5.1 implies that full cooperation will not arise in economies with 5, 6 or 7 agents. In fact, in this example, efficiency obtains only for economies in which $n = 2^\gamma$ for some non-negative integer γ (see Proposition 5.2).

This observation is consonant with the general intuition provided at the end of the Introduction. For the grand coalition to be viable, it must be that a deviation triggers off a "long" chain of subsequent deviations, with the final coalition structure fine enough to create a "large" degree of inefficiency. In such a case the initial deviation will not occur, guaranteeing the stability of the grand coalition. On the other hand, if the number of players is such that an initial deviation comes to rest with a "large" producing coalition, this will be enough to unsettle the grand coalition.

All of this is not present in the analysis so far. Proposition 5.1 embodies a negative finding on efficiency, but at the same time fails to give us a complete description of equilibrium outcomes. For instance, we cannot tell from the proposition if the grand coalition indeed forms along the sequence suggested by that proposition. Our next result provides an additional sufficient condition under which efficiency emerges precisely along this sequence. We will assume that for all n, $\alpha(n) \leqslant 2n$. Notice that our previous example satisfies this condition. So does any modification of that example where the utility functions are $u_i(x_i, y) = x_i + y^\delta$, with $0.5 \leqslant \delta < 1$.

Define a sequence of positive integers $n_0, n_1, n_2, \ldots$ by the conditions: $n_0 = 1$, and recursively, having defined n_k, $n_{k+1} = \alpha(n_k)$.

PROPOSITION 5.2. *Suppose that for all n, $\alpha(n) \leqslant 2n$. Then $\mathscr{B}(\{N\}) \neq \varnothing$ if and only if $n = n_k$ for some k.*

Proof. By Proposition 5.1, if $\mathscr{B}(\{N\}) \neq \varnothing$ for some N of cardinality n, then it must be the case that $\mathscr{B}(\{N'\}) = \varnothing$ for all N' such that $\alpha(n) > n' > n$. It suffices, therefore, to prove that $\mathscr{B}(\{N\})$ is nonempty whenever $n = n_k$ for some k. For $k = 0$ this is trivially true. Suppose, inductively, that $\mathscr{B}(\{N'\}) \neq \varnothing$ when $n' = n_k$, and consider a set of players N such that $n = n_{k+1}$. Let z be an *equal-division* best-response payoff allocation, i.e., $z_i = a(n)$ for all i. We claim that $z \in \mathscr{B}(\{N\})$.

Suppose not. Then there exists $(\mathscr{P}', z')$ that blocks $(\{N\}, z)$. For the leading perpetrator to gain, given the construction of $\alpha(n)$, it must be the case that the size of a maximal coalition in $\mathscr{P}'$ is greater than n_k. Since $n = n_{k+1} \leqslant 2n_k$ this means that there is a unique maximal coalition in $\mathscr{P}'$ of size greater than n_k and less than $n_{k+1} = \alpha(n_k)$. But then, by Proposition 5.1, such a coalition structure cannot admit a binding agreement, which contradicts the supposition that $(\mathscr{P}', z')$ blocks $(\{N\}, z)$. ∎

Thus Proposition 5.2 provides a complete characterization of those economies which can sustain efficient binding agreements. It should be

noted, however, that the proposition does not fully describe equilibrium coalition *structures*. Under a simplifying assumption, progress can be made in this direction:

A. *If $\mathscr{P}$ has more than one maximal coalition, permit only those best response strategy vectors such that only one of these coalitions bears the entire cost of production. That is, the payoff of a maximal coalition is restricted to the two values $\{a(s), a^f(s)\}$, and in any situation, assume that one and only one coalition will receive the lower payoff.*

PROPOSITION 5.3. *Suppose assumption (A) is satisfied. Consider a coalition structure $\mathscr{P}$ with at least one nonsingleton coalition in it. Then $\mathscr{P}$ is an equilibrium coalition structure if and only if it has a unique maximal coalition with cardinality equal to n_k for some k.*

Remark 5.2. Observe that Proposition 5.3 strengthens the conclusions of Propositions 5.1 and 5.2, yielding in addition a complete description of which coalition structures are immune to blocking. Proposition 5.3 also implies that, in general, several agents will free ride in equilibrium. If $z \in \mathscr{B}(\mathscr{P})$ and n_k is the size of the maximal coalition in $\mathscr{P}$, then $z_i = a^f(n_k)$ for all i not belonging to the maximal coalition; i.e., all agents who are not in the maximal coalition are free riders.

Proof of Proposition 5.3. We will proceed by induction on k. First we establish the Proposition for $k = 1$; that is, for a coalition structure $\mathscr{P}$ with maximal coalition(s) of size n_1 or less.

Step 1. Consider, first, the case in which the maximal coalition size in $\mathscr{P}$ is n', where $1 < n' < n_1$. If this maximal coalition is unique then the fact that $\mathscr{B}(\mathscr{P}) = \varnothing$ follows directly from Proposition 5.1. If there is more than one maximal coalition, then by Assumption A, only one of them carries out production. With this known, the result follows again by directly applying the proof of Proposition 5.1.

Step 2. Suppose maximal coalition(s) in $\mathscr{P}$ have cardinality exactly n_1. We will prove that $\mathscr{B}(\mathscr{P}) \neq \varnothing$ if and only if there is a unique maximal coalition. Suppose $\mathscr{P}$ does have a unique maximal coalition of size n_1. Let $z^e(\mathscr{P})$ denote the equal division, best response payoff,[28] i.e., $z^e(\mathscr{P})_i = a(n_1)$ for all i who belong to the maximal coalition and $z^e(\mathscr{P})_j = a^f(n_1)$ for all other j.

We claim that $z^e(\mathscr{P}) \in \mathscr{B}(\mathscr{P})$. Members of the maximal coalition each receive $a(n_1) \geqslant a^f(1)$, while members of other coalitions receive

[28] We will henceforth use $z^e(\mathscr{P})$ to denote the equal division, best response payoff for any coalition structure $\mathscr{P}$ that has unique maximal coalition.

$a^f(n_1) \geqslant a(n_1)$. Thus $z^e(\mathscr{P})_i \geqslant a^f(1)$ for all $i \in \mathscr{P}$. By the argument in Step 1, every deviation must culminate in $\mathscr{P}^*$. The best that a deviator can hope to achieve is, therefore, $a^f(1)$. Since $z^e(\mathscr{P})$ provides every agent no less than this, $z^e(\mathscr{P}) \in \mathscr{B}(\mathscr{P})$. Next, we prove that $\mathscr{B}(\mathscr{P}) = \varnothing$ if there are several maximal coalitions of cardinality n_1. Suppose there are exactly *two* such maximal coalitions. Call them S_1 and S_2. Consider some $z \in \beta(\mathscr{P})$ where S_1 bears all the cost of provision, so that its *per capita* payoff is $a(n_1)$. Now construct $\mathscr{P}'$ by breaking S_1 into a leading (singleton) perpetrator who was earning no more than the per capita payoff $a(s)$ and a residual. By our earlier argument, $\mathscr{P}'$ is an equilibrium coalition structure, and $z^e(\mathscr{P}') \in \mathscr{B}(\mathscr{P}')$. The leading (and only) perpetrator must receive $a^f(n_1) > a(n_1)$. So $(\mathscr{P}', z^e(\mathscr{P}'))$ blocks $(\mathscr{P}, z)$, completing the proof.

Inductively, suppose that every coalition structure that contains no more than m (and no less than two) maximal coalitions (each of size n_1) has an empty set of binding agreements. Moreover, suppose that for each such coalition structure $\mathscr{P}$ with maximal coalitions of the form $(S_1, ..., S_r)$ (where $2 \leqslant r \leqslant m$), there exists an allocation $z \in \beta(\mathscr{P})$ which is blocked by $(\mathscr{P}', z^e(\mathscr{P}'))$, where $\mathscr{P}'$ is the coalition structure formed by creating one perpetrator from each of $S_1, ..., S_{r-1}$. Notice that this hypothesis is satisfied for $m = 2$.

Now consider $\mathscr{P}$ with $m+1$ maximal coalitions, each of size n_1. Call them $(S_1, ..., S_{m+1})$.

Consider $z \in \beta(\mathscr{P})$ and suppose, without loss of generality, that S_1 bears all the cost of provision, so that its *per capita* payoff is $a(n_1)$. Now construct $\mathscr{P}'$ in the following way. Leaving S_{m+1} unchanged, break up $S_1, ..., S_m$ by taking from each of them one perpetrator such that the (leading) perpetrator from S_1 is a player j for whom $z_j \leqslant a(n_1)$. The resulting structure has a unique maximal coalition, S_{m+1}, of size n_1, so it is an equilibrium structure. The leading perpetrator gains just as in the previous paragraph. Finally, consider any re-merging of the other perpetrators. This leads to a coalition structure $\hat{\mathscr{P}}$ with r maximal coalitions of size n_1, where $2 \leqslant r \leqslant m$. By the induction hypothesis, $\mathscr{B}(\hat{\mathscr{P}}) = \varnothing$, and moreover, there is $\hat{z} \in \beta(\hat{\mathscr{P}})$ such that $(\mathscr{P}', z^e(\mathscr{P}'))$ blocks $(\hat{\mathscr{P}}, \hat{z})$. It also follows from the induction hypothesis that the leading perpetrator of this block is one of the perpetrators involved in the re-merging. This verifies all the blocking conditions, so that the proof is complete in the case $k = 1$.

Now proceed inductively to establish the proposition for all n_k. Suppose that the proposition is true for all $k = 1, ..., K$, for some $K \geqslant 1$. Now consider a coalition structure $\mathscr{P}$ with a maximal coalition S, of cardinality s.

First we study the case where $n_K < s < n_{K+1}$. By Proposition 5.1 and the additional argument using assumption (A), which we used in Step 1 of the case in which $k = 1$, it can again be shown that this coalition structure cannot be stable.

It remains to consider the case where $s = n_{K+1}$. We claim that $\mathscr{P}$ is stable if and only if it contains a unique maximal coalition. The argument is the same as the one we used in Step 2 of the case in which $k = 1$.

This completes the proof of the proposition. ∎

6. SYMMETRIC TU GAMES

In this section we consider a subclass of normal form games which we call symmetric TU games. These are games in which all players are identical and *within* a coalition players can carry out interpersonal transfers of utility without affecting the strategic environment for the other coalitions.

Of course, there is no presumption that the *equilibrium* coalition structures of such games will be symmetric. Indeed, in general there are symmetric TU games with asymmetric equilibrium coalition structures, as our analysis of a symmetric Cournot oligopoly will show.[29] The objective of this section, however, is to demonstrate that in a large class of symmetric TU situations, the set of equilibrium structures of a game in this class is identical to that obtained in an "artificial" game where each coalition is constrained to choose only strategies that yield equal utility to all its members. This represents an enormous computational simplification.

A game $\Gamma = (N, (X_i, u_i)_{i \in N}$ is said to be a *symmetric TU game* if it satisfies the following conditions:

(i) $X_i = X_j$ for all $i, j \in N$, and for any $x \in X$, $i \in N$ and a permutation $\rho : N \mapsto N$, $u_i(x) = u_{\rho(i)}(x_\rho)$, where x_ρ is induced in the obvious way by the permutation ρ.

(ii) Let $\hat{x} \in X$, and $S \in \mathscr{N}$. Then for each $v \in \mathbb{R}^S$ such that $\sum_{i \in S} v_i = \sum_{i \in S} u_i(\hat{x})$, there is $x_S \in X_S$ such that $v_i = u_i(x_S, \hat{x}_{-S})$ for all $i \in S$, and $u_i(x_S, y_{-S}) = u_i(\hat{x}_S, y_{-S})$ for all $y_{-S} \in X_{-S}$ and $i \notin S$.

Condition (i) describes symmetry: each player has the same set of available actions, and any permutation of an action vector leads to the same permutation of the payoff vector corresponding to that action vector. Condition (ii) formalizes transferability, stating that if some payoff vector is feasible for a coalition, then so is any other payoff vector with the same aggregate payoff to this coalition. Moreover, this can be achieved by changing *only* actions taken by the coalition concerned, and that too in a way that changes nothing for players external to the coalition.

For the class of symmetric TU games, we define the notion of binding agreements just as we did before.

[29] See Yi [36] for other applications of our solution concept to symmetric games.

A symmetric TU game is a game with *positive externalities* if, whenever $S \in \mathscr{P}$, $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$ and $S \in \mathscr{P}'$, then: $x \in \mathscr{B}(\mathscr{P}')$ implies that there exists $y \in \mathscr{B}(\mathscr{P})$ such that $\sum_{i \in S} u_i(y) \geqslant \sum_{i \in S} u_i(x)$, whenever $\mathscr{B}(\mathscr{P})$ is nonempty.

Thus, loosely speaking, symmetric TU games with positive externalities have the property that every coalition S enjoys positive externalities from the merger of coalitions other than S. It is important to emphasize that in our definition, the term positive externalities does not necessarily relate to the *strategies* of the other firms but to the act of their forming mergers. For example, as we shall see in the next section, the simple Cournot oligopoly defines a symmetric TU game with positive externalities, although the outputs/strategies of rival firms impose negative externalities on any given firm. Note that the definition, as stated, does not use the primitives of the model.[30] But it turns out to be an easy requirement to check, as our analysis of Cournot oligopolies later in this paper should make clear. Indeed, a number of economic examples satisfy this requirement of positive externalities, though it should be pointed out that other reasonable examples might not (see, e.g., Yi [36]).

We begin by showing that in a symmetric TU game with positive externalities equilibrium binding agreements can be characterized in a relatively simple form. In particular, $x \notin \mathscr{B}(\mathscr{P})$ if there exists *any* refinement of $\mathscr{P}$ with a binding agreement that yields a higher aggregate utility to some subcoalition of $\mathscr{P}$.

PROPOSITION 6.1. *Suppose $\Gamma = (N, (X_i, u_i))$ is a symmetric TU game with positive externalities. Suppose, moreover, that best respose allocations exist for every coalition structure.*[31] *Let $x \in \beta(\mathscr{P})$. Then $x \in \mathscr{B}(\mathscr{P})$ if and only if there does not exist $(\mathscr{P}', x')$ such that $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$ and $x' \in \mathscr{B}(\mathscr{P}')$, with $\sum_{i \in S} u_i(x') > \sum_{i \in S} u_i(x)$ for some $S \in \mathscr{P}'$, $S \notin \mathscr{P}$.*

Proof. Suppose $x \in \beta(\mathscr{P})$ and there is no subcoalition that can get a higher aggregate utility in any equilibrium binding agreement in any refinement of $\mathscr{P}$. Clearly then, there exists no $(\mathscr{P}', x')$ which can satisfy conditions (B.1) and (B.2) of blocking. This means that $x \in \mathscr{B}(\mathscr{P})$, and completes the proof of the "if" part of the proposition.

To prove the converse, consider $x \in \mathscr{B}(\mathscr{P})$ and suppose, contrary to the claim of the Proposition, that there exists $(\mathscr{P}', x')$ such that $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$, $x' \in \mathscr{B}(\mathscr{P}')$ and $\sum_{i \in S} u_i(x') > \sum_{i \in S} u_i(x)$, for some S such that $S \in \mathscr{P}'$ and $S \notin \mathscr{P}$. By condition (ii) of the definition of a symmetric TU game, we can

[30] A stronger version of this assumption, based on the primitives is as follows. For $\mathscr{P}$, $\mathscr{P}'$ and S as in the above definition, for every $x \in \beta(\mathscr{P}')$ and $y \in \beta(\mathscr{P})$, $\sum_{i \in S} u_i(y) \geqslant \sum_{i \in S} u_i(x)$. This version of the assumption is satisfied by the Cournot oligopoly game analyzed in the following Section.

[31] See Proposition 2.2 for sufficient conditions.

find a strategy vector $\hat{x}'_S$ such that if we define $\hat{x}'$ to be the strategy vector $(\hat{x}'_S, x'_{-S})$, then

$$u_S(\hat{x}') \gg u_S(x), \tag{6.1}$$

with the additional proviso that

$$\hat{x}' \in \beta(\mathscr{P}') \text{ and } u_{-S}(\hat{x}'_S, w_{-S}) = u_{-S}(x'_S, w_{-S}) \text{ for all } w_{-S} \in X_{-S}. \tag{6.2}$$

Now, either $\hat{x}' \in \mathscr{B}(\mathscr{P}')$ or $\hat{x}' \notin \mathscr{B}(\mathscr{P}')$. In the latter case, there exist $\mathscr{P}'' \in \mathscr{R}(\mathscr{P}')$ and $x'' \in \mathscr{B}(\mathscr{P}'')$ such that $(\mathscr{P}'', x'')$ blocks $(\mathscr{P}', \hat{x}')$. Fix a set of perpetrators and residuals in this move, and let T be a leading perpetrator. It must be the case that $T \subset S$. To see this, note from (6.2) that $\hat{x}'$ was constructed from x' by changing none of the payoffs and incentives to coalitions other than S in $\mathscr{P}'$. Moreover, $x' \in \mathscr{B}(\mathscr{P}')$. Consequently, if $\hat{x}' \notin \mathscr{B}(\mathscr{P}')$, the leading perpetrator T must be a subset of S.

It follows, using (6.1) and the claim above, that $u_T(x'') \gg u_T(\hat{x}') \gg u_T(x)$. Also note that $\mathscr{P}'' \in \mathscr{R}(\mathscr{P})$.

So in both cases, there exists a coalition structure $\hat{\mathscr{P}} \in \mathscr{R}(\mathscr{P})$, an allocation $\hat{x}$, and a coalition $\hat{S} \in \hat{\mathscr{P}}$ with $\hat{S} \subset S$, such that (a) $\hat{x} \in \mathscr{B}(\hat{\mathscr{P}})$, and (b) $u_{\hat{S}}(\hat{x}) \gg u_{\hat{S}}(x)$.

It will be convenient to require something more of $\hat{S}$. Choose $\hat{S}$ to be the *smallest* subset of S with the property that there exists $\hat{\mathscr{P}}$ with $(\hat{\mathscr{P}}, \hat{S})$ satisfying all the requirements of the previous paragraph.

With $(\hat{P}, \hat{S})$ chosen in this way, consider, now, any coalition structure $\mathscr{Q}$ such that $\hat{S} \in \mathscr{Q}$, and such that $\mathscr{Q} \in \mathscr{R}(\mathscr{P})$ and $\hat{\mathscr{P}} \in \mathscr{R}(\mathscr{Q})$. We claim that if $\mathscr{B}(\mathscr{Q}) \neq \varnothing$, then there exists an allocation $w \in \mathscr{B}(\mathscr{Q})$ such that $u_{\hat{S}}(w) \geqslant u_{\hat{S}}(\hat{x})$.

To prove this claim, observe first from the assumption of positive externalities that if $\mathscr{B}(\mathscr{Q}) \neq \varnothing$, then there exists $w' \in \mathscr{B}(\mathscr{Q})$ such that $\sum_{i \in \hat{S}} u_i(w') \geqslant \sum_{i \in \hat{S}} u_i(\hat{x})$. By part (ii) of the definition of a symmetric TU game, it is possible to find a strategy vector $w_{\hat{S}}$ such that if we define w to be the strategy vector $(w_{\hat{S}}, w'_{-\hat{S}})$, then

$$u_{\hat{S}}(w) \geqslant u_{\hat{S}}(\hat{x}),$$

with the additional proviso that

$$w \in \beta(\mathscr{Q}) \text{ and } u_{-\hat{S}}(w_{\hat{S}}, y_{-\hat{S}}) = u_{-\hat{S}}(w'_{\hat{S}}, y_{-\hat{S}}) \text{ for all } y_{-\hat{S}} \in X_{-\hat{S}}. \tag{6.3}$$

Because of the construction (6.3), it follows that if $w \notin \mathscr{B}(\mathscr{Q})$, it *must* be blocked with some leading perpetrator $T \subset \hat{S} \subset S$, using some binding agreement w'' in some refinement of $\mathscr{Q}$ and therefore $\mathscr{P}$. Note also that $u_T(w'') \gg u_T(w) \geqslant u_T(\hat{x}) \gg u_T(x)$. But then T contradicts the construction of $\hat{S}$ as the *smallest* coalition with these properties.

This proves the claim.

The claim allows us, therefore, to choose a coalition structure $\mathscr{Q}$ such that (a) $\hat{S} \in \mathscr{Q}$, (b) there exists $w \in \mathscr{B}(\mathscr{Q})$ with $u_{\hat{S}}(w) \gg u_{\hat{S}}(x)$, and (c) all other coalition structures $\mathscr{Q}'$ such that $\mathscr{Q}' \in \mathscr{R}(\mathscr{P})$ and $\mathscr{Q} \in \mathscr{R}(\mathscr{Q}')$, and such that $\hat{S} \in \mathscr{Q}'$, have no equilibrium binding agreements.

Our final claim is that $(\mathscr{Q}, w)$ blocks $(\mathscr{P}, x)$. To prove this, choose any set of perpetrators and residuals in the move from $\mathscr{P}$ to $\mathscr{Q}$ such that $\hat{S}$ is the leading perpetrator. By construction, conditions (B.1) and (B.2) of the blocking condition are satisfied. It remains to verify (B.3).

Consider, first, the re-merging of a single perpetrator T with its residual R. By construction of $\mathscr{Q}$, this leads to a coalition structure $\mathscr{Q}'$ "between" $\mathscr{Q}$ and $\mathscr{P}$, which has no binding agreement. Consider any $y \in \beta(\mathscr{Q}')$, and the coalition $T \cup R$. Using condition (ii) of a symmetric TU game, construct another allocation $y' \in \beta(\mathscr{Q}')$ such that $u_T(w) \gg u_T(y')$. But then it follows that $(\mathscr{Q}, w)$ blocks $(\mathscr{Q}', y')$, using the perpetrator T as the leading (and unique) perpetrator.

Inductively, consider some re-merging $\mathscr{Q}'$ of $\mathscr{Q}$ (that excludes the leading perpetrator $\hat{S}$), and suppose that for all re-mergings $\mathscr{Q}$" "between" $\mathscr{Q}$ and $\mathscr{Q}'$, the blocking condition (B.3) has already been verified. Again, by the construction of $\mathscr{Q}$, it must be the case that $\mathscr{B}(\mathscr{Q}') = \varnothing$. Let T be any perpetrator involved in the remerging and let W be the coalition in which T lies after re-merging. Consider any $y \in \beta(\mathscr{Q}')$. Using condition (ii) of a symmetric TU game, construct another allocation $y' \in \beta(\mathscr{Q}')$ (by changing only the strategy vector on W) such that $u_T(w) \gg u_T(y')$. But then it follows that $(\mathscr{Q}, w)$ blocks $(\mathscr{Q}', y')$, using the perpetrator T as the leading perpetrator. To see this, assign perpetrators and residuals in the move from $\mathscr{Q}'$ to $\mathscr{Q}$, by simply using the same perpetrators involved in the re-merging, and by dubbing T as the leading perpetrator. Conditions (B.1) and (B.2) are met right away. To check (B.3), simply use the induction hypothesis at the beginning of this paragraph. Thus the proof of the final claim is complete: $(\mathscr{Q}, w)$ blocks $(\mathscr{P}, x)$.

But this contradicts the supposition that $x \in \mathscr{B}(\mathscr{P})$, and we are done. ∎

We are now going to introduce a solution concept *with each coalition restricted to equally divide its worth* under any structure. Becuase we do not really make this assumption as a behavioral description in this model, the concept is an artificial one: its use will lie in the fact that it is easy to use in examples and yields exactly the equilibrium coalition structures of the original model. On the other hand, a model where equal division is assumed may be of independent conceptual interest. It is then of some value to know that there is a correspondence between such a model and the unrestricted case, as we shall see.

Some additional definitions will help the exposition. Let $\mathscr{P}$ be a coalition structure, and $S \in \mathscr{P}$. An *S-equal allocation for* $\mathscr{P}$ is $x \in X$ such that for all

$i, j \in S$, $u_i(x) = u_j(x)$. An *equal allocation* (*for* $\mathscr{P}$) is an allocation that is S-equal for $\mathscr{P}$, for all $S \in \mathscr{P}$. Now define

$$\beta_e(\mathscr{P}) = \{x \in X \mid x \text{ is an equal allocation and } \nexists S \in \mathscr{P} \text{ and an } S\text{-equal allocation } (y_S, x_{-S}) \text{ such that } u_S(y_S, x_{-S}) \gg u_S(x)\}.$$

Thus $\beta_e(\mathscr{P})$ puts together only those equal allocations that satisfy the best response property with respect to a restricted class of deviations: each coalition S is permitted only equal division among its members. Recall that $\beta(\mathscr{P})$ denotes the set of strategies satisfying the (unrestricted) best response property with respect to $\mathscr{P}$.

We now proceed to define a restricted notion of *equal binding agreements*. This is different in two ways from the original definition. First, as already mentioned, coalitions are restricted to equal division. Second, the notion of blocking is rudimentary and very easy to check.

Proceed recursively. For $\mathscr{P}^*$, the coalition structure of singletons, all allocations in $\beta_e(\mathscr{P}^*)$ (which is trivially the same as $\beta(\mathscr{P}^*)$) are equal binding agreements. Now suppose that for all refinements of some coalition structure $\mathscr{P}$, equal binding agreements have been defined.

Let $x_e \in \beta_e(\mathscr{P})$ be an equal allocation satisfying the (modified) best-response property, and let $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$. Say that $(\mathscr{P}', x'_e)$ *e-blocks* $(\mathscr{P}, x_e)$ if

(E.1) x'_e is an equal binding agreement for $\mathscr{P}'$.

(E.2) For some coalition $S \in \mathscr{P}'$, which is a strict subset of some coalition in $\mathscr{P}$, $u_S(x'_e) \gg u_S(x_e)$.

To complete the recursion, say that an equal allocation $x_e \in \beta_e(\mathscr{P})$ is an *equal binding agreement*, written $x_e \in \mathscr{B}_e(\mathscr{P})$, if it is not e-blocked.

The conditions (E.1)–(E.2) have been described to facilitate direct comparison with (B.1)–(B.2) in the original definition of blocking. The main simplification (apart from using equal allocations) is that we require no analogue to (B.3). This makes the present set of conditions much easier to check than (B.1)–(B.3).

To state our main results, it will be useful to introduce a particular class of equal allocations that we may associate with any allocation. For $x \in X$ and $\mathscr{P} \in \Pi$, we define $e(\mathscr{P}, x)$ to be the set of equal allocations that give the same aggregate utility to each coalition in $\mathscr{P}$ as x does, *and* such that the options of each individual coalition remain unchanged. Formally, $y \in e(\mathscr{P}, x)$ if:

(1) y is an equal allocation,

(2) $\sum_{i \in S} u_i(x) = \sum_{i \in S} u_i(y)$ for all $S \in \mathscr{P}$,

(3) $u_S(z_S, y_{-S}) = u_S(z_S, x_{-S})$ for all $z_S \in X_S$ and $S \in \mathscr{P}$.

It follows from the definition of a symmetric TU game that for all $\mathscr{P} \in \Pi$ and $x \in X$, $e(\mathscr{P}, x)$ is nonempty.[32]

PROPOSITION 6.2. *Suppose $\Gamma = (N, (X_i, u_i))$ is a symmetric TU game. Then*

(a) $\beta^e(\mathscr{P}) \subseteq \beta(\mathscr{P})$;

(b) *If $x \in \beta(\mathscr{P})$, then $y \in \beta^e(\mathscr{P})$ for all $y \in e(\mathscr{P}, x)$.*

Proof. (a) Suppose the claim is false, i.e., suppose there exists $x \in \beta_e(\mathscr{P})$ such that $x \notin \beta(\mathscr{P})$. Then there exist $S \in \mathscr{P}$ and $y_S \in X_S$ such that $u_S(y_S, x_{-S}) \gg u_S(x)$. Since x is an equal allocation, it follows that

$$\frac{\sum_{i \in S} u_i(y_S, x_{-S})}{|S|} > u_i(x) \qquad \text{for all} \quad i \in S. \tag{6.4}$$

From the definition of a symmetric TU game it follows that there exists an S-equal allocation (z_S, x_{-S}) such that $\sum_{i \in S} u_i(z_S, x_{-S}) = \sum_{i \in S} u_i(y_S, x_{-S})$. Since (z_S, x_{-S}) is an S-equal allocation, (6.4) implies that $u_S(z_S, x_{-S}) \gg u_S(x)$, which contradicts the supposition that $x \in \beta_e(\mathscr{P})$.

(b) Suppose the claim is false. Then there exists $x \in \beta(\mathscr{P})$ and $y \in e(\mathscr{P}, x)$ such that $y \notin \beta_e(\mathscr{P})$. Since $y \notin \beta_e(\mathscr{P})$, there exists $S \in \mathscr{P}$ and an S-equal allocation (z_S, y_{-S}) such that $u_S(z_S, y_{-S}) \gg u_S(y)$. Since $y \in e(\mathscr{P}, x)$, it follows that $u_S(z_S, y_{-S}) = u_S(z_S, x_{-S})$ (part (3) of the definition of $e(\mathscr{P}, x)$).

Thus,

$$u_S(z_S, x_{-S}) \gg u_S(y).$$

Of course,

$$\sum_{i \in S} u_i(y) = \sum_{i \in S} u_i(x),$$

which implies that there exists $w_S \in X_S$ such that $\sum_{i \in S} u_i(w_S, x_{-S}) = \sum_{i \in S} u_i(z_S, x_{-S})$ and $u_S(w_S, x_{-S}) \gg u_S(x)$, which contradicts $x \in \beta(\mathscr{P})$. ∎

We may now state the main result of this section.

[32] Here is the outline of a proof. Let $\mathscr{P} = (S^1, ..., S^m)$. Replace x_{S^1} by y_{S^1} to obtain the S^1-equal allocation (y_{S^1}, x_{-S^1}), satisfying (ii) of the definition of a symmetric TU game. That condition implies, in particular, that for every $S^i \neq S^1$, and every $w_{S^i} \in X_{S^i}$, $u_{S^i}(y_{S^1}, w_{S^i}, x_{-[S^1 \cup S^i]}) = u_{S^i}(w_{S^i}, x_{-S^i})$. Now start with the allocation (y_{S^1}, x_{-S^1}) and proceed to S^2, obtaining in exactly the same way the allocation $(y_{S^1}, y_{S^2}, x_{-[S^1 \cup S^2]})$. In this recursive manner, we arrive at the allocation y. Clearly $y \in e(\mathscr{P}, x)$, as desired.

PROPOSITION 6.3. *Suppose* $\Gamma = (N, (X_i, u_i))$ *is a symmetric TU game with positive externalities. Suppose, moreover, that best response allocations exist for every coalition structure. Then*

(a) $\mathscr{B}_e(\mathscr{P}) \subseteq \mathscr{B}(\mathscr{P})$;

(b) *If* $x \in \mathscr{B}(\mathscr{P})$, *then* $x_e \in \mathscr{B}_e(\mathscr{P})$ *for all* $x_e \in e(\mathscr{P}, x)$.

In particular, the set of equilibrium coalition structures is unchanged by restricting attention to equal division.

Proof. We will prove this proposition by (backward) induction on the cardinality of coalition structures. For the coalition structure of singletons the result obviously holds, because for this coalition structure all Nash equilibria are equal allocations. Suppose, then, that the result is true for all coalition structures with at least $m+1$ coalitions in them, for some $1 \leqslant m \leqslant N-1$. Consider some $\mathscr{P}$ with m coalitions in it.

We first claim that if $x_e \in \mathscr{B}_e(\mathscr{P})$, then $x_e \in \mathscr{B}(\mathscr{P})$ as well. Obviously, $x_e \in \beta_e(\mathscr{P})$ and so, by Proposition 6.2, $x_e \in \beta(\mathscr{P})$. So, if $x_e \notin \mathscr{B}(\mathscr{P})$ there must exist $\mathscr{P}' \in \mathscr{R}(\mathscr{P})$ and $x' \in \mathscr{B}(\mathscr{P}')$ such that $(\mathscr{P}', x')$ blocks $(\mathscr{P}, x_e)$. Pick a set of perpetrators and residuals, and let S be the leading perpetrator. Now pick some allocation $x'_e \in e(\mathscr{P}', x)$. By the induction hypothesis, $x'_e \in \mathscr{B}_e(\mathscr{P}')$. Moreover, because $u_S(x') \gg u_S(x_e)$, it follows that $u_S(x'_e) \gg u_S(x_e)$ as well. But this means that $(\mathscr{P}', x'_e)$ satisfies (E.1) and (E.2) in relation to $(\mathscr{P}, x_e)$, and so e-blocks $(\mathscr{P}, x_e)$, a contradiction.

Next, we prove the converse: if $x \in \mathscr{B}(\mathscr{P})$, then for all $x_e \in e(\mathscr{P}, x)$, $x_e \in \mathscr{B}_e(\mathscr{P})$. Observe from Proposition 6.2 that $x_e \in \beta_e(\mathscr{P})$. Thus, if $x_e \notin \mathscr{B}_e(\mathscr{P})$, there exists $(\mathscr{P}', x'_e)$ which e-blocks $(\mathscr{P}, x_e)$. Of course, $x'_e \in \mathscr{B}_e(\mathscr{P}')$ which, by the induction hypothesis, means that

$$x'_e \in \mathscr{B}(\mathscr{P}'). \tag{6.5}$$

Let S be a coalition in $\mathscr{P}'$ satisfying (E.2) of the e-blocking condition. Let $V \in \mathscr{P}$ be the coalition from which S defected. By symmetry, we can assume without loss of generality that S consists of $|S|$ of the least well-off members of V under the strategy vector x. Certainly,

$$\sum_{i \in S} u_i(x'_e) > \sum_{i \in S} u_i(x).$$

But then, given (6.5), we can appeal to Proposition 6.1 and obtain a contradiction to the supposition that $x \in \mathscr{B}(\mathscr{P})$. ∎

We reiterate that this result may be of interest at two levels. At one level, we may think of equal division as a simple computational device which, by this proposition, gives rise to exactly the same set of coalition structures.

At another level, the behavioral assumption of equal division may be of interest in itself. In that case, the proposition tells us that this assumption buys us no more (and no less) when we describe equilibrium coalition structures.

As an independent consequence of Proposition 6.1, we can provide sufficient conditions under which every equilibrium binding agreement for the grand coalition belongs to $C^b(N)$—the β-core. Recall that

$$C^b(N) = \{x \in X \mid \nexists S \subseteq N \text{ such that for every } z_{-S} \in X_{-S} \text{ there exists } y_S \in X_S \text{ and } u_S(y_S, z_{-S}) \gg u_S(x)\}.$$

PROPOSITION 6.4. *Suppose $\Gamma = (N, (X_i, u_i))$ is a symmetric TU game with positive externalities. Suppose, moreover, that best response allocations exist for every coalition structure. Then $\mathscr{B}(\{N\}) \subseteq C^b(N)$.*

Remark 6.1. Since the β-core is a subset of the α-core, this also implies that, under the assumptions of Proposition 6.4, $\mathscr{B}(\{N\})$ is contained in the α-core.

Proof of Proposition 6.4. Suppose $x \in \mathscr{B}(\{N\})$ but x is not in the β-core. Then there exists a coalition S such that

$$\text{for every } z_{-S} \in X_{-S} \text{ there exists } y_S \in X_S \text{ such that } u_S(y_S, z_{-S})) \gg u_S(x). \tag{6.6}$$

Now consider the coalition structure $\mathscr{P}' = \{\{i\}_{i \notin S}, \{S\}\}$. Let $\hat{x} \in \beta(\mathscr{P})$. From (6.6) and the definition of a symmetric TU game, we know that $\sum_{i \in S} u_i(\hat{x}) > \sum_{i \in S} u_i(x)$. We can, therefore, construct x' such that

$$x' \in \beta(\mathscr{P}') \qquad \text{and} \qquad u_S(x') \gg u_S(x). \tag{6.7}$$

Since $x \in \mathscr{B}(\{N\})$, it follows from Proposition 6.1 that $x' \notin \mathscr{B}(\mathscr{P}')$. Thus, there exists $(\mathscr{P}'', x'')$ that blocks $(\mathscr{P}', x')$. Of course, given the construction of $\mathscr{P}'$, it must be the case that all possible perpetrators in such a blocking belong to S. Thus, there exists $S'' \subset S$ such that $S'' \in \mathscr{P}''$ and $u_{S''}(x'') \gg u_{S''}(x') \gg u_{S''}(x)$. Since $x'' \in \mathscr{B}(\mathscr{P}'')$, we can now appeal to Proposition 6.1 to obtain a contradiction to $x \in \mathscr{B}(\{N\})$. ∎

We end this section by showing that Propositions 6.1 and 6.3 rely crucially on the game being one with positive externalities. We construct an example of a symmetric TU game without positive externalities in which the set of equilibrium coalition structures do not coincide with those obtained by restricting attention to equal division.

As a first step in defining this game consider the following normal form. There are three players each having three strategies. Player 1 chooses rows, player 2 chooses columns and player 3 chooses matrices.

x_{3a}

	x_{2a}	x_{2b}	x_{2c}
x_{1a}	20, 20, 20	0, 0, 0	0, 22, 0
x_{1b}	0, 0, 0	4, 4, 10	16, 5, 25
x_{1c}	22, 0, 0	5, 16, 25	18, 18, 0

x_{3b}

	x_{2a}	x_{2b}	x_{2c}
x_{1a}	0, 0, 0	10, 4, 4	25, 5, 16
x_{1b}	4, 10, 4	0, 0, 0	0, 0, 0
x_{1c}	5, 25, 16	0, 0, 0	0, 0, 0

x_{3c}

	x_{2a}	x_{2b}	x_{2c}
x_{1a}	0, 0, 22	25, 16, 5	0, 18, 18
x_{1b}	16, 25, 5	0, 0, 0	0, 0, 0
x_{1c}	18, 0, 18	0, 0, 0	15, 15, 15

Now define a normal form game in which player i's strategy set is

$$X_i = \{x_a, x_b, x_c\} \times \triangle,$$

where $\triangle$ is the unit simplex in $\mathbb{R}^3$. The interpretation is that a player can choose either x_a, x_b or x_c and a distribution of his/her gross payoff among all the players. Let s_{ij} denote the share that i allocates to j. Let $\hat{u}_i(.)$ denote the payoff functions corresponding to the matrices. The actual payoff to player i is then specified as:

$$u_i((x_i, s_i)) = \sum_{j=1}^{3} s_{ji}\hat{u}_j(x).$$

The three matrices indicate the payoffs when $s_{ii} = 1$ for all i, i.e., when all transfers are 0. For example, $u_i((x_i, s_i)) = 20$ for all i if $x_i = a$ for all i and $s_{ii} = 1$ for all i. It is straightforward to check that this defines a symmetric TU game. Moreover it is easy to see that if $(x, s) \in \beta(\mathscr{P})$, then for any for any $S \in \mathscr{P}$ and $i \in S$, $s_{ij} = 0$ for all $j \notin S$. In particular, there are no transfers in any Nash equilibrium. The unique Nash equilibrium of this game is $x_i = x_{ic}$ and $s_{ii} = 1$ for all i. The corresponding payoffs are (15, 15, 15). The aggregate payoff to a best response of the grand coalition is 60, involving all players choosing $x_i = x_{ia}$. For the coalition structure $\mathscr{P} = (\{1, 2\}, \{3\})$, if $(x, s) \in \beta(\mathscr{P})$, then x is either (x_{1a}, x_{2b}, x_{3c}) or (x_{1b}, x_{2a}, x_{3c}). The aggregate payoffs to the two coalitions are 41 and 5 respectively. Corresponding to the

best response strategies, therefore, we have the following aggregate payoffs in the three kinds of coalition structures:

$$v(N) = 60, \quad v(\{i, j\}, \{k\}) = (41, 5), \quad v(\{1\}, \{2\}, \{3\}) = (15, 15, 15). \quad (6.8)$$

Now it should be obvious that the game is not one with positive externalities.

Consider the case in which each coalition divides the aggregate payoff equally among its members. A two player coalition can assure each member a payoff of 20.5. No e-blocking of this allocation is possible. But this also implies that the grand coalition cannot contain any *equal* binding agreement. On the other hand, it can be shown that the grand coalition does contain some (unrestricted) binding agreement.

To see this we begin by observing that if $\mathscr{P} = (\{i, j\}, \{k\})$, then there exists $(x, s) \in \mathscr{B}(\mathscr{P})$. And any such (x, s) must satisfy

$$u_i \geqslant 15, \quad u_j \geqslant 15, \quad u_i + u_j = 41, \qquad \text{and} \qquad u_k = 5. \qquad (6.9)$$

This observation is simply based on (6.8). Construct $(x, s) \in \beta(\{N\})$ such that $u = (5.5, 27, 27.5)$. From (6.8), this can be done. Now, we claim that $(x, s) \in \mathscr{B}(\{N\})$.

Suppose $(\mathscr{P}', x')$ blocks $(\mathscr{B}(\{N\}), (x, s))$. If the leading perpetrator is a singleton, $\{k\}$, then $\mathscr{P}'$ cannot be $\mathscr{P}^*$ because there does exist a binding agreement in the intermediate coalition structure. Thus $\mathscr{P}' = (\{k\}, \{i, j\})$. But then, by (6.9), $u_k(x') = 5 < u_k(x, s)$ which means that (B.2) of the blocking condition is not satisfied. The only other possibility is that the leading perpetrator is a two-player coalition $\{i, j\}$ and $\mathscr{P}' = (\{i, j\}, \{k\})$. By (6.9), $u_i(x') \geqslant 15$, $u_j(x') \geqslant 15$ and $u_i(x') + u_j(x') = 41$. But this means that either $u_i(x') \leqslant 27$ or $u_j(x') \leqslant 27$. Since at least one of these players was receiving a payoff of at least 27 under the strategy profile (x, s), we get a contradiction to (B.2) of the blocking condition. Thus $(x, s) \in \mathscr{B}(\{N\})$, and this completes the proof that $\{N\}$ does admit a binding agreement.

This example also shows that Proposition 6.1 depends on the game being one with positive externalities. As we have see, there exists an equilibrium binding agreement for the grand coalition yielding the payoff profile (5.5, 27, 27.5). Yet there exists $x' \in \mathscr{B}(\mathscr{P}^*)$, the Nash equilibrium, which yields a higher payoff to player 1.

7. A COURNOT OLIGOPOLY

The purpose of this section is to illustrate the nature of equilibrium binding agreements through another relatively simple but important

economic example. We will see that despite the simplicity of the model, the structure of equilibrium binding agreements can be fairly complex.

Consider a Cournot oligopoly. Assume a linear demand curve for a homogeneous product, and suppose there are n identical firms with constant average cost of production. It should be obvious that the only coalition structure that yields a Pareto optimal allocation is the grand coalition. Moreover, the grand coalition can ensure each player a higher payoff relative to the Cournot–Nash payoff. Thus, if binding agreements can be written, and if the *only* alternative to the grand coalition is Nash behavior, then it is clear that an equilibrium binding agreement will emerge in the grand coalition. However, Nash behavior is not the only alternative to the grand coalition if the number of firms exceeds 2. So it will be of particular interest to see if the grand coalition emerges as an equilibrium coalition under these conditions. Somewhat surprisingly, we find that this is not necessarily the case.

These are also the cases in which our theory yields a very different outcome from that predicted by the naive behavioral assumptions underlying either the β-core or the α-core. In the simple Cournot game with a linear demand curve, the β-core and the α-core both consist of all individually rational, Pareto optimal payoffs (since no coalition can do better if the complement produces enough to drive profits to zero). In other words, these theories *always* predict the formation of the grand coalition in Cournot oligopoly.

To be more specific, let the demand curve given by $p = a - by$, where p is the price and y is the aggregate demand. Denote the total cost function of firm i by cx_i, where c is a positive constant and x_i refers to firm i's output. Let $\mathscr{P} = \{S^1, ..., S^m\}$ be a coalition structure and s^i denote the number of firms in coalition S^i. Now, assuming that each coalition divides its profit equally, it is easy to see that if $x \in \beta(\mathscr{P})$, then the profit of a firm in coalition S^i is

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

This, coupled with the characterization result in Proposition 6.3, leads to considerable simplification in computing equilibria.

Salant, Switzer, and Reynolds [30] analyzed the best response equilibria of this model to show that if some firms merge while the others remain as singletons, this does not necessarily imply that the average profit of the merged firms is higher than the Nash profit.[33] In the language of this paper,

[33] Since the average monopoly profit is higher than the Nash profit, this phenomenon must relate to the merger of some strict subset of N.

they show, in particular, that if $x \in \beta(\mathscr{P})$, where $\mathscr{P} = S^K$ and $s < 0.8n$, then the average profit of coalition S is lower than the Nash profit. While they do not formalize a complete model of coalition formation, this result does turn out to be useful in Bloch's analysis of coalition formation [5] in the Cournot model. As we shall see, it can also be useful in characterizing equilibrium binding agreements.

It is easy to check that this model of Cournot oligopoly defines a symmetric TU game with positive externalities. By Proposition 6.3, the equilibrium coalition structures of this game can be computed by restricting attention only to those allocations that provide equal payoff to all members of each coalition, and using only the e-blocking conditions (E.1)–(E.2).

Observe that the only coalition structure that is efficient is the grand coalition. This follows from the fact that the aggregate utility in a coalition structure with m coalitions (with $m > 1$) is simply $m(a-c)^2/b(m+1)^2$, whereas in the grand coalition it is $\frac{1}{4}(a-c)^2/b$. The latter expression is larger than the former. Efficiency can, therefore, be checked simply by analyzing the stability of the grand coalition assuming equal division.

We begin with two simple observations regarding equilibrium coalition structures. First, the coarsest equilibrium structures cannot be too "fine." Second, if the grand coalition is *not* an equilibrium structure, then any coalition structure that blocks it must be asymmetric. These insights are quite general and go beyond the particular example studied here.

PROPOSITION 7.1. (i) *For each $n \geqslant 2$, there exists at least one equilibrium coalition structure with no more than $2\sqrt{n} - 1$ coalitions in it.*

(ii) *The grand coalition cannot be blocked by a coalition structure that contains coalitions of equal size.*

Proof. (i) Fix n, the number of firms. Suppose that the grand coalition is an equilibrium coalition structure. In that case we are done. If not, there exists a coalition T with t firms in it and an equilibrium coalition structure with m coalitions in it (one of which is T), such that

$$\frac{1}{t}\,\frac{(a-c)^2}{b(m+1)^2} > \frac{(a-c)^2}{4bn}.$$

Rearranging this expression, and using the fact that $1 \leqslant t$, we see that $m < 2\sqrt{n} - 1$, which completes the proof.

(ii) In a coalition structure with coalitions of equal size, all coalitions receive the same aggregate profit. Assuming equal division, a potential leading perpetrator can, therefore, gain by deviating from the grand coalition if and only if the aggregate payoff to all firms increases. As we have already seen, this is impossible. ∎

The first part of the proposition provides an *upper* bound on the number of coalitions in at least one equilibrium structure. It states that there cannot be "too much" competition. If the grand coalition does not have an equilibrium binding agreement, then it means that there must be some *intermediate sized* coalition structure which is stable, destroying the grand coalition. Thus the grand coalition survives if there exist "large" zones of instability in intermediate coalition structures. As pointed out in the introduction and again in Section 5, this suggests a cyclical pattern (in the number of players) in the viability of the grand coalition. We explore this in what follows. The second part of the proposition shows that the grand coalition must be stable if the only other equilibrium coalition structures are symmetric. Thus, all cases of inefficiency imply the existence of at least one asymmetric equilibrium coalition structure despite the obvious symmetry of the game itself.

Let $\hat{\Pi}$ denote the set of equilibrium coalition structures. Consider a coalition structure $\mathscr{P} = \{S^1, ..., S^m\}$. By Proposition 6.3, $\mathscr{P} \notin \hat{\Pi}$ if there exists a coalition T (a subset of S^i for some i) which belongs to an equilibrium coalition structure $\mathscr{Q} \in \mathscr{R}(\mathscr{P})$ in which the average payoff to T exceeds that to S^i. Letting t denote the number of firms in T and m^j denote the number of coalitions in $\mathscr{Q}$, this average payoff is $(a-c)^2/tb(m^j+1)^2$. Since this must exceed the original average payoff, it follows that

$$t(m^j+1)^2 < s^i(m+1)^2. \tag{7.1}$$

Of course, $\mathscr{P} \notin \hat{\Pi}$ even if there exists a singleton leading perpetrator who can block $\mathscr{P}$. Now consider (7.1) in the special case where $\mathscr{Q} = \mathscr{P}^*$, the coalition structure of singleton coalitions. Assume without loss of generality, that $\mathscr{P}$ is arranged such that $s^j \geqslant s^{j+1}$ for all $1 \leqslant j \leqslant m-1$. Then (7.1) is satisfied in this special case if

$$s^1(m+1)^2 > (n+1)^2. \tag{7.2}$$

In other words, $\mathscr{P} \notin \hat{\Pi}$ if a single firm can do better at the Nash equilibrium, i.e., if (7.2) is satisfied.

We can now use these conditions, along with the results of Section 6, to identify equilibrium coalition structures. Consider the following steps.

Step 1. Discard all coalitions structures for which (7.2) holds. Note that the Salant, Switzer, and Reynolds [30] result can be quite useful here; it implies discarding all coalition structures of the form S^K where the size of S is less than 80% of n. Let Π^1 be the set of coalition structures that remain.[34] Certainly, Π^1 will contain $\{N\}$ and $\mathscr{P}^*$. And in looking for

[34] Notice that if Π^1 contains only symmetric coalition structures, then by Proposition 7.1 (ii) we can immediately conclude that N is stable.

equilibria we can now restrict attention to Π^1, the set of all the coalition structures such that no coalition could gain by inducing the finest partition $\mathscr{P}^*$.

Step 2. In this step we will partition Π^1 in a certain manner. Let

$$\hat{\Pi}^2 = \{\mathscr{P} \in \Pi^1 \mid \mathscr{R}(\mathscr{P}) \cap \Pi^1 = \mathscr{P}^*\}.$$

These are those coalition structures of Π^1 in which coalitions can induce only the finest partition. Since no subcoalition can gain by doing so, it follows that $\hat{\Pi}^2 \in \hat{\Pi}$.

Step 3. Having obtained $\hat{\Pi}^2$, define Π^3 as the set of coalition structures $\mathscr{P}$ satisfying the following:

(i) $\mathscr{P} \in \Pi^1 \backslash (\hat{\Pi}^2 \cup \mathscr{P}^*)$;

(ii) there does not exist $S \in \mathscr{P}$, $T \subset S$ and $\mathscr{P}' \in \hat{\Pi}^2$ such that $T \in \mathscr{P}'$ and the average utility of coalition T is higher in $\mathscr{P}'$ than in $\mathscr{P}$.

The calculation in (ii) above is based on (7.1). Since $\hat{\Pi}^2 \in \hat{\Pi}$, according to Proposition 6.3, any coalition structure satisfying (i) but not (ii) cannot be in $\hat{\Pi}$. Now define

$$\hat{\Pi}^3 = \{\mathscr{P} \in \Pi^3 \mid \mathscr{R}(P) \cap \Pi^3 = \varnothing\}.$$

Clearly, $\hat{\Pi}^3 \subseteq \hat{\Pi}$.

This recursive procedure defines a way of computing $\hat{\Pi}^i$. Since the number of coalition structures is finite, there exists i such that $\Pi^{i+1} = \varnothing$. In that case,

$$\hat{\Pi} = \mathscr{P}^* \cup \hat{\Pi}^2 \cup \cdots \cup \hat{\Pi}^i.$$

We shall now illustrate this process for the Cournot oligopoly for n ranging from 2 to 9. We find that $n = 2$, 3, 4, and 9 the coarsest stable partition is the grand coalition but for $n = 5$, 6, 7 and 8 it is not. This suggests a recurring pattern of efficiency just as in Section 5. A general verification of this pattern remains an open question.

$\boldsymbol{n} = 2$. In this case both coalition structures are stable and the coarsest one is the grand coalition.

$\boldsymbol{n} = 3$. There are three kinds of coalition structures to consider. If there are two coalitions, one with 2 firms then it is easy to see that (7.2) holds so this is not stable. Thus $\mathscr{P}^*$ and $\{N\}$ are the only ones which are stable.

Henceforth we shall find it more convenient to represent the process through a table. Coalition structures which survive step 1 are denoted by

a dot in the column which indicates Π^1. The next column indicates a blocking structure, if any. The last column indicates those coalition structures that belong to $\hat{\Pi}$.

$\boldsymbol{n} = 4$.

$$(n+1)^2 = 25$$

	s^1	s^2	s^3	s^4	$s^1(m+1)^2$	Π^1	deviations if any	$\hat{\Pi}$
$\mathcal{P}^1$	4				$4 \times 4 < 25$	•		•
$\mathcal{P}^2$	3	1			$3 \times 9 > 25$		$\downarrow \mathcal{P}^*$	
$\mathcal{P}^3$	2	2			$2 \times 9 < 25$	•		•
$\mathcal{P}^4$	2	1	1		$2 \times 16 > 25$		$\downarrow \mathcal{P}^*$	
$\mathcal{P}^*$	1	1	1	1	25	•		•

$$(n+1)^2 = 36$$

	s^1	s^2	s^3	s^4	s^5	$s^1(m+1)^2$	Π^1	deviations if any	$\hat{\Pi}$
$\mathcal{P}^1$	5					$5 \times 4 < 36$	•	$\downarrow \mathcal{P}^5$	
$\mathcal{P}^2$	4	1				$4 \times 9 = 36$	•	$\downarrow \mathcal{P}^5$	
$\mathcal{P}^3$	3	2				$3 \times 9 < 36$	•	$\downarrow \mathcal{P}^5$	
$\mathcal{P}^4$	3	1	1			$3 \times 16 > 36$			
$\mathcal{P}^5$	2	2	1			$2 \times 16 < 36$	•		•
$\mathcal{P}^6$	2	1	1	1		$2 \times 25 > 36$			
$\mathcal{P}^*$	1	1	1	1	1	36	•		•

In this case $\hat{\Pi}^2 = \mathcal{P}^5$. As the column $\hat{\Pi}$ indicates, all the other partitions in Π^1 (except for $\mathcal{P}^*$) are unstable. In particular, $\mathcal{P}^3$ is not stable because 1 firm in the three-firm coalition can induce $\mathcal{P}^5$ and gain (the comparison here is between 3×9 and 16). The same is true of $\mathcal{P}^1$ and $\mathcal{P}^2$.

For $n = 6$, 7, and 8 also it can be shown that the grand coalition does not correspond to an equilibrium structure. We shall consider the case where $n = 9$, and show that the grand coalition is an equilibrium structure.

In this example $\mathcal{P}^{27}$ is the only element of $\hat{\Pi}^2$. Moreover,

$$\Pi^3 = \{\mathcal{P}^1, \mathcal{P}^8, \mathcal{P}^{19}\}.$$

In all other elements of $\Pi^1 \backslash (\hat{\Pi}^2 \cup \mathcal{P}^*)$, there is a leading perpetrator[35] that gets a higher payoff in $\mathcal{P}^{27}$. Since there exist no refinements of $\mathcal{P}^8$ and $\mathcal{P}^{19}$ that belong to Π^3, it follows that

$$\hat{\Pi}^3 = \{\mathcal{P}^8, \mathcal{P}^{19}\}.$$

[35] Typically, this is a single player from the largest coalition. The only exceptions are $\mathcal{P}^7$ and $\mathcal{P}^{15}$ — in these cases there is a two-player leading perpetrator from the largest coalition that can block.

Clearly, then $\Pi^4 = \mathscr{P}^1$. It is also easy to see that no coalition can gain by moving from $\mathscr{P}^1$ to $\hat{\Pi}^3$. Thus $\hat{\Pi}^4 = \mathscr{P}^1$ and

$$\hat{\Pi} = \{\mathscr{P}^1, \mathscr{P}^8, \mathscr{P}^{19}, \mathscr{P}^{27}, \mathscr{P}^*\}.$$

Thus $\{N\}$ remains an equilibrium coalition structure.

$\boldsymbol{n} = 9$.

$$(n+1)^2 = 100$$

	s^1	s^2	s^3	s^4	s^5	s^6	s^7	s^8	s^9	$s^1(m+1)^2$	Π^1	deviations if any	$\tilde{\Pi}$
$\mathscr{P}^1$	9									$9 \times 4 < 100$	•		•
$\mathscr{P}^2$	8	1								$8 \times 9 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^3$	7	2								$7 \times 9 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^4$	7	1	1							$7 \times 16 > 100$			
$\mathscr{P}^5$	6	3								$6 \times 9 < 100$	•	$\downarrow \mathscr{P}^{19}$ or $\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^6$	6	2	1							$6 \times 16 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^7$	6	1	1	1						$6 \times 25 > 100$			
$\mathscr{P}^8$	5	4								$5 \times 9 < 100$	•		•
$\mathscr{P}^9$	5	3	1							$5 \times 16 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{10}$	5	2	2							$5 \times 16 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{11}$	5	2	1	1						$5 \times 25 > 100$			
$\mathscr{P}^{12}$	5	1	1	1	1					$5 \times 36 > 100$			
$\mathscr{P}^{13}$	4	4	1							$4 \times 16 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{14}$	4	3	2							$4 \times 16 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{15}$	4	3	1	1						$4 \times 25 = 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{16}$	4	2	2	1						$4 \times 25 = 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{17}$	4	2	1	1	1					$4 \times 36 > 100$			
$\mathscr{P}^{18}$	4	1	1	1	1	1				$4 \times 49 > 100$			
$\mathscr{P}^{19}$	3	3	3							$3 \times 16 < 100$	•		•
$\mathscr{P}^{20}$	3	3	2	1						$3 \times 25 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{21}$	3	3	1	1	1					$3 \times 36 > 100$			
$\mathscr{P}^{22}$	3	2	2	2						$3 \times 25 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{23}$	3	2	2	1	1					$3 \times 36 > 100$			
$\mathscr{P}^{24}$	3	2	1	1	1	1				$3 \times 49 > 100$			
$\mathscr{P}^{25}$	3	1	1	1	1	1	1			$3 \times 64 > 100$			
$\mathscr{P}^{26}$	2	2	2	2	1					$2 \times 36 < 100$	•	$\downarrow \mathscr{P}^{27}$	
$\mathscr{P}^{27}$	2	2	2	1	1	1				$2 \times 49 < 100$	•		•
$\mathscr{P}^{28}$	2	2	1	1	1	1	1			$2 \times 64 > 100$			
$\mathscr{P}^{29}$	2	1	1	1	1	1	1	1		$2 \times 81 > 100$			
$\mathscr{P}^*$	1	1	1	1	1	1	1	1	1	100	•		•

REFERENCES

1. R. Aumann, The core of a cooperative game without sidepayments, *Trans. Amer. Math. Soc.* **98** (1961), 539–552.

2. R. Aumann and R. Myerson, Endogenous formation of links between players and of coalitions: An application of the Shapley value, *in* "The Shapley Value: Essays in Honor of Lloyd Shapley" (A. Roth, Ed.), pp. 175–191, Cambridge Univ. Press, Cambridge, UK, 1988.
3. D. Bernheim, Rationalizable strategic behavior, *Econometrica* **52** (1984), 1007–1028.
4. D. Bernheim, B. Peleg, and M. Whinston, Coalition-proof Nash equilibria. I. Concepts, *J. Econ. Theory* **42** (1987), 1–12.
5. F. Bloch, Sequential formation of coalitions in games with externalities and fixed payoff division, *Games Econ. Behav.*, in press.
6. F. Bloch, Endogenous structures of associations in oligopolies, *Rand J. Econ.* **26** (1995), 537–556.
7. C. Carraro and D. Siniscalco, Strategies for the international protection of the environment, *J. Public Econ.* **52** (1993), 309–328.
8. B. Chakravorti and C. Kahn, Universal coalition proof equilibrium, mimeo, University of Illinois, Champaign, 1991.
9. P. Chander and H. Tulkens, "The Core of an Economy with Multilateral Environmental Externalities," CORE Discussion Paper No. 9550, 1995.
10. M. Chwe, Farsighted coalitional stability, *J. Econ. Theory* **63** (1994), 299–325.
11. R. Coase, The problem of social cost, *J. Law and Econ.* **3** (1960), 1–44.
12. G. Debreu, A social equilibrium existence theorem, *Proceedings of the National Academy of Sciences* **38** (1952), 886–893.
13. B. Dutta and D. Ray, A concept of egalitarianism under participation constraints, *Ecoometrica* **59** (1989), 615–636.
14. B. Dutta and D. Ray, Constrained egalitarian allocations, *Games Econ. Behav.* **3** (1991), 403–422.
15. B. Dutta, D. Ray, K. Sengupta, and R. Vohra, A consistent bargaining set, *J. Econ. Theory* **49** (1989), 93–112.
16. B. Dutta and K. Suzumura, "On the Sustainability of R&D through Private Incentives," Indian Statistical Institute Discussion Paper No. 93–13, 1993.
17. J. Greenberg, "The Theory of Social Situations," Cambridge Univ. Press, Cambridge, UK, 1990.
18. T. Groves and J. Ledyard, Optimal allocation of public goods: a solution to the free rider problem, *Econometrica* **45** (1977), 783–810.
19. S. Hart and M. Kurz, Endogenous formation of coalitions, *Econometrica* **51** (1983), 1047–1064.
20. T. Ichiishi, A social coalitional equilibrium existence lemma, *Econometrica* **49** (1981), 369–377.
21. W. Lucas, "On Solutions to *n*-Person Games in Partition Function Form," Ph.D. dissertation, University of Michigan, Ann Arbor, 1963.
22. W. Lucas and J. Maceli, "Discrete Partition Function Games," Publication TR–344, School of Operations Research and Industrial Engineering, Cornell University, 1977.
23. A. Mas-Colell, An equivalence theorem for a bargaining set, *J. Math. Econ.* **18** (1989), 129–138.
24. P. Milgrom and J. Roberts, "Economics, Organization and Management," Prentice–Hall, Englewood Cliffs, NJ, 1992.
25. D. Pearce, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52** (1984), 1029–1050.
26. D. Ray, Credible coalitions and the core, *Internat. J. Game Theory* **18** (1989), 185–187.
27. D. Ray and R. Vohra, "Equilibrium Binding Agreements," Brown University Working Paper No. 92–8, 1992.
28. D. Ray and R. Vohra, A theory of endogenous coalition structure, mimeo, Boston University, 1996.

29. R. Rosenthal, Cooperative games in effectiveness form, *J. Econ. Theory* **5** (1972), 88–101.
30. S. Salant, S. Switzer, and R. Reynolds, Losses from horizontal merger: the effects of an exogenous change in industry structure on Cournot–Nash equilibrium, *Quart. J. Econ.* **93** (1983), 185–199.
31. H. Scarf, On the existence of a cooperative solution for a general class of *n*-person games, *J. Econ. Theory* **3** (1971), 169–181.
32. W. Shafer and H. Sonnenschein, Equilibrium in abstract economies without ordered preferences, *J. Math. Econ.* **2** (1975), 345–348.
33. P. Shenoy, On coalition formation: A game theoretic approach, *Internat. J. Game Theory* **8** (1979), 133–164.
34. R. Thrall and W. Lucas, *n*-person games in partition function form, *Naval Res. Logist. Quart.* **10** (1963), 281–298.
35. M. Walker, A simple incentive compatible scheme for attaining Lindahl allocation, *Econometrica*, **49** (1981), 65–73.
36. S.-S. Yi, "Endogenous Formation of Customs Unions under Imperfect Competition," Dartmouth College Discussion Paper, 1994.
37. J. Zhao, The hybrid solutions of an *N*-person game, *Games Econ. Behav.* **4** (1992), 145–160.