Coalition Formation[*]

Debraj Ray[†]

Rajiv Vohra[‡]

January 2013

[*] Preliminary draft, prepared as a chapter for *Handbook of Game Theory, Volume 4*, ed. by Peyton Young and Shmuel Zamir, North-Holland, forthcoming.

[†] New York University; debraj.ray@nyu.edu.

[‡] Brown University; rajiv_vohra@brown.edu.

Contents

1. Introduction

This chapter surveys the sizable and growing literature on coalition formation. We refer to theories in which one or more groups of agents ("coalitions") deliberately get together to jointly determine their actions. The defining idea of a coalition, in this chapter, is that of a group which can coordinate agreements among its members, while it interacts noncooperatively with other non-member individuals and the outside world in general.

It is hard to overstate the importance of coalition formation in economic, political and social analysis. Ray (2007) gives several examples in which such a framework comes to life: cartel formation, lobbies, customs unions, conflict, public goods provision, political party formation, and so on. Yet as one surveys the landscape of this area of research, the first feature that attracts attention is the fragmented nature of the literature. The theories that bear on our questions range from collusive behavior in repeated games, to models of bargaining, to cooperative game-theoretic notions of the core, or notions of coalition-proofness in noncooperative games. To unravel the many intricacies of this literature would take far more than a survey. To prevent our terms of inquiry from becoming unmanageably large, we impose a basic restriction.

Note that two fundamental notions are involved (usually separately) in the many theories that coexist. One has to do with the *formation* of groups, the "process" through which a coalition comes together to coordinate its subsequent actions. The other aspect involves the *enforcement* of group actions, say as an equilibrium of an appropriate game. In this survey, we deliberately omit the latter issue. We presume that the problem of enforcement is solved, once coalitions have chosen to form and have settled on the joint actions they intend to take. Of course, that does not necessarily mean that we are in a frictionless world such as the one that Coase (1960) envisaged. After all, the negotiations that lead up to an agreement are fundamentally noncooperative. In addition, it is entirely possible that a negotiation once concluded will be renegotiated. Such considerations place enough analytical demands that we exclude the question of *implementing* an agreement, perhaps via dynamic considerations, from this survey. On the other hand, we are centrally interested in the former issue, which also involve "no-deviation" constraints as action maintenance would, but a different set of them. Just because a coalition — once formed — is cooperative, does not mean that the creation of that coalition took place in fully cooperative fashion.

An example will make this clear. Suppose that a market is populated by several oligopolists, who contemplate forming a monopolistic cartel. Once formed, the cartel will charge the monopoly price and split the market. A standard question in repeated games (and in this particular case, in the literature on industrial organization) is whether such an outcome can be sustained as an "equilibrium". If a player deviates from the cartel arrangement by taking some other action, there will be punitive responses. One asks that all such play (the initial action path as well as subsequent punishments) be sustainable as an equilibrium of noncooperative play. This is the problem of enforcement. Ignore it by assuming that an agreement, *once made*, can be implemented. In contrast, we emphasize the question of formation. Say that the formation of the cartel is being negotiated. A firm might make proposals to the others about cost and revenue sharing. Other firms might contemplate alternative courses of action, such as the possibility of standing alone. In that case they would have to predict what their compatriots would do; for instance, whether they would form a smaller cartel made up of the remaining firms, thus effectively converting the situation into a duopoly. This sort of reasoning in the negotiations process is as non-cooperative as the enforcement problem, but it is a different problem. It can exist *even if* we assume that an agreement, once made, can be enforced without cost. And indeed, that is just what we do here.

So questions of enforcement may be out of the way, but a variety of models and theories remain. The literature on coalition formation embodies two classical approaches that essentially form two parts of this chapter:

(i) *The blocking approach*, in which we require the immunity of a coalitional arrangement to "blocking", perhaps subcoalitions or by other groups which intersect the coalition in question. Traditionally, blocking has been employed in a negative way, as undermining or destroying proposed arrangements. As we shall show, blocking can also be viewed as part of the "negotiation process" that leads up to an agreement.[1]

There is, of course, an entire area of cooperative game theory devoted to such matters, beginning with the monumental work of von Neumann and Morgenstern (1944). This literature includes notions such as the stable set, the core and the bargaining set (von Neumann and Morgenstern (1994), Gillies (1953), Shapley (1953), Aumann and Maschler (1964)). Extensions of these ideas to incorporate notions of farsighted behavior were introduced by Harsanyi (1974), and later Aumann and Myerson (1988). The farsightedness notion — one that is central to this chapter — was furthered developed by Chwe (1994), Ray and Vohra (1997), Diamantoudi and Xue (2007) and others.

(ii) *Noncooperative bargaining*, in which individuals make *proposals* to form a coalition, which can be accepted or rejected. Rejection leads to fresh proposal-making. The successive rounds of proposal, acceptance and rejection take the place of blocking. Indeed, on the noncooperative front, theories of bargaining have served as the cornerstone for most (if not all) theories of coalition formation. The literature begins with the celebrated papers of Ståhl (1977), Rubinstein (1982) and Baron and Ferejohn (1989), the subsequent applications to coalition formation coming from Chatterjee et al. (1993), Okada (1996), Seidmann and Winter (1998)

[1] While we recognize that the term "block" has not been in favor since Shapley (1973), 'the blocking approach' seems to us to be preferable to 'the improve upon approach', or to the coining of a new term.

and several others. There is also a literature on coalition formation that has primarily concerned itself with situations that involve pervasive externalities across coalitions, such as Bloch (1996), Yi (1996), Ray and Vohra (1999). In several of these papers bargaining theory explicitly meets coalition formation.

We mean to survey these disparate literatures. This is no easy task. After all, the basic methodologies differ — apparently at an irreconcilable level — over cooperative and noncooperative game approaches, and even with each methodological area there is a variety of contributions. Moreover, this is by no means the first survey of this literature: Bloch (2003), Mariotti and Xue (2003), and Ray (2007) are other attempts and while the survey component varies across these references, all of them provide a perspective on the literature. It is therefore important to explain how we approach the current task, and in particular why our survey is so very different from these other contributions.

We proceed by suggesting a unifying way of assessing the literature and perhaps taking it further. We propose a framework for coalition formation that has the following properties:

A. It nests the blocking and bargaining approaches under one umbrella model, and in particular it permits a variety of existing contributions to be viewed from a single perspective.

B. It allows for players to be farsighted or myopic, depending on the particular model at hand.

C. It deals with possible cycles in chains of blocking, a common problem in cooperative game theory.

D. It allows for the expiry or renegotiation of existing agreements, or insists that all deals are irreversible, depending on the context at hand.

The chapter is organized as follows. In the next Section we present an abstract, dynamic model of coalition formation, followed by a definition of an equilibrium process of coalition formation (EPCF). This framework will be shown to be general enough to unify the various strands of the literature, and to suggest interesting directions for further research in the area. Section 3 concerns the blocking approach to coalition formation. Here we review some of the basic concepts in classical cooperative game theory that are based on notions of coalitional objections or "blocking". We show how some of the standard notions of coalitional stability, such as the core of characteristic function games, can be subsumed under our general notion of an EPCF, despite the fact that these standard cooperative models are static while our general framework is a dynamic one. We then illustrate some of the limitations of the blocking approach in environments with externalities and argue that our explicitly dynamic model provides a way to resolve some of these difficulties.

Section 4 is devoted to a review of coalitional bargaining in noncooperative games. Here "coalitional moves" are replaced by individual proposals to a group of agents, and acceptance of such a proposal signifies the formation of a coalition. Externalities can be incorporated by considering partition functions rather than characteristic functions, and an equilibrium of a bargaining game now describes an equilibrium formation of coalitions. There are three distinct branches of this literature depending on whether agreements are permanently binding or temporary and whether renegotiation of existing agreements is possible. A suitable specialization of our model seems well suited to encompass all of the bargaining models. We show that in general the process of coalition formation corresponding to an equilibrium of a bargaining game conforms to our notion of an EPCF. We then describe some results on coalition formation from this literature.

The general framework that we use lays bare a large degree of incompleteness in the literature, something that's evident in the asymmetry of exposition between Sections 3 and 4. In terms of the general framework that we lay out, the existing literature falls short on several counts, and in particular, different aspects receive disparate attention in the bargaining and blocking approaches. For instance, the blocking approach has been concerned with questions of "chains of coalitional objections" so that farsightedness has appeared as a natural component, as in Harsanyi (1974) or Chwe (1994). In contrast, the bargaining approach, with its insistence on a well-defined game-theoretic structure has invariably been more explicit about the structure of moves, which then lends itself more naturally to considerations such as the renegotiation of agreements (or the impossibility thereof). Unfortunately, these uneven developments are mirrored in the varying emphasis we lay on these matters at different points of the text, and one can only hope that a survey ten years hence would be far more balanced.

Finally, Section 5 concerns one of the most important questions in coalition formation; namely, the possibility of achieving efficiency when there is no impediment to the formation of coalitions. For the clearest understanding of this issue we assume away the two most commonly recognized sources of inefficiency. First, decentralized or non-cooperative equilibria may quite naturally yield inefficient outcomes, e.g. Nash equilibria in games or competitive equilibria in the context of "market failure" resulting from incomplete markets or externalities. These problems are explicitly assumed away when we allow for coalitions to make binding agreements.[2] Second, inefficiency may arise due to incompleteness of information, which we have assumed away. Thus there may be a presumption that in our framework Pareto efficiency will obtain in equilibrium. Indeed, much of cooperative theory is built on this presumption, and the Coasian idea that efficiency is inevitable in a world of complete information and unrestricted contracting is very much part of the economics folklore. As the recent literature shows, however, efficiency in

[2] Note that in our framework inefficiency cannot arise in a two-player game even if there are externalities; full cooperation (formation of the grand coalition) must be the equilibrium outcome if it Pareto dominates the "non-cooperative" outcome represented by singleton coalitions.

the presence of externalities, even in a world with complete information and no restrictions on coalition formation, may be more elusive. If coalitional agreements are permanently binding, then the possibility of inefficiency in equilibrium is a robust phenomenon. The ability to renegotiate agreements can restore efficiency under certain conditions, e.g., in the absence of externalities, but not in general. A detailed examination of these issues is provided in Section 5.

2. The Framework

In this section, we describe a general framework for coalition formation that serves as an umbrella for a variety of different models in the literature. Central to our approach is a description of coalition formation in real time, one which allows for both irreversible and reversible agreements, and yields payoffs as the process unfolds.

2.1. Ingredients.

Our framework contains the following components.

[1] A finite set N of *players*, a compact set X of *states*, an infinite set $t = 0, 1, 2 \ldots$ of *dates*, and an initial state x_{-1} at the start of date 0.

[2] For each player i, a continuous one period payoff function u_i defined on X, and a (common) discount factor $\delta \in (0, 1)$.

[3] For every pair of states x and y, a collection of coalitions $\mathcal{E}(x, y)$ that are *effective* in moving the state from x to y, with $\mathcal{E}(x, x)$ being the collection of all coalitions.

[4] A *protocol*, ρ, which defines a probability over the choice of an "active coalition" S at each date.

[5] Along with S, a (possibly empty) set of *potential partners* $P \subseteq N \setminus S$, also chosen by the protocol. The interpretation is that S has the exclusive right to propose to a (possibly empty) set of partners $Q \subseteq P$ a move to a new state for which $S \cup Q$ is effective.

[6] An *order of responses*, again given by the protocol, for every set of partners Q included by S in its proposed move. If any individual in Q rejects the proposal, the state remains unchanged. The first such rejector is "responsible" for that move not occurring.

[7] At each date t, possible *histories* h_t that begin at x_{-1} and list all active coalitions, partners, and moves up to period $t-1$, as well as any individual "responsible" for the refusal of past moves. The protocol ρ is conditioned on this history.

The basic idea is both general and simple. Each date t begins under the shadow of a "going history" h_t, as in [7]. If a coalition S_t becomes active, as in [4], then it moves to a possibly new state x_t with the help of partners Q_t chosen from the potential set P_t (see [5]); by "move" we refer to $m_t = (x_t, S_t, Q_t)$. Of course, $S_t \cup Q_t$ must be *effective* in implementing the new state x_t.[3] Note that by [3] an active coalition always has the option not to call upon any partners and keep the state unchanged at the status quo (that, too, is a move, by convention). Each player receives a return at each state, with discounted returns added over time to obtain overall payoffs or values; see [2]. The one additional feature we need to embed into histories is a record of the "responsible" individual (if any) who rejected a proposed move in any previous period in which the state did not change; see [6]. There will be no such individual if the active coalition in that period suggested no change, or chose no partners, or if all its partners accepted the proposed move. (The reason we track the rejectors is that in some bargaining models, the choice of future active proposers may depend on the identity of past rejectors.)

It should be noted that implicit in the existence of a "pivotal" or "responsible" individual for each rejected move is that a proposed move must be *unanimously* accepted by partners who have been invited by the active coalition, and who respond sequentially given the protocol. As we discuss later in Section 4.1.4, there is little loss of generality in making the unanimity assumption, provided we redefine coalitional worths in an appropriate way.

The process formally continues *ad infinitum*, but all movements in states may or may not cease at a particular date. That will depend on the specification of the model. For instance, it is possible that for some states x, $\mathcal{E}(x, y) = \emptyset$ for all $y \neq x$. If such an *end state* is reached, the process ends, though our formalism requires payoffs to be received from that final state "forever after". The notion of end states will be useful in the blocking approach. Another possibility is that after certain histories the protocol may cease to choose active coalitions. This is useful in models in which the renegotiation of an existing agreement is not permitted.

The two leading applications of this general framework are, of course, to theories of coalition formation based on cooperative game theory, or what we will refer to here as the *blocking approach*, and a parallel theory based on noncooperative bargaining, what we will call here the *bargaining approach*. In the blocking approach, active coalitions do not make proposals to partners; formally, this is captured by having an empty partner set at every history. In the bargaining approach, active coalitions are invariably *individuals*, and proposals made by them to other players (the "partners") will occupy center stage. This allows us to nest a variety of solution concepts under a common descriptive umbrella.

A schematic representation of the model is shown in Figure 1. Time goes from left to right. The squares in the upper panel depict implemented states $x_{-1}, \ldots, x_t$, while the ovals in the lower panel denote various individuals. Following every history, the dark ovals denote members of an active coalition and the light ovals

[3] The importance of employing a notion of effectiveness in cooperative games was emphasized by Rosenthal (1972).

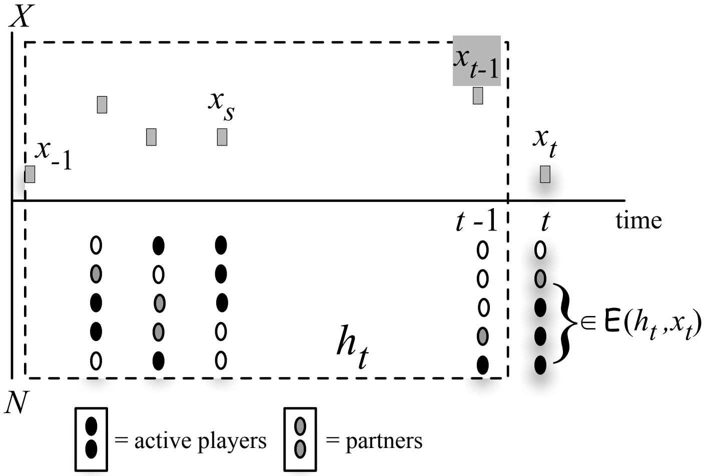

FIGURE 1. TIMELINE

are partners. Together, these create histories for later dates. The process continues, possibly indefinitely, and payoffs are received at every date.

2.2. **Process of Coalition Formation.** A *process of coalition formation* (PCF) is a stochastic process λ that takes existing histories (beginning with $h_0 \equiv \{x_{-1}\}$) to subsequent histories, consistent with the following restrictions:

(i) At every history h_t, λ is consistent with the probability measure induced by the protocol in selecting an active coalition S_t and potential partner set P_t, and

(ii) Every move $m_t = (x_t, S_t, Q_t)$ generated by λ has the properties that $Q_t \subseteq P_t$ and $S_t \cup Q_t \in \mathcal{E}(x_{t-1}, x_t)$.

Think of the active coalition S_t as "moving" the process from state x_{t-1} to x_t with the consent of its partners Q_t.

Why might the process be stochastic? One obvious reason is that the protocol that selects an active coalition may be stochastic; for instance, it may choose an active coalition equiprobably from the set of all possible coalitions. In addition, agents and coalitions may randomize their decisions; e.g., the choice of partners or moves.

A PCF defines *values* for every person i at every history h_t:

$$V_i(h_t, \lambda) \equiv \mathbb{E}_\lambda \left(\sum_{s=t}^{\infty} \delta^{s-t} u_i(x_s) | h_t \right),$$

where all states from t onwards are generated by λ conditioned on the history h_t. Note that the expectation is taken prior to the choice of active coalition, which is "about to happen" following that history.

2.3. **Equilibrium Process of Coalition Formation.** We now define an equilibrium process of coalition formation. Suppose that a PCF λ is in place and a particular coalition S is active at some date, along with a set of potential partners. Both S and its chosen set of partners must "willingly participate in the law of motion" of the process. We judge such participation by two criteria: (a) the move to a new state — which includes the possibility of no move — must be "profitable" for both S and any partners that are called upon to implement the move; and (b) of all the profitable moves that can be entertained, there is no other move that S can make that would be better for all the members of S.

Formally, a move $m = (x, S, Q)$ at history h_t and going state x_{t-1} is *profitable* for $S \cup Q$ if $S \cup Q \in \mathcal{E}(x_{t-1}, x)$ and

$$V_i(h_{t+1}, \lambda) \geq V_i(h^i_{t+1}, \lambda) \text{ for every } i \in S \cup Q,$$

where for each $i \in S \cup Q$, h^i_{t+1} is a subsequent history consistent with the state remaining unaltered ($x_t = x_{t-1}$). For $i \in Q$ it is the history resulting from i's rejection of the proposed move, under the presumption that all earlier respondents accepted the proposal. For $i \in S$ it is the history resulting from the status-quo move. That is, a move is profitable if no member of the active coalition would have been better off with the status-quo move and no member of the partner set would have been better off rejecting the move.

Given active coalition S and potential partner set P, a move $m = (x, S, Q)$ at history h_t with state x_{t-1} is *efficient* for S if there is no other move, say $m' = (z, S, Q')$, that is profitable for $S \cup Q'$ for some $Q' \subseteq P$, such that

$$V_i(h'_{t+1}, \lambda) > V_i(h_{t+1}, \lambda) \text{ for every } i \in S,$$

where, as before, h_{t+1} is the history created by implementing m, and h'_{t+1} is the history created by the alternative profitable move m'. The interpretation is that S can always form the coalition $S \cup Q'$ and make the move m', so no such move must strictly dominate what S actually does under the going PCF.

Note well that profitability and efficiency are defined relative to an ongoing process: the value functions will vary with the process in question.

An *equilibrium process of coalition formation* (EPCF) is a PCF with the property that at every history, every active coalition, faced with a given set of potential partners, makes an efficient and profitable move. That is, all movers (weakly) prefer their "prescribed" move to inaction, and that move is not dominated for members of S by some other profitable move that S can engineer.

Our formulation is related to the equilibrium definition introduced by Konishi and Ray (2003), and extended by Hyndman and Ray (2007). It is also related to the solution concept used in Gomes and Jehiel (2005) and Gomes (2005). As in these papers, the definition allows for a fully dynamic model of widespread externalities in which coalitions are farsighted in "blocking" a status-quo. But there are some important differences. The current approach is designed to allow for a level of generality substantial enough to encompass not only the blocking approach but also the bargaining approach. We include the possibility that an active coalition can enlist the help of others to form an "approval committee" that is capable of carrying the proposed move. As we shall see by example, bargaining is a special case of our formulation in which S is always a singleton asking for acceptance of a proposal from other players (the "partners"). On the other hand, the blocking approach corresponds to the polar case in which the partner set is always empty. This formulation also opens up intermediate and unexplored possibilities, in which a (non singleton) coalition can choose partners.

Setting this innovation aside, Konishi and Ray also define efficient and profitable moves. However, there is no protocol that determines which coalition is active at a given date. Konishi and Ray require that if there is *some* coalition for which a *strictly* profitable move exists from the current state, then the state *must* change. That is, some such coalition must perforce become active. Under our definition, it is possible that some coalition has a profitable move, but is not active. This distinction is blurred if different "dates" are very closely bunched together, for then (under a natural full-support assumption) *every* coalition must sooner or later become active. Moreover, as we shall see, having an explicit protocol has the distinct advantage of overcoming coordination problems that can arise with the blocking approach in a dynamic context.[4]

Since we will be showing how various solution concepts in the literature correspond to an EPCF of a suitably specialized version of our general model, existence of an EPCF in each of these cases will follow from corresponding existence results in the literature. While it would be desirable to provide a general result on the existence of an EPCF, we shall not attempt to do so here.

2.4. **Some Specific Settings.**

The framework presented so far is quite general and potentially amenable to a wide range of applications. It will be useful to present some concrete illustrations of how this general framework can be specialized to cover various models of coalition formation. In particular, we shall discuss state spaces, protocols and effectively correspondences in a more explicitly game-theoretic setting.

2.4.1. *State Spaces.*

So far we have let the state space be an entirely abstract object. Given our specific interest in coalition formation, however, it should be the case that *at a minimum*, a state must describe the coalitions that have formed as well as the (one-period) payoffs to each of the agents.

[4] Both our definition and that of Konishi and Ray's share the feature that they implicitly rely on dynamic programming; in particular, on the use of the one-shot deviation principle. Because coalitions have vector-valued payoffs, it is possible that the one-shot-deviation principle fails when every member of a coalition is required to be better off from a deviation. This issue is discussed in Konishi and Ray (2003) and Ray (2007).

But states can be suitably "tagged" to encode more information. For instance, in a situation in which all previous actions are irreversible, it is important to keep track of precisely which agents are "free" to engage in further negotiation. In such cases it suffices to tag every agent as "committed" or "uncommitted". When agreements are temporary or renegotiable, every agent is open in principle to further negotiation. However, the concept of effectivity will need to be suitably altered; for instance, with renegotiation, existing signatories to an agreement will need to be part of an effective coalition in any move to a new state even if they are physically not required in the implementation of that state.

In other models a description of the state in terms of ongoing *action vectors* may be necessary. This will be the case if such actions (rather than the formation of coalitions) are viewed as irreversible, at least for a certain length of time. In yet others there might be limits on the number of times a particular coalition is permitted to move, so that a state will need to keep track of this information.

Finally, in other settings that we do not emphasize in this survey, a state might describe a network structure, in which the links of each agent to every other agent are fully described.[5]

2.4.2. *Characteristic Functions and Partition Functions.* A particularly important object that is used to create the state space in a variety of situations is a mapping that assigns each coalition structure (or partition of the player set) to a set of payoffs for each of its member coalitions. The simplest version of such a mapping comes from a cooperative game in characteristic function form as defined by von Neumann and Morgenstern (1944). Such a game is defined as (N, V), where N denotes a finite set of players and for every coalition S (a non-empty subset of N), the set of feasible payoff profiles is denoted $V(S) \subseteq \mathbb{R}^S$.[6] In this setting a state may consist simply of a coalition structure and a feasible payoff profile. Typically, for a state x we will denote by $\pi(x)$ the corresponding coalition structure and $u(x)$ the profile of payoffs, where $u_S = (u_i)_{i \in S} \in V(S)$ for each $S \in \pi(x)$. In some applications we will find it more convenient to restrict payoffs to be efficient, i.e., $u(S) \in \bar{V}(S)$ for each $S \in \pi(x)$ where $\bar{V}(S, \pi) = \{w \in V(S, \pi) \mid \text{there is no } w' \in V(S, \pi) \text{ with } w' \gg w\}$.[7]

The presumption that the feasible payoffs for coalition S can be described independently of the players in $N \setminus S$ is easily justified if there are no external effects across coalitions. In many interesting models of coalition formation, however, the feasible payoffs to a coalition depend on the behavior of outsiders. As we shall see,

[5] There is a literature on networks that we do not address in this survey, but it is worth pointing out that our general framework incorporates the case of network formation as well. See Jackson (2010) for a comprehensive review of the literature on networks. For network formation in particular, see, for example, Jackson and Wolinsky (1996), Dutta, Ghosal and Ray (2005), and Page, Wooders and Kamat (2005).

[6] $\mathbb{R}^S$ denotes the $|S|$ dimensional Euclidean space with coordinates indexed by members of S.

[7] We use the convention $\gg, >, \geq$ to order vectors in $\mathbb{R}^S$.

externalities can be very effectively incorporated into the analysis by generalizing the notion of a characteristic function to a *partition function* (see Thrall and Lucas (1963)), and this is the formulation we adopt for this paper.[8] In such a game, the feasible payoff profile for a coalition depends on how the complementary players are organized into coalitions of their own. For a partition function game (N, V), the feasible set for S is therefore written as $V(S, \pi) \subseteq \mathbb{R}^S$, where π is a coalition structure that contains S. The interpretation is that a coalition S embedded in an ambient structure π can freely choose from the set of payoffs $V(S, \pi)$. In this setting, again, at a minimum, a state x will refer to a pair consisting of a coalition structure $\pi(x)$ and a payoff profile $u(x)$ such that $u(x)_S \in \bar{V}(S, \pi(x))$ for every $S \in \pi(x)$.

Note, however, that all externalities in a partition function are fed through the existing coalition structure rather than on specific *actions* that non-coalitional members might take. How might a partition function be then compared to the more familiar setting of a game in which players choose actions from their respective strategy sets and the payoff for players in a coalition depends on the actions of all players? The answer is that the partition function, while a primitive object for our purposes, is often derived from just such a normal form setting. The idea is that each coalition in the structure (never mind for the moment how they have formed) can freely coordinate the actions of their members (they can write binding agreements), but they cannot commit to a binding course of play *vis-à-vis* other coalitions and so play non-cooperatively "against" them. More precisely, given a coalition structure for a normal form game, equilibrium non-cooperative play across coalitions can be defined by first defining coalitional best responses as those which generate vector-valued maximal payoffs for each coalitions, and then imposing Nash-like equilibrium conditions; see, e.g., Ichiishi (1981), Ray and Vohra (1997), Zhao (1992) and Haeringer (2004). The resulting set of equilibrium payoffs may then be viewed as a partition function.[9] In this way, a partition function can be constructed from a normal form game, with the understanding that an equilibrium concept is already built into the function to begin with.

Here are four examples of partition functions; for others, see Ray (2007).

EXAMPLE **1** (Cournot Oligopoly)**.** *A given number, n, of Cournot oligopolists produce output at a fixed unit cost, c. The product market is homogeneous with a linear demand curve: $p = A - bz$, where z is aggregate output. Standard calculations*

[8] For an extended discussion of the historical background on characteristic functions and partition functions see Chapter 11.2.1 of Shubik (1982) and Chapter 2.2 of Ray (2007).

[9] In general, when the payoffs to members of a coalition depend on the actions of outsiders, this procedure will yield a partition function rather than a characteristic function. It is possible, though, in our view, not advisable to make enough (heroic) assumptions on the behavior of outsiders to go all the way from a normal form to a characteristic function. For example, assuming the worst in terms of outsiders' actions leads to the α-characteristic function. See Section 3.9 for a critique of this approach.

tell us that the payoff to a single firm in an m-firm Cournot oligopoly is

$$\frac{(A-c)^2}{b(m+1)^2}.$$

If each "firm" is actually a cartel of firms, the formula is no different as long as each cartel attempts to maximize its total profits (and then freely allocated these profits among its members). Define

$$v(S,\pi) = \frac{(A-c)^2}{b\left[m(\pi)+1\right]^2},$$

where $m(\pi)$ is the number of cartels in the coalition structure π. Then it is easy to check that $V(S,\pi)$ is the collection of all payoff vectors for S that add up to no more than $v(S,\pi)$.

This partition function is particularly interesting, in that it does not depend on S at all, but only on the ambient coalition structure π.[10]

EXAMPLE **2** (Public goods)**.** *There are n people. Person i gets utility $c + h(g)$, where c is private consumption and g is a global public good, produced from the sum of individual contributions. Assume that coalitions can freely make transfers to members and can coordinate member actions without cost. To calculate the best response of coalition S, simply maximize, for given aggregate contributions $\overline{T}$ from the complement $-S$,*

$$\sum_{i\in S} c_i + sh(g)$$

subject to $g = g(T)$ and $T = \sum_{k\in S}(w_k - c_k) + \overline{T}$, where g is the production function for the public good.

It is easy to check that this formulation gives rise to a partition function.[11]

EXAMPLE **3** (Customs Unions)**.** *There are n countries, each specialized in the production of a single good. There is a continuum of consumers equally dispersed through these countries. They all have identical preferences. Impose the restriction that no country or coalition can interfere with the workings of the price system except via the use of import tariffs. Then for each coalition structure — a partition of the world into customs unions — there is a coalitional equilibrium, in which each customs union chooses an optimal tariff on goods imported into it.*

In particular, the grand coalition of all countries will stand for the free-trade equilibrium: a tariff of zero will be imposed if lump-sum transfers are permitted within unions.

[10] For literature related to this formulation, see Salant, Switzer and Reynolds (1983), Bloch (1996) and Ray and Vohra (1997).

[11] See Ray and Vohra (2001).

A specification of trade equilibrium for every coalition structure generates a partition function for the customs union problem.[12]

EXAMPLE **4** (Conflict)**.** *There are several individuals. Each coalition in a coalition structure of individuals expends resources to obtain a reward (perhaps the pleasures of political office). Resources may be spent to lobby, finance campaigns, or engage in cross-coalitional conflict, depending on the particular application. Suppose that the probability p_S that coalition S wins depends on the relative share of resources r expended by it:*

$$p_S = \frac{r_S}{r_S + r_{-S}}. \tag{1}$$

The per-capita value of the win will generally depend on the characteristics of the coalition (for instance, coalitional size, s); write this value as w_S. The coalition then chooses resource contributions from its members to maximize

$$s p_S w_S - \sum_{i \in S} c(r_i),$$

where c is the individual cost function of contributions, $r_S = \sum_{i\in S} r_i$, and r_{-S} is taken as given.

This generates a well-defined transferable-utility partition function.[13]

2.4.3. *Remarks on Protocols and Effectivity Correspondences.* Protocols, along with the specification of coalitions that are effective for each move, can capture various levels of complexity and detail.

When a protocol chooses a nonsingleton coalition to be active, we are typically in the world of cooperative game theory, in which that coalition takes the opportunity to "block" an existing state: here, to be interpreted here in more positive light as moving the current state to a new one. The reinterpretation is important. Under the classical view, the issue of "what happens after" a state is blocked is sidestepped; blocking is viewed more as a negation, as the imposition of a constraint that must be respected (think of the notion of the core, for instance). Here, a "block" is captured by a physical move to a new state made by the active coalition in question. Partner sets, while formally easy enough to incorporate in the definition, generally do not exist in the theory of cooperative games.

In contrast, in the world of noncooperative coalition formation, as captured (for instance) by Rubinstein bargaining, the protocol simply chooses a proposer: a singleton active coalition.

Here are some examples of protocols.

[12] For literature that relies on a similar formulation, see Krugman (1993), Krishna (1998), Ray (1998, Chapter 18), Aghion, Antràs and Helpman (2007) and Seidmann (2009).

[13] For related literature, see Esteban and Ray (1999), Esteban and Sákovics (2004), and Bloch, Soubeyran and Sánchez-Pagés (2006).

Uniform Protocol. A coalition (or proposer) is randomly selected to be active from the set of all possible coalitions (or proposers) at any date.

Rejector Proposes. The first rejector of the previous proposal is chosen to be the new proposer, while if the previous proposal passed, a new proposer is chosen equiprobably from the set of all "uncommitted" players.

Later, we adopt a generalization of both these well-known protocols.

Protocols that choose active coalitions must also address the question of potential partner selection. As we've discussed, in the blocking approach the potential partner set is typically empty, or so it is taken to be in existing literature. An active coalition must move on its own. In the bargaining approach, the potential partner set is typically the set of all individuals who are free to receive proposals.

This last item depends intimately on the situation to be analyzed. In models of irreversible agreements, every player who has previously moved (including one who has committed to do so on her own) is no longer free to entertain new offers, or to make them. In models of reversible agreements, *every* player is a potential partner at every date. To be sure, what they can achieve will depend on whether their pre-existing agreements have simply expired or are binding-but-renegotiable. That is a matter dealt with by the effectivity correspondence, and some remarks on this are in order.

There are two broad sets of considerations that are involved in specifying effectivity. The first has to to do with which coalitions can move at a particular state. For instance, as we have already seen, issues of renegotiation (or its absence) can preclude coalitions from moving twice, or perhaps moving only with the blessings of ancillary players who must then be incorporated into the specification of effectivity.

Other conditions that determine which coalitions can move might come from the rules that describe the situation at hand. For instance, in multilateral bargaining with unanimity, as in the Rubinstein model or its multi-player extensions, only the grand coalition of all players can be effective in making a move. On the other hand, in models in which a majority of players is needed in determining if a new proposal can be implemented, effective coalitions must be a numerical majority.

The second broad consideration that determines effectivity is an understanding of just what a coalition can implement when it does move. After all, the new state specifies payoffs not just for the coalition in question, but for *every* player. For instance, even when the game is described by the relatively innocuous device of a characteristic function, a coalition is presumably "effective" over moves that preserve existing payoffs to other coalitions untouched by its formation, while implementing for itself any payoff vector in its own characteristic function. But there still remains the question of what happens to members of the "coalitional fragments" left behind as the coalition in question forms. Often, the issue is settled by presuming that one can assign any payoff vector in the characteristic function of those fragments, which are after all, coalitions in their own right.

In summary, the effectivity correspondence describes the power that each set of players possesses over the implementation of a new state. This entails the use of several considerations that we discuss in greater detail as we go along. Effectivity specifies both physical feasibility as well as the rules of the game: legalities, constitutions and voting rules. In addition, and provided the state is described carefully, effectivity correspondences allow easily for situations with irreversible, temporary or renegotiable agreements. The protocol and the effectivity correspondence play complementary roles.

Various combinations of protocol and effectivity correspondence can be used to serve different descriptive needs. For instance, one might wish to capture the possibility that a coalition which does not move at a particular state remains inactive until the state changes:

A Single Chance at Any State. For any history h_t, let m be the maximal date at which the state last changed; i.e., m is the largest date τ such that $x_\tau \neq x_{\tau-1}$. Then, if $m < t-1$, exclude all active coalitions chosen between $m+1$ and t, and choose a coalition at random from the set of remaining players (no coalition is ever chosen again if the remaining set is empty). This restriction guarantees that at any state, a chosen coalition that does not "actively move" is debarred from being active again, until the state changes.

As another example, the protocol might only choose *doubleton* coalitions to be active, as in Jackson and Wolinsky (1996) on networks:

Link Formation in Networks. The protocol chooses a *pair* of players (perhaps randomly, and perhaps with restrictions depending on what has already transpired); such a pair can form a link in a network. In addition, the protocol can be augmented to choose singleton players to unilaterally sever existing links.

3. The Blocking Approach: Coalitions in Cooperative Games

The classical concepts in cooperative game theory are based on coalitions as the primary units of decision making. These concepts rely on the notion of coalitional "objections" or "blocking". A proposed allocation is *blocked* by some coalition if there is an alternative allocation, feasible for that coalition, that improves the payoffs of each of its members. Blocking is used as a test to rule *out* certain allocations from membership in the solution. What happens "after" a block is typically not part of the solution, which consists only of those allocations that are *not* blocked. The notion of the core is the leading example of the use of blocking as a negation.[14] Various notions of the bargaining set (for example, Aumann and

[14] The formal concept of the core for characteristic function games was introduced by Gillies (1953) and Shapley (1953). See Chapter 6 of Shubik (1982) for additional background.

Maschler (1964) and Dutta et al. (1989)), look beyond the immediate consequence of an objection by considering the possibility of counter objections. However, these considerations are really meant to refine the logic of negotiations underlying the stability of an allocation in the grand coalition rather than to describe coalition formation. While we will not be reviewing this literature, the interested reader is referred to Maschler (1992) for an extensive survey.

But the concept of blocking can also be used in a more "positive" way, as a generator of actual moves that actively involve the formation of coalitions or coalition structures. In these approaches, it becomes especially necessary to explicitly model what follows a blocking action. Often, there will be repercussions: the blocking of an allocation may be followed by additional moves. In such contexts the question of farsightedness becomes focal: do players only derive payoffs from the immediate results of their blocking activities, or must they consider the ongoing implications of their initial actions as further blocks are implemented in turn?

When taken to their logical limit, considerations such as these naturally provoke a view of coalition formation as one that occurs in real time, on an ongoing basis, with blocking translated into moves, and with the discount factor as a yardstick for judging just how farsighted the players are.

With these notions in mind, we review the blocking approach to coalition formation. First, we discuss several solution concepts that rely on blocking. We then show how some of these solutions can be usefully subsumed in the general concept of an EPCF we introduced in the previous section. Finally, we show how the EPCF serves to both illustrate and deal with some of the pitfalls of the blocking approach, in which a sequence of moves is viewed more as an abstract shorthand rather than an actual course of actions. In particular, we will argue that the simplicity afforded by an abstract notion of blocking becomes too restrictive in more general settings; for instance, in the presence of externalities, where deviations become hard to handle without an explicitly dynamic process of coalition formation. In summary, in many special cases of interest, the findings from a fully dynamic model parallel those from the classical, "timeless" theory of coalitional blocking. In addition, in the more general setting, the dynamic model yields a resolution to some of the difficulties that arise with the traditional approach.

3.1. **The Setting.** Much of the analysis that follows can be conducted in the setting of our general framework in Section 2. However, both from an expositional perspective and in an attempt to link directly to existing literature, it will be useful to work with partition functions. (See Section 2.4.2 where we've already introduced them.)

Our state space, then, will explicitly track two objects. A state x contains information about the going coalition structure $\pi(x)$ and a vector of payoffs $u(x)$, one for each player. It is natural to suppose that these arise from an underlying partition function $V(S, \pi)$ defined for every π and every coalition $S \in \pi$. That is, if we denote by u_S the restriction of u to coalition S, then $u(x)_S \in \bar{V}(S,\pi)$ for each $S \in \pi(x)$.

We will augment this traditional setting with an effectivity correspondence, $\mathcal{E}(x,y)$, that specifies the set of coalitions that have the power to change x to y. It is sometimes more convenient to use the notation $x \to_S y$ to denote $S \in \mathcal{E}(x,y)$. Denote by $\Gamma = (N, V, \mathcal{E})$ the (extended) partition function game.

While coalition structures directly generate externalities and affect payoffs in partition function games, they are of interest (as outcomes) even in characteristic function games. For an extensive treatment of various solution concepts in the context of exogenously given coalition structures, see Aumann and Dreze (1974). Our interests are closer to contributions such as Shenoy (1979) and Hart and Kurz (1983) that study the *endogenous* formation of coalition structures. Although this literature is mostly concerned with characteristic function games, thereby assuming away externalities, the payoff to members of a coalition can still depend on the entire coalition structure for strategic reasons.[15] Greenberg (1994) provides an excellent review of this literature.

Note that the traditional cooperative-game setting has no explicit notion of time. Yet solution concepts abound that take stock of "farsightedness" and therefore implicitly involve time. We will argue below that our abstract dynamic setting allows us to naturally incorporate such farsightedness and thereby integrate different solution concepts in this literature.

3.2. **Blocking.** We begin with the standard notion of blocking applied to Γ. A pair (T, y), where T is a coalition and y a state, is an *objection* to state x (or equivalently, T *blocks* x with y) if $T \in \mathcal{E}(x,y)$ and $u(y)_T \gg u(x)_T$.

This notion of an objection quickly leads to the fundamental concept of the core: a state is in the *core* of Γ if there does not exist an objection to it.

It also leads to the equally fundamental von Neumann-Morgenstern stable set: a set $Z \subseteq X$ of states is *stable* if no state in Z is blocked by any other state in Z (internal stability), and if every state *not* in Z is indeed blocked by some state in Z (external stability).

In the context of a characteristic function game, the effectivity correspondence is typically replaced with the requirement that an objection (T, y) satisfy $u(y)_T \in V(T)$ (or $u(y)_T \in \bar{V}(T)$). The standard notion of the core is defined as $C(N, V) = \{u \in \mathbb{R}^N \mid \nexists S \subseteq N \text{ and } u' \in V(S) \text{ with } u' \gg u_S\}$. It is important, however, to emphasize that in what follows we invoke the generality of an extended partition function game.[16]

[15] See in particular the Owen value; Owen (1977) and Hart and Kurz (1983).

[16] Indeed, it is possible to define blocking in a still more general setting (see, for example, Lucas (1992)). An *abstract game* is defined as $(X, \succ)$, where X is the set of states and $\succ$, referred to as a dominance relation, is a binary irreflexive relation on X. Say that y blocks x if $y \succ x$. In this general

3.3. **Consistency and Farsightedness.** There are two (related) drawbacks of this notion of blocking, which have both been studied in the recent literature. The first is that of consistency. The traditional notion of the core has been criticized for not being consistent; see Ray (1989) and Greenberg (1990). While allocations for the grand coalition are tested against possible objections from subcoalitions, the objections are not similarly tested for further objections. In some situations, this turns out only to be a conceptual issue that doesn't affect the set of states ultimately judged to be stable (see the discussion following Proposition 2 below). However, in general, the issue becomes impossible to ignore. We illustrate this with the following example.

EXAMPLE **5** (Cournot Oligopoly and Farsightedness)**.** *Consider Example 1 with three identical firms, each with a constant average cost of 2. Suppose the inverse demand function is $p = 14 - z$, where z denotes aggregate output. Suppose that all firms within a coalition are required to share profits equally. We will generally use π_N to denote the coalition structure containing the grand coalition alone, π_i the coalition structure in which i is a singleton and the other two are together, and π_0 the finest partition of three singletons. With minor abuse of notation, we will use i, j... to denote singleton coalitions as well as agents, and ij, ijk... to denote multi-agent coalitions. Thus, $\pi_N = \{123\}$, $\pi_i = \{i, jk\}$ and $\pi_0 = \{1, 2, 3\}$. Standard computation yields the following partition function:*

$$\begin{array}{lllllll} x_N: & \pi_N & = & \{123\}, & u(x_N) & = & (12,12,12) \\ x_1: & \pi_1 & = & \{1,23\}, & u(x_1) & = & (16,8,8) \\ x_2: & \pi_2 & = & \{2,13\}, & u(x_2) & = & (8,16,8) \\ x_3: & \pi_3 & = & \{12,3\}, & u(x_3) & = & (8,8,16) \\ x_0: & \pi_0 & = & \{1,2,3\}, & u(x_0) & = & (9,9,9) \end{array}$$

Clearly, any single player i can move from the grand coalition to π_i and, in turn, either j or k can move from π_i to π_0. So every singleton coalition i has a myopic objection to the state corresponding to the grand coalition; it can get 16 rather than 12 by unilaterally moving to π_i. However, this is not a sustainable gain. The intermediate structure π_i is itself unstable: either player in the two-player coalition jk will do better by moving to the finest coalition structure. The myopic blocking notion fails to take such repercussions into account.

The von Neumann-Morgenstern stable set resolves the consistency issue by only taking seriously those objections which are themselves stable. But that brings us to the second drawback of the traditional blocking notion: it is myopic. Such myopia creates problems with the notion of the stable set, as Harsanyi (1974) first pointed out. The following example, due to Xue (1998), provides a simple illustration of this problem.

formulation, the restrictions imposed by an effectivity correspondence are implicit in the definition of dominance.

EXAMPLE **6** (Stability and Farsightedness)**.** *Consider two players and three states. Suppose that only player 1 is effective in moving from state a to b and only player 2 is effective from b to c. The payoffs to the two players in each of the states are in parentheses in Figure 2.*

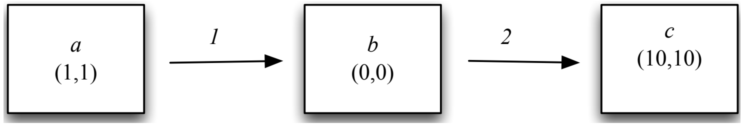

FIGURE 2. ILLUSTRATION OF EXAMPLE 6.

The core consists of states a and c. These two states also constitute the unique stable set. The stability of a, however, is based on myopic reasoning. Farsightedness on the part of player 1 would surely cause her to move to b, anticipating that player 2's self-interest will then lead to c as the final outcome. Clearly, stability in the sense of von Neumann and Morgenstern does not imply farsightedness.

This discussion motivates the concept of "farsighted blocking". A coalition moves, not necessarily because it has an immediate objection, but because its move can trigger further changes, leading *eventually* to a favorable outcome.

(T, y) is a *farsighted objection* to x if there is a collection of states $y_1, \ldots, y_m$ and a corresponding collection of coalitions, $T_1, \ldots, T_m$, where $T_1 = T$, such that $x \to_{T_1} y_1 \to, \ldots, y_{m-1} \to_{T_{m-1}} y_m$, and $u(y_m)_{T^k} \gg u(y_{k-1})_{T_k}$ for all $k = 1, \ldots, m$. We will often refer to y as farsightedly blocking x, leaving T implicit.

A farsighted objection pays no attention to what might transpire "immediately" after the objection is made. The first coalition to move may induce an "intermediate" state, in the anticipation that there may well be other states on the way to the "final" state. The definition asks that the objecting coalition be better off at the "end" of this process. Furthermore, it is required that every participant at every intermediate step, namely the coalitions T_k for $k \geq 2$, be better off "pushing" the process a step further at the state y_{k-1}, once again with the "final state" y in mind. Note that the coalition that initiates the sequence of moves has an optimistic view of the ensuing path. After all, there may be multiple potential continuations from the first step, but it is enough to find *some* sequence of moves that makes all the participating coalitions better off at the "final state".

The notion of farsighted blocking was suggested by Harsanyi (1974) in his critique of the stable set. It was formalized by Chwe (1994) in developing his notion of the largest consistent set, and introduced as "sequential blocking" in the the context of equilibrium binding agreements by Ray and Vohra (1997).

As we will discuss in Section 3.10, this definition of farsighted blocking is not without its own drawbacks. For now, it is imperative to note that the definition cannot make sense unless it is intimately tied to consistency. A farsighted objection y to x has not much meaning unless matters indeed terminate at y. A state that acts as a "credible" farsighted objection must itself have immunity with respect to the same kind of objections that can in turn be leveled at it. The "farsighted stable set" comes close to addressing these issues.

3.4. **The Farsighted Stable Set.** The marriage of farsightedness and consistency leads us to investigate the notion of a farsighted stable set (Harsanyi (1974)).[17] Say that a set Z^* of states is a *farsighted stable set* if no state in Z^* is farsightedly blocked by any other state in Z^* (internal stability), and if every state *not* in Z^* is farsightedly blocked by some state in Z^* (external stability). Put another way, if we attach a description to the states in (or not in) Z^* — call them stable (or unstable) – then no stable state has a farsighted objection that terminates in another stable state, while every unstable state does have such an objection.

Observe that at first sight, this appears to add very little at a conceptual level, simply replacing the blocking relation used for von-Neumann-Morgenstern stability by its farsighted analogue. But that is not the case. Recall from Example 6 that a coalition might "exploit" the von-Neumann-Morgenstern stable set by blocking some element in it, while profiting from that block even if the initial objection is "counter-objected" to by some other element in the stable set. In other words, *a coalition could be better off even by moving to an unstable state.*

That exploitation is not possible any more in the farsighted stable set. Suppose that a coalition T replaces a "stable point" $x \in Z^*$ by a new state w. If $w \in Z^*$, then w is "stable" and, in addition, cannot serve as a farsighted objection to x, so T cannot be better off by the internal stability property. If $w \notin Z^*$, then by external stability, there is a farsighted objection to w that leads to some $y \in Z^*$, which is "stable". But then T cannot be better off under y, for if it were, the entire sequence of states starting with w and terminating in y would act as a farsighted objection to x, which is ruled out by internal stability.

Thus, as we wrote above, the farsighted stable set captures the joint imposition of consistency and farsightedness. But there is a certain degree of bootstrapping implicit in the notion of a stable set. In the discussion above, a farsighted objection was treated as "credible" if it terminates in a "stable" state; i.e., a state in Z^*. But "stability" does not automatically guarantee that no further farsighted objection exists, only that such an objection must itself terminate in an "unstable" state, defined to be a state *not* in Z^*. Thus stability and instability need to be simultaneously defined.

[17] It is of interest to note that Harsanyi originally provided a definition that does not conform to the one given here, insisting in addition to farsightedness that each step of the blocking chain result in an instant improvement. However, in the last section of his paper, Harsanyi eliminates — correctly, in our opinion — this extraneous requirement; see also Chwe (1994) which makes this point.

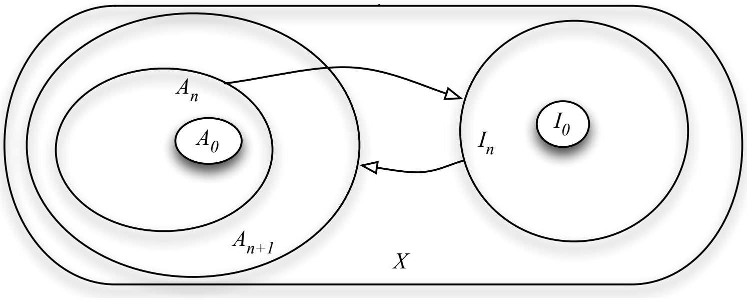

FIGURE 3. RECURSIVE DEFINITION FOR CREDIBILITY OF FARSIGHTED OBJECTIONS

In contrast, consider a recursive definition of stability. Say that a state is "admissible" if there exists no farsighted objection to it and let A_0 be the set of all such states. It is hard to quarrel with this as a sufficient requirement for the stability of a state, but as a definition it is incomplete. However, we should certainly label as "inadmissible" all states to which there exists a farsighted objection terminating in A_0. Let the set of such states be denoted I_0. Clearly $A_0 \cap I_0 = \emptyset$. We may now recursively proceed with the concept, labeling additional states as admissible, if any farsighted objection from such a state terminates in I_0. Label all such states at A_1. Note that $A_0 \subseteq A_1$ and $A_1 \cap I_0 = \emptyset$. Next, define I_1 to be the set of all states to which there exists a farsighted objection terminating in A_1. We can continue in this way to recursively define A_n and I_n given A_{n-1} and I_{n-1}. This process must widen the scope of admissibility and inadmissibility at every step of the recursion; see Figure 3 for a graphical depiction. In other words, for all n, $A_{n-1} \subseteq A_n$, $I_{n-1} \subseteq I_n$, and $A_n \cap I_n = \emptyset$.

Define $A^* = \cup_{n=0}^{\infty} A_n$ and $I^* = \cup_{n=0}^{\infty} I_n$. A^* may be viewed as the collection of "unambiguously stable" states, while every state in I^* is "unambiguously unstable". We will refer to A^* as the *farsighted core*.[18]

Note that the farsighted core is nonempty if and only if A_0 is nonempty. A sufficient condition for A_0 to be nonempty is the existence of end states, from which no further change is possible. (Recall that x is an end state if $\mathcal{E}(x, y) = \emptyset$ whenever $y \neq x$.) However, even if the farsighted core is nonempty, it is possible that the admissible and inadmissible sets do not cover the entire set of states. If they do, then, as the next Proposition shows, the bootstrapping inherent in the definition of a stable set may be entirely avoided.

PROPOSITION **1.** *Suppose $A^* \cup I^* = X$. Then the farsighted core, A^*, is the unique farsighted stable set.*

[18] Note that Béal et al. (2008) define the farsighted core to be the set of states immune to *any* farsighted objection, which is the same as A_0.

Proof. It is easy to check that A^* is a farsighted stable set. Now suppose that K is a farsighted stable set. Certainly $A_0 \subseteq K$, which implies, from the internal stability of K, that $I_0 \cap K = \emptyset$. This in turn means that $A_1 \subseteq K$. A recursive application of this argument shows that, for all n, $A_n \subseteq K$ and $I_n \cap K = \emptyset$. Thus, if all the admissible and inadmissible sets cover X it must be the case that $A^* = K$. ■

As we shall see in the next Section, the recursive approach works fully when the effectivity correspondence only permits subsets of existing coalitions to form.[19] Such situations may be important when communication becomes impossible across already formed coalitions.

3.5. Internal Blocking: A Recursive Definition.

The potential circularity of farsighted stability is easily avoided when only subsets of existing coalitions can precipitate new states. We refer to this as *internal blocking*.

For a partition π and a subcoalition T of some $S \in \pi$, denote by $\pi_{|_T}$ the partition obtained from π by dividing S into T and $S \setminus T$, leaving all other elements of π unchanged. If $\pi = (S, S_1, \ldots S_m)$, and $T \subset S$, then $\pi_{|_T} = (T, S \setminus T, S_1, \ldots S_m)$. A partition π' is said to be a *refinement* of π if every $T \in \pi'$ is a subset of some $S \in \pi$, and at least one is a strict subset. It is said to be an *immediate refinement* of π if $\pi' = \pi_{|_T}$ for some $T \subset S \in \pi$.

Now we can describe internal blocking by placing a necessary restriction on the effectivity correspondence: if $T \in \mathcal{E}(x, y)$ and $x \neq y$, then $T \subset S$ for some $S \in \pi(x)$, $T \in \pi(y)$ and $\pi(y)$ is a refinement of $\pi(x)$. Thus, only a subset of an existing coalition is effective in making a non-trivial change to an existing state.

We will now describe a recursive procedure for constructing the farsighted core which exploits the assumption of internal blocking.[20]

With the restriction to internal blocking (and the use of efficient payoffs) the state corresponding to the finest partition, π_0, is an end state, and therefore belongs to A_0. Consider a state x such that the only refinement of $\pi(x)$ is π_0. Since payoffs in each state are required to be efficient, and objections can only make the coalition structure finer, any objection to x must lead to π_0. Now, $x \in I_0$ if there is an objection to it and $x \in A_1$ otherwise. Recursively, suppose that all states with associated partitions of cardinality $k + 1$ or greater (where $k < n$) have been labeled either admissible or inadmissible. Consider any state x with associated coalition structure of cardinality k. Since objections can only lead to states that

[19] This is by no means the only situation in which the recursive approach works. More generally, if bounds are placed on the number of times a coalition can move at any node, then one can carry out the same recursive procedure provided that states are appropriately defined to keep track of those bounds.

[20] The definition will resemble that of a coalition proof Nash equilibrium of Bernheim et al. (1987). But that concept is purely non-cooperative and is not directly related to the theme of this chapter.

have already been labeled admissible or inadmissible, x can now be assigned a unique label: admissible if no farsighted objection terminates in an admissible state and inadmissible otherwise. Continuing in this, all the way to the coalition structure π_N, every state can be identified as either admissible or inadmissible. By Proposition 1, all admissible states defined through this recursive procedure constitute the farsighted core as well as the unique farsighted stable set.

As we shall see, the farsighted core corresponds to different (familiar) solution concepts depending on the context. To explore those connections it will be useful to first draw a general connection between the farsighted core and the process of coalition formation.

3.6. **An Equilibrium Process and the Farsighted Core.** An important test of the versatility of our dynamic process of coalition formation is its connection to the (static) notion of the farsighted core. To study this connection we consider a process of coalition formation that inherits some of the simplicity of the static model. In particular, states will refer to a partition and feasible payoffs to coalitions within the partition, and the partner set for any active coalition will be taken to be empty. In additional to the effectivity correspondence, the only additional ingredient that will be necessary in defining a process of coalition formation is a protocol.

Recall that a protocol chooses an active coalition depending on the history. Here we will seek to condition the protocol on the history in a minimal way and require that at any given state a coalition has at most one chance to make a move. Suppose $x_{t-1} \neq x_{t-2}$. In other words, the immediate history at date t is not a result of inaction on the part of some coalition. In such a case, the protocol will depend only on the current partition, i.e., on $\pi(x_{t-1})$. It will choose a subcoalition T of some coalition $S \in \pi(x_{t-1})$ to be active with probability $\rho(T \mid x_{t-1})$. Given the restriction to internal blocking $\rho(T' \mid x_{t-1}) = 0$ for any T' that is not a subset of a coalition in $\pi(x_{t-1})$.[21] Suppose, on the other hand, $x_{t-1} = x_{t-2}$. Let $T \subset S \in \pi(x_{t-2})$ be the active coalition at $t-1$, i.e., the one that chose to keep the state unchanged, or effectively gave up an opportunity to make a real change. The protocol will then assign positive probabilities only to sub coalitions of S that have not (unlike T) already declined to change the state. If there are no such subcoalitions of S that remain, the protocol will choose a subcoalition of some other $S' \in \pi(x_{t-1})$ that hasn't similarly exhausted all chances to change x_{t-2}. If all coalitions in $\pi(x_{t-1})$ have already declined to move, the state remains unchanged, and becomes an absorbing state.

We will now explore some conditions under which the absorbing states of an EPCF coincide with the farsighted core, and thereby identify situations in which the dynamic process predicts the same set of stable outcomes as those emerging from the static model of internal blocking. To be consistent with the blocking definitions,

[21] Strictly speaking, this is not necessary. The protocol could assign a positive probability to such a coalition, but the only state this coalition would be effective for is x_{t-1}, and the only change would be that one unit of time would go by.

we will assume throughout the rest of this Section that any non-status-quo profitable move is *strictly* profitable for all members of the active coalition, i.e., there is a strict inequality in the corresponding definition of Section 2.3.

A state x is said to be an *absorbing state* of an EPCF λ if the process does not move from x whatever the history. Formally, x is an absorbing state of λ if $\lambda(h_{t+1} \mid h_t) = 1$ for all histories h_t, h_{t+1} such that $x_{t-1} = x_t = x$.

An EPCF is said to be absorbing if from any state it leads to an absorbing state in a finite number of steps.

Given our restrictions on the protocol, and because every change in a state only serves to refine the coalition structure, it follows that every process defined on $(N, V, \mathcal{E}, \rho)$ is absorbing.

We will now show that if the steps of a farsighted objection and those of a profitable move can be compressed into a single step, then there is a tight connection between the absorbing states of an EPCF and the farsighted core. This abstract result will be useful in Sections 3.8 and 3.9.

We say that an EPCF is *immediately absorbing* if whenever x is a transient state there exists an objection (T, y) to x such that y is an absorbing state.

We say that an extended partition function $(N, V, \mathcal{E})$ has the *one-step objection property* if whenever x is not in the farsighted core there exists an objection, (T, y) such that y belongs to the farsighted core. Thus, the initiating coalition achieves a higher payoff in the very first step.[22] These situations are important not only because of the added simplicity of the farsighted core but also because there is then no ambiguity stemming from the possibility of multiple continuation paths in a farsighted objection. The latter is crucial in drawing a connection with EPCFs. Otherwise, as we will see in Section 3.10, there may be good reason for not expecting the farsighted core to be related to an EPCF.

We can now present a result connecting the farsighted core to absorbing states of the dynamic model.

LEMMA **1.** *Suppose $(N, V, \mathcal{E})$ has the one-step objection property and λ is an immediately absorbing EPCF of $(N, V, \mathcal{E}, \rho)$. Then all absorbing states of λ coincide with the farsighted core of $(N, V, \mathcal{E})$.*

Proof. Consider an EPCF λ. The state corresponding to the finest coalition structure is clearly an absorbing state. It is also by definition in the farsighted core. Thus, the equivalence between absorbing states of an EPCF and the farsighted core holds for the finest coalition structure. We now use an induction argument to prove the result. Accordingly, assume that the result holds for all states with at least $k + 1$ coalitions in the coalition structure.

[22] Under this condition farsightedness reduces to consistency.

Suppose $\pi(x)$ consists of k coalitions and x is not an absorbing state. Since λ is immediately absorbing this means that there exists an objection (T, y) to x such that y is an absorbing state for T. Note that T must be a strict subset of some $S \in \pi(x)$ for it to have an objection. Thus $\pi(y)$ is a refinement of $\pi(x)$, and it follows from the induction hypothesis that y is in the farsighted core. This implies that x is not in the farsighted core and completes the proof that the farsighted core is contained in the set of absorbing states of λ.

Next, we show that if x is an absorbing state, with $\pi(x)$ consisting of k coalitions, it must be in the farsighted core. Suppose not. Then, by the one-step objection property, there exists y and a coalition $T \subset S \in \pi(x)$, where $T \in \mathcal{E}(x, y)$ and y is in the farsighted core. Since $\pi(y)$ is a strict refinement of $\pi(x)$, it follows from the induction hypothesis that y is an absorbing state. By hypothesis, no coalition moves from x regardless of the history. There must be some history for which the protocol chooses T when the current state is x. Coalition receives $u(x)_T$ in perpetuity by not moving and $u(y)_T$ in perpetuity by moving to y. Since $u(y)_T \gg u(x)_T$, this is a strictly profitable move, and a contradiction to the hypothesis that x is an absorbing state. ∎

It is of course important to identify assumptions on the primitive model that will allow us to appeal to Lemma 1. That we shall do in Sections 3.8 and 3.9.

3.7. **Characteristic Functions.** Suppose the partition function is actually a characteristic function, so there are no externalities. We shall now impose some restrictions on the effectivity correspondence which are natural, perhaps even implicit, in this setting. Throughout this Section, in addition to internal blocking, it is assumed that if $T \in \mathcal{E}(x, y)$ and $y \neq x$, then (i) $\pi(y) = \pi(x)_{|T}$, (ii) $u(x)_S = u(y)_S$ for all S that belong to both $\pi(x)$ and $\pi(y)$, (iii) $T \in \mathcal{E}(x, y')$ for any y' such that $\pi(y) = \pi(y')$ and $u(y')_i = u(y)_i$ for all $i \notin T$. Thus, whenever a coalition can change the state, it must move to an immediate refinement. While it may choose any efficient feasible utility profile for itself, it must leave undisturbed the payoff configuration in coalitions that remain unchanged as a result of this move.

In many situations, we may want to go a step further and allow *all* subsets to form. Say that $\mathcal{E}$ has *full support* if for every $T \subset S \in \pi(s)$ there exists a state $y \neq x$ with $T \in \mathcal{E}(x, y)$. Given internal blocking, it is easy to see that x is in the core of $(N, V, \mathcal{E})$ if and only if $u(x)_S \in C(S, V)$ for every $S \in \pi(x)$.[23] As our next result shows, under these conditions, the farsighted core coincides with the core.

PROPOSITION **2.** *Suppose $(N, V, \mathcal{E})$ is such that (N, V) is a characteristic function game and $\mathcal{E}$ is restricted to internal blocking and has full support. Then the farsighted core coincides with the core of $(N, V, \mathcal{E})$, i.e., x belongs to the farsighted core if and only if $u(x)_S \in C(S, V)$ for all $S \in \pi(x)$.*

[23] Recall that $C(N, V)$ denotes the standard notion of the core. For $S \subseteq N$, $C(S, V)$ refers to the core of the characteristic function (N, V) restricted to S.

Proof. Suppose x belongs to the core of $(N, V, \mathcal{E})$ but there exists a farsighted objection (T, y) to x. Given internal blocking, the coalition structure corresponding to y, $\pi(y)$, must contain a coalition $T' \subseteq T$. Since (T, y) is a farsighted objection to x, $u(y)_{T'} \gg u(x)_{T'}$. The full support assumption implies that T' is effective in moving from x to y' where $\pi(y') = \pi(x)_{|T'}$ and $u(y')_{T'} = u(y)_{T'} \gg u(x)_{T'}$. But this contradicts the hypothesis that x belongs to the core of $(N, V, \mathcal{E})$.

Suppose x is in the farsighted core. We now claim that it must be in the core of $(N, V, \mathcal{E})$. Suppose not. Then there exists $S \in \pi(x)$ such that $u(x)_S \notin C(S, V)$. Let (S_1, u_1) be an objection to $u(x)_S$ such that $u_1 \in C(S_1, V)$. This can always be assured by taking S_1 to be one of the smallest subcoalitions of S with an objection to $u(x)_S$. By internal blocking and full support, it follows that $S_1 \in \mathcal{E}(x, y_1)$ where $u(y_1)_{S_1} = u_1$ and $\pi(y_1) = \pi(x)_{|S_1}$. If $u(y_1)_{S \setminus S_1} \in C(S \setminus S_1, V)$, then (S_1, y_1) is a farsighted objection to which there cannot be any objection from a subset of S_1 or $S \setminus S_1$. If $u(y_1)_{S \setminus S_1} \notin C(S \setminus S_1, V)$ we can find some subcoalition in $S \setminus S_1$, say S_2, with an objection from the core of S_2. Continuing in this way it is possible to construct a partition $(S_1, \ldots S_m)$ of S and $(u_1, \ldots u_m)$ such that $u_i \in C(S_i, V)$ for every i. Clearly, there is no farsighted objection to $((S_1, \ldots, S_m), (u_1, \ldots, u_m))$ from any allowable coalition in $(S_1, \ldots, S_m)$. This procedure can be applied to any $S' \in \pi(x)$ for which $u(x)_{S'} \notin C(S', V)$. All such objections can be collected into one farsighted objection which culminates in x' where $\pi'(x)$ is a refinement of $\pi(x)$ and $u(x')_T \in C(T, V)$ for all $T \in \pi(x')$. Of course, this must mean that there is no further farsighted objection to x'. But this contradicts the hypothesis that x is in the farsighted core. ■

Note that one of the steps in the above proof relies on the property that if $u \notin C(N, V)$ then there exists an objection (S, u') such that $u' \in C(S, V)$. The only reason this doesn't imply the one-step objection property is because coalitions other than S may also need to "move" in order to arrive at a stable outcome. If we ignore the rest of the coalition structure, then farsighted blocking becomes equivalent to myopic blocking and Proposition 2 can we be seen as the coalition structure analog of Ray (1989) and Proposition 6.1.4, Greenberg (1990).

We can now turn to a formal connection between the core and absorbing states of an EPCF.

PROPOSITION **3.** *Suppose (N, V) is a superadditive characteristic function game such that $V(S)$ is convex for all $S \subseteq N$. If x is in the core of $(N, V, \mathcal{E})$, then x is an absorbing state of every EPCF corresponding to $(N, V, \mathcal{E}, \rho)$.*

Proof. Suppose x is in the core but is not an absorbing state of an EPCF λ. This means that for some history at least one subcoalition of some $S \in \pi(x)$ has a profitable move. Let T be the last coalition according to the protocol which would choose to move from x to y. Of course, if T were to choose not to move, it would be assured of $u(x)_T$ in perpetuity. If y is an absorbing state, the fact that it is a profitable move for T contradicts the hypothesis that x is in the core. Thus, with positive probability, there is a further move from y. All possible paths leading from y must reach an absorbing state in a finite number of steps. Let m be the maximum number of steps along any such path before an absorbing state is reached. For every step $i = 1, \ldots, m$, following T's move, let μ^i be the probability measure on the states generated by λ. Denote by u^i the corresponding expected utility profile at stage i: $u^i = \int u(x) d\mu^i$. Since T has a strictly profitable move,

$$u_T^1 + \delta u_T^2 + \delta^2 u_T^3 + \ldots + \frac{\delta^{m-1}}{(1-\delta)} u_T^m \gg \frac{1}{(1-\delta)} u(x)_T. \tag{2}$$

It follows from superadditivity and the convexity of $V(S)$ that

$$u_T^j \in V(T) \text{ for all } j = 1, \ldots, m.$$

Letting

$$\hat{u} = (1-\delta)u^1 + \delta u^2 + \delta^2 u^3 + \ldots + \frac{\delta^{m-1}}{(1-\delta)} u^m$$

(2) can be rewritten as:

$$\hat{u}_T \gg u(x)_T. \tag{3}$$

Note that $\hat{u}$ is a convex combination of $u^1, \ldots, u^m$. Since $u_T^j \in V(T)$ for all j, it follows that $\hat{u}_T \in V(T)$, but then (3) contradicts the hypothesis that x is in the core. ■

PROPOSITION **4.** *Suppose x is an absorbing state of an EPCF of $(N, V, \mathcal{E}, \rho)$. If $\mathcal{E}$ satisfies the full support property, then x belongs to the farsighted core of $(N, V, \mathcal{E})$.*

Proof. Suppose x is an absorbing state but does not belong to the farsighted core. By Proposition 2, there exists $S \in \pi(x)$ such that $u(x) \notin C(S, V)$. Moreover, there exists $T \subset S$ and $u' \in C(T, V)$ such that $u' \gg u(x)_T$. Since x is an absorbing state, T receives $u(x)_T$ in perpetuity by not moving. However, by the full support assumption, it could move to a state y in which it receives u'. Since $u' \in C(T, V)$, we know from the previous Proposition that no subcoalition of T can move to a higher payoff. The only possible moves from y must come from coalitions in $N \setminus T$. Since that has no affect on the payoff to T we conclude that by moving to y coalition T can receive $u' \gg u(x)_T$. But this contradicts the hypothesis that x is an absorbing state. ■

Combing Propositions 3 and 4 we have:

PROPOSITION **5.** *Suppose (N, V) is a superadditive characteristic function game such that $V(S)$ is convex for all $S \subseteq N$ and $\mathcal{E}$ satisfies the full support property. Then all absorbing states of every EPCF of $(N, V, \mathcal{E}, \rho)$ coincide with the set of core (or farsighted core) of $(N, V, \mathcal{E})$.*

There is also an earlier literature that studies processes converging to core allocations. Green (1972), Feldman (1972), and Sengupta and Sengupta (1996) show how in a characteristic function game a process of recontracting can be constructed to lead from any non-core allocation to a core allocation. Recontracting refers to a process in which every active coalition makes a (myopic) improving move, without any guarantee of gaining at the end of the process. In contrast, Proposition 5 applies to farsighted behavior.

Konishi and Ray (2003) show how for a farsighted dynamic process can be constructed so as to have any particular core allocation as its absorbing state. They also provide conditions under which a deterministic process converges to a core allocation. Proposition 5 provides a stronger connection between the core and absorbing states since it concerns a coincidence of the set of core allocations and absorbing states of *any* EPCF. The key features of our model that make this possible are internal blocking and the specification of a protocol.

3.8. **Effectivity without Full Support.** It is important to stress that the full support property is not always natural. One important class of restrictions emanate from the possibility that additional disintegration of a newly-formed coalition may be legally or politically impossible. While we will have more to say about such "irreversible agreements" in Sections 4 and 5, in the current Section we discuss a model due to Acemoglu et al. (2008) in which the full support property does not hold for a very different reason. This is a model of a political game of coalition formation in which coalitions are farsighted and their ability to make a non-trivial move depends in an important way on the current state.

The political power of player i is described as $\gamma_i > 0$. A coalition $T \subseteq S$ is said to be *winning within* S if $\gamma_T > \alpha\gamma_S$, where $\gamma_T = \sum_{i \in T} \gamma_i$ and $\alpha \in [0.5, 1)$ denotes the degree of weighted supermajority required to win.

The payoff to players depends on the ultimate ruling coalition (URC). If S is the ruling coalition, $w(S)$ denotes the unique profile of utilities for members of S. Players outside the ruling coalition receive 0. A specific functional form, which we assume for convenience, is $w(S) = (\gamma_i/\gamma_S)_{i \in S}$.

In this model the only coalition of interest at each state is the ruling coalition and it will be useful therefore to define a state as $x = (w(S), S)$ with the interpretation that S is the ruling coalition.[24] We will use $R(x)$ to refer to the ruling coalition at x. Of course, coalition S can enforce such a state only from states in which it is a

[24] Although we have departed from our earlier formulation in replacing a coalition structure with a ruling coalition, this difference is not substantive. In particular, notions of the farsighted core and EPCF, as well as Lemma 1 are easily translated into the present model. Alternatively, we could retain the original formalism by associating with each ruling coalition the coalition structure in which all other players are singletons and, in addition, keeping track of the ruling coalition corresponding to every coalition structure of this form. The latter consideration is important for the coalition structure consisting of all singletons because in that case whether a singleton gets 0 or 1 depends on the identity of the ruling coalition.

winning coalition.

$$\text{For } x \neq y, \quad \mathcal{E}(x,y) = \begin{cases} R(y) & \text{if } R(y) \text{ is winning in } R(x) \\ \emptyset & \text{otherwise} \end{cases}$$

The effectivity correspondence does not satisfy the full support property because only winning coalitions can effect a non-trivial move. The model implicitly assumes internal blocking because a winning coalition must necessarily be a subset of the current ruling coalition. Given internal blocking, the recursive procedure described in Section 3.5 can be applied to determine the farsighted core. We illustrate this with the next Example.

EXAMPLE **7.** *There are three players with $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (4, 5, 6)$. A coalition $T \subseteq S$ is said to be winning within S if $\sum_{i \in T} \gamma_i > 0.5 \sum_{j \in S} \gamma_j$. The payoff profile for a ruling coalition is described as follows:*

$$\begin{array}{lcl} w(123) & = & (4/15, 5/15, 6/15), \\ w(12) & = & (4/9, 5/9), \\ w(13) & = & (0.4, 0.6), \\ w(23) & = & (5/11, 6/11), \\ w(i) & = & 1, \quad \textit{for all } i. \end{array}$$

Each two-player coalition is winning in N, and can therefore move to become a ruling coalition, and improve upon the status-quo. However, within each two-player coalition the more powerful player can win to become a singleton ruling coalition and earn 1. (The only way for a singleton to earn 1 rather than 0 is to form a winning coalition. Player 3 can do this from any two-player coalition, player 2 can do this only if the current coalition is 12, but player 1 is unable to do this from any two-player coalition). Thus, although there exist objections to the state corresponding to the grand coalition, none of them is credible because the weaker of the two players in the objecting coalition will be abandoned by the more powerful player at the next stage, ultimately doing worse than at the grand coalition. It is easy to see that N belongs to the farsighted core even though it is not in the core. Note that the singleton winning coalition from the two-player coalition prefers the final outcome, a payoff of 1, even at the grand coalition state but is not effective in making that move in a single step.

Notice that since players who are not in a ruling coalition earn 0, it is always better to belong to a ruling coalition than not. This implies that every farsighted objection must immediately end in a stable state. Otherwise, there must be at least one member of the initiating coalition who is left out of the final ruling coalition, and any such player would have been better-off not participating in the objection. In other words, every farsighted objection that ends in the farsighted core must be a one-step objection. This of course means that the one-step deviation property holds.

Now consider an EPCF for this model and assume that the protocol is deterministic: at every state there is a fixed order in which eligible coalitions are called upon to move. (We will presently specialize the protocol even further).

Suppose x is a transient state. Thus there a coalition T, winning in $R(x)$, with a profitable move. Without loss of generality, let T be the last such coalition given the protocol. If T does not move it is guaranteed a payoff of $u(x)$ forever, which must be less than the discounted payoff from moving. Note that all members of T experience an instantaneous gain by forming a winning coalition. However, if there is a further move (from a subcoalition), those left out of the second move then receive 0. In other words, if T is not an absorbing state, there is some $i \in T$ who receives 0 in all subsequent periods. The discounted payoff for i is therefore $w_i(T)$. For the move to be profitable, it must be the case that

$$w_i(T) > \frac{1}{1-\delta} u_i(x).$$

Since $w_i(T) \leq 1$ and $u_i(x)$ is bounded below by $\min_{i \in N} w_i(N)$, δ can be chosen close enough to 1 so that this is impossible. Thus, for δ high enough, every EPCF in this model must be immediately absorbing. As we've already observed, the one-step deviation property holds. We can therefore appeal to Lemma 1 to assert:

PROPOSITION **6.** *In the model of political coalition formation, for δ sufficiently close to 1, all absorbing states of an EPCF coincide with the farsighted core.*

With some additional assumptions it becomes possible to provide a sharper characterization of an EPCF in this model. Observe that for a winning coalition which can induce a state in the farsighted core, there is never any advantage in forgoing such an opportunity; a profitable move is also an efficient and profitable move. Given the protocol, there is a unique move from any state that is not an absorbing state: it is the move by the first coalition according to the protocol which is both winning and moves immediately to a state in the farsighted core. A lot depends, therefore, on the protocol.

Let

$\mathcal{T}(S) = \{T \subset S \mid T$ is winning within S and $(w(T), T)$ is in the farsighted core$\}$.

and let the first coalition in $\mathcal{T}(S)$ according to the protocol be denoted $\mathcal{T}^*(S)$. We can now describe the equilibrium process by a mapping ϕ, where

$$\phi(S) = \begin{cases} S \text{ if } (w(S), S) \text{ is in the farsighted core} \\ \mathcal{T}^*(S) \text{ otherwise.} \end{cases}$$

Starting from the grand coalition as the initial state, the process moves, in at most one step, to $\phi(N)$ as the absorbing state.

Acemoglu et al. (2008) provide a characterization of the subgame perfect equilibria of their extensive form game of political coalition formation by assuming that the power mapping γ, is generic in the sense that $\gamma_S \neq \gamma_T$ for $S \neq T$. By adopting this assumption, and imposing a restriction on the protocol we can obtain precisely their characterization through our framework. Suppose the protocol is such that among all winning coalitions, relative to the current state, priority is given to those with lower aggregate power. In other words, winning coalitions are arranged in ascending order of aggregate power: if S and S' are both winning, $\gamma_S < \gamma_{S'}$ implies that the protocol chooses S before S'. Given the genericity assumption, this means that $\mathcal{T}^*(S) = \arg\min_{A \in \mathcal{T}(S)} \gamma_A$. Now $\phi(S)$ can be written inductively as follows. Suppose $\phi(.)$ has been defined for all coalitions with fewer than k players. Then, for S with k players let

$$\phi(S) = \arg\min_{A \in \mathcal{T}(S) \cup S} \gamma_A,$$

where,

$$\mathcal{T}(S) = \{T \subset S \mid T \text{ is winning within } S \text{ and } T \in \phi(T)\}.$$

This is precisely the mapping ϕ defined by Acemoglu et al. (2008). They prove that every subgame perfect equilibrium of their extensive form game leads to $\phi(N)$ as the ultimate ruling coalition.

3.9. **Internal Blocking in the Presence of Externalities.** The characteristic function has proved to be a very useful construction in studying coalitional behavior. It was derived by von Neumann and Morgenstern (1944) from a more general specification of a game by taking the feasible payoffs for a coalition to be those it can achieve by assuming (from its point of view) the worst possible strategy choices of the complementary coalition. While von Neumann and Morgenstern (1944) adopted this conversion to a characteristic function mainly to study zero-sum games, it was subsequently applied to more general (normal form) games by Aumann and Peleg (1960) and Aumann (1961).[25] For example, the α-characteristic function defines $V(S)$ for a coalition S as the set of payoffs S can achieve through some joint strategy *regardless of the actions of players outside* S.[26]

In some settings, such as exchange economies without externalities, or zero-sum games, this conversion involves no loss of generality. However, in the presence of externalities, the standard construction is *ad hoc*, if not unreasonable. For example, in the Cournot oligopoly, it is hard to see why a cartel should fear that the complementary coalition will flood the market and drive profits to zero, as is implicit in the extreme pessimism embodied in the α-core. It may even be argued that coalition formation should be studied directly through a normal form game. The essence of the problem, however, can usually be captured through a partition function game which makes explicit the manner in which the feasible set of payoffs for a coalition depend on other coalitions. To be sure, this does not completely eliminate the complexities stemming from externalities. A blocking coalition must now predict how players outside the coalition will organize themselves into coalitions. In what follows we will refer to this as the prediction problem. Of course, one could again

[25] See Chapter 2.2 of Ray (2007) for a historical background.

[26] The core of the α-characteristic function is referred to as the α-core. In the β-characteristic function $u \in V(S)$ if for every strategy of players outside S there is some joint strategy in S that yields at least u.

cut though these complexities by making an assumption about how outsiders will organize themselves in response to a move by a blocking coalition. For example, one could assume that all outsiders will immediately break-up into singletons, or that all players left behind by a deviating coalition will continue to stay together. The former is related to the notion of γ-stability in Hart and Kurz (1983) and the latter to δ-stability.[27] The question is how to replace such assumptions with predictions about the equilibrium behavior of outsiders. Cooperative game theory has traditionally eschewed such considerations, but it is hard to see how coalitional behavior in the presence of externalities can be studied without making the response of outsiders endogenous to the theory. As we shall discuss, considerable progress has been made in resolving the prediction problem.[28]

Restricting attention to internal blocking aids significantly in terms of tractability. To focus on the main problem we shall suppose, in addition, that the allocation of the surplus within a coalition embedded in a coalition structure is not an issue. This is the case if there is a fixed rule for choosing a point in $\bar{V}(S, \pi)$, or if $\bar{V}(S, \pi)$ is a singleton for every (S, π), where $S \in \pi$. This is an extension of hedonic characteristic functions to hedonic partition functions.[29] Now when a coalition makes a "move", all it needs to do is predict the eventual coalition structure that will be precipitated by its own move. The farsighted core can be seen as a natural first step in formalizing a suitable solution concept.

At this point we should be more explicit about the effectively correspondence to be applied to this model. The assumption of a hedonic partition function makes it unnecessary to describe how coalitional worth will be shared within each coalition in a given coalition structure (we'll return to this issue for the general case in the next Section). In fact, a state can simply be defined as a coalition structure. Given the assumption of internal blocking it seems reasonable that if $S \in \pi$ and a subsocoalition T of S splits from S, in the first step it induces an immediate refinement, $\pi_{|T}$, i.e., all the remaining players in S stay together and all the coalitions in π remain unchanged. Of course, there is no presumption that $\pi_{|T}$ will remain unchanged.

It is also possible to temper the extreme optimism in the notion of a farsighted objection by allowing for the other coalitions involved in a sequence of moves to move in a different order or in fact to move simultaneously. Specifying this in a precise way leads to the concept of equilibrium binding agreements (EBA) of Ray

[27] See also Carraro and Siniscalco (1993), Dutta and Suzumura (1993), Chander (2007) and Chander and Tulkens (1997).

[28] It bears mentioning that Aumann and Myerson (1988) tackled the prediction problem head on. We do not discuss this paper here only because its axiomatic emphasis does not fit either the blocking or the bargaining approach which we have confined this chapter to. Maskin (2003) is another important contribution to this issue which doesn't fall within the purview of the present chapter. See also de Clippel and Serrano (2008).

[29] See Banerjee et al. (2001), Barberà and Gerber (2003, 2007) and Bogomolnaia and Jackson (2002).

and Vohra (1997), which can be seen as a variant of the farsighted core. Applied to the special case of hedonic partition functions, we can now formally define EBA.

When a process of coalition formation refines a partition there are some coalitions that are active movers, or perpetrators, in splitting from a larger coalition while others can be thought of as residuals, players left behind. If a coalition breaks into k new coalitions, $k-1$ of them must be perpetrators.

Coalition structures that correspond to *equilibrium binding agreements* (EBA) are defined recursively. The finest coalition structure, made up of singletons, and denoted π_0, is an EBA. If π is a coalition structure made up of singletons and one coalition, S, with two players, it is said to be blocked by π_0 if one of the players in S prefers π_0 to π. We can now proceed recursively. Suppose EBA and the associated notion of blocking has been defined for all refinements of π. Then, is said to be blocked by π' if π' is an EBA and their exists a collection of perpetrators in π' such that one of then, a leading perpetrator, prefers π' to π. Moreover, any re-merging of the other perpetrators with their respective residuals is blocked by π', with one of these perpetrators as the leading perpetrator. Note that "blocking" is well defined in the previous sentence because any re-merging of the other perpetrators results in a coalition structure which is a refinement of π. A coalition structure is an EBA if it is not blocked. Given the emphasis on full and unrestrained negotiations, EBA for the game are defined as the coarsest partition(s) that are EBA.

The notion of equilibrium binding agreements turns out to be particularly simple in the special case of symmetric TU games with positive externalities. In our framework (of partition functions rather than normal form games) these are partition functions in which utility is transferable within each coalition. For coalition S in coalition structure π, the aggregate utility is denoted $v(S,\pi)$, so that $V(S,\pi)$ is the collection of all payoff profiles that sum to no more than $v(S,\pi)$. (We will sometimes use (N,v) to denote a TU partition function). In a symmetric game the worth of a coalition depends only on the number of players in the coalition and the numerical coalition structure (number of players in each of the coalitions). Suppose $S \in \pi = \{S_1, \dots S_k\}$. Now $v(S,\pi)$, can simply be denoted $v(s,q)$ where s is the cardinality of S and $q = (s_1, \dots s_k)$, where s_i is the cardinality of S_i. A game is said to have positive externalities if a coalition's worth is higher when the other coalitions are merged. It is worth noting that the symmetric Cournot oligopoly is one example that satisfies all of these assumptions.

As shown in Ray and Vohra (1997), in a symmetric TU game with positive externalities, a state is an equilibrium binding agreement if no coalition can obtain higher aggregate utility in a binding agreement of *some* refinement of the original coalition structure. In other words, EBA are simply all the states in the farsighted core if $\mathcal{E}$ allows a coalition to move to *any* refinement. Throughout this Section we assume that the effectivity correspondence has this form. Of course, this immediately implies that the one-step deviation property holds.

In this setting, additional simplicity comes from the fact that in identifying equilibrium coalition structures there is no loss of generality in assuming that coalitional worth is divided equally among all members of a coalition; see Proposition 6.3 in Ray and Vohra (1997). In other words, we can assume that states are restricted to satisfy the property that for any $S \in \pi(x)$, for all $i \in S$, $u_i(x) = a(S, \pi) = \frac{v(S,\pi(x))}{|S|}$. Thus, we can take the number of states to be finite.

Let $a(s, q)$ denote the average worth of a coalition of size s in a numerical structure q, where $q = (q_1, \ldots q_k)$ describes the sizes of the various coalitions in the partition. We will assume that the distribution of average worths satisfies the genericity assumption in the sense that $a(s, q) \neq a(s', q')$ if $s' \neq s$ or $q' \neq q$.

PROPOSITION 7. *Suppose (N, V) is a symmetric, TU, partition function game with positive externalities and the genericity assumption holds. Then, for δ close enough to 1, all absorbing states of every EPCF of $(N, V, \mathcal{E}, \rho)$ coincide with EBA.*

Proof. Suppose x is not an absorbing state of an EPCF. This means that for some history at least one coalition has a profitable move. Let T be the last coalition according to the protocol which would choose to move from x to y. Of course, if T were to choose not to move, it would be assured of $u(x)_T$ in perpetuity. Either y is an absorbing state or, with positive probability, there is a further move from y. Consider the latter case.

All possible paths leading from y must reach an absorbing state in a finite number of steps. Let m be the maximum number of steps along any such path before an absorbing state is reached. Let u^i denote the expected payoff at each step, $i = 1, \ldots m$, with $u^1 = u(y)$. Let Z be the set of states in the support of u^m and $p(z)$ be the probability of z being the state at stage m, i.e., $u^m = \sum_{z \in Z} p(z)u(z)$.

The fact that T has a strictly profitable move means that

$$u^1_T + \delta u^2_T + \delta^2 u^3_T + \ldots + \frac{\delta^{m-1}}{(1-\delta)} u^m_T \gg \frac{1}{(1-\delta)} u(x)_T. \tag{4}$$

Let $a^* = \max_{z \in Z, S \in \pi(z), S \subset T} a(S, \pi(z))$ be the maximum average worth across all subcoalitions of T in any of the coalition structures corresponding to states in Z. Let z^* be a state in which a^* is attained and S^* the corresponding subcoalition of T. Let $\bar{u}$ be the maximum utility that any player gets in any state. From (4) we obtain the following inequality for coalition S^*.

$$(1 + \delta + \ldots + \delta^{m-2})\bar{u} + \frac{\delta^{m-1}}{1-\delta}(a^* - u_i(x)) > 0 \text{ for all } i \in S^*. \tag{5}$$

Given that $\pi(z^*)$ is a refinement of $\pi(x)$ it follows that $u_i(x) = u_j(x)$ for all $i, j \in S^*$. We now claim that

$$a^* > u_i(x) \text{ for all } i \in S^*. \tag{6}$$

Suppose not. Then $a^* < u_i(x)$ for all $i \in S^*$ because, by the genericity assumption, $a^* \neq u_i(x)$. Since the number of states is finite, there exists $\epsilon > 0$ which denotes the minimum (absolute) difference, between $a(s, q)$ and $a(s', q')$ for $q \neq q'$. This means that if $a^* < u_i(x)$, then $a^* - u_i(x) \leq -\epsilon$. Substituting this in (5) we have:

$$(1 + \delta + \ldots + \delta^{m-2})\bar{u} - (\frac{\delta^{m-1}}{1 - \delta})\epsilon > 0,$$

which is impossible if δ is close enough to 1. This establishes (6).

To summarize, we have shown that if x is a transient state, one of the following must be true:

(i) there is a move by T to y which is an absorbing state and $u(y)_T \gg u(x)_T$,
(ii) there is a coalition $S^* \subset T$ and an absorbing state z^* such that $S^* \in \pi(z^*)$ and $u(z^*)_{S^*} \gg u(x)_{S^*}$.

Since a coalition is effective in moving to any refinement of the current coalition structure, it follows that in case (ii) S^* can move directly from x to z^*. Thus, in any event, there is a profitable move to an absorbing state, and the process must be immediately absorbing. The proof now follows from Lemma 1 and the fact that the farsighted core is the set of EBA. ■

Note that in the dynamic model it is possible for a strictly profitable move to be one in which the absorbing state results in the same utility profile as the current state, with all the gains being reaped in the intervening, transitory periods. Such a move in the dynamic model cannot possibly serve as an objection in the static model. To tie the dynamic model to the static one, therefore, we have to impose additional assumptions. In Proposition 7 this is achieved through the genericity assumption.

3.10. **Beyond Internal Blocking.** To extend the theory beyond internal blocking a recursive definition of farsighted stability will not suffice. Despite the sometimes obscure nature of abstract stable sets, they do offer a promising approach at this level of generality. For the farsighted stable set to be meaningful, though, it's important to specify the effectivity correspondence with some care. In this respect, as we observed in the previous Section, things are relatively straightforward with hedonic partition functions. We can follow Diamantoudi and Xue (2007) and Ray and Konishi (2003) in assuming that the *immediate* consequence of a coalitional move is that unaffected coalitions remain intact and any residuals left behind by the perpetrator remain together.[30]

In more general partition function games, when the payoff division within a coalition is endogenous, a blocking coalition induces a state which consists of both a coalition structure and a feasible payoff for each coalition: $x \rightarrow_S y$ means that

[30] This may seem appear to be similar to the δ-stability approach but the important difference is that here this is only meant to describe the *immediate* consequence of a coaltional move, not the final outcome.

S is effective in imposing the partition $\pi(y)$ and payoffs $u(y)_T$ in $\bar{V}(T)$ for each $T \in \pi(y)$. The restriction on $\pi(y)$ as described in the previous paragraph seems reasonable enough, but it is not immediately obvious what restrictions (besides feasibility), if any, ought to be imposed on $u(y)_T$.

Indeed, this question is relevant even in the simpler setting of characteristic functions. With the standard notion of (myopic) blocking in a characteristic function game, what transpires outside the coalition does not affect the blocking coalition, and therefore the definition of dominance does not depend on what is assumed about the payoff distribution for outsiders. In fact, the stable set takes the set of states to be imputations (individually rational payoffs in $\bar{V}(N)$) and S can move from x to y provided as $u(y)_S \in \bar{V}(S)$. In other words, $\mathcal{E}(x, y) = \{T \mid u(y)_T \in \bar{V}(T)\}$. In effect, S is assumed to have the power to choose *any* feasible payoff profile for outsiders. It so happens that in a characteristic function with myopic blocking this doesn't make a difference to the stable set.

But with farsighted blocking this is not so. A coalition may be able to engineer a sequence of moves that constitute farsighted blocking only by arranging the payoffs to outsiders in some particular way. This is clearly unreasonable. To see a rather dramatic example of this consider a four player TU game with the following characteristic function. $v(i) = 0$ for all i, $v(12) = v(13) = v(23) = 2$, $v(123) = 6$ and for all S, $v(S \cup 4) = v(S)$. Although Player 4 is a dummy player, if we assume the effectivity correspondence to be unrestricted, as in the previous paragraph, there is a singleton farsighted stable set in which the dummy player receives a positive payoff, for example, $\{y\}$, where $y = (1, 1, 0, 4)$. Consider $x = (2, 2, 2, 0)$, a core allocation. There is a farsighted objection to x leading to y. It begins with a move by player 4 to $y^1 = (6, 0, 0, 0)$, followed by a move by player 2 to $y^2 = (0, 0, 0, 6)$ and, finally, by coalition 12 to y. The logic depends crucially on player 4 assigning 0's to two of the other three players, and then another one assigning 0's to all players other than 4.

There have been a number of recent papers showing that the farsighted stable set generally exists in characteristic function games. This is remarkable because the existence of the stable set is not guaranteed in characteristic function games, as shown by Lucas (1968). Diamantoudi and Xue (2005) showed that in Lucas's example the farsighted stable set does exist. Béal et al. (2008) proved existence for TU games and Bhattacharya and Brosi (2011) for NTU games.[31] All of these results, however, assume that a blocking coalition has complete power to choose the payoff allocation for outsiders. It is not known if these positive results can be extended to a notion of farsighted stable sets with a more reasonable specification of the effectivity correspondence.

For hedonic partition functions, where payoff division is not an issue, Diamantoudi and Xue (2007) show that the notion of blocking used in defining EBA can be

[31] Mauleon et al. (2011) prove existence in two-sided matching models which don't have this complication as they are a special case of hedonic characteristic functions.

reformulated in terms of farsighted blocking, where the extreme optimism implicit in farsighted blocking is modified to make it robust to the precise manner in which perpetrators move. And this makes it possible, in the context of internal blocking, to interpret EBA as a stable set with an appropriate notion of dominance. They then go on to argue that a suitable extension of EBA to the general case, extended equilibrium binding agreements (EEBA), is the stable set with farsighted blocking in its standard (optimistic) form. Although neither the existence of EEBA nor its efficiency when it does exist can be guaranteed in general, extending the notion of blocking beyond internal blocking can help sustain efficiency. Their positive result on efficiency includes the important example of a Cournot oligopoly; see Section 5.

As Greenberg (1990) has shown, it is possible to formulate various stable standards of behavior along the lines of the stable set by assuming conservative rather than optimistic behavior on the part of a blocking coalition. Many of the solution concepts based on farsightedness can be comprehensively viewed through Greenberg's theory of social situations. As a complement to the current section we refer the reader to Mariotti and Xue (2003) for an excellent review of this approach.

The *Largest Consistent Set* of Chwe (1994) is a prime example of the blocking approach with farsightedness assuming pessimism on the part of a blocking coalition. Chwe considers a more general game than an extended partition function in the sense that the set of states, X, is an abstract set, not necessarily based on a partition function. A set $Z \subset X$ is said to be *consistent* if $x \in Z$ if and only if for all y, S such that $x \rightarrow_S y$, there exists $z \in Z$ such that $z = y$ or z is a farsighted objection to y and there exists some $i \in S$ such that $u(z)_i \leq u_i(x)$. Chwe proves that there is a unique consistent set, the *largest consistent set* (LCS), which contains all consistent sets.[32]

At this stage one may be tempted to leave well enough alone and accept the idea, as in Knightian uncertainty, that stable outcomes can be modeled either with optimistic beliefs or conservative beliefs or perhaps some combination of the two. However, this is serious drawback of the blocking approach. It is no less ad hoc than making some exogenous assumption about remaining players will organize themselves into a coalition structure in response to a coalitional deviation. The problem is easily shown through the following Example.[33]

EXAMPLE **8.** *Modify Example 6 so that player 2 can move to either state c or d.*

The unique farsighted stable set is $\{c, d\}$. State a is not in the farsighted stable set because player 1 expects player 2 to replace b with c. But now this is too optimistic a prediction. From b player 2 has the choice to move to either c or d, and

[32] Chwe also shows how LCS can be related to a conservative stable standard of behavior in Greenberg's theory of social situations.

[33] While the Examples in this Section do not represent hedonic partition functions, they can all be transformed (with the addition of players) to meet this property and still retain their message.

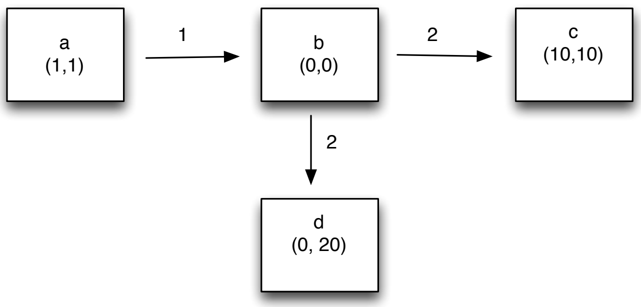

FIGURE 4. ILLUSTRATION OF EXAMPLE 8.

obliviously the rational decision for player 2 is to move to d. Thus, the (farsighted) reasoning for discarding a now appears to be flawed. In this example, Chwe's LCS resolves this problem. For a state to be in LCS it is enough that every initial move have *some* continuation that results in a stable outcome which would deter the initial "objection". By this criterion the LCS is $\{a, c, d\}$; unlike the farsighted stable set, LCS considers a to be stable. In this example, the path from b should lead only to d since that is the only rational move by coalition 2. It so happens that this corresponds to coalition 1 being conservative in its evaluation of the move from a to b. In general, however, a conservative forecast by the initiating coalition need not be "rational", as the next example shows.

EXAMPLE **9.** *Modify Example 8 by interchanging player 2's payoffs in c and d.*

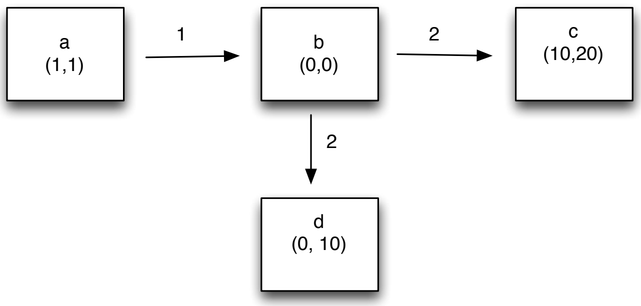

FIGURE 5. ILLUSTRATION OF EXAMPLE 9.

The farsighted stable set and LCS remain unchanged (the former is $\{c, d\}$ and the latter is $\{a, c, d\}$). But now, it should be clear that an "optimistic" view by player 1 is indeed the correct one; player 2 in his own interest will replace b by c, not d. Thus, in this example, it is the LCS which comes to the wrong conclusion.

It is easy to see that in both these examples, the notion of an EPCF provides a simple resolution. But how far can one go with the traditional blocking approach? Xue (1998) argues, persuasively, that a resolution to this issue requires us to define stability not for allocations, such as a, b, c and d in the above examples, but for *paths* such as (a, b, c) or (a, b, d). He then considers a stable standard of behavior for a situation with perfect foresight as a collection of paths satisfying internal and external stability.

For a stable standard of behavior σ, let $\sigma(a)$ denote all stable paths originating from a. If $\alpha \in \sigma(a)$, internal stability requires that there not exist a node $b \in \alpha$, a coalition S which is effective form moving from b to c such that S "prefers" $\sigma(c)$ to α. The same idea is used in defining external stability. The term "prefers" leads to two versions of stability: one based on optimism and other on conservatism.

It is easy to show that this notion of stability yields the correct answer for both Examples 8 and 9. However, in the next Example the problem reappears.

EXAMPLE **10.** *Modify Example 9 by addiing one more player and one more state.*

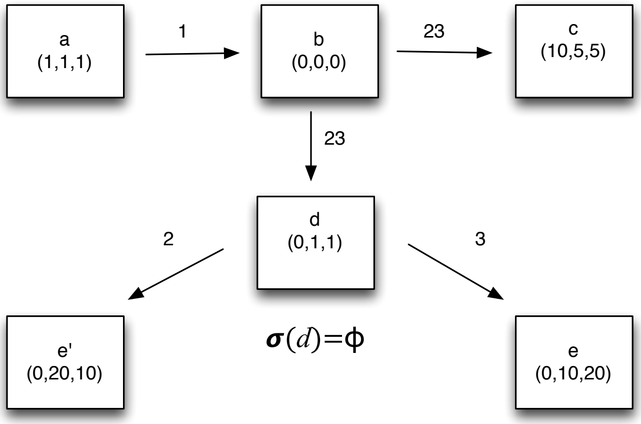

FIGURE 6. ILLUSTRATION OF EXAMPLE 10.

First consider the stable paths originating from d. Player 2 would like to move to e' while 3 would prefer to move to e. Neither can satisfy internal stability and therefore $\sigma(d) = \emptyset$. Note that while Xue (1998) shows that an acyclicity condition is sufficient for σ to be nonempty valued, this condition is violated at d. (This is similar to figure 4 in Xue (1998)). The problem arises because neither of the two objecting coalitions has a priority to make a move. Because $\sigma(d) = \emptyset$, the stability of (a, b, c) is vacuously satisfied; it is not possible to test if it can be deterred by a move from b to d and some $\beta \in \sigma(d)$. Indeed, (a, b, c) is a stable path. But this doesn't seem reasonable because for players 2 and 3 both e and e' dominate c, so coalition 23 should move from b to d, not c.

Unfortunately, this kind of situation is rather common and it seriously limits the applicability of this solution concept. Note that if there was a protocol, even a probabilistic one, to decide who has the first move, the problem wouldn't arise.[34] For example, if upon arriving at state d the protocol selects each of two permissible singleton coalitions with equal probability, then the expected payoff to players 2 and 3 (following state d) is 15 each. From state b, therefore, the move by coalition 23 to d is efficient (in the terminology of our dynamic model) but to c is not. And this implies that a is an absorbing state of every EPCF.

Taken together, these examples demonstrate a serious drawback of the blocking approach in dealing with farsightedness, at least in environment with externalities. Clearly, there are limits to how effectively one can capture farsightedness in a static concept of stability. The traditional approach in cooperative game theory has emphasized the virtues of abstracting away from the details of the negotiation process to highlight the essential features of cooperative behavior. In many situations that has indeed been a very fruitful approach, but in the present context it seems too confining not to introduce some details (as well as explicit dynamics). As the previous examples demonstrate, simply adding the notion of a protocol, and postulating rational behavior on the part of coalitions, can provide a way out of the conundrum that the standard approach produces. The dynamic process of coalition formation described in Section 2 is a direct way of studying farsighted coalitional behavior, especially in the presence of externalities. If this framework seems more complex than, say, the characteristic function form, it nevertheless seems to be necessary for the questions at hand. It could even be argued while there is additional structure on the model, the equilibrium concept is much more straightforward that ones we have discussed above, for example, EBA, LCS or farsighted stable sets. The fact that in some of the simpler cases, such as characteristic function games and symmetric TU games with positive externalities, we get the standard conclusions makes it possible to see this approach as a conceptual generalization rather than an alternative.

4. The Bargaining Approach: Coalitions in Noncooperative Games

In this Section, we study an approach to coalition formation based on noncooperative bargaining. Attention shifts from active *coalitions* to active *individuals*, and the notion of blocking is replaced by a direct emphasis on proposals and responses. In short, all negotiations are expressed formally as a *bargaining game*, for which we draw inspiration from Ståhl (1977), Rubinstein (1982), Chatterjee et al. (1993), Okada (1996) and several others.[35]

[34] Recall that one major difference between our dynamic model and that of Konishi and Ray (2003) is that the latter operates without a protocol. This can result in inefficiencies arising from a "coordination failure". Simply having a protocol of the kind we outlined in Section 3.6 can help avoid these kinds of inefficiencies.

[35] For related literature on bargaining, see Selten (1981), Binmore (1985), Baron and Ferejohn (1989), Gul (1989), Winter (1993), Perry and Reny (1994), Krishna and Serrano (1995), Moldovanu

Throughout, we regard the partition function as a primitive, with the idea that underlying this function is a game in strategic form. As already discussed, a partition function has the virtue of incorporating a number of different situations. Two-person or multi-person bargaining is, of course, included quite trivially. So is coalitional bargaining over a characteristic function, where different coalitions have access to different surpluses which may be divided. Finally, partition functions can also accommodate externalities across coalitions in the determination of coalitional surplus.

We define on the partition function a noncooperative bargaining game. A history-dependent protocol defines an active proposer at each date, perhaps stochastically. In the language of our general framework, active coalitions are now restricted to be singletons. A proposer is free to propose a particular (feasible) payoff allocation to any subset of a partner set specified by the protocol. For each proposal, the protocol also specifies the order in which the responders respond, either to accept or to reject the proposal.[36] Observe that a partner set was of no importance in the blocking approach (at least as it currently appears in the literature), but in the present context it is crucial, as we shall see from the special cases described below. Coalitions form through the course of this bargaining process as proposals are made and either accepted or rejected by responders.

The various ingredients of a bargaining model can be combined in different ways to generate distinct branches of the literature. In the next Section we show how a combination of the effectivity correspondence and protocol can be used to cover this diverse set of models.

4.1. **Essential Ingredients of a Coalitional Bargaining Model.**

4.1.1. *The Protocol.*

When a proposer is chosen by the protocol, she makes a proposal to a subset of eligible partners a division of their aggregate worth. If all respondents unanimously accept, the proposed coalition forms, and the process shifts to the set of eligible players remaining in the game. The *rejection* of a proposal creates a bargaining friction: payoffs are delayed by the passage of some time, which is discounted by everybody using the discount factor δ.

The probability with which different individuals are chosen to be proposers will generally depend on the history of events up to that date. Perhaps the simplest protocol is one in which an agent is chosen with uniform probability to be a proposer. This is known as the *uniform protocol*; see, e.g., Baron and Ferejohn (1989). Another simple protocol recognizes different individuals in a given order and anoints the first rejector of the previous proposal (if any) as the next proposer. This is the *rejector-proposes protocol*; see, e.g., Rubinstein (1982), Chatterjee et al. (1993).

and Winter (1995), Hart and Mas-Colell (1996), Bloch and Gomes (2006) and Compte and Jehiel (2010).

[36] This avoids the possibility of coordination failure, and is consistent with the rest of the literature.

One way to combine these two protocols while retaining each of them as special cases is to suppose that an active individual (from the set of free individuals) is chosen randomly at any stage if the previous proposal has been accepted. Otherwise, if the previous proposal has been rejected by some individual, that person gets to be the next proposer with probability μ. With probability $1 - \mu$, *another* eligible agent is equiprobably chosen to be the new proposer. Observe that if $\mu \simeq 1$, the rejector is likely to be the next proposer, while if $\mu \simeq 0$, the rejector is excluded from making a proposal in the next round. This class of protocols therefore accommodates a wide variation in rejector power. It encompasses all the ways a proposer is chosen that we know of in the literature.

A central feature of our protocol concerns the recognition of those players who are *eligible*, either to make new proposals or to entertain them from others. Which agent is eligible will depend, of course, on the situation to be modeled. In the case of irreversibly binding agreements, a player once included in some previous coalitional agreement is never again recognized by the protocol. When agreements are reversible, so that renegotiation is permitted, the partner set could include a player who belongs to a previously formed coalition, but only if all other members of that coalition are also included as responders. (Their agreement is needed to "free" the player to sign a new deal.) Such restrictions on the protocol can often be substituted with corresponding restrictions on the effectivity correspondence instead; recall footnote 21.

In short, our description of protocols can be construed as an attempt to model the essential elements of the situation at hand: how easy it is for a rejector to seize the initiative and make a fresh proposal, or how constrained previous signatories are in participating in further bargaining. In this survey, we do not concern ourselves with other approaches to "design" a proposal to deliberately "implement" some known solution concept. For instance, Gul (1989) shows how the Shapley value can be implemented as a stationary perfect equilibrium of a game with pairwise meetings. Hart and Mas-Colell (1996) do so in a more general context, through a bargaining game in which proposers are *required* to make proposals to the complete set of available players. Moreover, the rejection of a proposal leads, with some positive probability, to the proposer being entirely eliminated from future rounds of bargaining. These restrictions are clearly in the spirit of implementing a particular solution concept, as the description makes it clear that there is no particular attempt to identify the protocol with any observed bargaining situation.[37] Indeed, incorporating strategic coalition formation in these models remains an important direction for future work. For a review of this literature we refer the reader to Winter (2002).

[37] It should be added, however, that the above depiction makes more descriptive sense in a two-person bargaining context; see, Binmore, Rubinstein and Wolinsky (1986) and Sutton (1986). There these restrictions reduce, more realistically, to the possibility that the entire bargaining process might break down following any round of negotiation.

4.1.2. *Possibilities for Changing or Renegotiating Agreements.* Whether or not a coalitional agreement once made is subject to future revision is a fundamental issue.

Bargaining with Binding or Irreversible Agreements. In much of the literature, all agreements to form a coalition are "fully" binding, in the sense that they are irreversible. A coalition once formed cannot disintegrate or be subsequently absorbed into a larger group.[38] The protocol responds to histories by choosing only players (as proposers or respondents) who were not part of any previously formed coalition. It follows, therefore, that once all players are included in some formed coalition, the process of coalition formation must come to an end, though payoffs continue to be received as per the various agreements. In particular, the protocol ceases to choose new proposers. For examples of this sort of model, see Chatterjee et al. (1993) and Okada (1996) for characteristic function games, and Bloch (1996) and Ray and Vohra (1999) for partition function games.

Bargaining with Reversible Agreements. Situations in which agreements are only in force for a limited period of time can be modeled by a suitably specifying the protocol. For example, if agreements only last for one period, active proposers are chosen from the *entire* population of players, and partner sets are never restricted by the history of past coalition formation. For examples, see Stole and Zwiebel (1996), Gomes and Jehiel (2005), and Konishi and Ray (2003).

The possibility of renegotiating existing agreements is another case in which agreements may be reversible. In this case, an existing agreement may be changed, but only with the blessings of existing signatories. Such signatories must include all individuals who are party to *any* existing agreement that may need to be modified as a consequence of the new proposal. This is captured by an appropriate restriction on the effectivity correspondence or on the protocol. For examples, see Seidmann and Winter (1998), Okada (2000), Gomes and Jehiel (2005), Gomes (2005) , and Hyndman and Ray (2007).

4.1.3. *Payoffs in Real-Time or Not.* A substantial part of the literature studies models in which payoffs are only experienced after all coalitions have formed. This includes Rubinstein (1982), Bloch (1996), Chatterjee et al. (1993), Ray and Vohra (1999) for binding agreements and Seidmann and Winter (1998) and Okada (2000) for renegotiable agreements. A more recent literature, e.g., Konishi and Ray (2003), Hyndman and Ray (2007), Gomes (2005), Gomes and Jehiel (2005) and Xue and Zhang (2011), considers situations in which payoffs are realized continually, as coalitions can form and continue to renegotiate or discard previous agreements.

[38] In the special case of n-person bargaining, in which only the grand coalition has a surplus to divide, this is hardly an assumption as there is no collective incentive to alter any agreed-upon division of the surplus.

It is of course only natural for reversible agreements to be cast in a real-time framework. It should be clear that our general framework is well suited to cover such situations and, as we will see below, it can also encompass models in which payoffs are realized at the end of the coalition formation process.

4.1.4. *Majority Versus Unanimity.* There is also a distinction to be drawn using the rules of the game that determine when a proposal is to be passed. Two major candidates are unanimity, as in the Ståhl-Rubinstein model and its descendants, and majority vote, as in the Baron-Ferejohn model (and political economy models more generally). Once again, it is easy enough to accommodate a variety of such rules as part of the general framework. Without loss of generality, we can simply adopt the unanimity approach, incorporating all variants in the description of the partition function and the form of the effectivity function.

As an example, consider three-person bargaining with majority. The "true" function that describes this example sets the worth of the grand coalition equal to the surplus at stake (say 1 unit), while setting the worth of all subcoalitions to zero. Yet it is possible to use instead the characteristic function

$$v(S) = 1 \text{ if and only if } |S| > \frac{n}{2},$$

and use the unanimity protocol. What is altered is essentially a matter of interpretation: a proposal is never actually made to a subcoalition S, but it's *as if* it is: the proposal is in fact made to the grand coalition, with the implicit strategic presumption that the "targeted" majority subgroup S is effective for the change and will approve it.

In short, bargaining models that require majority approval can be easily embedded in coalitional bargaining models in which subcoalitions have power. In this sense there is little loss of generality in studying unanimity games, *provided* we are general enough to accommodate subcoalitional worths.

4.2. **Bargaining on Partition Functions.**

4.2.1. *Equilibrium in a Real-Time Bargaining Model.* We begin with a baseline model for bargaining in real time. As in the blocking approach, we consider a partition function (N, V) which assigns to each coalition S in a partition π a set of payoff allocations $V(S, \pi)$.

A state is denoted $x = (\pi, u, C)$ where $u_S \in V(S, \pi)$ for every $S \in \pi$ and C is the collection of "committed" players ($N \setminus C$ being the set of "uncommitted" players). We use the convention that all uncommitted players consist of singletons, while committed players belong to formed or committed coalitions (including possibly singletons). In other words, $i \notin C$ implies that $\{i\} \in \pi$ and every $j \in S \in \pi$ such that $|S| \geq 2$ implies that $j \in C$. Uncommitted players can be proposers or potential partners in all variations of the model; the others may be included to different extents if agreements are not completely irreversible.

At each stage of the bargaining process, we keep track of past proposers, proposals, rejectors (if any), and all committed as well as uncommitted coalitions.

A *history* at some stage is a list of such objects up to, but not including, the events that will occur at that stage. Such stages may be of various kinds: a proposer is about to be chosen, or a proposal about to be made, or a responder about to respond, or — such matters concluded — a state about to be implemented. Obvious nomenclature may be employed to distinguish between the histories leading up to different stages: “proposer histories,” “responder histories,” and “implementation histories”.

At proposer or responder histories players have to take actions. A full listing of a particular player’s actions — proposals and responses — for all such histories is a *strategy* for that player. To describe strategies more formally, consider an individual k. For a proposer history h at which k is meant to propose, she must choose a payoff vector and a coalition S that can implement the payoff in question. In standard bargaining theory such a vector would be given by a division of aggregate surplus among the individuals (see, e.g., Rubinstein (1982) and Baron and Ferejohn (1989)). In coalitional bargaining theory, it would be a division of *coalitional* surplus among the members of that coalition (see, e.g., Chatterjee et al. (1993), Seidmann and Winter (1998) and Okada (1996)). In bargaining theory with externalities, the payoff vector must come from a “conditional proposal”: if coalition structure π forms, we divide in this way, and if π' forms, we divide in that way, and so on (see, e.g., Ray and Vohra (1999)). In our real-time model the payoffs at each date are feasible given the going coalition structure. For formed coalitions they must also reflect the agreed upon payoff allocation corresponding to this coalition structure. This can be seen as a restriction on the effectivity correspondence.

An active agent proposes a new state to one or more of her available partners. She could employ a behavior strategy, which would be a probability distribution over (y, S), where y represents the new state and S a coalition containing i and a subset of available partners jointly capable of implementing that state. Denote by $\mu_k(h)$ the probability distribution that she uses at proposer history h.

Likewise, at a responder history h at which k is meant to respond, denote by $\lambda_k(h)$ the probability that k will accept the going proposal under that history. The full collection $\sigma = \{\mu_k, \lambda_k\}$ over all players k is a *strategy profile*.

A strategy profile σ induces *value functions* for each player. These are defined at all histories of the game, but the only ones that we will need to track are those just prior to the implementation of a fresh state (or the unaltered continuation of a previous state). Call these *implementation histories*. On the space of such histories, every strategy profile σ (in conjunction with the given protocol) defines a stochastic process P^{σ} as follows. Begin with an implementation history. Then a state is indeed "implemented". Subsequently, a new proposer is determined. The proposer proposes a state. The state is then accepted or rejected. (The outcome in each of these last three events may be stochastic.) At this point a new implementation history h' is determined. The entire process is summarized by the transition P^σ on implementation histories.

For each person i and given an implementation history h, the *value* for i at that date is given by

$$V_i^\sigma(h) = u_i(x) + \delta \int V_i^\sigma(h') P^\sigma(h, dh') \tag{7}$$

where x is the state implemented at h. Given any transition P^σ, a standard contraction mapping argument ensures that V_i^σ is uniquely defined for every i.

Say that a strategy profile σ is an *equilibrium* if two conditions are met for each player i:

(a) At every proposer history h for i, $\mu_i(h)$ has support within the set of proposals that maximize the expected value $V_i^\sigma(h')$ of i, where h' is the subsequent implementation history induced by i's actions and the given responder strategies.

(b) At every responder history for i, $\lambda_i(h)$ equals 1 if $V_i^\sigma(h') > V_i^\sigma(h'')$, equals 0 if the opposite inequality holds, and lies in $[0, 1]$ if equality holds, where h' is the implementation history induced by acceptance, and h'' the implementation history induced by rejection.

4.2.2. *Two Elementary Restrictions.* To ease the exposition, we impose two simple restrictions on equilibrium. First, equilibria might involve delay: a proposal could be rejected on the equilibrium path. To be sure, it is natural for delays to arise in bargaining with incomplete information.[39] But in complete information models such delays are more artificial, and stem from two possible sources. The first is a typical folk-theorem-like reason in which history dependent strategies are bootstrapped to generate inefficient outcomes, including equilibria with delay. More subtly, equilibria may involve delay because an unacceptable proposal is made to deliberately affect the identity of the rejector and subsequently the choice of the next proposer. For examples, see Chatterjee et al. (1993) and Ray and Vohra (1999). This will *only* happen for protocols that are sensitive to the identity of previous rejectors. In several models such as Rubinstein (1982) and random proposer models as in Baron and Ferejohn (1989) and Okada (1996, 2006, 2011) this phenomenon is impossible. Moreover, even in situations in which the protocol is sensitive to past rejections, the literature provides reasonable conditions under which all equilibria are no-delay equilibria; see in particular Chatterjee et al. (1993) and Ray

[39] For a recent example that attempts to get around the Coase conjecture in models of one-sided incomplete information; see Abreu, Pearce and Stacchetti (2012).

and Vohra (1999). Accordingly, in what follows we shall restrict ourselves to *no-delay equilibria*, in which at each proposer history, a proposer makes an acceptable proposal.

We will also restrict ourselves to equilibria which satisfy a minor additional restriction, which we call "compliance". Say that an individual is *compliant* if, whenever she responds to a proposal, she takes an action that makes the proposer better off, provided that this does not harm her in any way. The terms "better off" and "harm" are defined with respect to equilibrium value functions, in just the same way as equilibrium payoffs are. This refinement is of a lexicographic nature: it only applies when there is no danger to the payoff of the individual concerned. Alternatively, one could just as easily think of compliance as an equilibrium refinement rather than as a lexicographic restriction on individual preferences. Thus an equilibrium strategy profile is *compliant* if for no individual and no history is there a deviation by a responder which increases the payoffs of a proposer not decreasing the responder's payoff.

To the extent that we are aware, there is no serious departure from compliance in any part of the literature, so we have no hesitation in imposing this requirement.

4.2.3. *EPCF and Bargaining Equilibrium.* In this Section, we link up bargaining equilibrium (which will nest classical models of two-player noncooperative bargaining but contain much more) with the general solution concept of this chapter, that of an EPCF.

The central feature of such a connection is the link between an acceptable proposal on the one hand, and the notion of a profitable and efficient move on the other.

PROPOSITION **8.** *Consider any bargaining game. Then the equilibrium process P^σ corresponding to any no delay, compliant bargaining equilibrium σ is an EPCF.*

Proof. Pick any bargaining equilibrium σ. It generates a process P^σ on all implementation histories. This is a PCF, once we identify every implementation history with the notion of a "history" under the EPCF. Specifically, retain the list of all active coalitions, partners, moves, and previous rejectors up to period $t-1$.

Fix a history, as just identified, and consider the choice of any active (singleton) coalition — the proposer — together with a set of potential partners, as dictated by the protocol of the bargaining game. Consider the equilibrium proposal made. Since, by the no-delay hypothesis each partner accepts the proposal, doing so must yield a value that is at least as high as the value following a rejection. So it is immediate that the proposal is profitable for all the partners concerned. We now establish efficiency (and therefore profitability[40]). Suppose, on the contrary, that there is an alternative profitable proposal that makes the proposer strictly better

[40] With a singleton proposer, verification of efficiency suffices to guarantee profitability for the proposer.

off (in terms of value under P^σ). That proposal must involve at least one partner, otherwise the proposer can unilaterally achieve his alternative, which is impossible in equilibrium. Consider the equilibrium responses to this proposal in the order given by the protocol, up to all the respondents or the first equilibrium rejection, if any.[41] By compliance, each respondent (working back recursively from this point) will take the action that benefits the proposer, in case the respondent is indifferent between the equilibrium action and some other. Therefore the alternative move can be implemented by a proper deviation. This makes him better off relative to the putative equilibrium, a contradiction. ■

4.3. **Some Existing Models of Noncooperative Coalition Formation.** In this section, we describe some models of coalition formation, and embed these into the real-time setup developed in the previous section. In most cases, these existing models are not real-time theories, and the resulting embedding is perforce somewhat unnatural. That, to us, is a virtue: not only will we be able to describe the positive and useful features of these models, but — to the extent that a real-time description is called for in some situations — we will also be able to point out potential inadequacies in the existing literature.

4.3.1. *The Standard Bargaining Problem.* Much of what we do relies on the solution to a standard bargaining problem, due to Ståhl (1977) and Rubinstein (1982). In this section, we briefly recapitulate that problem. A group of n persons divide a cake of size 1; there are no subcoalitions of any value, and there are no externalities. A protocol chooses a proposer in every round, and everyone else responds sequentially to the proposal in some given order. If the proposal is rejected, a new round begins. Future rounds are discounted by a common discount factor.

To cover both the uniform proposer protocol and the rejector-proposes protocols, we consider a general protocol in which the first rejector of a proposal is chosen to be the next proposer with probability $\mu \in [0,1]$, and with probability $1-\mu$, *another* uncommitted agent is equiprobably chosen to be the new proposer.

As we will see, in the pure bargaining problem, the equilibrium will be one in which the grand coalition forms immediately. Since there are no intervening states before the end of the coalition formation process, there is no distinction between a real-time model and one in which payoffs are received at the end of the process.

If $n = 2$, we have two-person bargaining. A remarkable property of this two-person model is that subgame perfection fully pins down equilibrium payoffs. The proposition that follows is well-known from Rubinstein (1982), though we write it for a broader class of protocols:

[41] Our assumption of no-delay equilibrium does not rule out the possibility that a *deviating* proposal must be accepted even if the responder is indifferent between doing and not doing so. This is where compliance is used.

PROPOSITION **9.** *There is a unique subgame perfect equilibrium payoff vector in the two-person bargaining model.*

Proof. Existence will be shown below; assume it for now and prove uniqueness. Let M and m be the supremum and infimum equilibrium payoff to either player as a *responder*, conditional on her rejecting the current offer but before the next proposer has been decided.[42] Then, because a proposer can always assure herself an infimum of at least $1 - M$, and because a responder must be given at least m,

$$m \geq \delta[\mu(1-M) + (1-\mu)m],$$

where μ is the probability that a current rejector gets to propose next. That implies

$$m \geq \frac{\delta\mu(1-M)}{1-\delta(1-\mu)}. \tag{8}$$

But no proposer can obtain more than $1 - m$, so it is *also* true that

$$M \leq \delta[\mu(1-m) + (1-\mu)M].$$

or

$$M \leq \frac{\delta\mu(1-m)}{1-\delta(1-\mu)}. \tag{9}$$

Combining (8) and (9), it is easy to see that

$$m \geq \frac{\delta\mu\left(1 - \frac{\delta\mu(1-m)}{1-\delta(1-\mu)}\right)}{1-\delta(1-\mu)},$$

and simplifying this yields the inequality

$$m \geq \frac{\delta\mu}{1-\delta(1-2\mu)}. \tag{10}$$

Following an analogous line of reasoning,

$$M \leq \frac{\delta\mu}{1-\delta(1-2\mu)}. \tag{11}$$

and together (10) and (11) show that

$$M = m = \frac{\delta\mu}{1-\delta(1-2\mu)} \equiv m^*, \tag{12}$$

which establishes uniqueness.

Existence can now be shown by construction. Have each player accept an offer if it yields her at least m^* (defined in (12), and always make the proposal $(1-m^*, m^*)$ when it is her turn to propose. It is easy to verify that this strategy profile constitutes a perfect equilibrium. ■

[42] When discount factors are not the same, these values vary across the players but the proof follows exactly the same lines.

This proposition and its accompanying proof reveal that the equilibrium involves immediate agreement, with the proposer and the responder receiving

$$\frac{1-\delta(1-\mu)}{1-\delta(1-2\mu)} \quad \text{and} \quad \frac{\delta\mu}{1-\delta(1-2\mu)}$$

respectively. It is worth noting that no matter how small μ is, as long as it is strictly positive, the division of the cake must converge to an equal split as "bargaining frictions" vanish; i.e., as δ converges to 1. It is true that the first individual to propose may acquire a lot of power, especially if μ is small, but the value of that added power becomes negligible provided both players are extremely patient.

To extend this analysis of the bargaining problem to $n > 2$ we restrict attention to equilibria with stationary strategy profiles. It is easy to see that in equilibrium each player must make an acceptable proposal to the grand coalition. Let m_i be the amount that i will accept as a responder, provided that all responders *after* her in the response order are planning to accept that proposal.[43]

In equilibrium, (m_i) must be built from an expectation about payoffs conditional on rejection; these would be a probabilistic combination of i's payoff as a proposer $(1 - \sum_{j \neq i} m_j)$ and as a responder (m_i). Therefore, m_i must solve the following equation:

$$m_i = \delta\{\mu[1 - \sum_{j \neq i} m_j] + (1-\mu)m_i\},$$

or

$$(1-\delta)m_i = \delta\mu[1 - \sum_{j=i}^{n} m_j],$$

which tells us that $m_i = m$ for all i, and

$$m = \frac{\delta\mu}{(1-\delta)+\delta\mu n}. \tag{13}$$

This solution extends the two-person case and once again, convergence occurs to equal division as bargaining frictions disappear, provided that $\mu > 0$.

Unfortunately, the uniqueness result for two-person Rubinstein bargaining no longer survives with three or more players if we allow for non-stationary equilibrium strategies. The argument, due to Herrero (1985) and Shaked (see Sutton (1986)) can be generalized to the full class of protocols we consider here; see Ray (2007) for details).

Moreover, when we consider more general coalitional games next, the protocol has an important bearing on the actual equilibrium payoffs; see Example 12 below.

[43] It is unnecessary to describe here what happens if a later responder is planning to reject, as such a proposal will be rejected anyway and that is all that matters for our argument.

4.3.2. *Coalitional Bargaining With Irreversible Agreements.* In this section, we study a model of coalitional bargaining in which an agreement once made is irreversible. The protocol is therefore restricted so that no agent who has made a deliberate decision to join a coalition can participate in future negotiations.[44] This is indeed the framework for most of the bargaining literature that we have already cited. The real time model developed in Section 4.2, which in turn is a special case of EPCF, can be further simplified to set up a canonical example of irreversible coalitional bargaining with externalities. On still further specialization, the latter yields the bulk of the models used in the literature.

Recall that a proposal refers to a division of the worth of a coalition among its members. In a characteristic function game, for coalition S this is simply a point in $V(S)$. But in a partition function, the worth of a coalition will vary with the going coalition structure. Therefore a proposal must consist of a set of *conditional statements* that describe a proposed division of coalitional worth for every contingency; i.e., for every conceivable coalition structure that finally forms. More precisely, a proposal is a pair $(S, \mathbf{v})$, where $\mathbf{v}$ is a collection of allocations $\{v(\pi)\}$, one for each partition π that contains S, feasible in the sense that for every coalition S in π,

$$v_S(\pi) \in V(S, \pi).$$

We adopt the general version of the protocol in which the first rejector of a proposal becomes the next proposer with probability μ.

Strategies and equilibrium are exactly as defined in the general model.

The literature we will be describing in this Section concerns bargaining with payoffs assigned only after all coalitions have formed. However, it is easy enough to describe a general procedure for embedding such models into our real-time framework. The first step is to specify a way of assigning payoffs to the players each time a coalition gets formed (leading to a new state). We do this by presuming that at date 0 all agents are free and the initial coalition structure is one of singletons.[45] From then on, payoffs are received in every period in a perfectly well-defined fashion. For individuals i who have yet to form a coalition or have deliberately decided to stand alone, they are given by $V(\{i\}, \pi_t)$ at date t, where π_t is the coalition structure prevailing at that date. For individuals i who are part of some coalition S, they are given by $v(\pi_t)$ at date t, which comes from the agreement $\mathbf{v}$ that members of S have entered into at some earlier date.

Next, we need to show that an equilibrium of a standard bargaining game, say σ, is a also an equilibrium in the real-time framework with payoffs being defined at each step that a new coalition forms. We know that no player has a deviation that

[44] That includes agents who have deliberately made the decision to stand on their own.

[45] In the irreversible agreements model, this assumption is without any serious loss of generality, especially if we assume that already-formed coalitions at the start of the game have made their agreements to begin with.

can result in higher payoff *after all coalitions have formed*. So the only possibility of a profitable deviation must come from capturing some gains *before* the coalition formation process comes to an end. If the long-run payoff to the deviating player is *lower*, then for δ close to one, the transitory payoffs cannot make up for the long-run loss. The only possibility that remains for a profitable deviation is that it reaps some transitory gains but eventually results in the same payoff as σ. Moreover, because agreements are irreversible, any such deviation must result in a final coalition structure different from the one under σ. But this can be ruled out by a mild genericity assumption that a different coalition structure cannot yield the deviating player *exactly the same* payoff as the equilibrium payoff. In many models, the equilibrium payoff to a player is closely related to the average worth of her coalition, and this genericity assumption can be then be imposed directly on the partition function. Notice that this genericity argument is very similar to the one we employed in Proposition 7 to connect the static notion of an EBA to an EPCF.

It is well known that the perfect equilibria of such a model can generate a huge multiplicity of outcomes. We already know, in fact, that the n-person bargaining game is prey to multiplicity when $n \geq 3$, and that game is a special case of the irreversible agreements model described here. In what follows, then, we retreat to the use of *stationary Markovian strategies*. They depend on a small set of state variables, and do so in a way that's insensitive to the passage of calendar time. The current proposal or response (while permitted to be probabilistic in nature) is not permitted to depend on "past history". Of course, it must be allowed to depend on the current set of free players, on the coalition structure that is currently in place and — in the case of a response — on the going proposal; after all, these are all payoff-relevant objects.[46]

Note that under our initial condition, the coalition structure must steadily coarsen (or remain unchanged) as time wears on. Thus payoffs are received in real time, and they must finally settle down to a limit value for all concerned. We have therefore successfully, and at little cost, embedded an irreversible agreements model into the real-time setting of a PCF.

The existence of a stationary equilibrium (possibly with randomization) can be proved along the lines of Ray and Vohra (1999) and Okada (2011).

It is instructive to see how the protocol affects the equilibrium payoffs. We illustrate this with a simple three-player characteristic function.

EXAMPLE **11.** *(N, v) is a TU-characteristic function with $N = \{1, 2, 3\}$, $v(i) = 0$ for all i, $v(S) = 1$ for every non-singleton S.*

If the protocol chooses the first rejector to become the next proposer ($\mu = 1$), it is easy to see that in any stationary equilibrium the first proposer offers $\delta/1 + \delta$

[46] We will also permit proposers to condition their new proposals on the identity of the last rejector (in the current round of negotiations), and for respondents to condition their responses on the identity of the proposer.

to one of other player and obtains $1/1 + \delta$. The equilibrium payoffs are therefore exactly as in Rubinstein bargaining with two players; as δ approaches 1, the two players share the aggregate surplus approximately equally. If, however, a proposer is chosen with equal probability ($\mu = 1/3$), regardless of any previous rejection, again a two-player coalition forms in equilibrium. However, the surplus is not shared equally between the two players who form the 'winning' coalition. The proposer offers $\delta/3$ to one of the other players and any such offer is accepted. For δ approaching 1, the proposer receives approximately 2/3 and the other player in the winning coalition receives approximately 1/3. This conforms, of course, to the Baron and Ferejohn (1989) characterization of equilibrium with majority voting.

We will have more to say about the efficiency of equilibrium outcomes in Section 5.

4.3.3. *Equilibrium Coalition Structure.* The central question of interest is the prediction of equilibrium coalition structure. As the reader might imagine, this is an ambitious and complex undertaking, especially in an ambient environment which allows for a variety of strategic situations and alternative protocols. While we do not have a comprehensive understanding of this problem and highlight it as a fundamental open problem, it is possible to make progress in specific cases. We outline some available results.[47]

In what follows we consider only TU partition functions though the analysis extends to certain nontransferable payoffs under some restrictions.[48]

We begin with an additional restriction, which is that the (TU) partition function is *symmetric*. Such a function has the property that the worth of a coalition depends only on its size and the ambient *numerical* coalition structure it is embedded in; namely, the collection of coalition *sizes* in the coalition structure. Use the notation $\mathbf{n}$ to refer to both numerical coalition structures and substructures, the latter being collections of positive integers (including the "null collection" ϕ) that add up to any number strictly less than the total number of players. In the sequel, a substructure is to be interpreted as a collection of coalitions that has "already formed" in a subgame. Define the *size of a substructure* to be the number of all players in it; that is, the sum of coalition sizes in the substructure.

With some abuse of notation, then, $v(T, \pi)$ may be written as $v(t, \mathbf{n})$, where t is the size of coalition T and $\mathbf{n}$ is the numerical coalition structure corresponding to π. Define the *average worth* of t in $\mathbf{n}$ by

$$a(t, \mathbf{n}) \equiv \frac{v(t, \mathbf{n})}{t}.$$

We may interpret this number as the average worth of any coalition of size T embedded in a coalition structure π with associated numerical structure $\mathbf{n}$. In what

[47] More detail on the discussion that follows can found in Ray (2007).

[48] Specifically, we can extend results to symmetric, convex sets of payoffs to each coalition.

follows, we impose for expository convenience the genericity condition

$$a(t, \mathbf{n}) \neq a(s, \mathbf{n}') \text{ for all } t \neq s. \tag{14}$$

We now present an algorithm that calculates a particular coalition structure. Specifically, for each substructure $\mathbf{n}$ with size less than n, the algorithm assigns a positive integer $t(\mathbf{n})$, to be interpreted in the sequel as the size of the coalition that forms *next*. By applying this rule repeatedly starting from the "null structure" with no coalitions, we will generate a particular numerical coalition structure.

STEP 1. For all $\mathbf{n}$ of size $n-1$, define $t(\mathbf{n}) \equiv 1$.

STEP 2. Suppose that $t(\mathbf{n})$ is defined for all substructures $\mathbf{n}$ of size greater than m, for some $m \geq 0$. For all such $\mathbf{n}$, define

$$c(\mathbf{n}) \equiv \mathbf{n}.t(\mathbf{n}).t(\mathbf{n}.t(\mathbf{n}))\ldots,$$

where the notation $\mathbf{n}.t_1 \ldots t_k$ simply refers to the numerical coalition structure obtained by concatenating $\mathbf{n}$ with the integers $t_1, \ldots, t_k$.

STEP 3. For any $\mathbf{n}$ of size m, let $t(\mathbf{n})$ be the integer in $\{1, \ldots, n-m\}$ that maximizes the expression $a(t, c(\mathbf{n}.t))$.

STEP 4. Once the recursive definition is completed for all structures including the null substructure ϕ, define a numerical coalition structure (by

$$\mathbf{n}^* \equiv c(\phi).$$

This completes the description of the algorithm. Its connection to equilibria is striking and direct under the rejector-proposes protocol, through the links are wider-reaching as we shall see subsequently:

PROPOSITION **10.** *Assume the rejector-proposes protocol (that is, $\mu = 1$ in our class of protocols). Then under the genericity condition on average worths, there exists $\delta^* \in (0, 1)$ such that if $\delta \in (\delta^*, 1)$, every no-delay equilibrium must yield $\mathbf{n}^*$ as the numerical coalition structure.*

We omit the proof, but it is easy to construct one along the lines of Ray and Vohra (1999). The main argument is based on the following steps. To begin with, when the partition function is symmetric and the rejector gets to propose with probability one, every compatriot to whom an individual makes offers has the same options conditional on refusal as the proposer currently does. It follows that when the discount factor is close to 1, the same argument used to show equal division in Rubinstein bargaining applies, and any formed coalition must exhibit (roughly) equal division of their worth. It follows from an inductive argument that at every stage with a substructure already formed, it pays to form a coalition that maximizes "predicted" average worth, the qualifier "predicted" coming from the fact that "final" average worths are not fully defined until the coalition structure has fully formed. (That is where the induction is used.) That leaves a small item to be verified. Even though coalition formation is occurring in real time and these final payoffs won't be received until every equilibrium coalition has formed, this does not worry the early-forming coalitions provided that they are sufficiently patient, a requirement that will be picked up by the construction of the threshold δ^*. The real-time structure is of little importance in a model with irreversible agreements (our genericity condition (14) helps to substantially simplify matters here).

It should be noted that the proposition does not assert that an equilibrium implementing $\mathbf{n}^*$ actually exists. Sometimes it may not (see Ray and Vohra (1999)), but these cases are easily ruled out by a mild restriction on average worths. Define *algorithmic average worth* $\hat{a}(\mathbf{n})$ to be the maximized average worth $a(t, c(\mathbf{n}.t))$ achieved by choosing t at any stage of the algorithm (indexed by the substructure $\mathbf{n}$). Say that algorithmic average worth is *nonincreasing* (or that the partition function satisfies NAW, for "nonincreasing average worth") if $\hat{a}(\mathbf{n}) \geq \hat{a}(\mathbf{n}.t(\mathbf{n}))$ for every substructure $\mathbf{n}$ such that $\mathbf{n}.t(\mathbf{n})$ has size smaller than n.

NAW has bite *only* for partition functions. For all characteristic functions, in which the worth of a coalition depends only on the coalition itself, NAW must be trivially satisfied.[49] Whether or not NAW applies more generally (i.e., when externalities are present) is a less transparent question, and the answer will largely depend on the application at hand. But we haven't come across an interesting economic or political application where NAW isn't satisfied. Both the Cournot oligopoly and the public goods model satisfy NAW. It is also important to note that in both cases the algorithm yields an equilibrium coalition structure which is typically *not* the grand coalition, resulting in inefficient outcomes.

Now we apply NAW by showing that in its presence, $\mathbf{n}^*$ is achieved under a variety of protocols. Recall that in our class of protocols, a rejector counterproposes with probability μ, while another uncommitted player is chosen uniformly otherwise. Say that a protocol from this class is *rejector-friendly* if the rejector gets to counterpropose with better than even probability: $\mu > 1/2$. Of course, the familiar rejector-proposes protocol is a special case.

PROPOSITION **11.** *Under NAW and genericity, there is a discount factor* $\delta^* \in (0,1)$ *such that if* $\delta \in (\delta^*, 1)$*, there exists an equilibrium that yields the numerical coalition structure* $\mathbf{n}^*$.

We note once again that $\mathbf{n}^*$ is singled out, because the equilibrium behavior we identify is connected closely to *equal* division of the available worth as bargaining frictions go to zero. For a more nuanced discussion of related issues and some qualifications, see Ray and Vohra (1999) and Ray (2007).

[49] The reason is simple. Our algorithm involves the stepwise maximization of average worth, setting each maximizing coalition aside as the algorithm proceeds. If there are no externalities across coalitions, such a process must result in a sequence of (maximal) average worths that can never increase; for if they did, such coalitions would have been chosen *earlier* in the algorithm, not *later*. This simple observation also assures us that NAW does not demand restrictions such as superadditivity: *all* (symmetric) characteristic functions satisfy it.

An uncomfortably familiar feature of bargaining models is that their predictions are often sensitive to the finer points of procedure — to the *protocol*, in the language of this chapter. This is why we are taking care to present results that cover a broad class of protocols. One might worry, though, that the class isn't broad enough. For instance, the two propositions in this sections have been stated for the class of rejector-friendly protocols, procedures in which the rejector has quite a bit of power. One can show, however (see Ray 2007) that Proposition 11 can be extended to *all* the protocols we consider, provided that the rejector always has a strictly positive probability of making the next proposal, and can also unilaterally exit with no time delay. The condition that is needed is a strengthening of NAW to one in which average worth *strictly* declines along the path of the algorithm;[50] see Ray (2007, p.64, Proposition 5.3). It is also possible to provide additional restrictions on algorithmic average with that guarantee that $\mathbf{n}^*$ is the *unique* equilibrium structure under the rejector-proposes protocol (see Ray and Vohra (1999), Theorem 3.4). As shown in Ray and Vohra (1991, 2001), both the Cournot oligopoly and the public goods model satisfy this addition conditional for uniqueness. Thus $\mathbf{n}^*$ appears to be a focal prediction.

We end our discussion of the symmetric case by returning to a remark made at the start of this section: that it is possible to incorporate nontransferable utility. Suppose that we retain all the symmetry assumptions, but replace the TU worth $v(S, \pi)$ by some *symmetric* set of payoffs $V(S, \pi)$. Nothing of substance will change as long as we are willing to assume that each such set is convex. It is easy to obtain an intuition of why the same arguments go through. Average worth will now need to be replaced by the symmetric utility obtained "along the diagonal" for each $V(S, \pi)$, and the same algorithm may be written down with average worth replaced by this symmetric utility.[51]

4.4. **Reversibility.**

So far, we have assumed that a commitment to form a coalition, once made, cannot be undone. In many situations this isn't a bad assumption. But there are, of course, numerous scenarios in which agreements may be reversed freely (or at little cost). A free-trade area or customs union may initially exclude certain countries and later incorporate them. While two firms might merge, a multi-product firm may also spin off divisions into sub-firms. Political coalitions may form and reform.

Reversibility comes in two flavors. An agreement may be viewed as indefinitely in place, unless signatories to that agreement voluntarily agree to dissolve it. We will call this the case of *renegotiable agreements*. Or an existing agreement may

[50] That is, $a(\mathbf{n}) > a(\mathbf{n}.t(\mathbf{n}))$ for every substructure $\mathbf{n}$ such that $\mathbf{n}.t(\mathbf{n})$ has size smaller than n.

[51] Convexity is needed. There is no guarantee otherwise that "equal division" will be followed within all coalitions, and our techniques cease to apply.

simply come with an expiry date, after which new options can be freely explored. One might call this case of *temporary agreements*.[52]

What is the role played by a renegotiable commitment? Why would a commitment to form a group first be made, then reversed? Why not simply eschew the making of that commitment in the first place? As a concrete instance, consider the public goods model with three players. Assume that an initial proposer is drawn randomly, that proposals must be universally acceptable to the players involved, and that the first rejector of a going proposal gets to make a new proposal. If group formation is irreversible, it is easy to see that there is only one (numerical) equilibrium structure. Player i stands alone, and players j and k band together; see Example 14 below for details. The outcome is inefficient.

Now suppose that new proposals can always be made. Then there are two possibilities, both leading to an efficient outcome. First, if a player moves off on her own, the other two players disband as well, incurring a temporary loss of payoff but thereby getting into position to enforce a symmetric, efficient outcome with the all three players coming together. If this path indeed constitutes credible play, then no player will move off in the first place, and the outcome is efficient to begin with.

(Observe that this isn't even a possibility if commitments are irreversible. Once player i moves off, there is no bringing her back, so players j and k will never disband.)

The second possibility concerns a situation in which once player i moves off, players j and k do not find it worthwhile to disband. For instance, this could happen if player i can make a commitment which is irreversible for some length of time, a situation which can be readily modeled by lowering the discount factor of all players. In this case the outcome will still be efficient, but the path to efficiency as well as the final outcome will look very different. Some player i *must* initially move off. Thereafter, players j and k must cajole her back to the grand coalition with an offer that gives her more than what she gets in the structure $\{i, jk\}$. So we are ultimately at an efficient outcome, one that is "skewed" in favor of the individual who was lucky enough to be the first to make a commitment. Notice that *the commitment must have been made* for her to take advantage of it, and so the equilibrium path involves a transitory phase of inefficiency, followed by a Pareto-superior outcome.

This example may be easily modified to take account of temporary agreements. For instance, suppose that the signatories to the agreement to bring player i back into the fold cannot commit to honor this agreement in the future. If — in that future — some other player j were to unilaterally desert the agreement and take up the same stance as player i did, then there may be little in the situation to induce player i to

[52] To be sure, agreements may be both temporary and renegotiable (within the period for which the agreement is in place), but here we only look at the two features separately.

take up the conciliatory offer in the first place. That could result in a permanent failure to achieve an efficient outcome, a theme that we return to in Section 5.

Both renegotiable and temporary agreements (and several other variants) can be defined using the effectivity correspondence. Suppose that a pair $x = (\pi, u)$ represents a going state, embodying certain agreements, and a move is contemplated to a new state $y = (\pi', u')$, in which one or more of those agreements are disrupted. When agreements are renegotiable, we would like to describe the coalitions that are effective in moving the state from x to y. First, if a player's coalitional *membership* is affected as a consequence of a proposed move, the move *must* be disrupting some previous agreement to which that player was a signatory. That player must be included in any group that is effective for the proposed move. Second, the proposed move might affect the (ongoing) *payoff* to a particular agent, without altering her coalitional membership. Must consent be sought from that agent? The situation here is more subtle. It may be that the payoff is affected because a fellow-member of a coalition wishes to reallocate the worth of that coalition. In that case — given that the existing allocation is in force — it is only reasonable that our agent be on the approval committee for the move. On the other hand, our agent's payoff may be affected because of a coalitional change elsewhere in the system, which then affects our agent's coalition via an externality. Our agent is "affected", but need not be on the approval committee because she wasn't part of the agreement "elsewhere" in the first place.[53]

More formally, for any move from x to y, let $M(x, y)$ denote the set of individuals whose coalitional membership is altered by the move, and $W(x, y)$ the set of individuals j whose one-period payoffs are altered by the move: $u_j(x) \neq u_j(y)$. Say that agreements are *binding but renegotiable* if the following restrictions are met:

[B.1] For every state x and proposed move y, $M(x, y) \subseteq S$ whenever $S \in \mathcal{E}(x, y)$.

[B.2] Suppose that $T \cap W(x, y)$ is nonempty for some existing coalition T. Then if the proposed move involves no change in coalition structure, *or* if payoffs are described by a characteristic function, $T \cap W(x, y) \subseteq S$ whenever $S \in \mathcal{E}(x, y)$.

[B.1] is obvious. To understand [B.2], note that the payoffs in coalition T have been affected. But if there has been no change in the coalition structure, or if the situation is describable by a characteristic function to begin with, how could that happen? It can only happen if there is a deliberate reallocation within that coalition, and then [B.2] demands that all individuals affected by that reallocation must approve it. It is in this sense that [B.1] and [B.2] together formalize the notion of binding yet renegotiable agreements.

Restrictions such as [B.1] and [B.2] placed on the effectivity correspondence permit us to explore all sorts of other variants. For instance, a theory of "temporary

[53] Notice that we wouldn't insist that our player should *not* be on that approval committee; it's just that our definition is silent on the matter.

agreements" can characterized as follows: a coalition that's effective for moving x to y must contain all members of at least $m-1$ of the m new coalitions that form, and the coalition not included must be a subset of an erstwhile coalition.[54] As a second variant, allow a coalition to break up or change if some given fraction (say a majority) of the members in that coalition permit that change. Some political voting games or legislative bargaining would come under this category. Now any effective coalition must consist of at least a majority from *every* coalition affected by the move from one state to another. In the reverse direction, [B.1] and [B.2] could be further strengthened: for instance, one might require that a coalition once formed can never break up again. This would lead us back to the model with irreversible commitments.

We return to the efficiency properties of models with reversible agreements in Section 5 below.

5. The Welfare Economics of Coalition Formation

A central question in the theory of coalition formation has to do with the attainment of efficiency. The Coaseian idea that efficiency is inevitable in the absence of informational frictions and full contracting is deeply ingrained in the economics literature. Indeed, if there are no restrictions on contracting, the fundamental impediments to efficiency are generally seen as arising from adverse selection or moral hazard, these stemming from deeper asymmetries of information. That incentive compatibility constraints may rule out first-best efficiency is, of course, well understood, and points to the importance of second-best efficiency, as in the notion of incentive efficiency of Holmström and Myerson (1984).[55] First-best efficiency can sometimes be restored by cleverly designing mechanisms, or rules of the game, that align individual incentives with the social goal of efficiency.[56] In our complete information framework, of course, these complications do not arise; it is trivial to design an efficient mechanism. On the other hand, by granting agents full freedom to form coalitions of their choice we implicitly rule out certain kinds of mechanisms. In effect, every coalition is permitted to adopt an efficient mechanism of its own.[57]

As we shall seek to explain in this Section, there are many situations in which the very possibility that groups can form serves as an impediment to efficiency. It isn't

[54] It is to be interpreted as a "residual" left by the other "perpetrating coalitions".

[55] For cooperative theory, though, the problem runs even deeper. Restricting all coalitions to incentive compatible contracts does not necessarily yield core stability. Even in otherwise classical environments without externalities, such as exchange economies, the incentive compatible core may be empty; see for example, Vohra (1999) and Forges, Mertens and Vohra (2002). For reviews of this literature we refer the reader to Forges, Minelli and Vohra (2002) and Forges and Serrano (2013).

[56] See, for example, Palfrey (2002) for a review of the implementation literature.

[57] Recall the discussion in Section 4.1.1. See also Ray and Vohra (2001) for further elaboration on this point in the context of the free-riding problem.

a question of incomplete information, though of course there must be some limits to contracting. But what are these limits, and how precisely do these manifest themselves?

5.1. **Two Sources of Inefficiency.** There are actually two sources of inefficiency that we seek to make explicit in this Section. The first can cause inefficiency even when there are no externalities across coalitions. The second is fundamental to situations with inter-coalitional externalities, those typically captured by partition functions. The heart of the first inefficiency is that the "correct" coalition is often not formed, because the active set of players responsible for forming the group seeks to maximize its *own* payoff (or more accurately, to find a maximal payoff vector for itself), but in doing so it will generally need to enlist partners *who have to be suitably compensated*. The nature and amount of that compensation depends crucially on the protocol that governs the process of coalition formation. At the heart of the second inefficiency is a more classical concept: that of externalities. In a typical coalition formation problem, the two effects are often intertwined, but it is instructive to see them separately. We shall begin this discussion by assuming that agreements are irreversible.

EXAMPLE **12.** *Consider a three-player TU characteristic function with* $v(123) = 1 + \epsilon$*, where* $\epsilon \in (0, 0.5)$*,* $v(ij) = 1$ *for all* i*,* j*, and* $v(i) = 0$ *for all* i*.*

While efficiency requires that the grand coalition be formed, as is well know, the blocking approach does not yield this an equilibrium outcome. This is not a balanced game and its core is empty; for any division of $v(N)$ among the three players there is a two-player coalition with an objection. If only internal objections are permitted, the coarsest coalition structure that is an EBA consists of a two-player coalition and a singleton. One way to explain why no player will try to form the grand coalition and capture the additional ϵ is that this is too small relative to the power of any one player to prevent the other two from forming a two-person coalition of their own. Presumably, efficiency could be restored if ϵ is high enough (at least 0.5), or if some player were given more power. To examine the latter possibility, suppose coalitions form as follows: a person is chosen at random to be the "ringleader", and she chooses any coalition she pleases. Once the coalition forms, a fraction k of its worth must be equally divided among the members. The ringleader gets to keep the rest. At one extreme, $k = 1$ and all worths must be equally divided, a case emphasized by Farrell and Scotchmer (1988) and Bloch (1996). At the other extreme, $k = 0$ and the ringleader is a perfect dictator.

It is easy to see in this example that efficiency obtains if and only if k is *below* a certain threshold k^*, given by

$$k^* = \frac{6\epsilon}{1 + 4\epsilon} < 1.$$

The intuition is straightforward. When k is small, the ringleader picks up almost the entire worth and will therefore seek to maximize coalitional worth. The division of payoffs may be distasteful, but the outcome is efficient in the Pareto sense. On the other hand, when k exceeds the threshold k^*, the amount that the ringleader has to share with her chosen compatriots becomes an obstacle to efficiency. It is easy to see that in the equal division limit with $k = 1$, the ringleader will seek to maximize average worth, which results in a two-player coalition being formed.

To see how the presence of externalities creates a distinct source of inefficiency, we consider a partition function version of the ringleader example.

EXAMPLE **13.** *There are six players in a symmetric TU partition function game. The only (numerical) coalition structures that matter (all others result in 0 to each player) are $\pi(a) = (3,3), \pi(b) = (3,2,1)$ and $\pi(c) = (2,2,1,1)$, corresponding to three kinds of states, described as follows.*

$$\begin{array}{llllll} a: & \pi(a) & = & (3,3), & v(3,\pi(a)) = 4 & & \\ b: & \pi(b) & = & (3,2,1), & v(3,\pi(b)) = 2, & v(2,\pi(b)) = 5, & v(1,\pi(b)) = 0.5 \\ c: & \pi(c) & = & (2,2,1,1), & v(2,\pi(c)) = 3, & v(1,\pi(c)) = 0.5 & \end{array}$$

Suppose the player with the lowest index in any coalition is the ringleader who can capture the entire worth of the coalition ($k = 0$). Although $(3,3)$ is the only efficient coalition structure, it is not an 'equilibrium'. If a three-person coalition were to form, the next ringleader will form a two-player coalition rather than three. But this results in the first ringleader receiving only 2. The equilibrium strategy for the first ringleader will therefore be to form a two-player coalition, which leads to state c, an inefficient outcome.

In summary, we have shown two things so far. First, the ability to internalize the marginal gains from coalition formation is crucial to the formation of the "right" coalitions. Second, when there are externalities, even that may not be enough. We proceed now to examine a variety of models of coalition formation and their associated implications for efficiency.

5.2. **Irreversible Agreements and Efficiency.** We begin by studying irreversible agreements, and we draw on both the bargaining and blocking approaches as needed. We address both sources of inefficiency. For the first, it suffices to study characteristic functions.

We have already seen that the stationary equilibria of any reasonable bargaining game will impose some restrictions on how far the payoff to the proposer can diverge from the average worth of the coalition she seeks to form.[58] If this divergence is small enough (effectively giving very little extra power to the 'ringleader' of Example 12), the first coalition to form in equilibrium will be the one which maximizes average worth among all coalitions. (In Example 12, a two-player coalition

[58] Implicitly, the blocking approach also imposes similar restrictions.

will form.) Indeed, as we saw in Section 4.3.3, for a wide variety of protocols the unique equilibrium coalition structure in symmetric games is given by an application of the algorithm that recursively maximizes average worth. In Example 12 we can say more. For *any* general protocol in which μ, the probability of the first rejector being the next proposer, is greater than ϵ there is a unique equilibrium which results in the formation of a two-player coalition;[59] see Ray (2007), page 143 for details.

As the next proposition shows, inefficiency is even more pervasive than this simple example would suggest. The formation of the grand coalition in equilibrium implies that no other coalition has higher average worth, a condition that may not hold even in a balanced (but non-symmetric) game.

PROPOSITION **12.** *Suppose (N, V) is a characteristic function game and the first rejector of a proposal is chosen to be the next proposer with probability $\mu \in (0, 1]$. If, for all discount factors sufficiently close to one, there is an equilibrium in which grand coalition forms immediately, regardless of the identity of the first proposer, then $\frac{v(N)}{|N|} \geq \frac{v(S)}{|S|}$ for all $S \subseteq N$.*

Proof. Suppose in equilibrium each proposer makes an acceptable proposal to the grand coalition. This must mean that the proposer cannot do better by making an acceptable proposal to a subcoalition of N. Letting m_i denote the minimum amount that i will accept from a proposer, provided all remaining responders plan to accept, this implies

$$(15) \qquad v(N) - \sum_{j \in N, j \neq i} m_j \geq v(S) - \sum_{j \in S, j \neq i} m_j, \text{ for all } S \subseteq N.$$

Since the grand coalition forms immediately in equilibrium, regardless of the proper's identity, we can apply the same argument used in Section 4.3.1 to show that $m_i = m$ for all i, where

$$m = \frac{\mu v(N)}{(1-\delta) + \delta\mu|N|},$$

and rewrite (15) as:

$$v(N) - (|N| - 1)m \geq v(S) - (|S| - 1)m,$$

or

$$v(N) - (|N| - |S|)m \geq v(S).$$

[59] The division of the surplus between the proposer and her partner will, however, depend on the precise form of the protocol. In the rejector-proses protocol the proposer receives a little more than 1/2 whereas in the random proposer protocol she receives a little more than 2/3, exactly as in Example 11. The difference, of course, is that in the present example the equilibrium outcome is inefficient.

Substitutiing for m, this is equivalent to

$$v(N)\frac{(1-\delta)+\delta\mu|S|}{(1-\delta)+\delta\mu|N|} \geq v(S),$$

or

$$\frac{v(N)}{(1-\delta)+\delta\mu|N|} \geq \frac{v(S)}{(1-\delta)+\delta\mu|S|}.$$

As δ converges to 1, this yields $\frac{v(N)}{|N|} \geq \frac{v(S)}{|S|}$. ■

Thus, if agreements are irreversible, inefficiency cannot be ruled out even in simple characteristic function games for any reasonable bargaining process that doesn't artificially restrict the coalitions that can form.[60] [61] A converse of Proposition 12 holds for the rejector-proposer protocol (Chatterjee et al. (1993)) as well as the random proposer protocol (Okada (1996, 2011)). Since we are focusing on inefficiency, we refrain from trying to establish the converse for the more general protocol, but settling this remains an interesting open question.

We now turn to the question of externalities across coalitions. As we observed in the Section 4.3.3, both the Cournot oligopoly and the public goods model typically yield inefficient equilibria. To explain the nature of the problem we consider a simple three-player special case of the public goods model. The highest social surplus requires full cooperation, and the complete breakdown of cooperation, with each player acting as a singleton, is the worst outcome. However, it is highly profitable for a single player to break away from the grand coalition *if* the other two stay together. In this example, unlike the three-player Cournot oligopoly of Example 5, the coalition of two players would be worse off at the Nash outcome and so the expectation that the remaining two players will indeed stay together seems reasonable.

[60] Bargaining games that implement the Shapley value deliver efficiency *a fortiori*, but they do not allow a proposal to be made to any coalition of the proposer's choosing. For example, Hart and Mas-Colell (1996) require proposals to be made to the full set of available players, and Gul (1989) considers a process of pairwise meetings in which one player buys out the other's resources and continues to bargaining with the remaining players. In this respect Gul's model is similar in spirit to the models of renegotiation such as Okada (2000) and Seidmann and Winter (1998) in which coalitions form gradually.

[61] Xue and Zhang (2011) suggest another modification of the bargaining model to establish the existence of an efficient equilibrium for a partition function game with irreversible agreements. In their model, the choice of an individual proposer through a protocol is replaced by a bidding mechanism as in Perez-Castrillo and Wettstein (2002). *All players* bid simultaneously on each of the feasible moves from a given state, with the 'winning move' being one that attracts the maximum aggregate bid. In particular, player i has an influence on the move, even if the move involves a change in the coalition structure that leaves i's coalition unchanged. As they show, there is at least one stationary equilibrium in which this turns out to be sufficient to internalize the gains from forming the 'correct' coalition structure.

EXAMPLE **14** (Public goods revisited)**.** *The three-player partition function is defined as follows.*

$$\begin{array}{lllllll} x_N: & \pi_N & = & \{123\}, & u(x_N) & = & (12,12,12) \\ x_1: & \pi_1 & = & \{1,23\}, & u(x_1) & = & (16,7,7) \\ x_2: & \pi_2 & = & \{2,13\}, & u(x_2) & = & (7,16,7) \\ x_3: & \pi_3 & = & \{12,3\}, & u(x_3) & = & (7,7,16) \\ x_0: & \pi_0 & = & \{1,2,3\}, & u(x_0) & = & (6,6,6) \end{array}$$

The effectivity relations for a three-player hedonic game are as follows. For internal blocking, where $i \neq j \neq k$:

$$\pi_N \rightarrow_i \pi_i, \quad \pi_N \rightarrow_{jk} \pi_i, \quad \pi_i \rightarrow_j \pi_0.$$

When external blocking is permitted we also have:

$$\pi_0 \rightarrow_N \pi_N, \quad \pi_0 \rightarrow_{jk} \pi_i.$$

With internal blocking it is easy to see that each π_i is an EBA since none of the two players in the larger coalition would gain by precipitating π_0. In fact, these are the coarsest coalition structures corresponding to EBA. The grand coalition is not an EBA because player i gains by moving to π_i and taking a 'free-ride'.[62] Of course, this conclusion depends critically on the restriction to internal blocking. Indeed, with internal blocking the absorbing states of an EPCF yield precisely the same set of outcomes (Proposition 7). Which player is able to gain the free-riding advantage starting from the grand coalition depends on the protocol: the first player given an opportunity to move will decide to stand alone. It is natural now to ask how this may change if we depart from the assumption of internal blocking.

Consider the notion of farsighted blocking as embodied in the notion of EEBA introduced by Diamantoudi and Xue (2007). Recall that this is the set of farsighted stable sets in the present context. It is not difficult to see that in Example 14 π_N is an EEBA. Since it is a singleton set it obviously satisfies internal stability. All other coalition structures have farsighted objections culminating in π^N: $\pi_0 \rightarrow_N \pi_N$ and $\pi_i \rightarrow_j \pi_0 \rightarrow_N \pi_N$. Each π_i also constitutes an EEBA. Clearly, $\pi_N \rightarrow_i \pi_i$ and $\pi_0 \rightarrow_N \pi_N \rightarrow_i \pi_i$. Moreover, for $j \neq i$, $\pi_j \rightarrow_i \pi_0 \rightarrow_N \pi_N \rightarrow_i \pi_i$. Note that in the last step the (optimistic) presumption is that player i will have the opportunity to move from π_N to π_i. In our dynamic model this presumption will be justified only if the protocol selects player i to be the first potential mover from π_N.

The observation in Example 14, that (strong) efficiency can be supported through EEBA even when it's not possible to do so through EBA, can be extended to a class of hedonic partition function games in which π_N dominates π_0 and every

[62] In this simple example each π_i is Pareto efficient but not *strongly* Pareto efficient in the sense that the aggregate payoff to all three players is higher in the grand coalition. If transfers are possible, then only the grand coalition is efficient and it can be shown (Ray and Vohra (1997)) that the coarsest EBAs correspond to the intermediate coalition structures, which are inefficient.

other coalition structure π contains a player in a non-singleton coalition for whom π_N dominates π.

PROPOSITION **13.** *(Diamantoudi and Xue (2007)). Suppose π_N is a Pareto efficient coalition and Pareto dominates π_0. Then π_N is an EEBA if for all $\pi \neq \pi_N$ and $\pi \neq \pi_0$ there is a coalition $S \in \pi$ such that $|S| \geq 2$ and $u_i(\pi_N) > u_i(\pi)$ for some $i \in S$.*

This is an important positive result because its assumptions are satisfied in symmetric games with positive externalities, e.g. pure public goods economies and the Cournot oligopoly. Thus, efficiency can be restored in such games if we remove the restriction to internal blocking and adopt EEBA as the equilibrium concept. As Diamantoudi and Xue (2007) show, if the assumptions of Proposition 13 are not satisfied *all* EEBAs may be inefficient.

The intuition for Proposition 13 should be clear from our discussion of Example 14. If π_N Pareto dominates π_0 a coalition seeking to reach π_N only needs to engineer a chain of moves that lead to π_0 as the penultimate step in a far-sighted objection terminating in π_N. This argument relies of course on an optimistic view of the world which, as we saw in Example 8, can be problematic. Moreover, matters can be more complicated if we properly take account of payoffs on the path to equilibrium. This will be become clear from our continued discussion of Example 14 in the next Section.

5.3. **Reversible Agreements and Efficiency.** Agreements may be reversible either because they are temporary or because they can, in principle, be renegotiated. We consider each of these cases in turn.

5.3.1. *Temporary Agreements.* Suppose are agreements are only valid for one period of time (assuming them to last for some other fixed period of time would make no difference to the analysis). We shall illustrate the efficiency issue by reexamining Example 14 through our dynamic model. In doing so we assume that the protocol has the form described in Section 3.6. In particular, at each state a coalition is given at most one chance to make a move. For simplicity we also assume that when the state is π_N each of the singletons who have not yet been given a chance to move are chosen with equal probability. Also assume that at state π_0 the first coalition chosen to make a potential move is N.

For δ sufficiently high, there does exist an EPCF with π_N as the unique absorbing state. This EPCF has the grand coalition moving from π_0 to π_N and a player $j \neq i$ moving from π_i to π_0. The grand coalition represents an absorbing state because the only coalition that could possible gain by moving away from this is state is a singleton, say i, hoping to obtain 16 by moving to π_i. However, this is immediately followed by a payoff of 6 and then a return to 12 forever. For $\delta > 2/3$ this is not profitable since $12 + \delta 12 > 16 + \delta 6$. In fact, it can be shown that π_N is the only possible absorbing state for any EPCF in this example. In particular, π_i cannot be an absorbing state. Suppose it is. This means of course that π_0 is not an absorbing state because jk would then have a profitable move to the absorbing state π_i. Nor can π_N be an absorbing state because then jk has an efficient and profitable move to π_N via π_0. (Note that from π_0 player i cannot move to π_i). Now consider a move by player j from π_i to π_0 followed by a move by N to π_N. From here the process will either get absorbed into π_i or, move to π_j or π_k. The worst possible outcome for player j is that it moves immediately to π_i. Thus the worst that this yields to player j for these steps is $6 + 12\delta + \delta^2 7$ compared to $7 + \delta 7 + \delta^2 7$. For $\delta > 1/5$ this is a profitable move, contradicting the hypothesis that π_i is an absorbing state. Thus, for $\delta > 2/3$, the unique absorbing state is π_N. In particular, unlike EEBA, it is no longer possible to sustain π_i as an absorbing state of an EPCF. This difference results from the explicit accounting of temporary gains which can make it worthwhile for players to move even if there is an eventual return to the status quo. As we shall next show, this reasoning can make it impossible to achieve efficiency through the dynamic process even under the assumptions of Proposition 13

Now change Example 14 so that in the intermediate coalition structure π, the singleton, i, receives 19 rather than 16 (all other payoffs remain unchanged). Proposition 13 continues to apply and π_N therefore remains an EEBA. However, we will now show that π_N cannot be an absorbing state of an EPCF. In fact, there are no absorbing states and it is impossible to achieve Pareto efficiency in any EPCF.[63] We claim that π_N cannot be an absorbing state. Consider the case in which player 1 is selected to make a move at π_N and she moves to π_1. The worst that can happen for player 1 from that state is a move to π_0 followed by a move to π_N. This yields player 1 the payoff $19 + \delta 6$ compared to $12 + \delta 12$ in the next two periods (with no change in future periods). Since $\delta < 1$, this is a profitable move, contradicting the hypothesis that π_N is an absorbing state. In fact, the unique EPCF is one in which there is no absorbing state and the transitions between states are the following:

$$\begin{aligned} &\pi_0 \to_N \pi_N, \\ &\pi_i \to_j \pi_0 \text{ whenever } j \text{ is selected to move,} \\ &\pi_N \to_i \pi_i \text{ whenever } i \text{ is selected to move.} \end{aligned}$$

The equilibrium process immediately moves from π_0 to π_N; from π_N to each of the intermediate coalition structures with probability 1/3 and then immediately to π_0. Thus, the process visits π_0 and π_N one-third of the time and the remainder is equally divided between the other three intermediate states. The expected payoff to each player, ignore discounting, is therefore $(1/3)12 + (1/3)(11) + (1/3)6 = 9.67$ which is clearly inefficient.

5.3.2. *Renegotiation.* We've seen that proposer incentives are often distorted by the potential loss of control that accompanies a rejected proposal. In short, a proposer must always give some fraction of the surplus away, and a wedge is driven

[63] Konishi and Ray (2003) provide other examples of abstract games with similar features in their model.

between socially and privately optimal actions. On the other hand, intuition suggests that if outcomes can be renegotiated, then the already-agreed-upon arrangements safeguard *existing* payoffs against any loss of control from making a fresh proposal. This suggests two things. First, if there is surplus left on the table, then that surplus should eventually be seized and divided in some way among all parties. Second — and somewhat in contrast to the first point — the seizure of that surplus won't generally happen at the very first round. The safeguards may have to be put in place in earlier rounds, necessitating step-by-step progress towards efficiency (and hence a sacrifice of full dynamic efficiency). These ideas lie at the heart of contributions by Seidmann and Winter (1998), Okada (2000), Gomes and Jehiel (2005), Gomes (2005) and Hyndman and Ray (2007).

To make the point about gradualism completely explicit, recall Example 8. It describes a three-player symmetric characteristic function with $v(123) = 1+\epsilon$, where $\epsilon \in (0, 0.5)$, $v(ij) = 1$ for all i, j, and $v(i) = 0$ for all i. Apply to this the rejector-proposes protocol. Then, if only irreversible arrangements are possible, and the discount factor is close enough to unity, an (inefficient) two-person coalition forms and a valuable third player is omitted, for reasons already discussed. With renegotiation, matters are different. A two-person coalition will still form at first, but the *eventual* outcome is efficient, as the third person can be taken in without any fear of dilution to the already committed players in the two-person coalition. The formation of an "intermediate coalition" essentially protects the parties to that agreement. The agents included in the intermediate coalition can block any attempt by the excluded agent to undercut them, because they are already signatory to a binding agreement that can only be abolished with the consent of both players. That reduces the power of the excluded agent to extract surplus, and the grand coalition can finally form.

One feature of this example is that ultimately all renegotiation ceases and the economy "settles down". More importantly, are those limit payoffs efficient? Consider the following example of a four-person characteristic function, with $v(S) = 3$, if $S = N$, $v(12) = v(34) = 1$ and $v(S) = 0$ otherwise. Suppose that the protocol is "rejector proposes". Provided that the discount factor is close enough to unity, there is an equilibrium in which the coalition structure $\{12, 34\}$ forms but no further progress is made: all proposals to the grand coalition are rebuffed and the rejector demands the entire surplus net of existing payoffs to the other three agents).

Notice, however, that the failure to achieve efficiency is based on rather knife-edge considerations. An efficiency-enhancing proposal may be rejected, true, but events post-rejection cannot hurt our existing players by too much, *because ongoing agreements are binding*. If a proposer does not mind being rejected as long as subsequent play benefits others *and* does not hurt her, such history-dependent inefficiencies can be broken provided that the status quo agreements are binding.That motivates the following concept: say that an individual is *benign* if she prefers an outcome in which some other individuals are better off, provided that she (and every individual) is just as well off. The benignness "refinement" is of a lexicographic nature. Our individual first and foremost maximizes her own payoff, and benignness only kicks in when comparisons are made over outcomes in which her payoff is unaffected. There is no danger to the payoff of the individual concerned.

Benignness has found support in a number of different experimental settings (including bargaining); see, e.g., Andreoni and Miller (2002), Charness and Grosskopf (2001) and Charness and Rabin (2002), among others. Indeed, these studies suggest something stronger: people are sometimes willing to *sacrifice* their own payoff in order to achieve a socially efficient outcome. Given its lexicographic insistence on maximizing one's own payoff, benignness certainly doesn't go that far.

Under benignness, asymptotic efficiency must be attained under every possible equilibrium:

PROPOSITION **14.** *Assume* [B.1] *and* [B.2]. *In characteristic function games all equilibria are absorbing. Moreover, if the set of states is finite, every pure strategy benign equilibrium is asymptotically efficient: every limit payoff is static efficient.*

This proposition, taken from Hyndman and Ray (2007), is to be contrasted with the folk-theorem-like results obtained in Herrero (1985) and Chatterjee et al. (1993) for the case of irreversible agreements. With repeated negotiation, no amount of history-dependence in strategies can hold players away from an (ultimately) efficient outcome. In this sense, Proposition 14 represents a substantial extension of Okada (2000) and Seidmann and Winter (1998), who showed that renegotiation achieves efficiency in superadditive characteristic functions when equilibria are restricted to be Markovian. (Neither superadditivity nor the Markovian assumption is needed here.) The issue of how far these results can be pushed by restricting attention to Markovian equilibria is addressed by Proposition 15 below.

While we omit a formal proof, it is easy to see the intuition behind this result. Suppose, contrary to our assertion, that convergence occurs to an inefficient limit. Then a proposer will have the incentive to propose a payoff vector that Pareto-dominates this payoff. This follows from two observations. First, because agreements are binding, the proposer cannot be hurt by making such a proposal. She can always continue to enjoy her going payoff.[64] Second, the proposer is benign. She certainly gains from the proposal *if it is accepted*, and there is no reason to invoke benignness. But the point is that she prefers to make the proposal *even if it is rejected*. For rejection must entail that all the rejectors are better off by *not* accepting the proposal, while the assumption that agreements are binding ensures that no one is strictly hurt (see previous paragraph). A benign proposer would therefore prefer the resulting outcome to the presumed equilibrium play, which is continued stagnation at the inefficient payoff vector.

[64] To make this argument work, we must "already" be at the limit payoff, otherwise the proposer may do some (small, but positive) damage to her own prospects by the very act of making the proposal. This is why we assume a finite set of states, though the finiteness can be dropped; see Hyndman and Ray (2007).

It must be reiterated, though, that the ability to write binding agreements cannot guarantee *full* efficiency in the dynamic sense. As we've seen, absorption will generally require time — i.e., the formation of intermediate coalition structures — before a final outcome is finally settled upon. These intermediate outcomes may well be inefficient. So the path taken as a whole cannot be dynamically efficient.[65]

Most importantly, with externalities across coalitions, matters can be very different. To be sure, renegotiation might restore efficiency, as is easily illustrated by allowing for renegotiation in Example 14. Once the two partners j and k in π_i willingly break up to precipitate π_0, and all three join forces to go to π_N, no further changes can occur because player i can no longer trigger π_i without the consent of the other two. Thus, renegotiation yields efficiency in this Example. Unfortunately, this is not generally the case. The ubiquitous absorption results reported for characteristic functions break down when externalities are present. Equilibrium payoffs may cycle, and even if they don't, inefficient outcomes may arise and persist. Finally — and in sharp contrast to characteristic functions — such outcomes are not driven by the self-fulfilling contortions of history-dependence. They occur even for Markovian equilibria.

It is tempting to think of inefficiencies as entirely "natural" equilibrium outcomes when externalities exist. Such an observation is generally true, of course, for games in which there are no binding agreements. When agreements can be costlessly written, however, no such presumption can and should be entertained. These are models of *binding* agreements, a world in which the so-called "Coase theorem" is relevant. For instance, all two-player games invariably yield efficiency, quite irrespective of whether there are externalities across the two players. This is not to say that the "usual intuition" plays no role here. It must, because the process of negotiation is itself modeled as a noncooperative game. But that is a very different object from the "stage game" over which agreements are sought to be written.

Consider a three-player example.

EXAMPLE **15** (*The Failed Partnership*)**.** *There are three agents, any two of whom can become "partners". The outsider to the partnership gets a "low" payoff: zero, say. A three-player partnership is assumed not to be feasible (or has very low payoffs).*

$$\begin{array}{llcllcl}
x_N: & \pi_N & = & \{123\}, & u(x_N) & = & (0,0,0) \\
x_1: & \pi_1 & = & \{1,23\}, & u(x_1) & = & (0,10,10) \\
x_2: & \pi_2 & = & \{2,13\}, & u(x_2) & = & (5,0,5) \\
x_3: & \pi_3 & = & \{12,3\}, & u(x_3) & = & (5,5,0) \\
x_0: & \pi_0 & = & \{1,2,3\}, & u(x_0) & = & (6,6,6)
\end{array}$$

and use any effectivity correspondence that satisfies (B.1) and (B.2), yet allows all free players to get together without any further consultation or clearance.

[65] There is another reason for the failure of dynamic efficiency: such efficiency may necessitate ongoing cycles across different states. See Hyndman and Ray (2007) for an extended discussion.

The crucial feature of this example is that player 1 is a bad partner, or — for the purposes of better interpretation — a *failed partner*. Partnerships between him and any other individual are dominated — both for the partners themselves and for the outsider — by all three standing alone. In contrast, the partnership between agents 2 and 3 is rewarding (for those agents).

In this example, and provided δ lies sufficiently close to 1, the outcomes x_2 and x_3 — which are inefficient — must be absorbing states *in every equilibrium*.

A formal proof of this observation isn't needed; the discussion to follow will suffice. Why might x_2 and x_3 be absorbing? The reason is very simple. Despite the fact that x_2 (or x_3) is Pareto-dominated by x_0, player 1 won't accept a transition to x_0. If she did, players 2 and 3 would initiate a further transition to x_1. Player 1 *might* accept such a transition if she is very myopic and prefers the short-term payoff offered by x_0, but if she is patient enough she will see ahead to the infinite phase of "outsidership" that will surely follow the short-term gain. That is why it is impossible — in the game as described — to negotiate one's way out of x_2 or x_3. This inefficiency persists in *all* equilibria, history-dependent or otherwise.

This example raises four important points:

1. *The Nature of Agreements*. Notice that the players *could* negotiate themselves out of x_2 or x_3 if 2 and 3 could credibly agree never to write an agreement while at x_0. Are such promises reasonable in their credibility? It may be difficult to imagine that from a legal point of view, player 1, who has voluntarily relinquished all other contractual agreements between 2 and 3, could actually hold 2 and 3 to such a meta-agreement. Could one interpret the stand-alone option (x_0) as an *agreement* from which further deviations require universal permission? Or does "stand-alone" mean freedom from all formal agreement, in which case further bilateral deals only need the consent of the two parties involved? It is certainly possible to take the latter view. One might even argue that this is the only compelling view. In that case efficiency will need to be sacrificed.

But there are situations in which the former view might make sense. No-competition clauses that require an ex-employee not to join a rival firm, at least for some length of time, may be interpreted as an example. Such clauses allow a firm to let a top-level executive or partner go when it is in the interest of both of them (and therefore efficient) to do so, but at the same time prevent — perhaps temporarily — another move in which the executive gets absorbed by a third party. While that latter move is, in itself, also efficient, it might prevent the former move from even being entertained.

2. *The Efficiency Criterion: From Every State, or Some?* Observe that the inefficient states x_2 or x_3 wouldn't be *reached* starting from any other state. This is why the interpretation, the "failed partnership", is useful. The example makes sense in a situation in which players have been locked in with 1 on a past deal, on expectations which have failed since. The game "begins" with the failed partnership, so

to speak. Nevertheless, that raises the question of whether there invariably exists *some* initial condition for which efficiency obtains. (That answer is trivially yes in the current example.)

For all three-player games, and provided we are willing to make some minimal assumptions, the answer is in the affirmative. That is, for all δ close enough to 1, there exists an initial state and a stationary Markov equilibrium with efficient absorbing payoff limit from that state (see Proposition 4 in Hyndman and Ray (2007)). But it is also true that a general result in this direction is elusive: there is a *four*-player partition function such that for δ sufficiently close to 1, *every* stationary Markov equilibrium is inefficient starting from *any* initial state (see the four-player example in Hyndman and Ray (2007)).

3. *More on Transfers*. Recall that upfront transfers are not permitted in the failed partnership. Were they allowed in unlimited measure, players 2 and 3 could reimburse player 1 for the present discounted value of his losses in relinquishing his partner. Depending on the discount factor, the amounts involved may be considerable, and might strain the presumption of deep pockets or perfect credit markets needed to carry such transfers out. But they would break the deadlock.

How does the ability to make transfers feed into efficiency? It is important to distinguish between two kinds of transfers. Coalitional or partnership worth could be freely transferred between the players within a coalition. Additionally, players might be able to make large upfront payments in order to induce certain coalitions to form. In all cases, of course, the definition of efficiency should match the transfer environment.[66]

Within-coalition transferability often does nothing to remove inefficiency. For instance, nothing changes in the failed partnership of Example 15. On the other hand, upfront transfers *across* coalitions have an immediate and salubrious effect in that example. Efficiency is restored from every initial state. The reason is simple. If player 1 is offered any (discount-normalized) amount in excess of 5, he will "release" player 2. In view of the large payoffs that players 2 and 3 enjoy at state x_1, they will be only too pleased to make such a payment. The final outcome, then, from any initial condition is the state x_1, and we have asymptotic efficiency.

But transfers can be a double-edged sword. The discussion that follows is based on Gomes and Jehiel (2005).

[66] For instance, if transfers are not permitted, it would be inappropriate to demand efficiency in the sense of aggregate surplus maximization.

EXAMPLE **16** (*Ubiquitous Bad Partnerships*)**.** *Consider the following three-player game:*

$$\begin{array}{lllllll}
x_N : & \pi_N & = & \{123\}, & u(x_N) & = & (0,0,0) \\
x_1 : & \pi_1 & = & \{1,23\}, & u(x_1) & = & (0,a,a) \\
x_2 : & \pi_2 & = & \{2,13\}, & u(x_1) & = & (a,0,a) \\
x_3 : & \pi_3 & = & \{12,3\}, & u(x_3) & = & (a,a,0) \\
x_0 : & \pi_0 & = & \{1,2,3\}, & u(x_0) & = & (b,b,b)
\end{array}$$

Assume that $b > a > 0$.

As in Example 15, use any effectivity correspondence that satisfies (B.1) and (B.2), yet allows all free players to get together.

In this symmetric example, there is a unique efficient state by any criterion. It is state x_0. It Pareto-dominates every other state. In particular, every two-player partnership is an unambiguous disaster. It is obvious that in any reasonable description of this game that precludes upfront transfers, there is a unique absorbing state, which is the state x_0. But the introduction of upfront transfers changes this rather dramatically. Under the uniform proposer protocol, *every stationary Markov equilibrium is inefficient starting from any initial state: the state x_0 can never be absorbing*.

This remarkable observation highlights very cleanly the negative effects of upfront transfers. The usefulness of a transfer is that it frees agents from inefficient outcomes, as in the case of the failed partnership in Example 15. But its potential danger lies in the possibility that individuals may *deliberately* generate inefficient outcomes to seek such transfers. This notion of transfers as ransom turns out to be particularly vivid in this example. Whenever it is the turn of an agent to move at state x_0, she creates a bad partnership with another agent and then waits for a transfer to unlock the partnership. That situation recurs again and again. Can we still be sure that transfers will actually be paid, and that all of these actions properly lock together as an equilibrium? The answer is yes; for details, see Gomes and Jehiel (2005) and the detailed exposition in Ray (2007).

In short, the deviating players do suffer a loss in current payoff when they move away from the efficient state. But the prospect of inflicting a *still* greater loss on the outsider raises the possibility that the outsider will pay to have the state moved back — albeit temporarily — to the efficient point. Thus the presumption that unlimited transfers act to restore or maintain efficiency is wrong.

Notice how the example stands on the presumption (just as in the failed partnership) that two players can always form a partnership starting from the situation in which all three players stand alone. If this contractual right can be eliminated in the act of making an upfront transfer, then efficiency can be restored: once state x_0 is regained, there can be no further deviations from it. This line of discussion is exactly the same as in the failed partnership and there is nothing further to add here.

More generally, the efficient state in this example has the property that a subset of agents can move away from that state, leaving other agents worse off in terms of current payoffs. Whenever this is possible, there is scope for collecting a ransom, and the potential for a breakdown in efficiency. Gomes and Jehiel (2005) develop this idea further.

4. *What About Superadditive Games?* Both the failed partnership, as well as the four-player game mentioned above, involve situations that are not superadditive. If, for instance, the grand coalition can realize the Pareto-improvement then player 1 can control any subsequent shenanigans by 2 and 3 (he will need to be part of any coalition that is effective for further change), and he will therefore permit the improvement, thereby restoring efficiency.

There are subtle issues that need to be addressed here. First, in games with externalities superadditivity is generally not to be expected. For instance, in Example 1 (the Cournot oligopoly), it is easy to see that if there are just three firms, firms 1 and 2 do worse together than apart, provided that firm 3 stands separately in both cases. At the same time, this argument does not apply to the grand coalition of all firms. Indeed, it is not hard to show that every partition function derived from a game in strategic form must satisfy *grand coalition superadditivity* (GCS):

[GCS] For every state $x = (u, \pi)$, there is $x' = (u', \{N\})$ such that $u' \geq u$.

Is GCS a reasonable assumption? It may or may not be. One possible interpretation of GCS is that it is a “physical” phenomenon; e.g., larger groups organize transactions more efficiently, or share the fixed costs of an enterprise such as a business or public good provision. But such superadditivities are often the exception rather than the rule. After all, the entire doctrine of healthy competition is based on the notion that physical superadditivity, after a point, is not to be had. In general, too many cooks do spoil the broth: competition among groups can lead to efficiency gains not possible when there is a single, and perhaps larger, group attempting to act cooperatively. To be sure, in all of the cases, the argument must be based on some noncontractible factor, such as the creativity or productivity created by the competitive urge, or ideological differences, or the possible presence of standalone players who are outside the definition of our set of players but nevertheless have an effect on their payoffs (such as the standalone third player in the Cournot example).

But there is a different notion of GCS, summarized in the notion of the *superadditive cover*. After all, the grand coalition can write a contract which exactly replicates the payoffs obtainable in some other coalition structure. For instance, companies do spin off certain divisions, and organizations do set up competing R&D groups. In principle, the grand coalition can agree not to cooperate, if need be, and yet write agreements that bind across all players. For instance, in the failed partnership of Example 16, the ability to insert no-compete clauses at will effectively converts the game into its superadditive cover. To the extent that such clauses cannot be enforced for an infinite duration, the model without grand-coalition superadditivity can be viewed as a simplification of this, and other, real-world situations.

In fact, Gomes and Jehiel (2005) demonstrate that in many situations efficiency can be restored if there exists an efficient state that is negative-externality free (ENF) in the sense that no coalition, or collection of coalitions, can move away from it and hurt a player who is not party to such a change.

[ENF] There exists an efficient state, $x = (u, \pi)$, such that for all $i \in N$ and $x \to_{S_1} x_1 \to_{S_2} x_2 \ldots \to_{S_k} y$, such that $i \notin S_k$ for all k, $u_i(y) \geq u_i(x)$.

Note that ENF is not satisfied in Example 14. It is easy to see that ENF is weaker than GCS since a move from the grand coalition requires unanimous consent. And it holds trivially in characteristic function games since there are no externalities.

PROPOSITION **15.** *Consider a TU partition function. Every Markovian equilibrium is asymptotically efficient under either one of the following set of conditions:*

(i) *Upfront transfers are permitted, the set of states is finite (modulo transfers) and there exists an ENF state; Gomes and Jehiel (2005).*
(ii) *There are no upfront transfers and GCS is satisfied; Hyndman and Ray (2007).*

Gomes (2005) establishes a related efficiency result under GCS, without assuming a finite number of states but by assuming that coalitions once formed cannot become smaller. He also provides an example to show that GCS cannot be weakened to ENF without allowing upfront transfers across coalitions.

REFERENCES

ABREU, D., D. PEARCE, AND E. STACCHETTI (2012), "One-Sided Uncertainty and Delay in Reputational Bargaining," mimeo, Princeton University.

ACEMOGLU, D., G. EGOROV, AND K. SONIN (2008), "Coalition Formation in Non-Democracies," *Review of Economic Studies*, **75**, 987–1009.

AGHION, P., P. ANTRÀS, AND E. HELPMAN (2007), "Negotiating Free Trade," *Journal of International Economics*, **73**, 1–30.

ANDREONI, J. and J. MILLER (2002), "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism", *Econometrica*, **40**, 737–753.

AUMANN, R. (1961), "The Core of a Cooperative Game Without Sidepayments," *Transactions of the American Mathematical Society* **98**, 539–552.

AUMANN, R. AND J. H. DREZE (1974), "Cooperative Games with Coalition Structures," *International Journal of Game Theory*, **3**, 217–237.

AUMANN, R. and M. MASCHLER (1964), "The Bargaining Set for Cooperative Games," in *Advances in Game Theory* (M. Dresher, L. Shapley, and A. Tucker, eds.), Annals of Mathematical Studies No. 52; Princeton, NJ: Princeton University Press.

AUMANN, R. and R. MYERSON (1988), "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., . 175–191. Cambridge: Cambridge University Press.

AUMANN, R. AND B. PELEG (1960), "von Neumann-Morgenstern Solutions to Cooperative Games without Side Payments," *Bulletin of the American Mathematical Society*, **66**, 173–179.

BANERJEE, S., KONISHI, H. and T. SÖNMEZ (2001), "Core in a Simple Coalition Formation Game, *Social Choice and Welfare* **18**, 135–153.

BARBERÀ, S. AND A. GERBER (2003), "On Coalition Formation: Durable Coalition Structures," *Mathematical Social Sciences*, **45**, 185–203.

——— (2007), "A Note on the Impossibility of a Satisfactory Concept of Stability for Coalition Formation Games," *Economics Letters*, **95**, 85–90.

BARON, D. and J. FEREJOHN (1989), "Bargaining in Legislatures," *American Political Science Review*, **83**, 1181–1206.

BÉAL, S., J. DURIEU AND P. SOLAL (2008), "Farsighted Coalitional Stability in TU-Games," *Mathematical Social Sciences*, **56**, 303–313.

BERNHEIM, D., PELEG, B. and M. WHINSTON (1987), "Coalition-Proof Nash Equilibria. I. Concepts," *Journal of Economic Theory*, **42**, 1–12.

BHATTACHARYA, A. AND V. BROSI (2011), "An Existence Result for Farsighted Stable Sets of Games in Characteristic Function Form," *International Journal of Game Theory*, **40**, 393–401.

BINMORE, K. (1985), "Bargaining and Coalitions," in *Game-Theoretic Models of Bargaining* (A. Roth (ed.)), Cambridge: Cambridge University Press.

BINMORE, K., A. RUBINSTEIN AND A. WOLINSKY (1986), "The Nash Bargaining Solution in Economic Modelling," *The RAND Journal of Economics*, **17**, 176–188.

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

——— (2003), "Non-cooperative Models of Coalition Formation in Games with Spillovers," in *The Endogenous Formation of Economic Coalitions*, ed. by C. Carraro, and D. Siniscaclo, 311–352. Edward Elger.

BLOCH, F. and A. GOMES (2006), "Contracting with Externalities and Outside Options," *Journal of Economic Theory*, **127**, 172–201.

BLOCH, F., SOUBEYRAN, R. and SÁNCHEZ-PAGÉS (2006), "When does Universal Peace Prevail? Secession and Group Formation in Conflict," *Economics of Governance* **7**, 3–29.

BOGOMOLNAIA, A. AND M. O. JACKSON (2002), "The Stability of Hedonic Coalition Structures," *Games and Economic Behavior*, **38**, 201–230.

CARRARO, C. and D. SINISCALCO (1993), "Strategies for the International Protection of the Environment, *Journal of Public Economics*, **52**, 309–328.

CHANDER, P. (2007), "The Gamma-Core and Coalition Formation," *International Journal of Game Theory*, **35**, 539–556.

CHANDER, P, AND H. TULKENS (1997), "The Core of an Economy with Multilateral Environmental Externalities," *International Journal of Game Theory*, **26**, 379–401.

CHARNESS, G. and B. GROSSKOPF (2001), "Relative Payoffs and Happiness: An Experimental Study," *Journal of Economic Behavior and Organization*, **45**, 301–328.

CHARNESS, G. and M. RABIN (2002), "Understanding Social Preferences With Simple Tests," *Quarterly Journal of Economics*, **117**, 817–869.

CHATTERJEE, K., B. DUTTA, D. RAY and K. SENGUPTA (1993), "A Noncooperative Theory of Coalitional Bargaining," *Review of Economic Studies*, **60**, 463–477.

CHWE, M. (1994), "Farsighted Coalitional Stability," *Journal of Economic Theory*, **63**, 299–325.

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

COMPTE, O. AND P. JEHIEL (2010), "The Coalitional Nash Bargaining Solution," *Econometrica*, **78**, 1593–1623.

DE CLIPPEL, G. AND R. SERRANO (2008), "Bargaining, Coalitions and Externalities: A Comment on Maskin," Brown University, Working Paper 2008-16.

DIAMANTOUDI, E. AND L. XUE (2003), "Farsighted stability in hedonic games," *Social Choice and Welfare*, **21**, 39–61.

——— (2005), "Lucas' Counter Example Revisited," Discussion paper, McGill University.

——— (2007), "Coalitions, Agreements and Efficiency," *Journal of Economic Theory*, **136**, 105–125.

DUTTA, B., D. RAY, K. SENGUPTA and R. VOHRA (1989), "A Consistent Bargaining Set, *Journal of Economic Theory* ,**49**, 93–112.

DUTTA, B. and K. SUZUMURA (1993), "On the Sustainability of R&D Through Private Incentives," Indian Statistical Institute Discussion Paper No. 93-13.

DUTTA, B., GHOSAL, S. and D. RAY (2005), "Farsighted Network Formation," *Journal of Economic Theory*, **122**, 143–164.

ESTEBAN, J. and RAY, D. (1999), "Conflict and Distribution", *Journal of Economic Theory* **87**, 379–415.

ESTEBAN, J. and J. SÁKOVICS (2004), "Olson vs. Coase: Coalitional Worth in Conflict," *Theory and Decision* **55**, 339–357.

FARRELL, J. AND S. SCOTCHMER (1988), "Partnerships", *Quarterly Journal of Economics*, **103**, 279–297.

FELDMAN, A. (1974), "Recontracting Stability," *Econometrica*, **42**, 35–44.

FORGES, F., J.-F. MERTENS AND R. VOHRA (2002), "The Ex Ante Incentive Compatible Core in the Absence of Wealth Effects," *Econometrica*, **70**, 1865–1892.

FORGES, F., E. MINELLI AND R. VOHRA (2002), "Incentives and the core of an exchange economy: a survey," *Journal of Mathematical Economics*, **38**, 1–41.

FORGES, F. AND R. SERRANO (2013), "Cooperative Games with Incomplete Information: Some Open Problems," *International Game Theory Review*, forthcoming.

GILLIES, D. (1953), "Locations of Solutions" in *Report of an Informal Conference on the Theory of N-Person Games*, ed. H.W. Kuhn, mimeo, Princeton University.

GOMES, A. (2005), "Multilateral Contracting with Externalities," *Econometrica*, **73**, 1329–1350.

GOMES, A. and P. JEHIEL (2005), “Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies,” *Journal of Political Economy*, **113**, 626–667.

GREEN, J. (1974), “The Stability of Edgeworth’s Recontracting Process,” *Econometrica*, **42**, 21–34.

GREENBERG, J. (1990), *The Theory of Social Situations*, Cambridge, MA: Cambridge University Press.

——— (1994), “Coalition structures,” in *Handbook of Game Theory, Volume 2*, ed. by R. J. Aumann, and S. Hart, 1305–1337.

GUL, F. (1989), “Bargaining Foundations of Shapley Value,” *Econometrica*, **75**, 81–95.

HAERINGER, G. (2004), “Equilibrium Binding Agreements: A Comment,” *Journal of Economic Theory*, **117**, 140–143.

HARSANYI, J. (1974), “An Equilibrium-Point Interpretation of Stable Sets and a Proposed Alternative Definition,” *Management Science*, **20**, 1472–1495.

HART, S. and M. KURZ (1983), “Endogenous Formation of Coalitions,” *Econometrica*, **51**, 1047–1064.

HART, S. AND A. MAS-COLELL (1996), “Bargaining and Value,” *Econometrica*, **64**, 357–380.

HERINGS, P. J.-J., A. MAULEON AND V. VANNETELBOSCH (2009): “Farsightedly Stable Networks,” *Games and Economic Behavior*, 67(2), 526–541.

HERRERO, M. (1985), “n-player Bargaining and Involuntary Unemployment,” Ph.D. Dissertation, London School of Economics.

HOLMSTRÖM, B. and R. MYERSON (1983), “Efficient and Durable Decision Rules with Incomplete Information,” *Econometrica*, **51**, 1799–1819.

HYNDMAN, K. and D. RAY (2007), “Coalition Formation with Binding Agreements,” *Review of Economic Studies*, **74**, 1125–1147.

ICHIISHI, T. (1981), “A Social Coalitional Equilibrium Existence Lemma,” *Econometrica*, **49**, 369–37.

JACKSON, M. (2010), *Social and Economic Networks*, Princeton University Press.

JACKSON, M. and A. WOLINSKY (1996), “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, **71**, 44–74.

KONISHI, H. and D. RAY (2003), “Coalition Formation as a Dynamic Process,” *Journal of Economic Theory*, **110**, 1–41.

KRISHNA, P. (1998), “Regionalism vs Multilateralism: A Political Economy Approach,” *Quarterly Journal of Economics*, **113**, 227–250.

KRISHNA, V. and R. SERRANO (1995), “Perfect Equilibria of a Model of n-Person Non-Cooperative Bargaining,” *International Journal of Game Theory*, **24**, 259–272.

KRUGMAN, P. (1993), “Regionalism versus Multilateralism: Analytical Notes,” in *New Dimensions in Regional Integration* (J. de Melo and A. Panagariya (eds.)), Cambridge, UK: Cambridge University Press.

LUCAS, W. (1968), “A Game with No Solution,” *Bulletin of the American Mathematical Society*, **74**, 237–239.

LUCAS, W. F. (1992), “von Neumann-Morgenstern Stable Sets,” in *Handbook of Game Theory, Volume 1*, ed. by R. J. Aumann, and S. Hart, 543–590. North-Holland.

MARIOTTI, M. AND L. XUE (2003), “Farsightedness in Coalition Formation,” in *The Endogenous Formation of Economic Coalitions*, ed. by C. Carraro and. D. Siniscalco, 128–155. Edward Elger.

MASKIN, E. (2003), “Bargaining, Coalitions and Externalities,” Presidential address of the Econometric Society.

MAULEON, A., V. VANNETELBOSCH AND W. VERGOTE (2011), “von Neumann-Morgenstern farsightedly stable sets in two-sided matching,” *Theoretical Economics*, **6**, 499–521.

MOLDOVANU, B. AND E. WINTER (1995), “Order Independent Equilibria,” *Games and Economic Behavior*, **9**, 21–34.

OKADA, A. (1996), “A Noncooperative Coalitional Bargaining Game with Random Proposers,” *Games and Economic Behavior*, **16**, 97–108.

——— (2000), “The Efficiency Principle In Non-Cooperative Coalitional Bargaining”, *Japanese Economic Review*, **51**, 34–50.

——— (2011), “Coalitional Bargaining Games with Random Proposers: Theory and Application,” *Games and Economic Behavior*, **73**, 227–235.

OWEN, G. (1977), “Values of Games with A Priori Unions,” in *Essays in Mathematical Economics and Game Theory* (R. Henn and O. Moschlin (eds.)), New York, NY: Springer Verlag.

PAGE, F., M. WOODERS AND S. KAMAT (2005), “Networks and Farsighted Stability,” *Journal of Economic Theory*, **120**, 257–269.

PALFREY, T. (2002), “Implementation Theory,” in *Handbook of Game Theory, Volume 3*, ed. by R. J. Aumann, and S. Hart, 2271–2326. North-Holland.

PEREZ-CASTRILLO, D. AND D. WETTSTEIN (2002), “Choosing Wisely: A Multibidding Approach,” *The American Economic Review*, **92**, 1577–1587.

PERRY, M. and P. RENY (1994), “A Non-Cooperative View of Coalition Formation and the Core,” *Econometrica*, **62**, 795–817.

RAY, D. (1989), “Credible Coalitions and the Core,” *International Journal of Game Theory*, **18**, 185–187.

——— (1998), *Development Economics*, Princeton, NJ: Princeton University Press.

——— (2007), *A Game-Theoretic Perspective on Coalition Formation*, Oxford University Press.

RAY, D. and R. VOHRA (1997), “Equilibrium Binding Agreements,” *Journal of Economic Theory*, **73**, 30–78.

——— (1999), “A Theory of Endogenous Coalition Structures,” *Games and Economic Behavior*, **26**, 286–336.

——— (2001), “Coalitional Power and Public Goods,” *Journal of Political Economy*, **109**, 1355–1384.

ROSENTHAL, R. (1972), “Cooperative Games in Effectiveness Form,” *Journal of Economic Theory*, **5**, 88–101.

RUBINSTEIN, A. (1982), “Perfect Equilibrium in a Bargaining Model,” *Econometrica*, **50**, 97–109.

SALANT, S., S. SWITZER and R. REYNOLDS (1983), “Losses from Horizontal Merger: The Effects of an Exogenous Change in Industry Structure on Cournot–Nash Equilibrium,” *Quarterly Journal of Economics*, **93**, 185–199.

SEIDMANN, D. (2009), “Preferential Trading Arrangements as Strategic Positioning,” *Journal of International Economics*, **79**, 143–159.

SEIDMANN, D. and E. WINTER (1998), “Gradual Coalition Formation,” *Review of Economic Studies*, **65**, 793–815.

SELTEN, R. (1981), “A Non-Cooperative Model of Characteristic Function Bargaining,” in *Essays in Game Theory and Mathematical Economics in Honor of Oscar Morgenstern* (V. Böhm and H. Nachtkamp (eds.)), Mannheim: Bibliographisches Institut.

SENGUPTA, A. and K. SENGUPTA (1996), “A Property of the Core,” *Games and Economic Behavior* ,**12**, 266–73.

SHAPLEY, L. (1953), “Open Questions” in *Report of an Informal Conference on the Theory of N-Person Games*, ed. H.W. Kuhn, mimeo, Princeton University.

——— (1973), “Let’s Block “Block”,” *Econometrica*, **41**, 1201–1202.

SHENOY, P. (1979), “On Coalition Formation: A Game Theoretic Approach,” *International Journal of Game Theory*, **8**, 133–164.

SHUBIK, M. (1983), *Game Theory in the Social Sciences*, Cambridge, MA: MIT Press.

STÅHL, I. (1977), “An n-Person Bargaining Game in the Extensive Form,” in *Mathematical Economics and Game Theory* (R. Henn and O. Moeschlin (eds.)) Berlin: Springer-Verlag.

STOLE, L. AND J. ZWIEBEL (1996), “Intra-firm Bargaining under Non-binding Contracts,” *Review of Economic Studies*, **63**, 375–410.

SUTTON, J. (1986), “Non-Cooperative Bargaining Theory: An Introduction,” *Review of Economic Studies*, **53**, 709–724.

THRALL, R. and W. LUCAS (1963), “n-Person Games in Partition Function Form,” *Naval Research Logistics Quarterly*, **10**, 281–298.

VOHRA, R. (1999), “Incomplete Information, Incentive Compatibility, and the Core,,” *Journal of Economic Theory*, **86**, 123–147.

VON NEUMANN, J. and O. MORGENSTERN (1944), *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press.

WINTER, E. (1993), “Mechanism Robustness in Multilateral Bargaining,” *Theory and Decision*, **40**, 131–47.

——— (2002), “The Shapley Value,” in *Handbook of Game Theory, Volume 3*, ed. by R. J. Aumann, and S. Hart, 2025–2054. North-Holland.

XUE, L. (1998), “Coalitional Stability under Perfect Foresight,” *Economic Theory*, **11**, 603–627.

XUE, L. AND L. ZHANG (2011), “Bidding and sequential coalition formation with externalities,” *International Journal of Game Theory*, **41**, 49–73.

YI, S-S. (1996), “Endogenous Formation of Customs Unions under Imperfect Competition: Open Regionalism is Good,” *Journal of International Economics*, **41**, 153–177.

ZHAO, J. (1992), “The Hybrid Solutions of an n-Person Game,” *Games and Economic Behavior*, **4**, 145–160.