Coalitional Power and Public Goods

Journal of Political Economy (2001)

Debraj Ray and Rajiv Vohra

The problem of inefficiency associated with the provision of public goods has traditionally been viewed as stemming from lack of information.

The theory of mechanism design attempts to resolve this problem by constructing mechanisms that provide agents the appropriate incentives to implement efficient allocations.

But the construction of a satisfactory mechanism solves the free-rider problem only if we assume that the agents are bound by such a mechanism.

Can it solve the free rider problem if coalitions of agents can reject a proposed (efficient) mechanism and use mechanisms of their own?

To give cooperation and efficiency the best chance of emerging in equilibrium, assume

- Complete information - (mechanism design is trivial)
- Agents can make binding agreements for contributions to the pure public good as well as side-payments to each other.

This is the context in which frictionless, Coasean bargaining is sometimes believed to lead to efficiency.

Suppose a Lindahl allocation is proposed for such an economy.

Is it possible that a coalition may be able to do better for its members by seceding from the grand coalition and selecting some feasible allocation for itself?

One (old) answer: NO
because a Lindahl allocation belongs to the 'α-core'. It is not possible for a coalition to make an improvement over the Lindahl allocation if this results in those outside the coalition to contribute 0. The latter assumption is implicit in the notion of the α-core.

Roberts (1974):

> "the simple adaptation of the definition of the core which has proven appropriate for private goods economies may not be suitable with public goods economies..... However, the task of developing an alternative core definition (or some other formalization of the intuitive notion of social stability) which better recognizes the structure of the public goods problem is a very delicate one. Guaranteeing that solution allocations will exist in a significant class of economies proves to be a particularly difficult problem".

The main difficulty in studying coalitional behavior in a model with a (pure) public good arises from the fact that the payoffs to a coalition depend critically on the behavior of agents outside the coalition – externalities.

Our notion of an equilibrium coalition structure is based on:

- The formation of a coalition implies a binding agreement among its members. But any such agreement must be voluntary - unanimous consent.

- Once a coalition is formed it cannot change its composition.

- Each coalition makes a prediction (which in equilibrium will be correct) of the contributions of agents in the complementary coalition. The actions of the complement will be based on similar considerations. Some or all of the remaining agents may form a coalition of their own, again predicting the behavior of the remainder, and so forth.

Our solution concept will imply that agents within a coalition maximize surplus and achieve efficiency. Inefficiency can, therefore, only arise if the grand coalition of all agents fails to form.

Full cooperation refers to the formation of the grand coalition.

No cooperation refers to the coalition structure of singletons – Nash equilibrium of a game of voluntary contributions.

Thus an equilibrium coalition structure describes an endogenously determined collection of cooperative agreements within each coalition in the coalition structure.

A Simple Model of Pollution Control

n identical players – interpreted as regions, firms or countries.

Each region chooses how much to contribute towards a pure public good - pollution control.

The production of z units of (public) benefit requires cost $c(z)$. Assume $c(.)$ is strictly increasing and strictly convex.

If Z is the total amount of pollution control produced by all regions, then the payoff to a region that produces z:

$$Z - c(z). \tag{1}$$

A region may make an offer to write a binding agreement with some other regions. If all the regions (to which the offer is made) agree on this arrangement, a coalition of regions forms, which is then bound to jointly decide on the extent of its control activity.

The agreement may specify not just production levels, but also transfers across coalitional members (these may be proposed, in principle, to induce some regions to sign up).

Since c is strictly convex each region produces the same level of pollution control.

Suppose each member of a coalition of size s produces z. Then the average payoff to each member of this coalition is:

$$sz - c(z) + Z_{-i}.$$

The optimization problem for a coalition of size s is equivalent to:

$$\max_z sz - c(z). \tag{2}$$

The solution is independent of the actions of players outside this coalition – in terms of production decisions each coalition has a dominant strategy. (Of course, the payoff to each coalition does depend on the actions of the regions outside the coalition.)

An Illustrative Example

For this example, assume that within each coalition the total surplus is divided equally. This is a common assumption in the literature, e.g. Bloch (1996) and Alesina and Spolaore (1997). But our main results will not rely on this – we will allow for transfers.

Given the symmetry of the model, equal division means that a coalitional agreement boils down to a proposal regarding the number of partners that a region seeks.

A numerical coalition structure specifies the sizes of the various coalitions in a partition or coalition structure of the n regions, e.g. $(s_1, s_2, \ldots, s_m)$, where s_i is a positive integer for each $i = 1, \ldots m$ and $\sum_{i=1}^{m} s_i = n$.

Suppose the numerical coalition structure is $\mathbf{n} = \{s_1, s_2, \ldots, s_m\}$. Let $z(s_j)$ denote the per member production level in coalition S_j. Then the average payoff to players in coalition S_i is:

$$\alpha(s_i, \mathbf{n}) = s_i z(s_i) - c(z(s_i)) + \sum_{j \neq i} s_j z(s_j).$$

Suppose the cost function is quadratic: i.e.,

$$c(z) = .5z^2.$$

The maximization problem for a coalition of size s is now

$$\max_z sz - .5z^2$$

and the solution is

$$z(s) = s.$$

Thus, the average payoff is

$$\begin{aligned}\alpha(s_i, \mathbf{n}) &= s_i z(s_i) - c(z(s_i)) + \textstyle\sum_{j \neq i} s_j z(s_j) \\ &= .5s^2 + \textstyle\sum_{j \neq i} s_j^2.\end{aligned}$$

$$\alpha(s_i, \mathbf{n}) = .5s^2 + \sum_{j \neq i} s_j^2.$$

This describes the outcome for a given coalition structure.

Our main aim is to predict the equilibrium coalition structure.

To see what is involved, we now consider economies of varying size.

$$\alpha(s_i, \mathbf{n}) = .5s^2 + \sum_{j \neq i} s_j^2.$$

coalition size: s	2	3	4	5	6	7	8
direct payoff: $.5s^2$	2	4.5	8	12.5	18	24.5	32

$n = 2$. In the grand coalition, the per capita payoff is 2; $\alpha(2, (2)) = 2$.
If the two regions do not form one coalition (and play a Nash equilibrium),

$$\alpha(1, (1, 1,)) = 0.5 + 1 = 1.5$$

Since we are assuming that cooperation is feasible, it is in everyone's interest to form the grand coalition – first best efficiency.

Note: this is a 'dominant strategy' for 2 regions even if there are others that have left the game.

$$\alpha(s_i, \mathbf{n}) = .5s^2 + \sum_{j \neq i} s_j^2.$$

coalition size: s	2	3	4	5	6	7	8
direct payoff: $.5s^2$	2	4.5	8	12.5	18	24.5	32

$n = 3$. $\alpha(3, (3)) = 4.5$.

Now, if a single regions goes alone it must predict what the remaining two regions will do.

$$\alpha(1, (1, 2)) = .5 + 4 = 4.5$$
$$\alpha(1, (1, 1, 1)) = .5 + 2 = 2.5$$

But we know from the previous step that the former prediction is the right one, resulting in the payoff 4.5.

What about forming a two-region coalition?

$$\alpha(2, (1, 2)) = 2 + 1 = 3 < 4.5.$$

The average payoff cannot be higher than that in the grand coalition.

Again, the equilibrium is (n).

$$\alpha(s_i, \mathbf{n}) = .5s^2 + \sum_{j \neq i} s_j^2.$$

coalition size: s	2	3	4	5	6	7	8
direct payoff: $.5s^2$	2	4.5	8	12.5	18	24.5	32

$n = 4$. $\alpha(4, (4)) = 8$.
By the previous step, if one region decides to stay alone, it should predict that the remainder will stay together, and

$$\alpha(1, (1, 3)) = .5 + 9 = 9.5.$$

This means that the equilibrium coalition structure is not the grand coalition.

It is also easy to see that it is not $(2, 2)$.

It is $(1, 3)$.

The equilibrium coalition structure in this case is $(1, 3)$.

$$\alpha(n, (n)) = 8, \qquad \alpha(1, (1, 3)) = 9.5, \qquad \alpha(3, (1, 3)) = 5.5.$$

While the singleton earns 9.5, the average payoff in the three-region coalition is 5.5 - less than that in the grand coalition. The equilibrium outcome is not first best efficient.

Note: the singleton's (rational) decision to stay alone is based on the (correct) prediction that the rest will stay together. For example, $(\alpha(1, (1, 1, 1, 1)) = 3.5$.

$$\alpha(s_i, \mathbf{n}) = .5s^2 + \sum_{j \neq i} s_j^2.$$

coalition size: s	2	3	4	5	6	7	8
direct payoff: $.5s^2$	2	4.5	8	12.5	18	24.5	32

$n = 5$. $\alpha(5, (5)) = 12.5$.

By the previous step, if one region were to stand alone, it would not be able to ensure the stability of the remaining four; they will configure themselves into $(1, 3)$.

$$\alpha(1, (1, 1, 3)) = 0.5 + 1 + 9 = 10.5.$$

Forming a 2, 3 or 4 region coalition also yields less than the average worth of the grand coalition.

Thus the equilibrium is (n) – the grand coalition reappears!

After this, a direct computation becomes intractable.

But our main characterization result allows us to conclude that in this example, after 5, full cooperation does not appear until $n = 8$. And then not again until $n = 13$.

So, efficiency prevails in economies of size 2, 3, 5, 8, 13, . . . – not always.

There are two features that are quite general:

- A coalition structure of singletons — and more generally, a symmetric inefficient coalition structure — is never an equilibrium. The reason is that the symmetric per-region payoff to the grand coalition would always dominate the payoffs to such a structure. Thus if there is inefficiency (as in the four-region case of the example), the equilibrium structure must be asymmetric.

- As n increases, the degree of inefficiency cannot become too large. So while there may be some inefficiency as the number of regions grows, there can't be too much.

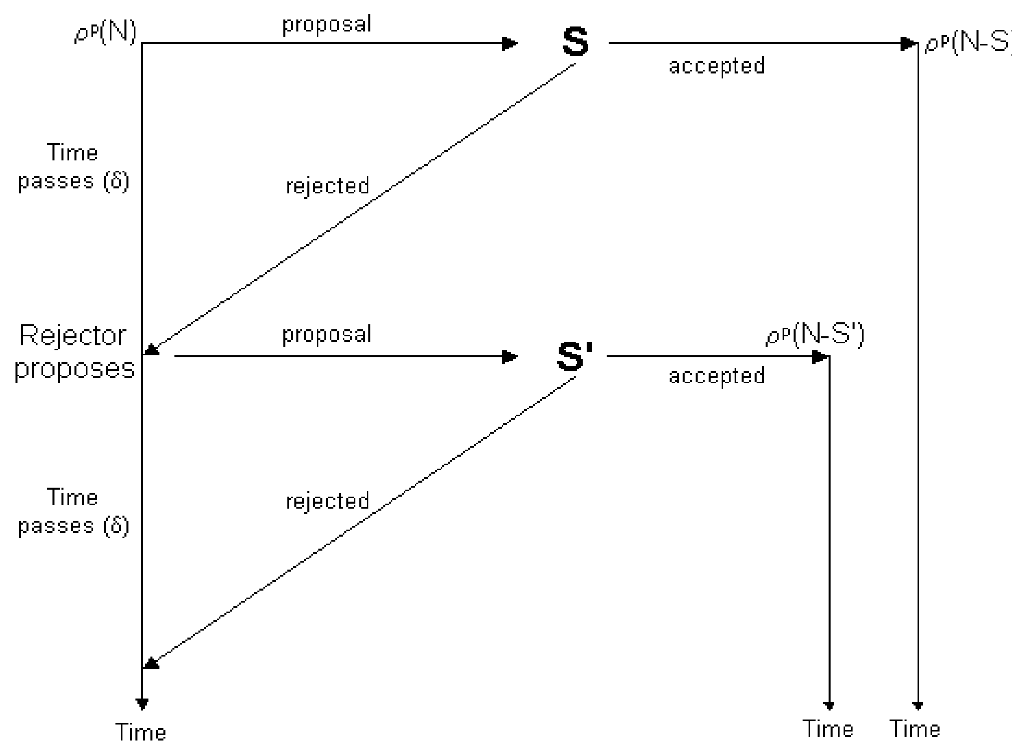
The General Model of Coalitional Bargaining

As in Ray and Vohra (1999): coalitional bargaining over a partition function (externalities).

Earlier literature:

Chatterjee, Dutta, Ray and Sengupta (1993): coalitional baragining without externalities.

Rubinstein (1982): two-person bargaining.



$\rho^P(N)$
proposal
S
accepted
$\rho^P(N-S)$
Time passes (δ)
rejected
Rejector proposes
proposal
S'
$\rho^P(N-S')$
accepted
Time passes (δ)
rejected
Time
Time
Time

A (stationary) strategy for a region will condition its (possibly probabilistic) proposal only on the current state of the game — the current set of negotiating regions and the coalitions that have already formed. It also requires that the accept-reject decision for proposals made by other regions to it not depend on anything else but the current set of regions, the coalitions that have already left, as well as the identity of the proposer and the nature of the proposal.

A stationary (perfect) equilibrium is defined to be a collection of stationary strategies such that there is no history at which a region benefits by a deviation from its strategy.

Our question is: can we describe the coalition structure(s) that must arise under a stationary perfect equilibrium?

Notation and assumptions:

Assume that there is a unique solution to (2) for all s and strictly increasing in s.

Denote by $z(s)$ the solution to (**??**) and let

$$\begin{aligned} f(s) &= sz(s), \\ h(s) &= c(z(s)), \\ g(s) &= f(s) - h(s). \end{aligned}$$

The average worth of a coalition of size s in a coalition structure $\mathbf{n}$ is:

$$\alpha(s, \mathbf{n}) = \sum_j f(s_j) - h(s), \qquad (1)$$

Assume that $f(.)$ is convex.

Decompositions of Positive Integers

Let $\mathcal{T} = \{m_1, m_2, \ldots\}$ be an ordered collection of increasing positive integers, where $m_1 = 1$. For any integer $n \geq 2$ define the $\mathcal{T}$-decomposition of n as a collection $\mathrm{s}(n) \equiv (t_1, \ldots, t_k)$ of (possibly repeated) elements of $\mathcal{T}$ satisfying the following properties:

(1) t_k is the largest integer in $\mathcal{T}$ which is strictly smaller than n.

(2) For any $i \in \{1, \ldots k-1\}$, t_i is the largest integer in $\mathcal{T}$ no greater than $(n - \sum_{j=i+1}^{k} t_j)$.

Since $1 \in \mathcal{T}$, $\mathrm{s}(n)$ is well-defined and unique for any $n > 1$.

For example, if $\mathcal{T} = \{1, 2, 3, 5, 8, 13, ..\}$, then

$$s(10) = (2, 8), \qquad s(12) = (1, 3, 8).$$

Now consider a special collection $\mathcal{T}^* = \{m_1, m_2, \ldots\}$ of positive integers with the property that $m_1 = 1$, and for each $i \geq 1$, m_{i+1} is the smallest integer n, with the property that $n > m_i$ and

$$g(n) \geq \alpha(t_1, \mathbf{s}(n)) = \sum_{i=1}^{k} f(t_i) - h(t_1) \qquad (4)$$

where $(t_1, \ldots, t_k)$ is the $\mathcal{T}^*$ decomposition of n.

$\mathcal{T}^*$ is unique and computable recursively.

For any positive integer n, define its strict decomposition to be just its $\mathcal{T}^*$-decomposition, and its decomposition to be either its strict decomposition if $n \notin \mathcal{T}^*$, or the singleton set $\{n\}$ if $n \in \mathcal{T}^*$. The notation $\mathbf{d}(n)$ (resp. $\mathbf{s}(n)$) will refer to the decomposition (resp. strict decomposition) of n. For notational convenience, the decomposition of 0 is empty.

Observation 1 *If* $\mathbf{d}(n) = (n_1, \ldots n_k)$*, where* $k \geq 2$*, then* $n_i \neq n_j$ *for all* $i, j \in \{1, \ldots k\}$*,* $i \neq j$*. If* $\mathcal{T}^* = \{m_1, \ldots m_i, m_{i+1}, ..\}$*, then* $m_1 = 1$*,* $m_2 = 2$ *and* $m_{i+1} < 2m_i$ *for all* $i \geq 2$*.*

We characterize the equilibrium (numerical) coalition structures of our bargaining game for discount factors "close to" unity.

Theorem 1 *Fix a set of regions n. Then there exists $\delta^* \in (0, 1)$ such that for all $\delta \in (\delta^*, 1)$, the bargaining process yields a unique numerical coalition structure, which is just the decomposition of n.*

The proof involves an application of a uniqueness result from the more general model of partition function bargaining developed in Ray and Vohra (1999).

Combining Observation 1 and Theorem 1 we see that the equilibrium of this (ex-ante) symmetric game typically displays a high degree of asymmetry:

Corollary 1 *Suppose the number of regions is* n *and the grand coalition does not form in equilibrium, i.e., the equilibrium coalition structure is* $\mathbf{d}(n) = (n_1, \ldots, n_k)$*, where* $k \geq 2$*. Then* $n_i \neq n_j$ *for all* $i, j \in \{1, \ldots k\}$ *and* $n_k > n/2$*.*

How Much Inefficiency?

Efficient outcomes are to be had only through the formation of the grand coalition; only a binding agreement among the set of all regions can cause all the effects of provision to be fully internalized.

Yet it should also be clear that there cannot be too much inefficiency. If a high degree of inefficiency were anticipated in equilibrium, then some region would move to make a suitable proposal to the grand coalition, and all the regions, fearing a huge loss in the event that the proposal falls through, would accept.

Corollary 1 has an obvious implication for computing a lower bound on the degree of inefficiency: in the decomposition of n there must be one coalition of size at least $n/2$. This coalition generates a per-capita surplus that is no less than $g(n/2)$, while every other coalition enjoys a per-capita surplus of at least $f(n/2)$ (the output of the largest coalition). This, in turn, is at least as large as $g(n/2)$, so that the ratio of equilibrium to potential surplus is at least

$$\frac{g(n/2)}{g(n)}.$$

Theorem 1

[1] *For each n, let $e(n)$ denote the ratio of equilibrium to potential surplus. Then*

$$e(n) > \frac{4}{3}\frac{g(n/2)}{g(n)}.$$

[2] *If k is the number of equilibrium coalitions, then*

$$k < \log_2 n + 1.$$