# When Transaction Costs Restore Efficiency: Coalition Formation with Costly Binding Agreements

Zsolt Udvari *

JOB MARKET PAPER

October 29, 2018

For the most recent version please click here

**Abstract**

Establishing binding agreements is often costly in real world economies. The costly nature of these agreements decreases the gains from cooperation and affects which agreements form by changing the incentives of agents, potentially leading to different equilibrium outcomes. However, economic theory often assumes away from these costs or associates them with a negative impact on the surplus of agents as they reduce the gains from cooperation. In this paper I explore the implications of costs associated with binding agreements on equilibrium agreement structures. Using an alternating offers bargaining model of coalition formation I show that surprisingly, the presence of transaction costs can lead to an efficient outcome in situations where inefficiency arises in equilibrium without these costs. These results provide new insights for policies targeting transaction costs.

**Keywords**: Contracts and Organizations, Transaction Costs, Externalities, Coalition Formation, Bargaining

**JEL Classification**: C71, C78, D23, D62, D71, D86

*Department of Economics, Boston University, 270 Bay State Road, Boston, MA, USA, 02215.

Email: zsoltudv@bu.edu, Web site: https://sites.google.com/view/zsoltudvari

# 1 Introduction

Binding agreements are widespread in everyday life. Economic activities requiring the collaboration of multiple people are often regulated by legally enforceable contracts between the participants, such as employment contracts or agreements between firms specifying a transaction. The purpose of these agreements is to ensure that the collaborating parties act in a way that is collectively beneficial for them as a group, preventing situations where agents seek to maximize their own benefits, disregarding the interests of others. There are numerous situations where these contracts are not available for free: for instance, the contracting parties have to hire and pay a lawyer to ensure that the correct legally enforceable contract is written.

Various fields of Economics have different approaches regarding binding agreements and the costs associated with them. Economic theory in general considers the presence of contracting costs to be harmful for the overall surplus of agents. Non-cooperative game theory and its applications are usually based on the assumption that parties cannot even make binding agreements. In cooperative game theory and in the theory of coalition formation, while the fundamental assumption is that binding agreements are feasible, typically the costs of establishing them are not modeled explicitly. In this paper I build a model where binding agreements are feasible and costly, and I show that in a wide range of situations the presence of agreement costs *improves* the total surplus of all agents.

Costs associated with binding agreements have a significant impact on the formation of agreements as agents' incentives change, resulting in potentially different agreement structures in equilibrium compared to a costless environment. This paper investigates how the costs of establishing binding agreements influence the negotiation about entering into contracts and the efficiency of the resulting outcomes. In such a setting a natural hypothesis is that contracting costs lead to efficiency problems. Although the efficient contracts would be written in an environment where agreements are free, as these costs reduce the gains from cooperation, agents fail to reach the efficient outcome in a costly environment. While this hypothesis is correct in some settings, surprisingly the opposite phenomenon is also possible: *costly binding agreements may help reaching the efficient outcome when efficiency is not reached in the absence of these costs.* These interesting cases are the focus of this paper.

Basic economic intuition suggests that the presence of costs related to establishing or enforcing binding agreements has a negative impact on the economy, as these costs decrease the gains from the economic activity specified by the agreement. Since the costs of establishing binding agreements are not directly related to production or any kind of economic activity, these costs are essentially *transaction costs*. The "Coase Theorem" (originating from Coase (1960)), one of the best known ideas in Economics, states that in the absence of transaction costs agents always reach an efficient outcome - an outcome that maximizes the total surplus across all agents - via negotiation. According to this argument, transaction costs serve as an obstruction to negotiation, and if they are sufficiently high, parties may fail to reach the surplus-maximizing outcome through bargaining. There is a large literature analyzing the effect of transaction costs on two-player Coasean bargaining and the consensus is that transaction costs reduce efficiency (see Anderlini and Felli (2001, 2006), Bolton and Faure-Grimaud (2010) and Lee and Sabourian (2007) among others).

When the logic of the Coase Theorem is applied to the formation of agreements among agents, it is expected that players reach the surplus-maximizing outcome when they negotiate without transaction costs. In addition, sufficiently high transaction costs impede negotiating and does not allow parties to reach efficiency. While this Coasean logic is accurate in some settings, as suggested above, the opposite phenomenon can also happen.

In situations with more than two agents, even in the absence of transaction costs, it is possible that agents fail to establish the contracts leading to the highest overall surplus. This phenomenon is already known in the coalition formation literature, see for example Ray and Vohra (1997, 2001), Diamantoudi and Xue (2007) or Hyndman and Ray (2007). This paper shows that paradoxically, the presence of transaction costs may restore the surplus-maximizing outcome when agents do not reach it in a costless environment. This phenomenon is quite the opposite of the spirit of the Coase Theorem. In a Coasean world, the only effect transaction costs can have is to obstruct negotiating partners from reaching the efficient outcome that would arise in a frictionless setting. I show that this is not always the case: in some situations, the presence of transaction costs leads to the efficient outcome which would never be reached in a setting free of transaction costs.

The existence of surplus-increasing transaction costs has important policy implications. Despite the common belief among economists, in some situations the welfare-improving action regarding transaction costs is to keep them high. In this paper I show that under some conditions, an environment with lower transaction costs is not necessarily desirable, as it can lead to social welfare loss when the formation of small groups is a potential issue. Therefore, when policy-makers decide about policies targeted at the reduction of transaction costs, a more careful approach is necessary and industry structures should be taken into account.

The intuition and the mechanism through which the presence of costs associated with establishing agreements restores efficiency is different based on the source of the inefficiency arising without transaction costs. I present two conceptually different types of situations where coalition formation leads to an inefficient outcome via costless bargaining, and I show how the presence of transaction costs helps restoring efficiency.

The first source of inefficiency I consider is when agents establish agreements to maximize their joint payoff disregarding the payoffs of others outside of their agreement. Even if the efficient outcome - where the combined surplus of all agents is maximal - is a single contract among all agents, it is possible that the surplus per capita is higher for a specific contract within a smaller group. In this situation agents have an incentive to form that smaller group and ensure themselves higher payoffs than they could expect in the efficient outcome. In the presence of transaction costs this incentive is weaker as the transaction cost "taxes" the gains from excluding others from the agreement, therefore it can help reaching the efficient outcome. It is important to note that the efficient outcome is also subject to the same transaction cost. However, since the total surplus is higher in the efficient outcome, the same cost results in a lower relative loss.

The second possible source of inefficiency occurs in settings with externalities among contracting groups. In these situations the well-being of an agent does not only depend on the contract she establishes, but also on what agreements others, who are not part of the agent's group, form. A notable example is free-riding in public good provision, as analyzed in Ray and Vohra (2001). Inefficiency due to free-riding arises because, even if it is known that some players will be free-riders and do not contribute to public good provision, the rest of the players are still better off if they make a binding agreement specifying a high level of contribution in order to maximize their own payoff. Due to the non-excludable consumption of public goods, the contributing players increase the free-riders' payoff even more than their own as a side effect. Free-riders, in some sense, are "forcing" other players to form these binding agreements on high contribution by declaring that they will not contribute. Introducing transaction costs makes the formation of agreements harder for the contributing players, therefore free-riders can no longer expect others to contribute. For this reason, when deciding about whether to free-ride or join a contributing group, the potential free-riders rather choose to contribute. That is, the presence of transaction costs can prevent free-riding despite that free-riders themselves are not subject to these costs.

This paper is organized as follows. Section 2 introduces some important concepts and provides a review of the related literature. Section 3 defines the coalition formation game I use in my analysis. Then I turn to the two different situations described above where transaction costs can restore efficiency. First, in Section 4 I analyze games without externalities among coalitions; then in Section 5 I study situations with externalities. In Section 6 I discuss the influence of some important assumptions on my results. Section 7 concludes the paper.

# 2 Background

In this section I start by introducing some important concepts related to binding agreements among groups and I provide different possible interpretations for the general transaction cost term used throughout the paper. Then I summarize the related literature.

## 2.1 Coalition formation

This paper studies how transaction costs affect the formation of agreements among agents. Agents establish these agreements to formalize the conditions of an economic activity that is beneficial to every participant. A natural framework to analyze such situations is cooperative game theory.

Cooperative game theory focuses on how the members of coalitions divide the coalitional value - the surplus of the coalition - taking the coalition structure as given. In coalition formation models agents negotiate with each other about entering into binding agreements, therefore the resulting coalition structure is an equilibrium of an explicitly or implicitly modeled bargaining game. Coalitions are groups of players that maximize the joint surplus of the entire group. A binding agreement among the members of a coalition ensures that players will indeed act in a way that increases the joint surplus, and not seek to maximize their own benefits instead.

Traditionally, coalition formation models do not explicitly separate the costs of establishing or enforcing coalitional agreements from the value derived from them. However, in economic or political applications of coalition theory it is very important to decompose the value achieved via cooperation from the costs of establishing or maintaining the coalition. The reason why an explicit modeling of these costs is important is that there are situations where the costs associated with agreements change while the economic or political activity specified by the agreements are unaffected. Moreover, governments or other institutions are often capable of influencing transaction costs via regulation.

To decide whether a policy that modifies transaction costs associated with establishing agreements is desirable or not, we need to understand how people adjust their decisions about entering into contracts in the new environment. This paper provides a method to predict the consequences of these policies.

## 2.2 Costly binding agreements

Costs associated with binding agreements have several potential interpretations. First, they can be interpreted as *contracting costs* which are simply the monetary costs of writing the contract. Another possible interpretation is that the costs of binding agreements are *enforcement costs* related to enforcing the actions specified by the agreement, such as the distribution of surplus. These costs can be also interpreted as fees of a supervisor actively monitoring that the contracting parties keep their end of the bargain. Alternatively, these costs can arise due to difficulty of coordination among agents. Depending on the analyzed situation, the search for potential coalition partners can also be a source of agreement costs.

Although the possible interpretations are numerous, the costs associated with binding agreements have two defining properties in my model. First, these costs arise only if there is cooperation among at least two agents. An agent who acts on its own without cooperating with anyone else - in the terminology of cooperative game theory, a player in a singleton coalition - is never affected by these costs. The second property is that these costs are not directly related to the economic activity specified by the agreement. For example, consider a contract between an upstream and downstream firm specifying the delivery of goods of a given quantity and quality. The fee of the lawyer writing the contract does qualify as an agreement cost, while the price of fuel, wage of the truck driver and any costs directly related to the delivery are not agreement costs. In summary, costs associated with binding agreements reduce the gains from potential agreements without directly affecting the economic activity specified by these agreements.

## 2.3 Related literature

I use a cooperative game theory framework to build a model that shows the potential efficiency benefits of transaction costs in the formation of agreement structures among groups. The theory of cooperative games originates from von Neumann and Morgenstern (1944). Games in partition function form, are a specific class of cooperative games first introduced by Thrall and Lucas (1963). The model introduced in the next chapter is based on this class of games.

There is a growing literature on coalition formation, see Ray and Vohra (2015) for a survey. The papers most closely related to my model are Bloch (1996) and Chatterjee et al (1993) as they also use an extensive form bargaining game to determine the outcome coalition structures. Similarly to this paper, Ray and Vohra (1997), Hyndman and Ray (2007) and Diamantoudi and Xue (2007) also emphasize the efficiency dimension of coalition formation outcomes and find that the coalition formation negotiation process does not always lead to an efficient outcome in the absence of transaction costs.

Several other papers use a coalition formation framework to analyze public good provision games, most notably Ray and Vohra (2001), Furusawa and Konishi (2011). Dixit and Olson (2000) and Ellingsen and Paltseva (2016) use a somewhat different negotiation framework but has a similar inefficiency result as Ray and Vohra (2001) and this paper.

The effect of transaction costs on Coasean negotiation is extensively discussed in the literature. Anderlini and Felli (2001, 2006), Bolton and Faure-Grimaud (2010) and Lee and Sabourian (2007) assume different types of transaction costs affecting negotiation and find that the presence of transaction costs generally causes efficiency problems. White and Williams (2009), Mackenzie and Ohndorf (2013) and Robson and Skarpedas (2008) shows that costly enforcement of property rights leads to potential inefficiency.

There are multiple core ideas in the Organization Economics literature that are closely related to this paper. Legros and Newman (1996, 2013) are based on the idea that due to some non-contractible production decisions, firms that are willing to cooperate with each other have to hire a professional manager to ensure the efficiency of the cooperation.

This argument uses the same logic as the starting point of this paper: binding agreements are not available by default, there is a cost associated with them, possibly due to a third party not participating in the economic activity for which the coalition is formed. Legros et al (2018) uses the organizational framework described above to provide a model of endogeneous market structure.

# 3 Coalition Formation with Costly Binding Agreements

In this section I introduce the model I use to analyze which binding agreements arise in an environment with transaction costs. I use a sequential coalition formation game similar to Chatterjee et al (1993) and Bloch (1996). I define the formal model I use in my analysis, then I discuss the main assumptions imposed.

The presence of externalities between coalitions is an important feature of my model. Games of partition function form are cooperative games where a value of a coalition depends on how the rest of the players are partitioned into coalitions. For example in the case of three players, the value of a singleton coalition can be different when the other two players are in a two player coalition or are in separate singleton coalitions. Therefore, in order to capture externalities between players in my model, the values of coalitions are given by a partition function. Throughout the paper I assume that all players are symmetric in a sense that all coalitions of the same size have the same value, that is, the payoffs only depend on the size of the coalition but not on the identity of its members.

**Definition 1.** Let be $N = \{1, .., n\}$ the set of players with $n > 2$. The cooperative game $(V, N)$ of partition function form is a function $V$ defined on pairs of $S \subseteq N$ and $\pi \in \Pi(N)$ where $\Pi(N)$ denotes the set of possible partitions of $N$. The value of each coalition $S$, if the current partition is $\pi$, is given by $V(S, \pi) \in \mathbb{R}$, $S \in \pi$.

$V$ is symmetric if for all $\pi$ ,$\pi'$ and $S \in \pi$, $S' \in \pi'$ we have $V(S, \pi) = V(S', \pi')$ as long as $|S| = |S'|$ and $|\pi| = |\pi'|$ where $|\pi| = (|S_1|, ..., |S_k|) = (s_1, ..., s_k)$ with numbers arranged into a descending order.

The collection $(s_1, ..., s_k)$ is referred as *numerical coalition structure* (Ray and Vohra, 1999).

The assumption that $n > 2$ is maintained for all games analyzed in this paper as in the framework I use two player coalition formation problems are trivial.

In this paper I focus on a specific class of partition function games where the total surplus is the highest when the grand coalition forms. This property is called *cohesiveness*. That is, since I focus on cohesive games, the efficient outcome - where the total surplus of all players is maximal - is always the grand coalition.

An important special case of the cooperative game defined above is the *characteristic function* game where there are no externalities between coalitions. That is, the value of a coalition depends only on the members - or in the symmetric case only on the size of the coalition - and does not depend on the coalition structure formed by the rest of the players. In this case the value function reduces to $v : 2^N \to \mathbb{R}$ and cohesiveness is equivalent to superadditivity (that is, for all disjoint $S, T \subset N$ we have $v(S \cup T) \geq v(S) + v(T)$). Section 4 will focus on this special case.

The outcome coalition structure $\pi$, that assigns values to coalitions using the function $V$ defined above is determined by a noncooperative alternating offer bargaining game in the spirit of Rubinstein (1982), which is a common framework in the coalition formation literature using the non-cooperative approach, see Ray and Vohra (2015).

The process of the game is the following. At the beginning of the game no players are assigned to any coalitions. At the start of the game, a *protocol* randomly selects a player to be the first proposer. The role of the protocol is the same as the role of Nature in games with incomplete information. The selected player $i$ makes an offer to a set of players not yet assigned to any coalition to form coalition $S$. All players in $S$ (not including player $i$) answer this offer in a randomly determined order by accepting or rejecting it. If everyone accepts the offer, coalition $S$ forms and the players in $S$ leave the game. Then the game continues with $N \setminus S$ as the set of players, and a new proposer is picked randomly.

If there is a player in $S$ that rejects the offer, then $S$ does not form and all players return to the pool of players without coalitions. The game then continues with the protocol randomly selecting a new proposer from this pool. The game ends when every player is assigned to coalitions. The difference between a player in a singleton coalition and a player not yet assigned into coalitions is important.

Similarly to Bloch (1996), I assume that there is no discounting between the rounds of bargaining, and if the bargaining game continues infinitely players receive a payoff of zero. The intuition behind these assumptions is the following: I model economic situations where binding agreements are necessary to engage in a long-term economic activity creating the coalitional value. Once binding agreements are formed, the activity continues indefinitely, making the time spent on bargaining negligible as long as the bargaining process ends in finite rounds.

Note that coalitional agreements are assumed to be irreversible in a sense that once a player is assigned to a coalition, she can no longer receive another offer to be a part of a different coalition instead. This irreversibility assumption is crucial for the results presented in Section 4 and Section 5. I discuss the implications of relaxing this assumption in Section 6.

Another important assumption is that the coalitional surplus is divided equally. If the surplus is divided equally, then the offers made by players during the bargaining game simply contain the proposed coalition, there is no need to specify a distribution of coalitional surplus in the offer.

The formal definition of the coalition formation game is given below:

**Definition 2.** The coalition formation game $(V, N, N^*, \pi_{-N^*}, \Sigma, \rho)$ consists of the following:

- $N = \{1, ..., n\}$ is the set of players, $n > 2$
- $N^*$ is the set of players not yet assigned to a coalition, $N^* = N$ at the beginning of the game
- $\pi_{-N^*}$ is a partition of $N \setminus N^*$
- $V$ is a symmetric partition function
- $\sigma_P \in \Sigma_P : (N^*, \pi_{-N^*}) \to \Pi(N^*)$ is a strategy of the proposing player
- $\sigma_R \in \Sigma_R : (\Pi(N^*), \pi_{-N^*}) \to \{\text{Accept, Reject}\}$ is a strategy of a responding player
- $\rho$ is a protocol selecting a random player in $N^*$ to be the proposer if currently there is no proposing player

When the protocol selects a player to be the proposer, the player chooses a subset $S$ of $N^*$ including the player herself according to her strategy $\sigma_P$. If there are other players in this selected subset, they have to choose whether to Accept or Reject the offer to form coalition $S$. If all players choose Accept, $S$ is formed and $N^* \setminus S$ becomes the new $N^*$.

If $N^* = \emptyset$, all players receive payoffs according the following rule: for all players $i \in S \in \pi$, $u_i(S,\pi) = \frac{V(S,\pi)}{|S|}$.

The outcome of a game is a coalition structure $\pi$ that is a partition of $N$. Throughout the paper I will focus on outcomes rather than equilibrium strategies as I am interested in what coalition structures form. Note that the strategies of players are *stationary* as they do not depend on histories, only on payoff-relevant information such as the coalitions already formed, the set of players that are still in the game and the current proposal.

Due to the externalities captured by the partition function, when players decide whether to form a specific coalition $S$ they have to consider how the remaining players are going to organize themselves into coalitions. This is modeled by having $\sigma_P$ and $\sigma_R$ dependent on both $\pi_{-N^*}$ and $N^*$, the coalitional structure formed by players that are already in coalitions and the set of players yet to form into coalitions, respectively.

Similarly to Bloch (1996), Ray and Vohra (1997) and Kóczy (2007), I assume that the players can make a rational prediction about the other players' actions, therefore the equilibrium concept is (stationary) subgame perfect equilibrium in the sequential game defined above. In the rest of the paper I will use the notation $\sigma(V,N)$ to denote the sequential coalition formation game where the payoffs are given by the cooperative game $(V,N)$.

This paper modifies the framework defined above by introducing *transaction costs* to the model. These costs are assigned to coalitions and they simply decrease the value of the given coalition.

It is possible that larger coalitions are subject to higher transaction costs, therefore transaction costs are non-decreasing in the size of the coalition. Outside of this monotonicity, no further structure is assumed about the costs in this paper. Similarly to the value of the coalition, the transaction cost depends only on the size of the coalition and it is independent of which players are in that given coalition. Singleton coalitions are not subject to transaction costs since they do not need a binding agreement ensuring that they maximize the coalition's surplus instead of their own personal profit as the coalitional surplus coincides with the individual benefit.

A coalition formation game with costly binding agreements adds one more element to the game defined in Definition 2: a vector $\tau = \{t_1, t_2, ..., t_n\}$ with $t_i \leq t_j$ for all $i \leq j$. The $i$-th element of the vector represents the transaction cost that has to be paid by any coalition with $i$ players in it. The transaction cost for singleton coalitions, $t_1$ is always equal to zero. Due to transaction costs, the payoff of player $i$ in coalition $S$, when partition $\pi$ is formed, changes to the following:

$$u_i(S,\pi) = \frac{V(S,\pi) - t_{|S|}}{|S|}.$$

The assumption that transaction costs are non-decreasing in coalition size implies that $t_i \leq t_j$ for all $i < j$. I will use the notation $(V_t, N)$ and $\sigma(V_t, N)$ to refer to games $(V, N)$ and $\sigma(V, N)$ augmented with the vector of transaction costs $t$.

In the next two sections I show how introducing transaction costs changes the equilibrium outcome of coalition formation games and how it can help restore efficiency.

# 4 Games without externalities

There are many real-world situations where there are gains from cooperating with others. The problems I analyze in this section have the feature that cooperation is beneficial for all participating parties and the efficient outcome is the one where all players choose to cooperate, that is, the grand coalition of all players forms. Furthermore, the activity of a given coalition does not affect agents outside of that group. Examples of this type of games are situations where the joint value originates from technological synergies - such as economies of scale - or from the provision of excludable (for example, local) public goods.

This section focuses on situations where while the most efficient outcome is the grand coalition, it is not possible to divide the surplus of the grand coalition in a way that each possible combination of players gets at least as high payoff as they could ensure for themselves in a smaller coalition. These games represent the first possible reason why the formation of coalitions can lead to an inefficient outcome: a subset of players refuses to participate in the efficient grand coalition if they can achieve a higher payoff in a smaller coalition. However, this deviation reduces the total surplus of all players.

First I look at the equilibrium outcome of this type of games in the absence of transaction costs and show that for a class of games, bargaining without transaction costs leads to an inefficient equilibrium. Then I point out how introducing transaction costs can restore efficiency while increasing the total surplus of the players.

## 4.1 No transaction costs

As a higher degree of cooperation is beneficial for each player, it seems plausible that without transaction costs, players reach the most efficient outcome. This is certainly true when there are only two players because there is only one possible contract between agents. However, games with three or more players open the possibility to multiple potential contracts among players. In these situations the efficient outcome will be reached in equilibrium only if the marginal returns to cooperation are either constant or increasing as further players join the coalition. Instead, if there are decreasing marginal gains from cooperation, even in the absence of transaction costs, it is no longer guaranteed to reach the efficient outcome through a coalition formation game described in the previous section. This paper shows that in that case it is possible that the presence of transaction costs helps reaching the efficient outcome.

To demonstrate a situation where agents fail to reach the efficient outcome in the absence of transaction costs consider the following scenario. There are three manufacturers at the same location, operating in the same industry. The manufacturers can produce separately, but they can also choose to horizontally integrate with one or two other manufacturers. Integration is beneficial due to economies of scale: the total surplus of an industry structure consisting of two integrated firms and a single firm is higher than the combined surplus of three single firms; and the surplus produced by the three-firm integration is higher than the combined surplus of the two-firm integration and one single firm industry structure. However, the gains from integration are higher when moving from producing alone to operating as a two-manufacturer integration than the gains from moving to the full integration from the two-firm integration. Note that in this situation gains from integration, as there are no externalities among firms, are purely technological, there are no market power effects. This game can be captured by the following numerical example.

**Example 1.**
Consider a game with three players. The singleton coalition has a surplus of 20, the two-player coalition has a surplus of 70 and the grand coalition has a surplus of 102. The efficient outcome is the grand coalition since its total surplus, 102, is higher than 90 or 60, the total surplus when the numerical coalition structure is (2,1) and the combined surplus when all players are in singleton coalitions, respectively. However, the equilibrium of the bargaining game described in Section 3 leads to an outcome with a two-firm integration

and a single firm because the payoff players can expect from the grand coalition is 34, while in a two-player coalition they can get a payoff of 35. Therefore, when the first player makes her proposal, she offers the possibility of a two player coalition to one of the players with an equal split of the surplus, and the proposed player accepts it.

In this example the efficient outcome is not reached in equilibrium because the two players in the small coalition maximize their own benefits instead of the joint surplus, and they are better off when deviating from the efficient outcome.

Farrel and Scotchmer (1988) studies three-player games similar to the example above and proposes a solution to these kind of problems by promoting one of the players to a "ringleader" who has some power to capture a part of the surplus, without sharing it with the other players. According to their result, if the ringleader has enough power, the efficient outcome forms. Note that the distribution of the payoffs will be asymmetric as the ringleader takes a high portion of the total surplus. In this paper I propose a different solution to this problem that preserves the symmetry of players.

## 4.2 Introducing transaction costs

Now I introduce a transaction cost to Example 1. Running the horizontal integration of multiple manufacturing firms requires a professional manager who charges a fee for her services. Assume that this fee is equal to 9. Introducing this transaction cost changes the surplus available for players in the two and three firms coalitions to 61 and 93 respectively. Now the equal division of the surplus in the grand coalition gives 31 to each player, while the two player coalition gives only 30.5. Therefore, there is no incentive any more to form a two firm coalition as it is no longer possible to get higher payoff than in the efficient outcome, hence the resulting equilibrium outcome is the efficient grand coalition.

It is important to point out that the transaction cost restores efficiency despite that both the grand coalition and the frictionless equilibrium structure (2,1) are subject to the cost. As the same cost has a relatively higher effect on the players' payoff in the frictionless outcome compared to the efficient outcome, players' incentives change and it becomes desirable to form the grand coalition. As a result, the two-player coalition is no longer advantageous for the player making the first proposal.

In addition, the total payoff of all players is higher in the presence of transaction costs even if we account for the cost itself. This property implies that the expected payoff of a player is higher in that game as well, therefore ex ante every player is strictly better off when there are transaction costs. That is, if players can choose which game they want to play before the order of proposers is drawn, they all prefer the game where cooperation is costly compared to the one when forming coalitions is free. The interpretation of this result in previous the manufacturing industry example is the following: before the game starts and players can choose managers that operate the integrated firm and work for free or managers who work for a strictly positive wage, *they prefer to pay the manager instead of getting her services for free.*

Note that this result is *not* achieved in a setting where transaction cost eliminates the inefficient equilibrium outcome by discriminatively targeting it and making it more costly. Instead, the efficient outcome is subject to the same transaction cost. Moreover, even in cases where the transaction cost is slightly higher for the grand coalition (up to 12 compared to the 9 associated with the two player coalition), the same result still holds with the efficient outcome being the unique equilibrium and the presence of transaction costs is preferable by the players ex ante. Intuitively, the inefficient outcome is no longer an equilibrium because the total surplus is lower in that case, therefore the same transaction cost feels more costly from the point of view of a given player.

In summary, when cooperation is not associated with additional costs, the efficient outcome is not reached since the deviating two players can be better off than they would be in the efficient outcome at the expense of the third player. However, if transaction costs are introduced to the model and cooperation is costly enough, the advantage of forming the two-player coalition disappears. Contrary to the traditional perception, instead of hindering the economy from reaching the efficient state, transaction costs are pushing the economy towards efficiency.

## 4.3 Surplus improving transaction costs for games without externalities

Now I formalize a general result regarding the situations described above. First I state the conditions when the absence of transaction costs leads to an inefficient equilibrium of the coalition formation game. In the case of symmetric superadditive characteristic function games these conditions are equivalent to the emptiness of the core. Then I characterize the cases when there exists a vector of *surplus-increasing transaction costs* that ensures the formation of the grand coalition in equilibrium, while still low enough to make the sum of all payoffs higher than in the frictionless game. The formal definition of surplus-increasing transaction costs are the following:

**Definition 3.** Let $(V, N)$ be a cohesive game where $N$ is not an outcome of a stationary SPE in $\sigma(V, N)$. Then, if there exists a vector $t = (0, t_1, ..., t_n)$ of transaction costs such that there is a stationary SPE in $\sigma(V_t, N)$ with $N$ as an outcome and for all $\pi^*$ arising as an outcome of $\sigma(V, N)$, we have

$$V(N) - t_n \geq \sum_{S \in \pi^*} V(S),$$

then $t$ is a vector of surplus-increasing transaction costs.

Note that the transaction costs defined in Definition 3 do not include all possible transaction cost vector $t$ that increase the total surplus of players. Surplus-increasing transaction costs are defined as transaction costs that both restore the efficient outcome $N$ and increase the total surplus of the players. To find out what games have potential surplus-increasing transaction costs, the first step is to identify the set of games that do not reach the efficient outcome in an equilibrium without transaction costs.

Definition 3 has an important implication: if the efficient outcome is reached in the absence of transaction costs, then the Coasean argument is valid and transaction costs indeed hurt the economy. The potential surplus-improving effect of transaction costs is originating from the fact that the efficient outcome is not always reached in an environment free of these costs.

In Example 1 the efficient grand coalition does not form because it is impossible to divide the value 102 in a way that any two players get at least 70 combined. Using the terminology of cooperative game theory, this feature means that the game has an empty core. Below I provide a formal definition of the core of a characteristic function game.

**Definition 4.** Let $(v, N)$ be a characteristic function game. The core of the game is the set $C(v, N)$ of vectors $x \in \mathbb{R}^n$ such that $\sum_{i \in N} x_i = v(N)$ and for all $S \subseteq N$,

$$\sum_{i \in S} x_i \geq v(S).$$

That is, the core is the set of possible distributions of the value of the grand coalition that guarantees every subcoalition to have at least as high payoff as they could earn if they formed the given subcoalition instead. If the core is nonempty - it is possible to divide the grand coalition's worth in a desirable way - then the grand coalition is expected to be stable. A natural question to ask is whether the grand coalition arises as the equilibrium of the sequential bargaining game when the core of the game is nonempty. Chatterjee et al (1993) shows that the statement is not true if the players are not symmetric. Here I show that the statement is true in the case of symmetric games.

**Proposition 1.** *Let $(v, N)$ be a symmetric characteristic function form game and $\sigma(v, N)$ is a coalition formation game with value function $v$ and player set $N$ where the core of $(v, N)$ is nonempty. Then there is a stationary SPE of $\sigma(v, N)$ that gives $N$ as outcome.*

*Proof.* Since the core of $(v, N)$ is nonempty, there is no coalition $S$ such that

$$\frac{v(S)}{s} > \frac{v(N)}{n}. \tag{1}$$

Condition (1) means that there is no coalition $S$ that is able to ensure higher average payoff to its members than the grand coalition. Given that, when player $i$ proposes to form $N$, for all other players $j \neq i$ it is an equilibrium strategy to accept it. By (1) it is clear that if any player declines the formation of $N$, she cannot expect higher payoff than $\frac{v(N)}{n}$, therefore there is no profitable deviation from accepting the offer to form $N$.

In addition, for any proposing player, when the set of remaining players is $N$, it is an equilibrium strategy to propose $N$ if the responders accept it. If the proposer proposes $N$ and the proposal gets accepted, the proposer receives a payoff of $\frac{v(N)}{n}$. Due to condition (1), $\frac{v(N)}{n}$ is the highest possible payoff a player can receive in the game, so no profitable deviation form proposing the grand coalition. □

The converse of Proposition 1 is also true: for all symmetric game $(v, N)$ such that there is a stationary SPE in $\sigma(V, N)$ such that the grand coalition is formed, then the core of the game must be nonempty (this implies that equal split of $v(N)$ is in the core).

**Proposition 2.** *Consider a symmetric superadditive characteristic function game $(v, N)$ where there is a stationary SPE in $\sigma(v, N)$ with $N$ as equilibrium outcome. Then, the core of $(v, N)$ is nonempty.*

I prove this proposition in the Appendix A.2.

Below I formulate that for every symmetric superadditive characteristic function game $(v, N)$ there exists a vector $t$ of transaction costs such that the core of $(v_t, N)$ is nonempty.

**Lemma 3.** *For every symmetric superadditive characteristic function game with empty core there is a cost $t$ associated with each non-singleton coalition such that the game $(v_t, N)$ has a nonempty core.*

The proof can be found in Appendix A.3.

Combining the results of Propositions 1, 2 and Lemma 3 leads to the following result.

**Corollary 4.** *Let $(v, N)$ b a symmetric superadditive game with characteristic functions where $N$ is not an outcome in any stationary SPE of $\sigma(v, N)$. Then, there exist a vector of transaction costs $t$ such that $N$ is the outcome of a stationary SPE of $\sigma(v_t, N)$.*

Corollary 4 ensures that the vector of transaction costs that restore $N$ as the outcome of $(v_t, N)$. However, it does not imply anything about the total surplus of players. The next result characterizes the class of games for which *surplus-increasing transaction costs* exist.

**Proposition 5.** *Let $(v, N)$ be a symmetric superadditive game with an empty core and let $\pi^*$ be the SPE of $\sigma(v, N)$ and $S^*$ is the coalition with highest average value in $\pi^*$ with $|S^*| = s$. If*

$$v(N) - \sum_{S \in \pi^*} v(S) \geq \frac{n \cdot v(S^*) - s \cdot v(N)}{n - s},$$

*then there is a surplus-increasing $t$.*

Proposition 5 is a direct consequence of Corollary 4 and Definition 3.

# 5 Games with externalities

This section analyzes situations with externalities among coalitions. There are numerous examples of these situations. Cartels and non-excludable public good provision exhibit *positive externalities*. Cartels are able to raise market prices in order to increase their revenues by colluding. However, firms outside of the cartel also benefit from the high market price. In public good provision settings, as the consumption is non-excludable, every individual benefits from the public good even if they do not participate in its production. Externalities are positive in these settings because the larger the cartel is, or the larger the group providing the public good is, the higher is the surplus of individuals *outside* of these groups.

In case of *negative externalities* this mechanism works backwards: the larger a given coalition is, the lower is the surplus of agents outside of the coalition. Political competition is a good example of negative externalities among groups.

In the remainder of this section first I apply the coalition formation framework defined in Section 3 to a problem of public good provision introduced by Ray and Vohra (2001). I use public good provision games to demonstrate how the formation of coalitions can lead to inefficient outcomes in the absence of transaction costs when there are externalities among players. I start by summarizing the main findings of Ray and Vohra (2001), then I show how the introduction of costly binding agreements affects the outcome predicted by the model and restores efficiency.

Following the public good provision application, I introduce some general results characterizing the existence of surplus-improving transaction costs in settings with positive or negative externalities.

## 5.1 Public good provision

Traditionally public goods are viewed as goods that cannot be efficiently provided by competitive markets due to the problem of free-riding. The reasoning is the following: in markets involving public goods no one can be excluded from consuming them regardless whether the consumers paid for them or not, which leads to free-riding problems. While Lindahl (1919) and Samuelson (1954) characterized the prices based on individual valuations that lead to efficient public good provision, in practice there are several problems that makes the implementation of Lindahl-Samuelson prices difficult.

One of these problems is that the agents' true valuation for the public good is private information, and agents are not willing to disclose it if they expect to be charged based on them. The economic literature usually focuses on mechanisms that are able to provide public goods efficiently, usually by proposing solutions for extracting the private information about the true valuation of the public good (see Clarke (1971) and Groves (1973) among others).

However, even if the valuations are common knowledge and the correct Lindahl-Samuelson prices can be determined, implementing them is a completely different problem. Agents still have the incentive not to contribute and free-ride. The actual payment of the Lindahl-Samuelson prices has to be forced by a government or a binding agreement among agents that specifies the contribution levels (potentially based on the Lindahl-Samuelson prices). In the example analyzed in this section agents are able to implement efficient levels of public good provision if they form coalitions where the members enter into an agreement that specifies contribution levels maximizing the joint surplus of the coalition. Ray and Vohra (2001) shows that a coalition formation game similar to Bloch (1996) can lead to inefficient provision of public goods.

Below I summarize the model of Ray and Vohra (2001) to analyze public good provision games with no transaction costs and to illustrate the inefficiency problem[1] in this setting. Then I introduce transaction costs to this model to show how the equilibrium outcome and its properties change.

### 5.1.1 No transaction costs

Let $N = \{1, .., n\}$ be the set of players. Each player $i$ has access to a technology to produce $z_i$ amount of public good at $c(z_i)$ cost, which is assumed to be convex in $z_i$. Every unit of the public good contributes to the payoff of all players, regardless of who produced the public good. The payoff of player $i$ is given by

$$u_i = Z - c(z_i),$$

where $Z = \sum_{i \in N} z_i$. Players can form coalitions among each other, and within coalitions they can make binding agreements such that the members of the coalition maximize the payoff of the entire coalition, not just the payoff of the given player. While the members of a coalition cooperate with each other, the cross-coalition interaction is noncooperative: players ignore the payoffs of any other player outside of their coalition. That is, a coalition $S$ of $s$ players solves the following maximization problem:

$$\max \sum_{i \in S} z_i - c(z_i).$$

Since $c(\cdot)$ is convex, the coalition will choose a production plan where each member produces the same quantity $z_S$. Therefore the maximization problem is essentially simplifies to

$$\max s \cdot z_S - c(z_S).$$

[1]Note that Ray and Vohra (2001) define two versions of this public good provision game. Here I refer to the version they label as "restricted game".

After each coalition $S$ made its respective production decision, the next step is to calculate the payoff of each player. The payoff of player $i$ in coalition $S$ is given by

$$u_i = s \cdot z_S - c(z_S) + \sum_{S' \neq S} s' \cdot z_{S'},$$

where $s'$ is the number of players in coalition $S'$.

The coalitions are formed as a subgame-perfect equilibrium of a bargaining game similar to the coalition formation game described in Section 3. There is a random order in which players not yet assigned to coalitions make offers to other players to form a coalition. The proposed players respond to the offer in a random order. If everyone accepts the offer, the coalition forms and the players leave the game. If a player refuses the offer, she becomes the next proposer. The game continues until each player is assigned to a coalition. The resulting coalitions decide about the amount of public good to be produced, and these production decisions determine the payoffs. Players are assumed to have a payoff of zero if the bargaining never ends.

When players decide about what kind of offer to propose or whether to accept or reject a particular offer, they make a rational prediction about which coalition structure the remaining players will form in later stages of the game.

To illustrate how this coalition formation game leads to an inefficient outcome, consider a case with $n = 4$ and $c(z) = \frac{z^2}{2}$. Table 1 table below lists the possible numerical coalition structures and the payoffs associated with them.

**Table 1: Public good provision with no transaction costs**

| Numerical coalition structure | Payoffs of players | | | |
|---|---|---|---|---|
| (4) | 8 | 8 | 8 | 8 |
| (3,1) | 5.5 | 5.5 | 5.5 | 9.5 |
| (2,2) | 6 | 6 | 6 | 6 |
| (2,1,1) | 4 | 4 | 5.5 | 5.5 |
| (1,1,1,1) | 3.5 | 3.5 | 3.5 | 3.5 |

As the Table 1 shows, the total payoff is the highest in the case when the grand coalition of all the four players forms, therefore that is the efficient outcome. However, the equilibrium of the game is the coalition structure (3,1). If the player making the first offer chooses to form a singleton coalition, the remaining players cannot achieve higher payoff than 5.5 in any possible numerical coalition structure, hence these players have no incentive to deviate from forming the three-player coalition after the first player left the game. Since this outcome gives the highest possible payoff for the first proposer, she has no incentive to deviate from this strategy either. Therefore (3,1) is an equilibrium coalition structure but it is not efficient.

Note that the player in the singleton coalition is free-riding: she produces the least amount of the public good of all players and has the highest payoff due to enjoying the benefits of the high level of provision by others.

The intuition behind this outcome is the following. The player who has the opportunity to make the first offer realizes that even if she free-rides, the remaining players cannot do better than cooperating with each other and producing a large amount of public good. By declaring that she contributes only the minimum amount, the first player "forces" the remaining players to a situation when the best they can do is to produce the highest possible amount of public good to maximize their own payoffs. However, due to the presence of externalities, this high level of public good provision benefits the free-riding player as well. Since the player in the singleton coalition bears lower cost than players in the three person coalition, the free-rider has a higher payoff than the other players.

### 5.1.2 Introducing transaction costs

This section shows how the presence of transaction costs change the outcome described above. In the previous example it was possible to enter into binding agreements within coalitions without any additional costs. Now consider a case when establishing binding agreements costs 0.3 for any non-singleton coalitions. Singleton coalitions are not subject to this transaction cost as there is no need of binding agreements in this case. Introducing this transaction cost modifies the payoffs as presented in Table 2:

Notice that in this game the unique equilibrium outcome is the grand coalition. Why is this outcome different from the frictionless game? Now if the player making the first offer decides to form a singleton coalition, the remaining players no longer have any incentive to form the three player coalition, as they had in the game without transaction costs. When the second player makes her offer, she will realize that if she also decides to form

**Table 2: Public good provision with transaction costs**

| Numerical coalition structure | Payoffs of players | | | |
|---|---|---|---|---|
| (4) | 7.925 | 7.925 | 7.925 | 7.925 |
| (3,1) | 5.4 | 5.4 | 5.4 | 9.5 |
| (2,2) | 5.85 | 5.85 | 5.85 | 5.85 |
| (2,1,1) | 3.85 | 3.85 | 5.5 | 5.5 |
| (1,1,1,1) | 3.5 | 3.5 | 3.5 | 3.5 |

a singleton coalition she will be better off than if she is in the three person coalition provided that the last two players are going to form the two person coalition. Since the last two players are indeed better off forming the two player coalition, the player making the second offer anticipates this move and declares to form its own singleton coalition if the first player chose to do so. That is, if the first proposer forms a singleton coalition, the resulting numerical coalition structure is going to be (2,1,1) compared to the (3,1) outcome without transaction costs.

Now the first player foresees this scenario and realizes that she ensures the maximal payoff to herself when she offers everyone to form the grand coalition. Since all players are able to make the same prediction and conclude that this outcome maximizes their payoff, they accept the offer.

In the game without transaction costs the efficient outcome (the grand coalition) failed to form in equilibrium, however in the game with transaction costs the equilibrium outcome is the efficient one. Similarly to the partnership games in Section 4, efficiency is attained not despite but because of the presence of transaction costs.

Moreover, the combined payoff of all players is higher in the game with transaction costs. In the first game the combined payoff is 26, in the second game it is 31.7. That is, if players have the opportunity to choose between the two regimes ex ante (before the identity of the first proposer is revealed), they strictly prefer the situation when making binding agreements is costly to the one when they are completely free.

It is also important to note that a special feature of efficiency restoring-mechanism in the previous example is that transaction costs help eliminate free-riding not because the free-rider is punished by the transaction cost. The free-rider is completely unaffected by the transaction cost because the cost is only paid by the players that are not free-riding.

The intuition behind this result is that in the case without transaction costs, the free-rider is able to guarantee himself a higher payoff than she would get in the efficient outcome. The reason is that even when she declares herself to be a free-rider, the remaining players still maximize their payoff when they form the three player coalition. However, if forming the three-player coalition is less lucrative due to the transaction costs, the remaining three players will opt out of forming it, making the free-rider worse off than she would be in the grand coalition.

The existence of the transaction costs decreases the power of the potential free-rider to force the remaining players into a position where their best available option is to produce the most possible amount of public good, which also helps the free-rider. Hence the presence of transaction costs eliminates free-riding, despite that only the contributing players are directly affected by the cost. The *indirect* effect of the costs restores efficiency and improves the surplus of all agents.

## 5.2 Surplus improving transaction costs for games with externalities

Below I characterize the circumstances where there are surplus-improving transaction costs for coalition formation problems with externalities, then I discuss important differences between direct and indirect surplus-improving transaction costs and implications to situations with positive and negative externalities.

To identify the circumstances when there are surplus-increasing transaction costs for games with externalities I introduce the notion of responsible coalitions.

**Definition 5.** Let $(V, N)$ be a cohesive game where $N$ is not formed in any of the stationary SPE of $\sigma(V, N)$. Then, coalition $S \subset N$ is *responsible for not forming* $N$ if $S$ is formed after the first proposal in $\sigma(V, N)$.

It is easy to see that for any $S$ responsible for not forming $N$

$$\frac{V(S,\pi)}{s} > \frac{V(N)}{n}$$

where $\pi$ is the stationary SPE of $\sigma(V, N)$, that is, the average payoff of the responsible coalition has to be higher than the average payoff of the grand coalition. This is the only reason why the members of $S$ are not willing to join the grand coalition which maximizes the total surplus of all players.

In order to be surplus-increasing, a vector of transaction costs $t$ has to directly or indirectly decrease the value of any possible responsible coalition $S$ such that in the resulting game $\sigma(V_t, N)$ with the new equilibrium outcome $\pi_t$

$$\frac{V(S, \pi_t)}{|S|} \leq \frac{V_t(N, N)}{n}.$$

Proposition 6 characterizes necessary and sufficient conditions for existence of surplus-improving transaction costs.

**Proposition 6.** *Let $(V, N)$ be a cohesive partition function game where $N$ is not the outcome in any stationary SPE of $\sigma(V, N)$. Then, there exists a surplus-improving transaction cost vector $t$ if and only if for all potential responsible coalition $S$ and for all $\pi_S$ that is an outcome of a stationary SPE in the game $\sigma(V, N)$ conditional on $S$ is formed after the first offer, we have*

$$\frac{V(S, \pi_t) - t_{|S|}}{|S|} \leq \frac{V(N, N) - t_n}{n} \tag{2}$$

*for all $\pi_t$ outcome of $\sigma(V_t, N)$ conditional on $S$ is formed after the first offer and*

$$V(N, N) - t_n \geq \sum_{T \in \pi_S} V(T, \pi_S). \tag{3}$$

Proposition 6 is proven in Appendix A.4.

The nature of surplus-increasing transaction costs can be quite different in games with externalities compared to the case with no external effects. Proposition 5 characterizes the cases when it is possible to achieve the condition above in the case of games without externalities. In those situations the transaction cost $t$ is able to restore $N$ by imposing a high enough cost on the responsible $S$ such that in the game with transaction costs the average payoff in $S$ is no longer above the average payoff of $N$. This is a *direct* method of restoring $N$ as an outcome using transaction costs.

This method of restoring efficiency does not always work in games with externalities. In Section 5.1 the coalition responsible for not forming the efficient outcome was a singleton. Since singleton coalitions are never subject to transaction costs, it is impossible to restore the grand coalition directly using transaction costs.

However, free-riding was only profitable if the numerical coalition structure $\pi = (3, 1)$ is formed in equilibrium. If the outcome is $\pi_t = (2, 1, 1)$ for some transaction cost vector

$t$, then the free-rider no longer has incentive to stay out of the grand coalition. The transaction cost vector $t = (0, 0.3, 0.3, 0.3)$ is able to restore $N$ as the outcome of the coalition formation game *indirectly*, without imposing any costs on the responsible free-rider by "breaking up" the three-player coalition and hence decreasing the payoff of the free-rider.

In summary, in games with externalities transaction costs can restore efficiency directly by imposing costs on coalitions that are not willing to join the grand coalition or indirectly through external effects by changing the reaction of players outside the deviating coalition.

There are two important differences between direct and indirect surplus-improving transaction costs. First, the existence of direct surplus-improving transaction costs requires that every potential responsible coalition has at least two members. Indirect costs do not impose this restriction. The second difference is that direct costs do not rely on changing the coalition structure outside of the responsible coalition $S$, while indirect transaction costs do. Changing the coalition structure outside of $S$ is not always possible: for example if $|N \setminus S| = 1$, the coalition structure outside of $S$ is impossible to change.

In games without externalities it is impossible to have indirect surplus-increasing transaction costs due to the lack of external effects among coalitions. In the general case it is uncertain what are the conditions that determine whether direct or indirect transaction costs will be able to restore efficiency in a given game. However, for games with positive or negative externalities there are some rules that help determine which type of transaction costs we have to look for. In the rest of this section I analyze games with specific types of externalities.

Definition 6 of games with positive externalities captures the idea that for every coalition $S$, if players outside of $S$ organize themselves into bigger coalitions, the value of $S$ increases. The public good example presented above has this property.

**Definition 6.** A game $(V, N)$ is a game with positive externalities if for all $\pi, \pi' \in \Pi(N)$ and all $S \in \pi$,

$$V(S, \pi) \leq V(S, \pi')$$

when $\pi' = (\pi \setminus (S_i, S_j), S_i \cup S_j)$, $S_i, S_j \neq S$.

As we have seen in Section 5.1, if there are positive externalities, then it is possible to have a singleton coalition to be responsible for not forming the efficient grand coalition. As singleton coalitions are never subject to transaction costs, there are no surplus-improving direct transaction costs.

On the other hand, it is generally simple to have indirect surplus-improving transaction costs: if the coalition structure has smaller coalitions, then the values of coalitions (including the one responsible for not forming the efficient outcome) are lower. That is, the indirect transaction cost has to break up coalitions outside of the deviating one to suppress the incentives to deviate. Breaking up coalitions by adding transaction costs to them is simple. Hence in situations with positive externalities indirect transaction costs are more likely to succeed in restoring efficiency and increasing the surplus of all players.

Contrary to the case with positive externalities, the presence of negative externalities implies that the value of $S$ will be lower if players outside of $S$ are in bigger coalitions.

**Definition 7.** A game $(V, N)$ is a game with negative externalities if for all $\pi, \pi' \in \Pi(N)$ and all $S \in \pi$,

$$V(S, \pi) \geq V(S, \pi')$$

when $\pi' = (\pi \setminus (S_i, S_j), S_i \cup S_j)$, $S_i, S_j \neq S$.

As discussed above, with positive externalities it is not always possible to have direct suprlus-improving transaction costs as there are situations where singleton coalitions are deviating from the grand coalition. This is not the case with negative externalities. In Proposition 7 I show that in a game with negative externalities no singleton coalition can be responsible for not forming $N$.

**Proposition 7.** *Let $(V, N)$ be a cohesive symmetric game with negative externalities where $N$ is never the outcome in a stationary SPE of $\sigma(V, N)$. Then, no coalition $S$ with $|S| = 1$ can be responsible.*

*Proof.* Assume that in a stationary SPE of $\sigma(V, N)$ with outcome $\pi$, the coalition responsible for not forming $N$ is $S$ with $|S| = 1$. Then,

$$V(S, \pi) > \frac{V(N, N)}{n}.$$

Now consider the partition $\pi'$ of $n$ singleton coalitions. Due to negative externalities,

$$V(S, \pi) \leq V(S, \pi').$$

This implies

$$n \cdot V(S, \pi') > V(N, N),$$

which contradicts the cohesiveness of $V$. □

Proposition 7 shows that games with negative externalities share the feature that no singleton coalition can achieve higher payoff than its payoff in the grand coalition with games without externalities (in that case this feature is a simple consequence of superadditivity). For this reason, it is always possible to have transaction costs that directly restore the formation of the grand coalition and the condition stated in Proposition 5 for characteristic function games is sufficient for the existence of surplus-improving transaction costs in games with negative externalities.

However, in the case of negative externalities the condition in Proposition 5 is no longer a necessary condition for the existence of surplus-improving transaction costs. Due to externalities, there is another way of restoring the efficient outcome besides imposing a transaction cost on the "deviating" coalition such that the average payoff will be lower than in the grand coalition. The presence of negative externalities implies that a given coalition's payoff decreases when larger coalitions form outside of that coalition. As shown in Example 1, the existence of transaction costs can result in formation of larger coalitions. That is, if there is a coalition $S$ responsible for not forming the grand coalition, and the set $N \setminus S$ of the remaining players is not organized into a single coalition, then it is possible to have indirect surplus-improving transaction costs. However if $N \setminus S$ forms a single coalition, then unlike in the case with positive externalities, indirect transaction costs cannot work.

Generally, contrary to the case with positive externalities, in games with negative externalities it is more likely to have a direct surplus increasing transaction costs than an indirect one.

# 6 Robustness of results

In the previous sections I analyzed coalition formation games using an alternating offers bargaining framework defined in Section 3. One of the important assumptions of this model is that binding agreements are irreversible, and once a set of players forms a coalition they leave the game and other players cannot propose them to form another coalition. In this section I discuss how the choice of the coalition formation model and the irreversibility assumption influences the results presented in my paper.

## 6.1 Alternative models of coalition formation

The characterization of situations with surplus-improving transaction costs in Section 4 and 5 does not depend on the sequential structure of the bargaining game described in Definition 2.

It is possible to replace the coalition formation model with one of the "blocking" approach. This direction of the coalition formation research does not use dynamic bargaining models to predict what coalitions arise in equilibrium. Instead, the blocking approach focuses on stable coalition structures where there are no profitable deviations by any group of players. Due to externalities, when considering a deviation it is crucial what reaction they expect from the rest of the players. There are several different models in the literature that use the blocking approach, differing in their assumptions about the reactions to deviations.

In the recursive core of Kóczy (2007), the reactions to deviations are "rational" in a sense that when players decide about what coalitions to form in response to a deviation, they choose outcomes where their payoffs are maximized. If the sequential bargaining model replaced with the recursive core, all of the previous results would still hold.[2] Overall, the sequential structure of the model is not important, what is crucial is the rationality - best response property - of the reactions.

## 6.2 Reversible agreements

Agreements in this paper are assumed to be irreversible, that is, once a set players agrees to be in a coalition $S$, they can no longer receive offers to form another coalition and cannot be selected as proposers.

The model described in Section 3 is not suitable to deal with situations with reversible agreements. In this case the feature that no payoff is received by any player before the negotiation process ends can easily lead to agents bargaining forever in equilibrium and never reaching an agreement.

[2]This is not surprising, see Kóczy (2009) for relations between the recursive core and Bloch (1996).

However, my model is designed to analyze the formation of long-term, thus irreversible agreements. Reversible, temporary agreements need a different theoretical framework. It is an interesting question whether the phenomenon of surplus-increasing transaction costs is unique to long-term agreements.

One direction of the coalition formation research analyzes temporary agreements, see for example Ray and Vohra (1999), Diamantoudi and Xue (2007) or Hyndman and Ray (2007). While the complete characterization of situations with surplus-increasing transaction costs in the case of temporary agreements is outside of the scope of this paper, here I present an example of a coalition formation game with temporary agreements by Ray and Vohra (2015). This game yields inefficiency in equilibrium and I show how the presence of transaction cost helps restore efficiency while all players are better off ex ante.

**Example 2.**

Consider the following version of the public good game presented in Section 5.1.

**Table 3: Reversible agreements with no transaction costs**

| Numerical coalition structure | Payoffs of players | | |
|---|---|---|---|
| (3) | 12 | 12 | 12 |
| (2,1) | 7 | 7 | 19 |
| (1,1,1) | 6 | 6 | 6 |

Ray and Vohra (2015) shows that in the equilibrium outcome of this game players will cycle between the numerical coalition structures (3), (2,1) and (1,1,1) and the average payoff of the players will be 9.67. Now consider a transaction cost of 2.4 applied to any non-singleton coalition. The payoffs now change to

Now similarly to the example in Section 5.1, the unique equilibrium outcome of the coalition formation game is (3). The coalition structure (2,1) never forms, not even temporarily. The reason is that if one player announces that she forms a singleton coalition, the remaining players also chose to form singleton coalitions to maximize their own payoffs. Therefore no player wants to form any other coalition outside of the grand coalition and this outcome stays stable in every period. The average payoff for each player is 11.2, which is higher than the 9.67 from the game without transaction costs.

**Table 4: Reversible agreements with transaction costs**

| Numerical coalition structure | Payoffs of players | | |
|---|---|---|---|
| (3) | 11.2 | 11.2 | 11.2 |
| (2,1) | 5.8 | 5.8 | 19 |
| (1,1,1) | 6 | 6 | 6 |

Based on the analysis of this section it can be concluded that the existence of a surplus-increasing transaction cost is quite a robust phenomenon and is not simply the product of the specific modeling choices made in this paper.

# 7 Conclusion

In this paper I analyzed coalition formation games with the presumption that binding agreements are costly. I used a simple model of coalition formation to show that despite the common intuition, the costly nature of binding agreement can be beneficial. I demonstrated this argument in two different settings.

First I considered simple superadditive games without externalities among coalitions. In these games there are always gains from cooperation so the efficient outcome is the grand coalition, however, for a given range of payoffs all equilibria of the coalition formation game is different from the efficient one. I showed that in some cases there is a range of strictly positive transaction costs that helps players reach the efficient outcome and the combined payoff of players is higher than in a game without transaction costs. Proposition 5 characterizes the conditions for situations where the efficient outcome is not reached without transaction costs due to a subset of players seek to maximize their own payoff at the expense of the players outside of their subset, however adding a strictly positive transaction cost is able to suppress the incentives leading to the inefficient outcome and restores efficiency.

Next I analyzed situations with externalities among coalitions. Similarly to the case without externalities, bargaining without transaction cost can lead to inefficient equilibria, while adding transaction cost can help restoring efficiency. In this setting, the intuition behind this result is different than it was for games with no externalities. Inefficiency in public good provision arises when potential free-riders can abuse the situation that even after they declared that they free-ride, other players are forced to contribute the highest amount to public good production in order to maximize their own payoff, substantially increasing the free-riders' payoff as a side effect. The presence of transaction costs makes contribution harder for non-free-riders, therefore takes away some power from the free riders to "force" the rest of the players into a situation that benefits free-riders the most. Surprisingly, positive transaction costs can reduce the potential gains from free-riding despite that free-riders themselves are not subject to transaction costs.

Finally I investigated the consequences of using alternative coalition formation models or dropping the assumption of irreversible agreements. I concluded that the phenomenon of surplus-increasing transaction costs is robust to these modeling choices.

The most important implication of the results concerns the approach towards transaction costs in general. The belief of harmful nature of transaction costs is widespread in Economics, therefore most economists encourage policies reducing transaction costs. However, the results of this paper suggest that an environment with lower transaction costs is not necessarily desirable, as it can lead to social welfare loss when the formation of small groups is a potential issue. Therefore, when policy-makers decide about policies targeted at the reduction of transaction costs, a more careful approach is necessary and the industry structures should be taken into account.

The indirect surplus-improving effect of transaction costs has another interesting policy implication. The fact that the increased transaction cost affecting the *contributors* are able to stop *free-riders* and restore the efficient outcome suggests that it is possible to regulate markets *indirectly* if there are external effects among firms, organizations or other groups of agents, . Consider an undesirable situation on a given market - such as extreme market power or a plant causing heavy pollution - but it is prohibitively costly to mitigate it via direct regulation due to technical or legal constraints. If the policy-maker succesfully identifies that this situation is the result of the interaction of agents in a given market and manages to change the environment in a way that under the new circumstances a more desirable outcome emerges as a result of interaction between agents, it is possible to regulate the situation while avoiding the prohibitive costs of direct intervention.

There is plenty of room for future research on the topic of this paper. Besides investigating the effects of transaction costs under different coalition formation models, another interesting extension of the model is a version when the transaction costs are endogeneous. For example, there is a set of players - lawyers, professional managers or supervisors - that provide binding agreements and the price is determined by the supply and demand of these services. Studying coalition formation with endogeneous agreement costs would lead to a better understanding of situations where government interventions targeted at the reduction of transaction costs are justified.

# References


Anderlini L, Felli L (2001) Costly bargaining and renegotiation. Econometrica 69:377–411

Anderlini L, Felli L (2006) Transaction costs and the robustness of the coase theorem. Economic Journal 116:223–245

Bloch F (1996) Sequential formation of coalitions in games with externalities and fixed payoff division. Games and Economic Behavior 14:90–123

Bolton P, Faure-Grimaud A (2010) Satisficing contracts. Review of Economic Studies 77:937–971

Bondareva ON (1963) Some applications of linear programming methods to the theory of cooperative games (in Russian). Problemy Kybernetiki 10:119–139

Chatterjee K, Dutta B, Ray D, Sengupta K (1993) A noncooperative theory of coalitional bargaining. Review of Economic Studies 60:463–477

Clarke EH (1971) Multipart pricing of public goods. Public Choice 11:17–33

Coase RH (1960) The problem of the social cost. Journal of Law and Economics 3:1–44

Diamantoudi E, Xue L (2007) Coalitions, agreements and efficiency. Journal of Economic Theory 136:105–125

Dixit A, Olson M (2000) Does voluntary participation undermine the coase theorem? Journal of Public Economics 76:309–335

Ellingsen T, Paltseva E (2016) Confining the coase theorem: Contracting, ownership, and free-riding. Review of Economic Studies 83:547–585

Farrel J, Scotchmer S (1988) Partnerships. Quarterly Journal of Economics 103:279–297

Furusawa T, Konishi H (2011) Contributing or free-riding? Voluntary participation in a public good economy. Theoretical Economics 6:219–256

Groves T (1973) Incentives in teams. Econometrica 41:617–631

Hyndman K, Ray D (2007) Coalition formation with binding agreements. Review of Economic Studies 74:1125–1147

Kóczy L (2007) A recursive core for partition function form games. Theory and Decision 63:41–51

Kóczy L (2009) Sequential coalition formation and the core in the presence of externalities. Games and Economic Behavior 66:559–565

Lee J, Sabourian H (2007) Coase theorem, complexity and transaction costs. Journal of Economic Theory 135:214–235

Legros P, Newman A (1996) Wealth effects, distribution and the theory of organization. Journal of Economic Theory 70:312–341

Legros P, Newman AF (2013) A price theory of vertical and lateral integration. Quarterly Journal of Economics 128:725–770

Legros P, Newman AF, Udvari Z (2018) Competing for the quiet life: An organizational theory of market structure, working paper

Lindahl E (1919) Just taxation – a positive solution. In: Classics in Theory of Public Finance, St. Martin's Press, 1967

Mackenzie IA, Ohndorf M (2013) Restricted coasean bargaining. Journal of Public Economics 97:296–307

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

Peleg B, Sudhölter P (2003) Introduction to the Theory of Cooperative Games. Kluwer

Ray D, Vohra R (1997) Equilibrium binding agreements. Journal of Econommic Theory 73:30–78

Ray D, Vohra R (1999) A theory of endogenous coalition structures. Games and Economic Behavior 26:286–336

Ray D, Vohra R (2001) Coalitional power and public goods. Journal of Political Economy 109:1355–1384

Ray D, Vohra R (2015) Coalition formation. In: Handbook of Game Theory, Elsevier, pp 239–326

Robson A, Skarpedas S (2008) Costly enforcement of property rights and the coase theorem. Economic Theory 36:109–128

Rubinstein A (1982) Perfect equilibrium in a bargaining model. Econometrica 50:97–109

Samuelson P (1954) The theory of public expenditure. Review of Economics and Statistics 36:386–389

Shapley L (1967) On balanced sets and cores. Naval Research Logistics Quarterly 14:453–460

Thrall RM, Lucas WF (1963) $n$-person games in partition function form. Naval Research Logistic Quarterly 10:281–298

White L, Williams M (2009) Bargaining with imperfect enforcement. RAND Journal of Economics 40:317–339

# A Appendix

## A.1 Balanced collections and the Bondareva-Shapley Theorem

To prove Proposition 2 and Lemma 3 I use the concept of balanced collections and apply the Bondareva-Shapley Theorem. For the Theorem, see Shapley (1967) and Bondareva (1963). In this Appendix I use the notation of Peleg and Sudhölter (2003) for balanced collections.

**Definition 8.** Let $N$ be a set with $|N| = n$. Then for any $S \subseteq N$, the characteristic vector of $S$ is $\chi_S \in \mathbb{R}^N$ such that

$$(\chi_S)_i = \begin{cases} 1, & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}$$

**Definition 9.** A collection $\mathcal{B}$ of subsets of $N$ is balanced if there exist positive balancing weights $\delta_S$ such that

$$\sum_{S \in \mathcal{B}} \delta_S \chi_S = x_N$$

**Theorem** (Bondareva-Shapley Theorem)**.** *Let $(v, N)$ be a characteristic function game. Then the core of the game is nonempty if and only if there is no balanced collection $\mathcal{B}$ such that*

$$\sum_{S \in \mathcal{B}} \delta_S v(S) > v(N).$$

## A.2 Proof of Proposition 2

*Proof.* If there is a stationary SPE in $\sigma(v, N)$ where the outcome is $N$, then there is no coalition $S \subset N$ such that (1) holds. If there is, then $N$ cannot arise in a stationary SPE as the proposer has an incentive to propose $S$ instead of $N$ and in that subgame the responders' best reply is to accept it. Now I show that if there is no coalition $S$ such that 1 holds, the core of the game must be nonempty.

The condition 1 implies that there is no coalition $S$ with size $|S| = s$ such that for a balanced collection $\mathcal{B}$ with weights $(\delta_S)_{S \in \mathcal{B}}$ consisting only coalitions of size $s$, we have

$$v(N) < \sum_{S \in \mathcal{B}} \delta_S v(S). \tag{4}$$

The facts that the game is symmetric and 4 holds for all coalition size $s$ implies that there is no balanced collection $\mathcal{B}$ with weights $\delta_S$ such that 4 holds. By the Bondareva-Shapley Theorem it means that the core of $(v, N)$ is nonempty. $\square$

## A.3 Proof of Lemma 3

*Proof.* By the Bondareva-Shapley Theorem, we know that for all games with an empty core there is a balanced collection $\mathcal{B}$ with a weight system $(\delta_S)_{S\in\mathcal{B}} \geq 0$ such that

$$v(N) < \sum_{S\in\mathcal{B}} \delta_S v(S). \tag{5}$$

First I show that there must be a balanced collection $\mathcal{B}$ with weight system $(\delta_S)_{S\in\mathcal{B}}$ satisfying 5 such that $\delta_S = 0$ for all $|S| = 1$. Due to superadditivity, there is no $(\mathcal{B}, \delta_S)$ system satisfying 5 where only singleton coalitions have positive weights. As a consequence, if there exists $(\mathcal{B}, \delta_S)$ satisfying 5, it must have positive weight on a coalition $S$ with $|S| = k > 1$. Now it is easy to see that for a balanced system $(\mathcal{B}', \delta'_S)$, where all $S \in \mathcal{B}'$ has $|S| = k$, we have

$$\sum_{S\in\mathcal{B}'} \delta'_S v(S) \geq \sum_{S\in\mathcal{B}} \delta_S v(S) > v(N).$$

That is, to prove the lemma it is enough to show that there is a cost $t$ such that

$$v(N) - t \geq \sum_{S\in\mathcal{B}'} \delta_S (v(S) - t) \tag{6}$$

is true for all $S$ with coalition size $s$. Given that $\mathcal{B}'$ only contains coalitions with a size of $k$, the balanced system $(\mathcal{B}', \delta'_S)$ consists of $\binom{n}{k}$ coalitions of size $k$ and each player is exactly in $\binom{n-1}{k-1}$ different coalitions, therefore the weights are $\frac{1}{\binom{n-1}{k-1}}$. Condition 6 takes the form

$$v(N) - t \geq \binom{n}{k} \cdot \frac{1}{\binom{n-1}{k-1}} (v(S) - t)$$

$$v(N) - t \geq \frac{n}{k} v(S) - \frac{n}{k} t$$

Even if $v(N) < \frac{n}{k} v(S)$ due to emptiness of the core, if

$$t \geq \frac{nv(S) - sv(N)}{n - s}, \tag{7}$$

then the core of the game $(v_t, N)$ is nonempty. $\square$

## A.4 Proof of Proposition 6

*Proof.* Note that a vector of transaction costs $t$ is surplus-improving if it restores the grand coalition as outcome and the total surplus in the outcome of game $\sigma(V_t, N)$ is higher than in the outcome of game $\sigma(V, N)$. It is easy to see that condition (3) is necessary and sufficient for the increase in total surplus. It is enough to show that condition (2) is necessary and sufficient for restoring $N$ as the outcome of the coalition formation game $\sigma(V_t, N)$.

For necessity, assume that (2) does not hold, that is,

$$\frac{V(S, \pi_t) - t_{|S|}}{|S|} > \frac{V(N, N) - t_n}{n}$$

for some $\pi_t$ outcome. This means that there is a SPE in $\sigma(V_t, N)$ where the outcome is $\pi_t$ conditional on the formation of $S$ after the first proposal such that member of $S$ have higher payoff than they would get in $N$. Therefore the members of $S$ still have an incentive to form $S$ instead of the grand coalition and $t$ does not restore $N$ as the outcome of $\sigma(V_t, N)$. This proves necessity.

For sufficiency, I show that if (2) holds, then there is a $t$ such that $N$ is the outcome of $\sigma(V_t, N)$. Assume that (2) holds and consider the vector $t$ of transaction costs such that

$$t_i = \begin{cases} 0 & \text{if } i = 1 \\ t_{|S|} & \text{if } 1 < i \leq |S| \\ t_n & \text{if } |S| < i \leq n \end{cases}$$

Now I show that there is no coalition $T$ such that

$$\frac{V(T, \pi_t) - t_{|T|}}{|T|} > \frac{V(N, N) - t_n}{n}. \tag{8}$$

Assume there is a coalition $T$ such that (2) holds. Due to symmetry and (2), $|T|$ cannot be equal to $|S|$. It is also impossible that $|T| > |S|$ since

$$\frac{V(T, \pi_t) - t_n}{|T|} > \frac{V(N, N) - t_n}{n}$$

implies

$$\frac{V(T, \pi_t)}{|T|} > \frac{V(N, N)}{n} + \left(\frac{t_n}{|T|} - \frac{t_n}{n}\right).$$

Since the term in brackets is positive, that implies that the average payoff in $T$ is higher than in $N$, which means that $T$ is a potential responsible coalitions. That contradicts the assumption that (2) holds for all potential responsible coalitions. $|T| < |S|$ is also impossible due to similar reasoning as above.

Therefore (2) is sufficient for restoring $N$ as an outcome of $\sigma(V_t, N)$. $\square$