PUBLIC GOOD AGREEMENTS UNDER THE WEAKEST-LINK TECHNOLOGY

ALEJANDRO CAPARRÓS
MICHAEL FINUS

CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

CCHS

Centro de Ciencias
Humanas y Sociales

IPP

INSTITUTO DE POLÍTICAS
Y BIENES PÚBLICOS

INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS – CSIC

Instituto de Políticas y Bienes Públicos (CSIC-IPP)
Consejo Superior de Investigaciones Científicas
C/ Albasanz, 26-28
28037 Madrid (España)

Tel: +34 91 6022300
Fax: +34 91 3045710

http://www.ipp.csic.es

Public Good Agreements under the Weakest-link Technology

March 2016

Alejandro Caparrós

Consejo Superior de Investigaciones Científicas (CSIC),
Institute for Public Goods and Policies (IPP),
Albasanz 26, 28037 Madrid, Spain.
E-mail: alejandro.caparros@csic.es

Michael Finus

University of Bath
Department of Economics
Bath, BA2 7AY, UK
E-mail: m.finus@bath.ac.uk

Abstract: We analyze the formation of public good agreements under the weakest-link technology. Coordination of and cooperation on migration policies, money laundering measures and biodiversity conservation efforts are prime examples of this technology. Whereas for symmetric players, policy coordination is not necessary, for asymmetric players cooperation matters but fails, in the absence of transfers. In contrast, with an optimal transfer scheme, asymmetry may not be an obstacle but an asset for cooperation, with even the grand coalition being stable. We characterize various types and degrees of asymmetry and relate them to the stability of agreements and associate gains from cooperation.

Key words: public goods, weakest-link technology, coalition formation
JEL classification: C71, C72, H41

1 Introduction

There are many cases of global and regional public goods for which the decision in one jurisdiction has consequences for other jurisdictions and which are not internalized via markets. Reducing global warming and the thinning of the ozone layer are examples in case. As Sandler (1998), p. 221, points out: "Technology continues to draw the nations of the world closer together and, in doing so, has created novel forms of public goods and bads that have diminished somewhat the relevancy of economic decisions at the nation-state level." The coordination of migration policies, the stabilization of financial markets, the fighting of contagious diseases and the efforts of non-proliferation of weapons of mass destruction have gained importance through globalization and the advancement of technologies.

A central aspect in the theory of public goods is to understand the incentive structure that typically leads to the underprovision of public goods as well as the possibilities of rectifying this. In this paper, we pick up the research question already posed by Cornes (1983), namely how cooperative institutions develop under different aggregation technologies, also called social composition functions. Among the three typical examples, summation, best-shot and weakest-link, we focus on the latter.[1] Weakest-link means that the benefits from public good provision depends on the smallest contribution. Examples include the classical example in Hirschleifer (1983) of building dykes against flooding, but also coordination of migration policies within the EU, compliance with minimum standards in marine law or enforcing targets for fiscal convergence in a monetary union, measures against money laundering, fighting a fire which threatens several communities, curbing the spread of an epidemic and maintaining the integrity of a network (Arce 2001 and Sandler 1998). Also protecting species whose habitat covers several countries is best described as a weakest-link public good.

For our analysis, we combine approaches from two strands of literature, which have developed almost independently: the literature on non-cooperative or privately provided public goods with a focus on the weakest-link technology and the literature on international environmental agreements (IEAs), which focuses exclusively on the summation technology. The IEA literature is an application of a broader literature on coalition formation in the presence of externalities where we focus on approaches belonging to non-cooperative coalition theory. We subsequently review these two strands of literature in section 2, set out our model in section 3 and derive some general results regarding the second (section 4) and first stage (section 5) of our two-stage coalition formation model, according to the sequence of backward induction. Since it turns out that the most interesting results are obtained for

---

[1] Better shot (weaker link) is a modification of the best shot (weakest link) technology where the marginal effect of an individual contribution on the global provision level decreases (increases) with the level of the contribution. For a formal exposition, see for instance Cornes (1993) and Cornes and Hartely (2007a,b).

the assumption of asymmetric players in the presence of transfers, we devote Section 6 to a detailed analysis on the type and degree of asymmetry which fosters stability and how this relates to the welfare gains from cooperation. Section 7 briefly checks the sensitivity of our results to alternative assumptions of equilibrium selection and Section 8 concludes. Along the way, we will argue that the results for coalition formation and the weakest-link technology are far more general than and different from those which have been obtained for the summation technology.

2 Relevant Literature

2.1 Public Goods and Weakest-Link

The first strand of literature on public goods has taken basically three approaches in order to understand the incentive structure of the weakest-link technology.

The first approach is informal and argues that the least interested player in the public good provision is essentially the bottleneck, which defines the equilibrium provision and which is matched by all others who mimic the smallest effort (e.g. Sandler and Arce 2002 and Sandler 2006). Moreover, it is argued that either a third party or the most well-off players should have an incentive to support the least well-off through monetary or in-kind transfers in order to increase the provision level.

The second approach is a formal approach (Cornes 1993, Cornes and Hartley 2007a,b, Vicary 1990, and Vicary and Sandler 2002). It is shown that there is no unique Nash equilibrium for the weakest-link technology, though Nash equilibria can be Pareto-ranked. It is demonstrated that except if players are symmetric, Nash equilibria are Pareto-inefficient. Improvements to this outcome are not considered in the form of coalitions but only by allowing monetary transfers between individual players. Because this changes players' endowments, it may also change their Nash equilibrium strategies as income neutrality does no longer hold (as this is the case under the summation technology). For sufficiently different preferences, this may increase the weakest player's provision level which may constitute a Pareto-improvement to all players. In some models (e.g. Cornes and Hartley 2007b and Vicary and Sandler 2002), which allow for different prices across players (the marginal opportunity costs in the form of foregone consumption of the private good), this is reinforced if the recipients face a lower price than the donor. In Vicary and Sandler (2002) it is also investigated how the Nash equilibrium provision level changes if monetary transfers are either substituted or complemented by in-kind transfers.

Finally, the third approach considers various forms of formal and informal cooperative

agreements, established for instance through a correlation device implemented by a third party, leadership and evolutionary stable strategies (e.g. Arce 2001, Arce and Sandler 2001 and Sandler 1998).

Our paper differs from this literature because it focuses on institution formation, and it improves upon this literature in three respects. Firstly, we combine a coalition formation model with general payoff functions and continuous strategies. Hence, our analysis of cooperation is not based on examples or simple matrix games (e.g. prisoners' dilemma, chicken or assurance games) with discrete strategies like the third approach, for which the generality of results is in doubt. Instead, we continue in the rigorous tradition of the second approach but consider not only Nash but also coalition equilibria. Secondly, we can measure the degree of underprovision not only in physical but also in welfare terms, allowing us to go beyond physical measures, like Allais-Debreu measure of waste, as used by Cornes (1993). Admittedly, this is easier in our TU-framework as equilibrium strategies are not affected by monetary transfers. Thirdly, our model allows not only for different marginal costs but also non-constant marginal costs of public good provision. However, in order to remain at a high level of generality, we do not consider in-kind transfers as some papers have done as they basically transform the weakest-link into a summation technology for which general results are difficult to obtain in the context of coalition formation.

2.2   International Environmental Agreements

The second strand of literature on IEAs can be traced back to Barrett (1994) and Carraro and Siniscalco (1993). This literature has grown quite substantially (see e.g. Battaglini and Harstad (forthcoming) for one of the most recent papers) since then, and the most influential papers are collected in a recent volume by Finus and Caparrós (2015) with an extensive survey. Within this literature, the non-cooperative approach is an application of a general theory of non-cooperative coalition formation in the presence of externalities as summarized in Bloch (2003) and Yi (1997). A general conclusion is that the size and success of stable coalitions depends on some fundamental properties of the underlying economic problem. It has been shown that problems can be broadly categorized into positive versus negative externalities (Bloch 2003 and Yi 1997). In positive (negative) externality games, players not involved in the enlargement of coalitions are better (worse) off through such a move. Hence, in positive externalities games, typically, only small coalitions are stable, as players have an incentive to stay outside coalitions. Typical examples of positive externalities include output and price cartels and the provision of public goods under the summation technology. If an output cartel receives new members, other players benefit from lower output by the cartel via

higher market prices. This is also the driving force in price cartels where the price increases with the accession of new members. In a public good agreement, players not involved in the expansion of a coalition benefit from higher provision levels but lower costs. In contrast, in negative externality games, outsiders have an incentive to join coalitions and therefore most coalition models predict the grand coalition as a stable outcome. Examples include trade agreements, which impose tariffs on imports from outsiders or R&D-collaboration among firms in imperfectly competitive markets where members gain a comparative advantage over outsiders if the benefits from R&D accrue exclusively to coalition members.

Until now non-cooperative coalition theory has mainly assumed symmetric agents due to the complexity which coalition formation adds to the analysis (see the surveys by Bloch 2003 and Yi 1997). In the context of positive externalities, general predictions about stable coalitions are difficult. It is for this reason that most papers on IEAs assume particular payoff functions and despite symmetry have to rely on simulations. Also for asymmetric agents hardly any analytical results have been obtained and the few exceptions assume particular functional forms and typically restrict the analysis to two types of players (e.g. Caparrós et al. 2011, Fuentes-Albero and Rubio 2010 and Pavolova and de Zeeuw 2013). Our paper differs from this literature in two fundamental respects. Firstly, none of the papers has investigated the weakest-link technology. Secondly, we demonstrate that for this technology much more general but also very different results can be obtained compared to the summation technology. We are able to characterize precisely the type and degree of asymmetry that is conducive for larger stable coalitions, which includes the grand coalition. In our conclusions (Section 8), we will argue that the simple coalition game we employ in this paper is sufficient to derive all interesting results as more complicated games would not add much to the analysis.

3   Model and Definitions

We consider the following payoff function of player $i \in N$:

$$\begin{aligned} V_i(Q, q_i) &= B_i(Q) - C_i(q_i) & (1) \\ Q &= \min_{i \in N} \{q_i\} \end{aligned}$$

where N denotes the set of players and Q denotes the public good provision level, which is the minimum over all players under the weakest-link technology. The individual provision level of player i is q_i. Payoffs comprise benefits, $B_i(Q)$, and costs, $C_i(q_i)$. Externalities across players are captured through Q on the benefit side.

In order to appreciate some features of the weakest-link technology, we will occasionally relate results to the classical assumption of a summation technology. The subsequent description of the model and its assumptions are general enough to apply to both technologies. For the summation technology, only $Q = \min_{i \in N} \{q_i\}$ has to be replaced by $Q = \sum_{j \in N} q_j$ in payoff function (1).[2] All important results of the summation technology mentioned in the course of the discussion are summarized in Appendix A.

Regarding the components of the payoff function, we make the following assumptions where primes denote derivatives.

Assumption 1: *For all $i \in N$: $B_i' > 0$, $B_i'' \leq 0$, $C_i' > 0$, $C_i'' > 0$. Furthermore, we assume $B_i(0) = C_i(0) = 0$ and $\lim_{Q \to 0} B_i'(Q) > \lim_{q \to 0} C_i'(q) > 0$.*

These assumptions are very general. They ensure the strict concavity of all payoff functions and existence of an interior equilibria as explained below. For the following definitions, it is convenient to abstract from the aggregation technology and simply write $V_i(q)$, stressing that payoffs depend on the entire vector of contributions, $q = (q_1, q_2, ..., q_N)$, which may also be written as $q = (q_i, \ q^{-i})$ where the superscript of q^{-i} indicates that this is not a single entry but a vector, comprising all provision levels except of player i, q_i.

Following d'Aspremont et al. (1983), the coalition formation process unfolds as follows.

Definition 1 *Cartel Formation Game In the first stage, all players simultaneously choose a membership strategy. All players who choose to remain outside coalition S act as single players and are called non-signatories or non-members, and all players who choose to join coalition S form coalition $S \subseteq N$ and are called signatories or members. In the second stage, simultaneously, all non-signatories maximize their individual payoff $V_j(q)$, and all signatories jointly maximize their aggregate payoff $\sum_{i \in S} V_i(q)$.*

Note that due to the simple nature of the cartel formation game, a coalition structure, i.e. a partition of players, is completely characterized by coalition S as all players not belonging to S act as singletons. The coalition acts like a meta-player, internalizing the externality among its members. The assumption of joint welfare maximization of coalition members implies a transferable utility framework (TU-framework). The cartel formation game is solved by backwards induction, assuming that players play a Nash equilibrium in each stage and hence a subgame-perfect equilibrium with respect to the entire game. In order to save on notation, we assume in this section that the second stage equilibrium vector for every

---

[2]More precisely, we mean a summation technology with equal weights, which we assume throughout the paper and therefore will not stress anymore.

coalition $S \subseteq N$ (denoted by $q^*(S)$ in Definition 2 below) is a unique interior equilibrium, even though this will be established later in Section 3.

Definition 2 *Subgame-perfect Equilibrium in the Cartel Formation Game*
(i) First Stage:
a) Assuming no monetary transfers in the second stage, coalition S is called stable if

$$\text{internal stability:} \quad V_i^*(S) \geq V_i^*(S \setminus \{i\}) \; \forall \; i \in S \; \text{and}$$
$$\text{external stability:} \quad V_j^*(S) \geq V_j^*(S \cup \{j\}) \; \forall \; j \notin S \; \text{hold simultaneously.}$$

b) Assuming monetary transfers in the second stage, coalition S is called stable if

$$\text{internal stability} \quad : \quad V_i^{*T}(S) \geq V_i^{*T}(S \setminus \{i\}) \; \forall \; i \in S \; \text{and}$$
$$\text{external stability} \quad : \quad V_j^{*T}(S) \geq V_j^{*T}(S \cup \{j\}) \; \forall \; j \notin S \; \text{hold simultaneously.}$$

(ii) Second Stage:
For a given coalition S that has formed in the first stage, let $q^*(S)$ denote the (unique) simultaneous solution to

$$\sum_{i \in S} V_i(q^*(S)) \geq \sum_{i \in S} V_i(q^S(S), q^{-S*}(S))$$
$$V_j(q^*(S)) \geq V_j(q_j(S), q^{-j*}(S)) \; \forall \; j \notin S$$

for all $q^S(S) \neq q^{S*}(S)$ and $q_j(S) \neq q_j^*(S)$.
a) In the case of no monetary transfers, equilibrium payoffs are given by $V_i(q^*(S))$, or $V_i^*(S)$ for short.
b) In the case of monetary transfers, equilibrium payoffs, $V_i^{*T}(q^*(S))$, or $V_i^{*T}(S)$ for short, for all signatories $i \in S$ are given by $V_i^{*T}(S) = V_i^*(S \setminus \{i\}) + \gamma_i \sigma_S(S)$ with $\sigma_S(S) := \sum_{i \in S}(V_i^*(S) - V_i^*(S \setminus \{i\}))$, $\gamma_i \geq 0$ and $\sum_{i \in S} \gamma_i = 1$ and for all non-signatories $j \notin S$ by $V_j^{*T}(S) = V_j^*(S)$.

Let us first comment on the second stage. Note that the equilibrium provision vector is a Nash equilibrium between coalition S and all the single players in $N \setminus S$. Only because of our assumption of uniqueness, we are allowed to write $V_i^*(S)$ instead of $V_i(q^*(S))$. As we assume a TU-game, monetary transfers do not affect equilibrium provision levels. Transfers are only paid among coalition members, exhausting all (without wasting any) resources generated by the coalition. Non-signatories neither pay nor receive monetary transfers. The

"all singleton coalition structure", i.e. all players act as singletons, subsequently denoted by $\{\{i\}, \{j\}, ... \{z\}\}$, replicates the non-cooperative or Nash equilibrium provision vector known from games without coalition formation. It emerges if either only one player or no player announces to join coalition S. By the same token, the grand coalition, i.e. the coalition which comprises all players, is identical to the socially optimal provision vector, sometimes also called the full cooperative outcome. Hence, our coalition game covers these two well-known benchmarks, apart from partial cooperative outcomes where neither the grand coalition nor the all singleton coalition structure forms. Moreover note that the monetary transfer scheme which we consider is the "optimal transfer scheme" proposed by Eyckmans and Finus (2004).[3] Every coalition member receives his free-rider payoff plus a share γ_i of the total surplus $\sigma_S(S)$, which is the difference between the total payoff of coalition S and the sum over all free-rider payoffs if a player i leaves coalition S. In other words, $\sigma_S(S)$ is the sum of individual coalition member's incentive to stay in ($\sigma_i(S) \geq 0$) or leave ($\sigma_i(S) < 0$) coalition S, $\sigma_i(S) := V_i^*(S) - V_i^*(S \setminus \{i\})$, which must be positive for internal stability at the aggregate, i.e. $\sigma_S(S) = \sum_{i \in S} \sigma_i \geq 0$. Thus, the transfer scheme has some resemblance with the Nash bargaining solution in TU-games, though the threat points are not the Nash equilibrium payoffs but the payoffs if a player leaves coalition S. The shares γ_i can be interpreted as weights, reflecting bargaining power. They matter for the actual payoffs of individual coalition members, but do not matter for the stability (or instability) of coalition S because stability only depends on $\sigma_S(S)$. Henceforth, when we talk about transfers, we mean transfers included in the class defined by the optimal transfer scheme.

Let us have now a closer look at the first stage. Note that internal and external stability defines a Nash equilibrium in terms of membership strategies. All players who have announced to join coalition S should have no incentive to change their announcement to stay outside S (internal stability) and all players who have announced to remain outside S should have no incentive to announce to join S instead, given the equilibrium announcements of all other players. Due to the fact that the singleton coalition structure can always be supported as Nash equilibrium in the membership game if all players announce to stay outside S (as a change of the strategy by one player would make no difference), existence of a stable coalition is guaranteed. We denote a coalition which is internally and externally stable and hence stable by S^*. In the case of the monetary transfer scheme considered here, it is easy to see that, by construction, if $\sigma_S \geq 0$, then coalition S is internally stable and if $\sigma_S < 0$, then neither this transfer scheme nor any other scheme could make coalition S internally stable. Further note that internal and external stability are linked: if coalition S is not externally

---

[3]Similar notions have been considered by Fuentes-Albero and Rubio (2010), McGinty (2007) and Weikard (2009).

stable because player j has an incentive to join, then coalition coalition $S \cup \{j\}$ is internally stable regarding player j. Loosely speaking, the transfer scheme considered here is optimal subject to the constraint that coalitions have to be stable.[4]

In the following, we introduce some properties which are useful in evaluating the success and incentive structure of coalition formation.[5]

Definition 3 *Effectiveness of a Coalition* *A coalition S is (strictly) effective with respect to coalition $S \setminus \{i\}$, $S \subseteq N$, $|S| \geq 2$ if $Q^*(S) \geq (>) Q^*(S \setminus \{i\})$. The coalition game is (strictly) effective if this holds for all $S \subseteq N$ and all $i \in N$.*

Definition 4 *Superadditivity, Positive Externality and Cohesiveness*
(i) A coalition game is (strictly) superadditive if for all $S \subseteq N$, $|S| \geq 2$ and all $i \in S$:

$$\sum_{i \in S} V_i^*(S) \geq (>) \sum_{i \in S \setminus \{i\}} V_i^*(S \setminus \{i\}) + V_i^*(S \setminus \{i\})$$

(ii) A coalition game exhibits a (strict) positive externality if for all $\forall S \subseteq N$, $|S| \geq 2$ and for all $j \in N \setminus S$:

$$V_j^*(S) \geq (>) V_j^*(S \setminus \{i\}).$$

(iii) A game is (strictly) cohesive if for all $S \subset N$:

$$\sum_{i \in N} V_i^*(\{N\}) \geq (>) \sum_{i \in S} V_i^*(S) + \sum_{j \in N \setminus S} V_j^*(S)$$

(iv) A game is (strictly) fully cohesive if for all $S \subseteq N$, and $|S| \geq 2$:

$$\sum_{i \in S} V_i^*(S) + \sum_{j \in N \setminus S} V_j^*(S) \geq (>) \sum_{i \in S \setminus \{i\}} V_i^*(S \setminus \{i\}) + \sum_{j \in N \setminus S \cup \{i\}} V_j^*(S \setminus \{i\}).$$

Definition 3 allows us to evaluate provision levels of different coalitions, in particular

---

[4]Every coalition S which is internally stable without transfers will also be internally stable with optimal transfers. However, the reverse is not true. Thus, if we can show that the coalition game exhibits a property called full cohesiveness (see Definition 4), i.e. the aggregate payoff payoff over all players increases with the enlargement of a coalition, then the global payoff of the stable coalition with the highest global payoff among the set of stable coalitions under an optimal transfer scheme is (weakly) higher than without transfers (or any other transfer scheme). Hence, optimal transfers have the potential to improve upon the global payoff of stable coalitions. For details see Eyckmans et al. (2012).

[5]Note that for Definitions 3 and 4 transfers do not matter. Firstly, equilibrium provision levels are not affected by transfers in our setting. Secondly, we look either at the aggregate payoff over all players or the aggregate payoff over all coalition members and non-signatories neither pay nor receive transfers by assumption.

compared to the situation when there is no cooperation. Note that full cohesiveness is the counterpart to effectiveness in welfare terms.

In Definition 4 all four properties are related to each other. For instance, a coalition game which is superadditive and exhibits positive externalities is fully cohesive and a game which is fully cohesive is cohesive. Typically, a game with externalities is cohesive, with the understanding that in a game with externalities the strategy of at least one player has an impact on the payoff of at least one other player. The reason is that the grand coalition internalizes all externalities by assumption.[6] Cohesiveness also motivates the choice of the social optimum as a normative benchmark, and it appears to be the basic motivation to investigate stability and outcomes of cooperative agreements. A stronger motivation is related to full cohesiveness, as it provides a sound foundation for the search for large stable coalitions even if the grand coalition is not stable due to large free-rider incentives. The fact that large coalitions, including the grand coalition, may not be stable in coalition games with the positive externality property is well-known in the literature (e.g. see the surveys by Bloch 2003 and Yi 1997). The positive externality can be viewed as a non-excludable benefit accruing to outsiders from cooperation. This property makes it attractive to stay outside the coalition. This may be true despite superadditivity holds, a property which makes joining a coalition attractive. In the context of a public good game with summation technology, stable coalitions are typically small because with increasing coalitions, the positive externality dominates the superadditivity effect (e.g. see Finus and Caparrós 2015).[7] Whether this is also the case in the context of the weakest-link technology is one of the key research question of this paper.

We close this section with a simple observation, which is summarized in the following lemma.

Lemma 1 *Individual Rationality and Stability Let a payoff be called individually rational if $V_i^*(S) \geq V_i^*(\{\{i\}, \{j\}, ...\{z\}\})$ in the case of no transfers, respectively, $V_i^{*T}(S) \geq V_i^{*T}(\{\{i\}, \{j\}, ...\{z\}\})$ in the case of transfers. In a coalition game which exhibits a positive externality, a necessary condition for internal stability of coalition S is that for all $i \in S$ individual rationality must hold.*

Proof. Applying the definition of internal stability and positive externality, $V_i^*(S) \geq V_i^*(S \setminus \{i\}) \geq V_i^*(\{\{i\}, \{j\}, ...\{z\}\})$ follows with the obvious modification for transfers. ∎

---

[6] Cohesiveness could fail if there are diseconomies of scale from cooperation, e.g. due to transaction costs which increase in the number of cooperating players. Our model abstracts from such complications.

[7] This is quite different in negative externality games. In Weikard (2009) it is shown that in a coalition game with negative externalities and superadditivity the grand coalition is the unique stable equilibrium, using the optimal transfer scheme in the case of asymmetric payoff functions.

Note that in negative externality games, this conclusion could not be drawn. A player in coalition S may be worse off than in the all singleton coalition structure, but still better off than when leaving the coalition.

4   Results of the Second Stage

4.1   Equilibrium Public Good Provision Levels

Generally speaking, the equilibrium strategy vector $q^*(S)$ can have different entries. We now develop the arguments that all entries are the same. For coalition members, it can never be rational to choose different provision levels as any provision level larger than the smallest provision level within the coalition would not affect benefits but would only increase costs. Their optimal or "ideal" choice in isolation (Vicary 1990), or their "autarky" provision level, is given by q_S^A, which follows from $\max \sum_{i \in S} V_i(q_S) \implies \sum_{i \in S} B_i'(q_S^A) = \sum_{i \in S} C_i'(q_S^A)$ in an interior equilibrium which is ensured by Assumption 1. Non-signatories' autarky provision levels, q_j^A, follow from $\max V_j(q_j) \implies B_j'(q_j^A) = C_j'(q_j^A)$ for all $j \notin S$.

In order to determine the overall equilibrium, some basic considerations are sufficient. Neither the coalition nor the singleton players have an incentive to provide (strictly) more than the smallest provision level over all players, $Q = \min_{i \in N} \{q_i\}$, as this would not affect their benefits but only increase their costs. They also have no incentive to provide (strictly) less than Q as long as $Q \leq q_j^A$, respectively, $Q \leq q_S^A$, as they are at the upward sloping part of their strictly concave payoff function. Strict concavity follows from Assumption 1 about benefit and cost functions (which ensure existence of an equilibrium). In the case of the coalition, we just note that the sum of strictly concave functions is strictly concave. Finally, players can veto any provision level above their autarky level. Thus, all players match Q as long as this is weakly smaller than their autarky level.

The replacement functions, $q_i = R_i(Q)$ (which are a variation of best reply functions, $q_i = r_i(q_{-i})$), as introduced by Cornes and Hartley (2007a,b) as a convenient and elegant way of displaying optimal responses in the case of more than two players, look like the ones

drawn in Figure 1.[8,9] The figure assumes a coalition with replacement function R_S, and two single players 1 and 2 with replacement functions R_1 and R_2, respectively. All replacement functions start at the origin and slope up along the 45°-line up to the autarky level of a player. At the autarky level, replacement functions have a kink and become horizontal lines, as no player can be forced to provide more than his autarky level. Hence, public good provision levels are strategic complements from the origin of the replacement functions up to the point where replacement functions kink. Consequently, all points on the 45°-line up to the lowest autarky level qualify as second stage equilibria (thick bold line). Thus, different from the summation technology, the second stage equilibrium is not unique. However, due to the strict concavity of all payoff functions, the smallest autarky level strictly Pareto-dominates all provision levels which are smaller. Therefore, is seems natural to assume that players play the Pareto-optimal equilibrium. Consequently, we henceforth assume this to be the unique second stage equilibrium[10] We relax this assumption in section 7.

[**Figure 1 about here**]

Proposition 1 *Second Stage Equilibrium Provision Levels Suppose some coalition S has formed in the first stage. The second stage equilibrium provision levels are given by the interval $q_i^*(S) \in [0, Q^A(S)]$, $Q^A(S) = \min\{q_i^A, q_j^A, ..., q_S^A\}$ and $q_i^*(S) = q_j^*(S) = q_S^*(S)$ $\forall i \neq j$; $i, j \notin S$. Public good provision levels are strategic complements up to the minimum autarky level $Q^A(S)$. The unique Pareto-optimal second stage equilibrium among the set of equilibria is $q_i^*(S) = q_j^*(S) = q_S^*(S) = Q^A(S) = Q^*(S)$ $\forall i \neq j$; $i, j \notin S$.*

Proof. Follows from the discussion above, including footnote 10. ∎

---

[8]Exploiting the aggregative structure of Bergstrom et al.'s (1986) non-cooperative public good model with summation technology, and following the discussion in Bergstrom et al. (1992) and Fraser (1992), Corners and Hartley (2007a) greatly simplify the proof of existence and uniqueness for the summation technology by exploiting the replacement function. Essentially, their proof boils down to a graphical argument in the q_i-Q-space as drawn in Figure 1, though replacement functions would look very different. If the individual replacement functions start at a positive level on the abscissa, are continuous and downward sloping over the entire strategy space, the aggregate replacement function (which is the vertical aggregation of the individual replacement functions for the summation technology) will have the same properties and it will intersect with the 45°-line, only once, and hence is the location of the unique Nash equilibrium.

[9]Using the technique of the replacement function, Corners and Hartley (2007b) analyze Nash equilibria for the weakest-link public good technology, showing that any non-negative level of the public good which does not exceed any individually preferred level is an equilibrium (and thus the game has a continuum of Pareto ranked equilibria). As their argument only requires convex preferences, it also holds in our coalition game. We only need to interpret coalition S as a single player for whom the aggregate preferences (as the sum of individual members' preferences) are convex. Note that reaction functions would also be upward sloping.

[10]The discussion for selecting the Pareto-optimal equilibrium would be very similar as discussed in Hirschleifer (1983) and Vicary (1990) in the context of Nash equilibria without coalition formation.

Assumption 2 *Among the set of second stage equilibria, the unique Pareto-optimal equilibrium is played in the second stage.*

It is evident that the summation technology would have very different properties. Replacement and reactions functions would be downward sloping and hence strategies are strategic substitutes. Moreover, there is no need to invoke Pareto-dominance to select equilibria as the equilibrium would be unique.

A useful result for the following analysis of the weakest-link technology is summarized in the following lemma.

Lemma 2   Coalition Formation and Autarky Provision Level *Consider a coalition S with autarky level q_S^A and a player i with autarky level q_i^A. If coalition S and player i merge, such that $S \cup \{i\}$ forms, then for the autarky level of the enlarged coalition, $q_{S \cup \{i\}}^A$, $\max\{q_S^A, q_i^A\} \geq q_{S \cup \{i\}}^A \geq \min\{q_S^A, q_i^A\}$ holds, with strict inequalities if $q_S^A \neq q_i^A$.*

Proof. The maximum of the sum of two strictly concave payoff functions is between the maxima of the two individual payoff functions. ■

Lemma 2 is illustrated in Figure 1 with the replacement function of the enlarged coalition denoted by $R_{S \cup \{1\}}$, assuming player 1 merges with coalition S. Note that merging of several players can be derived as a sequence of single accessions to coalition S.

4.2   Properties of the Public Good Coalition Game

For many of the subsequent proofs but also in order to understand generally how coalition formation impacts on equilibrium provision levels, the following lemma is useful.

Lemma 3   Coalition Formation and Effectiveness *Coalition formation in the public good coalition game with the weakest-link technology is effective.*

Proof. Case 1: Suppose that $Q^*(S)$ is the autarky level of a player j who does not belong to $S \cup \{i\}$. Hence, $q_j^A \leq q_i^A, q_S^A$ and $q_j^A \leq q_{S \cup \{i\}}^A$ due to Lemma 2 so that $Q^*(S) = Q^*(S \cup \{i\})$. Case 2: Suppose that $Q^*(S) = q_i^A$ initially and hence $q_j^A \geq Q^*(S)$ for all $j \notin S$. Moreover, $q_i^A \leq q_S^A$ and hence $q_i^A \leq q_{S \cup \{i\}}^A$ due to Lemma 2. Thus, regardless whether $Q^*(S \cup \{i\})$ is equal to the autarky level of the enlarged coalition, $q_{S \cup \{i\}}^A$, or equal to the autarky level of some other non-signatory j, $q_j^A \geq q_i^A$, $Q^*(S \cup \{i\}) \geq Q^*(S)$ must be true. Case 3: Suppose that $Q^*(S) = q_S^A$ before the enlargement, then the same argument applies as in Case 2. ■

Lemma 3 is useful in that it tells us that the public good provision level never decreases through a merger but may increase. It will strictly increase if the enlarged coalition contains

the (strictly) weakest-link player (either the single player who joins the coalition or the original coalition) whose autarky level before the merger was strictly below that of any other player. Because not all expansions of a coalition are strictly effective, the following properties also only hold generally in its weak form.

Proposition 2 *Positive Externality, Superadditivity and Full Cohesiveness* *The public good coalition game with the weakest-link technology exhibits the properties positive externality, superadditivity and full cohesiveness.*

Proof. *Positive Externality*: From Lemma 3 we know that $Q^*(S \cup \{i\}) \geq Q^*(S)$. Let $j \notin S \cup \{i\}$. Player j can veto any provision level above his autarky level if $q_j^A \leq Q^*(S \cup \{i\})$, and if $q_j^A > Q^*(S \cup \{i\})$ he must be at the upward sloping part of his strictly concave payoff function. Hence, $V_j^*(S \cup \{i\}) \geq V_j^*(S)$ must be true. *Superadditivity*: If the expansion from S to $S \cup \{i\}$ is not strictly effective, weak superadditivity holds. If it is strictly effective, i.e. $Q^*(S \cup \{i\}) > Q^*(S)$, then either i or S must determine $Q^*(S)$ before the merger. Then after the merger $q_{S \cup \{i\}}^A(S \cup \{i\}) > Q^*(S)$ from Lemma 2. Since the enlarged coalition $S \cup \{i\}$ can veto any provision level above $q_{S \cup \{i\}}^A(S \cup \{i\})$, moving from level $Q^*(S)$ towards $Q^*(S \cup \{i\}) \leq q_{S \cup \{i\}}^A(S \cup \{i\})$ must imply a move along the upward sloping part of the aggregate welfare function of the enlarged coalition and hence the enlarged coalition as a whole must have strictly gained. *Full Cohesiveness*: Positivity externality and superadditivity together are sufficient conditions for full cohesiveness. ∎

Lemma 3 and Proposition 2 are interesting in themselves but can be even more appreciated when compared with the summation technology. For the summation technology, effectiveness (with $Q = \sum_{j \in N} q_j$) and the positive externality property would also hold, though for a very different reason. Even though an expansion of the coalition also implied that signatories increase their aggregate provision level, non-signatories would not increase but decrease their provision level.[11] Because slopes of the reaction functions would be larger than -1, the overall provision level (strictly) increased. In other words, there would be leakage but less than 100%. The positive externality would not hold because outsiders get closer to their autarky provision level but because they take a free-ride. Non-signatories' benefits increasd through a higher total provision but their costs decreased as they have would reduced their individual contribution (see previous footnote).

In contrast, superadditivity could not be established at a general level for the summation technology, would require very restrictive assumptions to establish it and may in fact fail for typical examples. This is particularly true if the slopes of reaction functions are steep and

---

[11]For the special case of $B_i' = 0$, the provision level of non-signatories would remain constant and hence also their costs. Benefits would strictly increase through a merger as the total provision level strictly increases.

coalitions are small so that free-riding is particularly pronounced. It is for this reason that is difficult to establish generally full cohesiveness for the summation technology, at least we are not aware of any proof which is not based on the combination of superadditivity and positive externalities.[12]

Considering all properties in Proposition 2 together with the view of predicting stable coalitions in the first stage, general conclusions are not straightforward. On the one hand, also for the weakest-link technology the coalition game exhibits positive externalities, which following the literature predicts small coalitions. On the other hand, superadditivity always holds and strategies are strategic complements and not substitutes which may provide some indication that agreements may be more successful for the weakest-link than for the summation technology.[13] At least one may hope that more analytical results can be obtained for the first stage, different from the summation technology for which the analysis mainly relies on simulations.

5   Results of the First Stage

5.1   Symmetric Players

In order to analyze stability of coalitions, it is informative to start with the assumption of symmetric players which is widespread in the literature due to the complexity of coalition formation (see e.g. Bloch 2003 and Yi 1997 for overviews on this topic). Symmetry means that all players have the same payoff function. This assumption, which is sometimes also called ex-ante symmetry because, depending whether players are coalition members or non-members, they may be ex-post asymmetric, i.e. have different equilibrium payoffs. We follow the mainstream assumption and ignore transfer payments for ex-ante symmetric players.[14]

---

[12]It is somehow disturbing that the non-cooperative coalition formation literature analyzes ways to establish large stable coalitions without clarifying whether full cohesiveness holds. This shortcoming is valid for positive and negative externality games.

[13]Note that convexity does not hold for the public good coalition game, neither for the summation technology nor for the weakest-link technology. Convexity is a stronger property than superadditivity and implies that the gains from cooperation increase at an increasing rate with membership. Hence, convexity facilitates cooperation, an assumption frequently made in cooperative coalition theory, though, obviously, not appropriate in our context.

[14]For most economic problems and ex-ante symmetric players, in equilibrium, all players belonging to the group of signatories and all players belonging to the group of non-signatories chose the same economic strategies in the second stage (though signatories and non-signatories choose typically different strategies). Thus, all signatories receive the same payoff, and the same is true among the group of non-signatories. Consequently, transfers among signatories would create an asymmetry, which, though in theory possible, would be difficult to justify on economic grounds.

Proposition 3 *Symmetry and Stable Coalitions* *Assume payoff function (1) to be the same for all players, i.e. all players are ex-ante symmetric, then all players (signatories and non-signatories) are ex-post symmetric if coalition S forms, $V_i^*(S) = V_j^*(S)$ for all $i \neq j$. Moreover, $q^*(S) = q^*(S^{\#})$ for all possible coalitions $S \neq S^{\#}$, S, $S^{\#} \subseteq N$ and hence $V_i^*(S) = V_i^*(S^{\#})$ for all $i \in N$. Therefore, all coalitions are Pareto-optimal, socially optimal and stable, and there is no need for cooperation.*

Proof. Follows directly from Lemma 2 and applying the conditions of internal and external stability. ∎

Admittedly, Proposition 3 is less interesting when relating it to the literature on Nash equilibria cited in the introduction for the weakest-link technology which already concludes that there is not need for coordination for symmetric players. It is more interesting as a benchmark for coalition formation and when relating it to the summation technology: there would be a need for cooperation despite all players ex-ante symmetric, though stable coalitions tend to be small. Thus, in order to render the analysis interesting for the weakest-link technology, we henceforth consider asymmetric players.

5.2 Asymmetric Players

In order to operationalize and to make the concept of ex-ante asymmetric players interesting, we assume that autarky levels can be ranked as follows: $q_1^A \leq q_2^A \leq ... \leq q_N^A$ with at least one inequality sign being strict.[15] Henceforth, when we talk about ex-ante asymmetry, we mean this definition, without mentioning this explicitly anymore. We start with the assumption of no transfers

Proposition 4 *Asymmetry, No Transfers and Instability of Effective Coalitions* *Assume ex-ante asymmetric players and no transfers. a) All coalitions are Pareto-optimal, i.e. moving from a coalition $S \subseteq N$ to any coalition $S^{\#} \subseteq N$, $S \neq S^{\#}$, it is not possible to strictly increase the payoff of at least one player without decreasing the payoff of at least one other player. b) All strictly effective coalitions with respect to the all singleton coalition structure are not stable and all non-strictly effective coalitions are stable.*

Proof. a) Consider a coalition S with $Q^*(S)$ and any change through a change of membership of a group of players which leads to coalition $S^{\#}$ with $Q^*(S^{\#})$. Case 1: If $Q^*(S) = Q^*(S^{\#})$, $V_i(S) = V_i(S^{\#})$. Case 2: Suppose $Q^*(S^{\#}) > Q^*(S)$ which requires that player 1 is a member of $S^{\#}$. Hence, $q_1^A < Q^*(S) < Q^*(S^{\#})$ must hold, and at least player 1 must be worse off if

---

[15]Hence, we rule out the possibility (though unlikely) that all players have different payoff functions but the same autarky level.

coalition $S^{\#}$ forms. Case 3: Suppose $Q^*(S^{\#}) < Q^*(S)$ which implies that there is a player j with $q_j^A > Q^*(S)$ who must be worse off if $S^{\#}$ forms, regardless whether he is a member in any of these coalitions. b) Firstly, a strictly effective coalition requires the membership of the player with the lowest autarky level who will be strictly worse off than in the all singleton coalition structure (Case 2 in a) above) and instability follows from Lemma 1. Secondly, leaving a not strictly effective coalition with respect to no cooperation means that $Q^*(S) = Q^*(S \setminus \{i\})$ and hence internal stability follows trivially. External stability follows because either joining S such that $S \cup \{j\}$ forms is ineffective with respect to S or if it is strictly effective, then $q_j^A < Q^*(S)$ must be true and hence j is worse off in $S \cup \{j\}$ than as a single player, as just explained above. Hence, S is externally stable. ∎

Interestingly, even though all coalition structures are Pareto-optimal, not a single coalition is stable in the absence of transfers which strictly improves upon the non-cooperative equilibrium. The reason is that a strictly effective coalition requires membership of the players with the smallest autarky level who are worse off than when staying outside and individual rationality is a necessary condition for internal stability in a positive externality game. This is also one of the reason why all coalitions are Pareto-optimal (though not socially optimal). Any move from a coalition S to some other coalition $S^{\#}$ which changes the provision level means either a lower payoff to those players with the smallest autarky level if the provision level increases or to those with the largest autarky provision level if the provision level decreases. Note that for the summation technology results would be more ambiguous. The set of Pareto-optimal coalitions would normally only be a subset of all coalitions. In particular, the all singleton coalition structure would usually not be Pareto-optimal. Moreover, depending on the degree of asymmetry and the particular payoff function, no, one or some coalitions could be stable.

Given this unambiguous negative result for the weakest-link technology, we consider transfers (always in the form of the optimal transfer scheme) in the subsequent analysis. At the most basic level, we can ask the question: will transfers strictly improve upon no transfers? The answer is affirmative.

Proposition 5 *Asymmetry, Transfers and Existence of a Strictly Effective Stable Coalition Assume ex-ante asymmetric players and transfers. Then there exists at least one stable coalition S which Pareto-dominates the all singleton coalition structure with a strictly higher provision level.*

Proof. Because of asymmetry, we have $q_1^A < q_S^A < q_n^A$, and hence a strictly effective coalition S exists. A strictly effective coalition S compared to the all singleton coalition structure must include all players i for whom $q_i^A = \min\{q_1^A, q_2^A, ..., q_n^A\}$ is true and a player j with $q_j^A > q_i^A$

such that $q_i^A < Q^*(S) \le q_S^A$ from Lemma 2. Because it is strictly effective, $q_k^A \ge Q^*(S) > q_i^A$, all $k \notin S$ must be strictly better off (strict positive externality holds). Let there be only one player j in S. Hence, for all $i \in S$, $S \setminus \{i\} = \{\{i\}, \{j\}, ... \{z\}\}$ (the all singletons coalition structure) regardless which coalition member leaves. Therefore, because $q_S^A \ge Q^*(S) > q_i^A$, $\sigma_S(S) := \sum_{i \in S}(V_i^*(S) - V_i^*(S \setminus \{i\})) > 0$ follows from the strict concavity of the aggregate payoff function of S and hence $V_i^{*T}(S) = V_i^*(S \setminus \{i\}) + \gamma_i \sigma_S(S) > V_i^*(\{\{i\}, \{j\}, ... \{z\}\})$. Hence, S constitutes also a strict Pareto-improvement for all players in S compared to the all singleton coalition structure and S is internally stable. Now suppose S is externally stable and we are done. If S is not externally stable with respect to the accession of an outsider l (which requires $q_l^A > Q^*(S)$), then coalition $S \cup \{l\}$ is internally stable. If it is also externally stable we are done, otherwise the same argument is repeated, noting that eventually one enlarged coalition will be externally stable because the grand coalition is externally stable by definition. Due to the strict positive externality, and Lemma 1, the eventually stable coalition must Pareto-dominate the all singleton coalition structure. ∎

Note that a general statement as in Proposition 5 would not be possible for the summation technology. Establishing existence of a non-trivial coalition with transfers requires superadditivity but this property does not hold generally as pointed out above. However, predicting which specific coalitions are stable for the weakest-link technology is also not straightforward at this level of generality, though it turns out that our results are much more general than those obtained for the summation technology.[16] In the next section, we analyze how the nature of asymmetry affects stability. We first lay out the basic analysis for determining stable coalitions and then look into the details.

6 Stable Coalitions and the Nature of Asymmetry

6.1 General Considerations

In the context of the provision of a public good, it seems natural to worry more about players leaving a coalition than joining it and hence one is mainly concerned about internal stability. This is even more true because if coalition S is internally stable with transfers, but not externally stable, then a coalition $S \cup \{j\}$ is internally stable, with a provision level and a global payoff strictly higher than before.[17] Hence, we focus on this dimension of

---

[16]Analytical results for the cartel formation game have only been obtained in Barrett (2001), Fuentes-Albero and Rubio (2010) and Pavlova, de Zeeuw (2013), but they assume a particular payoff function and only two types of players, severely limiting the type of asymmetry.

[17]External instability requires $Q^*(S) < Q^*(S \cup \{j\})$ and hence the move from S to $S \cup \{j\}$ would be strictly fully cohesive by Proposition 2.

stability. Moreover, we consider only strictly effective coalitions compared to the all singleton coalition structure because all other coalitions are internally stable even without transfers as stated in Proposition 4. Because of strict effectiveness, all players with $q_i^A = q_1^A$ must be members of S, $q_1^A \leq q_2^A \leq ... \leq q_n^A$ with at least one inequality being strict. In the presence of transfers, we know from Section 2 that internal stability of coalition S requires that $\sigma_S(S) = \sum_{i \in S} \sigma_i(S) \geq 0$, with $\sigma_i(S) = V_i^*(S) - V_i^*(S \setminus \{i\})$.

In principle, we need to distinguish only two cases which are illustrated in Figure 2. In case 1, coalition S determines the equilibrium provision level and hence $q_S^A = Q^*(S)$. Consequently, $q_m^A > q_S^A$ for all $m \notin S$. S may be a subcoalition or the grand coalition. In case 2, an outsider m determines the equilibrium provision, $S \subset N$, and hence $q_m^A = Q^*(S)$. Because S is assumed to be strictly effective compared to the all singleton coalition structure, we must have $q_1^A \leq q_i^A < q_m^A < q_S^A$ (with all players i with $q_i^A < q_m^A$ being members of S).

<div align="center">[Figure 2 about here]</div>

For both cases (which are identical if $q_m^A = q_S^A$), we distinguish three groups of players in coalition S. "Weak players" $i \in S_1$ for which $q_i^A = Q^*(S \setminus \{i\}) < Q^*(S) < q_{S \setminus \{i\}}^A$ after they leave coalition S, "strong players" $j \in S_2$ for which $q_{S \setminus \{j\}}^A = Q^*(S \setminus \{j\}) < Q^*(S) < q_j^A$ is true and "neutral" players $k \in S_3$ for which $Q^*(S \setminus \{k\}) = Q^*(S) \leq q_k^A$ holds.[18] Weak players have an autarky provision level below the equilibrium provision level when S forms and hence gain from leaving coalition S, i.e. $\sigma_i(S) = V_i^*(S) - V_i^*(S \setminus \{i\}) < 0$. For strong players this is reversed; they have an autarky level above $Q^*(S)$ and if they leave, the new equilibrium provision level is lower and hence they lose from leaving, $\sigma_j(S) = V_j^*(S) - V_j^*(S \setminus \{j\}) > 0$. For neutral players $\sigma_k(S) = V_k^*(S) - V_k^*(S \setminus \{k\}) = 0$. Their autarky provision level is equal to $Q^*(S) = q_S^A$ in case 1 and larger than $q_S^A > Q^*(S) = q_m^A$ in case 2 but not large enough ($q_k^A \leq \tilde{q}$ in Figure 2) so when they leave coalition S, $q_{S \setminus k}^A \geq Q^*(S) = Q^*(S \setminus \{k\}) = q_m^A$. That is, neutral players do not affect the provision level after they leave.[19] Clearly, $S = S_1 \cup S_2 \cup S_3$ noting that the set of players in different groups do not coincide in case 1 and 2, as it is evident from Figure 2. For a given distribution of autarky levels in coalition S, S_1 and S_2 will be smaller and S_3 will be larger in case 2 than in case 1.

We define $\tilde{S} = S_1 \cup S_2$ because only these two groups of players affect stability. Thus,

---

[18]We use these terms for easy reference, having in mind a weak, strong or neutral interest regarding the level of public good provision.

[19]Formally, we have: Case 1 with $q_m^A > q_S^A$ for all $m \notin S$: $S_1 = \left\{ i \mid i \in S \wedge q_1^A \leq q_i^A < q_S^A < q_{S \setminus \{i\}}^A \right\}$, $S_2 = \left\{ j \in S \mid q_{S \setminus \{j\}}^A < q_S^A < q_j^A \right\}$ and $S_3 = \left\{ k \in S \mid q_{S \setminus \{k\}}^A = q_k^A = q_S^A \right\}$; Case 2 with $q_m^A < q_S^A$ for some $m \notin S$, $S \subset N$: $S_1 = \left\{ i \mid i \in S \wedge q_1^A \leq q_i^A < q_m^A < q_S^A < q_{S \setminus \{i\}}^A \right\}$, $S_2 = \left\{ j \in S \mid q_{S \setminus \{j\}}^A < q_m^A < q_S^A < q_j^A \right\}$ and $S_3 = \left\{ k \in S \mid q_m^A \leq q_{S \setminus \{k\}}^A \right\}$.

coalition $S \subseteq N$ is internally stable if and only if:

$$\sigma_S(S) = \sum_{j \in \tilde{S} \setminus S_1} \left[V_j(q_y^A) - V_j(q_{S \setminus \{j\}}^A) \right] - \sum_{i \in S_1} \left[V_i \left(q_i^A \right) - V_i(q_y^A) \right] \geq 0 \qquad (2)$$

with $y = S$ in case 1 and $y = m$ in case 2. Condition 2 stresses that what strong players gain by staying inside the coalition (first term) must be larger than what weak players lose by staying inside the coalition (second term).

When is this condition likely to hold? Consider first the first term in (2) above. Intuitively, for the $\tilde{S} \setminus S_1$ group of strong players, a large difference between q_j^A and q_y^A implies a large drop from q_y^A to $q_{S \setminus \{j\}}^A$ when they leave the coalition. Hence, q_y^A and $q_{S \setminus \{j\}}^A$ are at the steep part of the upward sloping part of a strong players' strictly concave payoff function V_j. In other words, the difference $V_j(q_S^A) - V_j(q_{S \setminus \{j\}}^A)$ is large, i.e. the gain from remaining in the coalition is large if the distance between q_y^A and q_j^A is large. For the S_1 group of weak players, we require just the opposite for condition 2 to hold: the closer q_i^A to q_y^A, the smaller the second term in (2) and hence the smaller the gain from leaving coalition S. Thus, roughly speaking, we are looking for a positively skewed distribution of autarky levels of the players in coalition S with reference to q_y^A. The weak players should have an autarky level close to the autarky level of the coalition in case 1 and close to the autarky level of outsider m in case 2. In contrast, the strong players should have an autarky level well above the coalitional autarky level in case 1 and well above the autarky level of player m in case 2. In the next subsection, we have a closer look how this relates to the underlying parameters and structure of the benefit and cost functions.

6.2   Asymmetry and Stability

In this section, we want to substantiate the intuition provided above about distributions of autarky levels of coalition members which are conducive to internal stability of a coalition. Analytically, we cannot simply consider different distributions of autarky levels as they may be derived from different payoff functions. Therefore, we need to construct a framework which allows to relate autarky levels to the parameters of the payoff functions. Hence, we consider a payoff function which has slightly more structure than our general payoff function (1), but which is still far more general than what is typically considered in the literature on non-cooperative coalition formation in general and in particular in the context of public good provision with a summation technology.[20] We use the notation $v_i(Q, q_i)$ to indicate the

---

[20] All specifications used in the context of the summation technology are a special case of payoff function (3) assuming $Q = \sum_{i \in N} q_i$ instead of $Q = \min_{i \in N} \{q_i\}$ For instance, the "quadratic-quadratic" payoff function, which has been extensively used for the analysis of international environmental agreements, is obtained by

difference to our general payoff function (1) which was denoted by $V_i(Q, q_i)$:

$$v_i(Q, q_i) = b_i B(Q) - c_i C(q_i) \tag{3}$$
$$Q = \min_{i \in N}\{q_i\}$$

where the properties of B and C are those summarized in Assumption 1. That is, we assume that all players share a common function B and C but differ in the scalars b_i and c_i. In addition, in order to simplify the subsequent analysis, we assume $C''' \geq 0$ and $B''' \leq 0$ (or, if $B''' > 0$, then B''' is sufficiently small).[21]

The following lemma shows the key advantage of payoff function (3): it allows us to characterize the autarky provision of any trivial or non-trivial coalition S, based on a single parameter.

Lemma 4 *Autarky Provision Level and Benefit and Cost Parameters* *Consider payoff function (3). The autarchy abatement level of a coalition S is given by $q_S^A = h(\theta_S)$, where h is a strictly increasing and strictly concave function implicitly defined by $\frac{C'(q)}{B'(q)} = \theta_S$, with $\theta_S = \frac{\sum_{i \in S} b_i}{\sum_{i \in S} c_i}$.*

Proof: See Appendix B.1.

That is, players can be ranked based on their parameter θ_i through the function h. Players with higher parameters θ_i will have higher autarky levels. We say a player k is "stronger" than a player l if $\theta_k > \theta_l$ and "weaker" if the opposite relation holds. According to our general analysis above, in case 1, when coalition S determines the equilibrium provision level, i.e. $Q^*(S) = q_S^A$, weak players are coalition members for which $\theta_i < \theta_S$ holds, strong players for which $\theta_j > \theta_S$ holds and neutral players for which $\theta_l = \theta_S$ holds. In case 2, when an outsider player m determines the equilibrium provision level, i.e. $Q^*(S) = q_m^A$, weak players are coalition members for which $\theta_i < \theta_m$ holds, strong players for which $\theta_{S \setminus \{j\}} < \theta_m$ holds and neutral players for which $\theta_m \leq \theta_j, \theta_{S \setminus \{j\}}$ holds. Accordingly, condition (2) can be

---

setting $B(Q) = a_1(Q) - \frac{a_2}{2}(Q)^2$ and $C(q_i) = a_3 q_i + \frac{a_4}{2} q_i^2$, with $a_j \geq 0$ for $j = \{1, 2, 3, 4\}$. For example, Barrett (1994) and Courtois and Haeringer (2012) assume symmtric players and a particular case of this functional form. In order to replicate their payoff function, we would need to set $a_1 = a$, $a_2 = 1, a_3 = 0, a_4 = 1, b_i = b$ $\forall i \in N$, $c_i = c \ \forall \ i \in N$. McGinty (2007) analyzes, using simulations, a game with asymmetric players with similar functions. In order to retrieve his function, we would need to set a_j for $j = \{1, 2, 3, 4\}$ as in Barrett's game but $b_i = b\alpha_i$ and $c_i = c_i$. For other payoff functions, including the linear benefit function considered for instance in Ray and Vohra (2001) or Finus and Maus (2008) a similar link could be established. This is also true for Rubio and Ulph (2006) and Dimantoudi and Sartzetakis (2006) although they analyze the dual problem of an emission game.

[21]If $B''' > 0$, a sufficient condition for the subsequent results to hold is $B''' < -2B''C''/C'$. See Appendix B.1.

written as follows:

$$\sigma_S(S,\Theta) = \sum_{j \in S_2} \left[v_j(\theta_y) - v_j\left(\theta_{S\setminus\{j\}}\right) \right] - \sum_{i \in S_1} \left[v_i(\theta_i) - v_i(\theta_y) \right] \quad (4)$$

where $\theta_y = \theta_S$ in case 1 and $\theta_y = \theta_m$ in case 2, with $\sigma_S(S,\Theta)$ indicating that internal stability of coalition S depends on the distribution of θ_i-values of players in S, Θ. We now ask the question how $\sigma_S(S,\Theta)$ changes if we change the θ_i-values of some players in S, assuming the same θ_S, but considering different distributions Θ.[22] To simplify the exposition, we focus on the case where all players in S share a common $c_i = c$ and the changes affect only the parameters b_i. However, all the results shown in this section hold if coalitions members share a common $b_i = b$ and marginal changes affect the parameters c_i (in the opposite direction; $b_i + \epsilon$ corresponds to $c_i - \epsilon$) instead, and only minor adjustments are needed to accommodate the case where players differ in both parameters. We detail these adjustments in footnote 23.[23]

Proposition 6 *Asymmetry and Stability Consider payoff function (3), with $c_i = c$ $\forall$ $i \in S$, a strictly effective coalition S with respect to the all singleton coalition structure and two distributions Θ and $\tilde{\Theta}$ of players θ_i-values in S, where $\tilde{\Theta}$ is derived from Θ by a marginal change ϵ of two b_i-values of players in S, such that $b_k - \epsilon$ and $b_l + \epsilon$, implying $\theta_{k-\epsilon} < \theta_k$ and $\theta_{l+\epsilon} > \theta_l$. Then $\sigma_S(S,\tilde{\Theta}) \geq \sigma_S(S,\Theta)$ if:*

(i) $\theta_l < \theta_k \leq \theta_y$;
(ii) $\theta_y < \theta_k \leq \theta_l$;
(iii) $\theta_k \leq \theta_y < \theta_l$ and $\theta_{\{k-\epsilon\}} \geq \theta_{S\setminus\{l+\epsilon\}}$.

Proof: See Appendix B.2.

All three conditions are illustrated in Figure 3, noting that θ_S remains the same through the marginal changes.

In condition (i), among the set of players with θ-values below θ_y, the θ-value of the weaker player l becomes larger at the expenses of the θ-value of the stronger player k. The set of players involved in marginal changes belongs to the group of weak players S_1. At the margin, it includes the possibility that player k is a neutral player before the marginal change.

---

[22] We could conduct a similar comparative analysis of autarky and equilibrium provision levels as a function of a uniform change of all benefit or cost parameters as in Cornes and Hartley (2007b). As this does not affect stability, we consider only non-uniform changes below.

[23] Dropping the assumption $c_i = c$ $\forall$ $i \in S$, respectively, $b_i = b$ $\forall$ $i \in S$, all conditions in Proposition 6 remain unchanged except condition (ii) for which the additional condition $v'_l(\theta)|_{\theta_{S\setminus l}} \geq v'_k(\theta)|_{\theta_{S\setminus k}}$ would be needed. The subsequent Corollary 1 and 2 also continue to hold.

In condition (ii), among the set of players with θ-values above θ_y, the $\theta-$value of a (weakly) stronger player l is increased at the expense of the θ-value of a (weakly) weaker player k. In case 1 where the coalition determines the equilibrium provision level, the set of players involved in marginal changes of θ-values is the set of strong players S_2 (see also Figure 2). In case 2 where an outsider determines the equilibrium provision level, the set of involved players could be strong players in S_2 but also neutral players S_3 (see also Figure 2).

In condition (iii) the marginal changes affect one player with a θ-value above and one below θ_y. The marginal changes involve an increase of the θ-value of a stronger player l at the expenses of the θ-value of a weaker player k. The weaker player will typically belong to the group of weak players S_1, but could also belong to S_3 at the margin in the special case when $\theta_k = \theta_y$. The stronger player l will always belong to S_2 because otherwise $\theta_{\{k-\epsilon\}} \geq \theta_{S\setminus\{l+\epsilon\}}$ will be violated (see Appendix B.2 for details). Important is that the θ-value of one player involved in the change is relatively strong within its group because this ensures $\theta_{\{k-\epsilon\}} > \theta_{S\setminus\{l+\epsilon\}}$, where $\theta_{S\setminus\{l+\epsilon\}}$ is the θ-value of coalition S if player l leaves coalition S under distribution $\tilde{\Theta}$.

Note that the weak inequality sign in Proposition 6 in terms of stability only applies to the particular case where both players k and l belong to the set of neutral players S_3 (which is only possible in case 2 in condition (ii)) as all other changes imply $\sigma_S(S, \tilde{\Theta}) > \sigma_S(S, \Theta)$ (see Appendix B.2. for details).

[**Figure 3 about here**]

In the following, we consider different distributions to illustrate Proposition 6. For simplicity, we assume that coalition S determines the equilibrium provision level (case 1). The formal definitions of these distributions are given in Definition 5 and are illustrated in Figure 4.

[**Figure 4 about here**]

Definition 5 *Distributions Consider a coalition S with $Q^*(S) = q_S^A$ and the following distributions of θ-values (with $c_i = c \ \forall \ i \in S$), generating changes of θ_i-values through a marginal change of b_i-values as explained in Proposition 6, denoting the cardinality of S by s with s being sufficiently large and let $\Delta > 0$ be the result of a sequence of changes ϵ in Proposition 6 with $\Delta < \theta_S$.*

Sequence 1:
(a) Asymmetric distribution Θ^Λ with $\theta_i = \theta_S - \Delta$ and for all $j \neq i$, $\theta_j = \theta_S + \frac{\Delta}{s-1}$.
(b) Symmetric distribution Θ^Ψ with $\theta_i = \theta_S - \Delta$, $\theta_j = \theta_S + \Delta$ and for all $k \neq i, j$, $\theta_k = \theta_S$, generated from Θ^Λ by applying a sequence of changes in Proposition 6 using condition (ii).
(c) Asymmetric distribution Θ^Ω with $\theta_i = \theta_S - \frac{\Delta}{s-1}$ for all $i \neq j$, $\theta_j = \theta_S + \Delta$, generated

from Θ^{Ψ} by applying a sequence of changes in Proposition 6 using condition (i).

Sequence 2:

(a) Uniform distribution Θ^{Γ} with $\theta_i = \theta_S - \left(\frac{(s-1)}{2} + 1 - i \right) \Delta$, $i = 1, ..., s$ such that there are $\frac{(s-1)}{2}$ players in S_1 and $\frac{(s-1)}{2}$ players in S_2 and one player in S_3, assuming s to be an odd number.

(b) Asymmetric distribution Θ^{Φ} with $\theta_i = \theta_S - \frac{\Delta}{4}(s+1)$ for all $i = 1, ..., \frac{(s-1)}{2}$, $\theta_k = \theta_S$ for player $k = \frac{(s-1)}{2} + 1$ and $\theta_j = \theta_S - \left(\frac{(s-1)}{2} + 1 - j \right) \Delta$, $j = \frac{(s-1)}{2} + 2, ..., s$ for all $j \neq i, k$, generated from Θ^{Γ} by applying a sequence of changes in Proposition 6 using condition (i).

(c) Asymmetric distribution Θ^{Υ} with $\theta_i = \theta_S - \frac{\Delta}{4}(s+1)$ for all $i = 1, ..., \frac{(s-1)}{2}$, $\theta_k = \theta_S$ for all players $k = \frac{(s-1)}{2} + 1, ..., s - 1$ and $\theta_j = \theta_S + \Delta(\frac{s-1}{2}) + \Delta \frac{(s-1)(s-3)}{8} = \theta_S + \Delta \frac{(s^2-1)}{8}$ for player $j = s$, generated from Θ^{Φ} by applying a sequence of changes in Proposition 6 using condition (ii).

(d) Asymmetric distribution Θ^{Ξ} with $\theta_i = \theta_S - \frac{\Delta}{8}(s+1)$ for all $i = 1, ..., s - 1$ and $\theta_j = \theta_S + \Delta \frac{(s^2-1)}{8}$ for player $j = s$, generated from Θ^{Υ} by applying a sequence of changes in Proposition 6 using condition (i).

Corollary 1 Distributions and Stability *For the distributions defined in Definition 5, the following relations hold:*

Sequence 1: $\sigma_S(S, \Theta^{\Lambda}) < \sigma_S(S, \Theta^{\Psi}) < \sigma_S(S, \Theta^{\Omega})$; $\sigma_S(S, \Theta^{\Omega}) > 0$.

Sequence 2: $\sigma_S(S, \Theta^{\Gamma}) < \sigma_S(S, \Theta^{\Phi}) < \sigma_S(S, \Theta^{\Upsilon}) < \sigma_S(S, \Theta^{\Xi})$; $\sigma_S(S, \Theta^{\Xi}) > 0$.

Proof. Follows directly from Proposition 6. $\sigma_S(S, \Theta^{\Omega}) > 0$ and $\sigma_S(S, \Theta^{\Xi}) > 0$ can be proved along the lines of the proof in Proposition 5. ∎

In sequence 1, we are moving from a (very) negatively skewed distribution Θ^{Λ} to a symmetric distribution Θ^{Ψ} finally ending up in a (very) positively skewed distribution Θ^{Ω}. Along this sequence, the value of $\sigma_S(S)$ increases. Whether $\sigma_S(S)$ is positive or negative cannot be said at this level of generality, except that we know that finally $\sigma_S(S, \Theta^{\Omega}) > 0$. Since distribution Θ^{Ω} (like Θ^{Ξ}) can always be generated from any distribution, there always exists an asymmetric distribution for which the grand coalition is stable. The reason is simple. For this type of distribution, regardless which player leaves coalition S, the subsequent equilibrium provision will be the provision level in the Nash equilibrium and because there is a strictly positive aggregate gain for coalition members in S from moving from $\{\{i\}, \{j\}, ..., \{n\}\}$ to any non-trivial coalition S, $\sigma_S(S) > 0$ must be true.

In sequence 2, we move from a symmetric, in fact uniform distribution Θ^{Γ} to a positively skewed distribution Θ^{Φ}, imposing further changes, generating distribution Θ^{Υ} and Θ^{Ξ}, in-

creasing $\sigma_S(S)$ on this way, noting that Θ^Ξ is a very positively skewed distribution. It is clear that a similar sequence could have been generated starting from a normal distribution.

Both sequences suggest that asymmetric distributions of autarky levels which are positively skewed may be more conducive to the stability of coalitions than rather symmetric distributions if the asymmetric gains from cooperation can be balanced in an optimal way through a transfer scheme. However, a relative symmetric distribution is more conducive to stability than a negatively skewed distribution of autarky levels. Hence, asymmetry of interests as such is not an obstacle to successful cooperation but can be actually an asset depending on the type of asymmetry. It is conducive to stability if there is no outlier at the lower end (condition (i) in Proposition 6). At the upper end, this is reversed. Instead of having many strong players it is better for stability to have one outlier at the top (condition (ii) in Proposition 6). If there is only one strong player left, he would pay transfers to all other weak players.

Essentially, what we did in Corollary 1 is to relate distributions to stability so that Proposition 6 is less abstract. We note that there is no unique measure to compare different distributions. Corollary 1 suggests that skewness could be a good measure. This is indeed the case for most distributions even though we need to mention one caveat: there is not always a one to one correspondence between the marginal changes listed in Proposition 6 and skewness. In other words, not all marginal changes which increase $\sigma_S(S)$ increase skewness (though most do) as we detail in Appendix B.3, using the Fisher-Pearson coefficient of skewness.

It is also important to point out that different from the summation technology where it is usually easier to obtain stability for smaller than for larger coalitions, this may not be true for the weakest-link technology. As Proposition 6 highlights, stability only depends on the distribution of autarky levels of players in coalition S, i.e. the θ_i-values. By adding a player outside coalition S to S, a new distribution is generated for which $\sigma_s < \sigma_{s \cup \{l\}}$ is possible. [24]

6.3 Asymmetry and Welfare

From the previous subsection we know how distributions of autarky levels relate to stability. Now we want to relate this to the global gains from cooperation. To this end, we define the global payoff of a given distribution Θ of players θ_i values as $W = \sum_{i \in N} v_i(\Theta)$. We further define the absolute (respectively, relative) gain from forming coalition S by $\Delta W(\Theta) := W^S - W^{Na}$ (respectively, $\Delta W(\Theta) := \frac{W^S}{W^{Na}}$), with superscript Na for Nash equilibrium and

---

[24] Consider a game with four players and the following payoff function $v_i = b_i \left(aQ - \frac{1}{2}Q^2 \right) - \frac{c}{2}q_i^2$ with $a = 10$, $c = 1$ and $b_i = \{4, 5, 5, 10\}$ for $i = 1, 2, 3, 4$. Then $\sigma_S(\{1, 2, 3, 4\}) = \frac{8410}{6069} > 0$ whereas $\sigma_S(\{1, 2, 3\}, \{4\}) = -\frac{120}{2057} < 0$.

S for the coalition. These definitions cover the case where S is the grand coalition and therefore the social optimum.

Proposition 7 *Asymmetry and Welfare Gains Consider payoff function (3), a strictly effective coalition S with respect to the all singleton coalition structure and two distributions Θ and $\tilde{\Theta}$ as defined in Proposition 6. Then, $\Delta W^S(\Theta) > \Delta W^S(\tilde{\Theta})$ if*

$$\theta_{\min}(\Theta) = \min\{\theta_1, ..., \theta_n\} < \theta_{\min}(\tilde{\Theta}) = \min\{\tilde{\theta}_1, ..., \tilde{\theta}_n\}$$

Proof. Using payoff function (3), $W = \sum_{i \in N} v_i = \sum_{i \in N} b_i B(q) - \sum_{i \in N} c_i C(q)$ from which it is evident that the marginal changes of b_i (or c_i-values) described in Proposition 6 do not change W. We know that W is strictly concave in q with $W'(q_N^A) = 0$ and $q_s^A = h(\theta_S)$ for all $S \subseteq N$ from Lemma 4. By construction, marginal changes do not affect θ_S and θ_m but may affect the smallest autarky level $\theta_{\min}$ Therefore, $W^{Na}(\Theta) < W^{Na}(\tilde{\Theta})$ and $W^S(\Theta) = W^S(\tilde{\Theta})$. ∎

Thus, the smaller the smallest autarky level, the smaller the provision level in the Nash equilibrium and hence the larger are the gains from cooperation, keeping the socially optimal and equilibrium provision level if coalition S forms constant as assumed by the marginal changes in Proposition 6 and 7. Thus, by using the concept of a sequence of marginal changes of b_i-values (and/or c_i-values), as introduced in Proposition 6, and also assumed in Proposition 7, very different distributions can be compared in terms of their global payoff implications. Comparing again our distributions defined in Definition 5, we find that relations are (almost) reversed.

Corollary 2 *Distributions and Welfare Gains For the Distributions defined in Definition 5, the following relations hold:*

$$\text{Sequence 1} \quad : \quad \Delta W(S, \Theta^\Lambda) = \Delta W(S, \Theta^\Psi) > \Delta W(S, \Theta^\Omega).$$
$$\text{Sequence 2} \quad : \quad \Delta W(S, \Theta^\Gamma) > \Delta W(S, \Theta^\Phi) = \Delta W(S, \Theta^\Upsilon) > \Delta W(S, \Theta^\Xi).$$

A comparison of Corollary 1 and 2 reveals: distributions which favour stability maybe associated with a lower global gain from cooperation and vice versa. Thus, the "paradox of cooperation", a term coined by Barrett (1994) in the context of the summation technology, may also hold for the weakest-link technology. However, a detailed comparison between Proposition 6 about stability and Proposition 7 about global payoffs reveals that the message is not so simple. It is true for the marginal changes imposed in condition (i) in Proposition 6: the gains from cooperation decrease but the stability value $\sigma_S(S)$ increases. It is not true for the changes (ii) and (iii) which are payoff neutral but increase the stability value $\sigma_S(S)$. Also

the skewness of different θ_i-distributions is only of limited use in characterizing welfare gains, except when considering the extreme: jumping from a very negatively skewed distribution to a very positively skewed distribution of θ_i-values through a sequence of marginal changes decreases the gains from cooperation but increases stability, confirming the "paradox of cooperation".

7   Equilibrium Selection

We applied the criterion of Pareto-dominance to select the equilibrium provision level (Assumption 2). As pointed out by Hirschleifer (1983), and reiterated by Vicary (1990), this equilibrium would also emerge if players choose their provision levels sequentially (and disclose their bids). In our context, this would be the case if, say, the coalition would act as a Stackelberg leader and the non-signatories as Stackelberg followers as in Barrett (1994). This also points to the fact that in our setting, there is no difference between the Stackelberg and Nash-Cournot assumption, which would be different for the summation technology.[25]

In contrast, experimental evidence suggests that efficient outcomes may be difficult to achieve when groups are large. Harrison and Hirshleifer (1989) found that in small groups coordination on the efficient equilibrium may occur, but Van Huyck et al. (1990) showed that this result does not hold if the group size is increased. This negative impact of group size on coordination was confirmed by other experimental studies for different variations of the weakest-link game (Cachon and Camerer, 1996; Brandts and Cooper, 2006; Weber, 2006 and Kogan et al., 2011). Though differences across different institutional settings in experiments are interesting, in our context the most relevant finding is the observation that the larger the number of players, the smaller equilibrium provision levels will be compared to the Pareto-optimal Nash equilibrium provision level.

There have been some attempts to model these experimental observations. Unfortunately, all those papers of which we are aware of assume symmetric players, at least symmetric benefit functions and hence are not directly applicable to our general setting. Nevertheless, we briefly discuss them to motivate our analysis below. Cornes and Hartley (2007a) use a symmetric CES-composition function to model various forms of weaker-link technologies. They show that at the limit, when the weaker-link approximates the weakest-link, a unique Nash equilibrium will be selected, though it is not the Pareto-optimal Nash equilibrium; the Nash equilibrium provision level decreases with the number of players for their assumption.

---

[25] For the summation technology, the Stackelberg assumption leads to larger coalitions than the Nash-Cournot assumption (see Finus 2003 for an overview). Only if $B_i^{''} = 0$ will there be no difference. The reason is that Stackelberg leadership provides the coalition members with a strategic advantage compared to non-members.

Other approaches originate from the concept of risk-dominance where players assume that other players may make a (small) mistake when choosing their provision level. Monderer and Shapley (1996) use the concept of the potential function which yields the risk-dominant equilibrium for symmetric players, which is unique and decreases in the number of players. A similar result is obtained by Anderson et al. (2001) using the concept of logistic equilibrium and a stochastic potential function, again assuming symmetric players and a linear payoff function.

Extending those theoretical papers to the general case of asymmetric players and general payoff functions as used in our paper would be a paper in its own right. Therefore, we only take the main conclusions from these papers for the motivation to consider two simple alternative assumptions: a) $\widetilde{q}_S^A = \alpha(n)q_S^A$ and b) $\widetilde{\widetilde{q}}_S^A = \alpha(n - s + 1)q_S^A$ for all $S \subseteq N$ where n is the total number of players and s the number of players in coalition S. Hence, the equilibrium provision level if coalition S forms is α times the $Q^*(S)$ known from Proposition 1. We assume that $\alpha(n)$ and $\alpha(n - s + 1)$ decrease in n and the latter increases in s. For $\alpha(n - s + 1)$ we may think for simplicity that if $s = n$, then $\alpha(1) = 1$ and if no coalition forms, $s = 1$, then $\alpha(n)$. The difference between both assumptions is how we count players where the second assumption treats the coalition as one player. Hence, anything else being equal, coalition formation by itself leads to improved coordination for the second alternative assumption.

The question we pose now is whether our results would still hold or, if not, what would change. We note that both alternative assumptions about α imply de facto a kind of modest provision level as considered in the context of the summation technology by Barrett (2002) and Finus and Maus (2008) which could lead to larger coalitions. For a given coalition S, autarky and equilibrium provision levels depart from optimality, but this could be compensated by larger coalitions being stable. Conceptually, this is more interesting in the absence of transfers because then modesty serves as a compensation device. With transfer, transfers serve as a compensation device and hence it seems obvious to maximize the gains from cooperation by choosing Pareto-optimal equilibrium provision levels as we have done in previous sections. Hence for brevity, we restrict our analysis to the most important items captured in sections 4 and 5.

Proposition 8 *Alternative Equilibrium Selection Consider two alternative assumptions: a) $\widetilde{q}_S^A = \alpha(n)q_S^A$ and b) $\widetilde{\widetilde{q}}_S^A = \alpha(n - s + 1)q_S^A$ for all $S \subseteq N$ and hence the equilibrium provision level is $\alpha(\cdot)Q^*(S)$.*
(i) For both assumptions, the coalition formation game is effective and the properties positive externality, superadditivity, cohesiveness and full cohesiveness hold (confirming Lemma 3 and Proposition 2).

(ii) Ex-ante symmetric players:

a) For assumption a) all coalitions are stable, deliver the same provision level and payoff but fall short of the social optimum. The larger the difference between $1 - \alpha(n)$, the larger the difference between the equilibrium provision level (global payoff) of stable coalitions and the socially optimal provision level (global payoff) (slightly modifying Proposition 3).

b) For assumption b), the grand coalition is the unique stable coalition. The larger the difference between $1 - \alpha(n)$, the larger is the gain in the grand coalition compared to the non-cooperative equilibrium in terms of global payoffs and provision level (modifying Proposition 3).

(iii) Ex-ante asymmetric players and no transfers:

For both assumptions a strictly effective coalition with respect to the all singleton coalition structure may be stable (modifying Proposition 4). For assumption a) the grand coalition is stable provided that, for all $i \in N$ for which $q_i^A < q_N^A$, $V_i(\alpha(n)q_i^A) \leq V_i(\alpha(n)q_N^A)$ holds ($V_i(\alpha(2)q_i^A) \leq V_i(q_N^A)$ under assumption b)). That is, the smaller $\alpha(n)$ under assumption a) ($\alpha(2)$ under assumption b)), the more likely it is that the grand coalition will be stable. For assumption a), a sufficient condition for the grand coalition being stable is $\alpha(n)q_1^A \leq q_N^A$.

(iv) Ex-ante asymmetric players and transfers:

For both assumptions an effective coalition with respect to the all singleton coalition structure exists (confirming Proposition 5).

Proof. (i) Slight modifications of the proofs of Lemma 3 and Proposition 2 deliver the result. (ii) Symmetric provision levels and payoffs for every $S \subseteq N$ are obvious. Global payoffs are strictly concave in provision levels and $\widetilde{q}$ and $\widetilde{\widetilde{q}}$ increase in $\alpha(\cdot)$. For assumption a), stability of all $S \subseteq N$ is obvious. For assumption b), $V_{i \in S}(S) = V_{j \notin S}(S)$ for all $S \subseteq N$, and $V_{i \in S}(S)$ increases in $\alpha(n-s+1)$ which increases in s and hence $V_{i \in S}(S) > V_{j \notin S}(S \setminus \{i\})$ for all $S \subseteq N$ and $s > 1$. Hence, all coalitions are internally stable but only the grand coalition is externally stable. (iii) Obvious, noting that for all $j \in N$ for which $q_j^A \geq q_N^A$ holds, the incentive to leave is not positive. (iv) Slight modifications of the proof in Proposition 5 delivers the result. ∎

Hence, the alternative assumptions do not change the general incentive structure of the game, all properties established in Lemma 3 and Proposition 2 continue to hold. For assumption a) it is confirmed that ex-ante symmetric players do not render the analysis of coalition formation interesting for the weakest-link technology. For assumption b) this is different but almost by assumption because coalition formation helps to coordinate on provision levels. The larger coalition S, the larger will be $\alpha(n-s+1)$ and the equilibrium provision level, and hence the gain from cooperation compared to the non-cooperative provision level. Result (iii) relates somehow to the modesty effect. For assumption a) if $\alpha(n)$ is sufficiently small

and hence the provision level in a coalition is low, the grand coalition will be stable. This highlights a paradox because the smaller $\alpha(n)$, the smaller will be the provision level and global payoffs in the grand coalition compared to the social optimum. For assumption b) the grand coalition can be stable if the provision level in the grand coalition drops sufficiently when one player leaves, i.e. to a modest provision level. This requires that $\alpha(n - s + 1)$ drops sufficiently from $s = n$ to $s = n - 1$ (i.e. from $\alpha(1) = 1$ to $\alpha(2)$). For assumption b), the paradox disappears if we assume $\alpha(1) = 1$. The grand coalition corresponds to the social optimum, and the global gain from full cooperation compared to no cooperation increases with the distance between $\alpha(1) = 1$ in the grand coalition and $\alpha(n)$ in the all singleton coalition structure. Finally, result (iv) confirms Proposition 4 about the existence of a non-trivial stable coalition in the presence of transfers.

8 Summary and Conclusion

In this paper, we have analyzed the canonical coalition formation model of international environmental agreements (IEAs) under a weakest-link aggregation technology. This technology underlies a large number of important regional or global public goods, such as coordination of migration policies within the EU, compliance with minimum standards in marine law, protecting species whose habitat cover several countries, compliance with targets for fiscal convergence in a monetary union, fighting a fire which threatens several communities, air-traffic control or curbing the spread of an epidemic.

The analysis of IEAs under the summation technology has typically been conducted assuming identical players and highly specific functional forms. Moreover, very few papers analyzed the role of asymmetric players and those are mainly based on simulations. Changing the focus of the analysis to the weakest-link technology has proven fruitful, as we were able to establish a large set of analytical results for general payoff functions. For instance, superadditivity and full cohesiveness are important features of a game, which could generally be established for the weakest-link technology. In contrast, we are unaware of an equivalent proof for the summation technology.

The analysis of the common assumption of symmetric players turned out to produce rather trivial results for the weakest-link technology: policy coordination proved unnecessary as all coalitions are stable and lead to the same Pareto optimal outcome. Hence, the bulk of the paper was devoted to the analysis of the role of asymmetric players. We showed that without transfers, though all coalitions are Pareto-optimal, no coalition is stable which departs from non-cooperative provision levels. However, if an optimal transfer is used to balance asymmetries, a non-trivial coalition exists, associated with a provision level strictly

above the non-cooperative level. We analyzed the kind and degree of asymmetry that is conducive to cooperation: a set of (weak) players, who prefer a similar provision below the average and one (strong) player with a preference for a provision level well above the average. This ensures that there is no weakest-link outlier at the bottom and one player with a very high benefit-cost ratio well above all other signatories, who compensates all other signatories for their contributions to an efficient cooperative agreement. For such an extremely positively skewed distribution of interests regarding optimal provision levels, we could show that even the grand coalition is stable. Unfortunately, such a distribution also implied that the "paradox of cooperation" continues to hold for this technology: asymmetries which are conducive to stability of coalitions yield low welfare gains from cooperation, and vice versa.

As monetary transfers play a crucial role in enabling successful cooperation in the light of asymmetric players, it seems suggestive to analyze the role of in-kind transfers in future research. However, as argued above, general analytical results will be much more difficult to obtain if at all. It is also clear that we focused on the most widespread coalition model and stability concept used in the literature on IEAs, and hence other concepts could be considered (Bloch 1997, Finus and Rundshagen 2009 and Yi 1997). Internal and external stability implies that after a player leaves the coalition, the remaining coalition members remain in the coalition. In the context of a positive externality game, this is weakest possible punishment after a deviation and hence implies the most pessimistic assumption about stability. This appears to be a good benchmark, also because we could show that even for this assumption the grand coalition can be stable with transfers. Without transfers, other stability concepts would come to similar negative conclusion as individual rationality is a necessary condition for almost all sensible equilibrium concepts and without transfers we could show that this condition is violated. Also the assumption of open membership is a pessimistic assumption regarding stability in a positive externality game as shown in Finus and Rundshagen (2009). In other words, those coalitions which we have identified as being stable would also be stable under exclusive membership. And, again, the grand coalition is anyway immune to external accession.

References

[1] Anderson, S.P., J.K. Goree and C.A. Holt (2001), Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium. *Games and Economic Behavior,* vol. 34(2), pp. 177-199.

[2] Arce, M. (2001), Leadership and the aggregation of international collective action. *Oxford Economic Papers,* vol. 53, pp. 114-137.

[3] Arce, M. and T. Sandler (2001), Transnational public goods: strategies and institutions. *European Journal of Political Economy,* vol. 17, pp. 493–516.

[4] Barrett, S. (1994), Self-enforcing international environmental agreements. *Oxford Economic Papers*, vol. 46, pp. 878-894.

[5] Barrett, S. (2001), International cooperation for sale. *European Economic Review*, vol. 45, pp. 1835-1850.

[6] Barrett, S. (2002), Consensus treaties. *Journal of Institutional and Theoretical Economics*, vol. 158(4), pp. 529-547.

[7] Battaglini, M. and B. Harstad (forthcoming), Participation and Duration of Environmental Agreements. *Journal of Political Economy.*

[8] Bergstrom, T.C., L. Blume and H. Varian (1986), On the private provision of public goods. *Journal of Public Economics,* vol. 29, pp. 25–49.

[9] Bergstrom, T.C., L. Blume and H. Varian (1992), Uniqueness of Nash equilibrium in private provision of public goods: an improved proof. *Journal of Public Economics,* vol. 49, pp. 391–392.

[10] Bloch, F. (2003), Non-Cooperative Models of Coalition Formation in Games with Spillovers. In: Carraro, C. (ed.), *The Endogenous Formation of Economic Coalitions.* Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 35–79.

[11] Brandts, J. and D.J. Cooper (2006), A change would do you good. An experimental study on how to overcome coordination failure in organizations. *American Economic Review*, vol. 96(3), pp. 669–693.

[12] Cachon, G. P. and C.F. Camerer (1996), Loss-avoidance and forward induction in experimental coordination games. *Quarterly Journal of Economics*, vol. 111(1), pp. 165–194.

[13] Caparrós, A., E. Giraud-Héraud, A. Hammoudi, and T. Tazdaït (2011), Coalition stability with heterogeneous agents. *Economics Bulletin,* vol. 31, pp. 286-296.

[14] Carraro, C. and D. Siniscalco (1993), Strategies for the international protection of the environment. *Journal of Public Economics*, vol. 52(3), pp. 309-328.

[15] Cornes, R.C. (1993), Dyke maintenance and other stories: some neglected types of public goods. *Quarterly Journal of Economics,* vol. 107, pp. 259–271.

[16] Cornes, R.C. and R. Hartley (2007a), Weak links, good shots and other public good games: building on BBV. *Journal of Public Economics*, vol. 91, pp. 1684-1707.

[17] Cornes, R.C. and R. Hartley (2007b), Aggregative public good games. *Journal of Public Economic Theory*, vol. 9(2), pp. 201–219.

[18] Courtois, P. and G. Haeringer (2012), Environmental cooperation: ratifying second-best agreements. *Public Choice*, vol. 151, pp. 565–584.

[19] d'Aspremont, C., A. Jacquemin, J. J. Gabszewicz and J. A. Weymark (1983), On the stability of collusive price leadership. *Canadian Journal of Economics,* vol. 16(1), pp. 17–25.

[20] Diamantoudi, E. and E.S. Sartzetakis (2006), Stable international environmental agreements: an analytical approach. *Journal of Public Economic Theory*, vol. 8(2), pp. 247-263.

[21] Eyckmans, J. and M. Finus (2004), *An Almost Ideal Sharing Scheme for Coalition Games with Externalities.* CLIMNEG Working Paper No. 62, University of Leuven (K.U.L.), Belgium.

[22] Eyckmans, J., M. Finus and Lina Mallozzi (2012), *A New Class of Welfare Maximizing Sharing Rules for Partition Function Games with Externalities.* Bath Economics Research Paper 6-2012.

[23] Finus, M. (2003), Stability and Design of International Environmental Agreements: The Case of Global and Transboundary Pollution. In: Folmer, H. and T. Tietenberg (eds.), *International Yearbook of Environmental and Resource Economics 2003/4.* Edward Elgar, Cheltenham, UK et al., ch. 3, pp. 82-158.

[24] Finus, M. and A. Caparrós (2015), *Game Theory and International Environmental Cooperation: Essential Readings.* Edward Elgar, Cheltenham.

[25] Finus, M. and S. Maus (2008), Modesty may pay! *Journal of Public Economic Theory*, vol. 10(5), pp. 801-826.

[26] Finus, M. and B. Rundshagen (2009), Membership rules and stability of coalition structures in positive externality games. *Social Choice and Welfare*, vol. 32, pp. 389-406.

[27] Fraser, C.D. (1992), The uniqueness of Nash equilibrium in the private provision of public goods: an alternative proof. *Journal of Public Economics,* vol. 49, pp. 389–390.

[28] Fuentes-Albero, C. and S.J. Rubio (2010), Can international environmental cooperation be bought? *European Journal of Operational Research,* vol. 202, 255–264.

[29] Harrison, G. W. and J. Hirshleifer (1989). An experimental evaluation of weakest link/best shot models of public goods. *Journal of Political Economy*, vol. 97(1), pp. 201–225.

[30] Hirshleifer, J. (1983), From weakest-link to best-shot: The voluntary provision of public goods. *Public Choice*, vol. 41, pp. 371-386.

[31] Kogan, S., A.M. Kwasnica and R.A. Weber (2011), Coordination in the presence of asset markets. *American Economic Review*, vol. 101(2), pp. 927–947.

[32] McGinty, M. (2007), International environmental agreements among asymmetric nations. *Oxford Economic Papers*, vol. 59, pp. 45–62.

[33] Monderer, D. and L. S. Shapley (1996), Potential Games. *Games and Economic Behavior*, vol. 14, pp. 124-143.

[34] Pavlova, Y. and A. de Zeeuw (2013), Asymmetries in international environmental agreements. *Environment and Development Economics*, vol. 18, pp. 51-68.

[35] Ray, D. and R. Vohra (2001), Coalitional power and public goods. *Journal of Political Economy*, vol. 109, pp. 1355-1384.

[36] Rubio, S.J. and A. Ulph (2006), Self-enforcing international environmental agreements revisited. *Oxford Economic Papers*, vol. 58, pp. 233-263.

[37] Sandler, T. (1998), Global and regional public goods: a prognosis for collective action. *Fiscal Studies,* vol. 19, pp. 221-247.

[38] Sandler, T. (2006), Regional public goods and international organizations. *Review of International Organizations*, vol. 1, p. 5-25.

[39] Sandler, T. and D. Arce (2002), A conceptual framework for understanding global and transnational public goods for health. *Fiscal Studies,* vol. 23, pp. 195-222.

[40] Van Huyck, J. B., R. C. Battalio and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, vol. 80(1), pp. 234–248.

[41] Vicary, S. (1990), Transfers and the weakest link: an extension of Hirshleifer's analysis. *Journal of Public Economics,* vol. 43, pp. 375–394.

[42] Vicary, S. and T. Sandler (2002), Weakest-link public goods: giving in-kind or transferring money. *European Economic Review,* vol. 46, pp. 1501–1520.

[43] Weber, R. A. (2006), Managing growth to achieve efficient coordination in large groups. *American Economic Review,* vol. 96, pp. 114–126.

[44] Weikard, H.-P. (2009), Cartel stability under an optimal sharing rule. *Manchester School,* vol. 77(5), pp. 575–593.

[45] Yi, S.-S. (1997), Stable coalition structures with externalities. *Games and Economic Behavior,* vol. 20, pp. 201-237.

A   Summation Technology: Main Results

The first order conditions of non-signatories assuming payoff function (1), with $Q = \sum_{j \in N} q_j$, gives $B_i' - C_i' = 0$. Total differentiation gives $B_i'' dQ - C_i'' dq_i = 0$ or $\frac{dq_i}{dQ} = \frac{B_i''}{C_i''} < 0$ and hence the replacement functions of non-signatories are downward sloping. For signatories, we have $\frac{dq_i}{dQ} = \frac{\sum_{i \in S} B_i''}{C_i''} < 0$ and hence taken together the aggregate replacement function is downward sloping, implying a unique second stage equilibrium for every $S \subseteq N$. Alternatively for reaction functions, we derive for non-signatories $B_i'' dq_i + B_i'' dq_{-i} - C_i'' dq_i = 0$ or $\frac{dq_i}{dq_{-i}} = \frac{B_i''}{B_i'' - C_i''} < 0$ with $\frac{B_i''}{B_i'' - C_i''} > -1$. For signatories, we have $\frac{dq_i}{dq_{-i}} = \frac{\sum_{i \in S} B_i''}{\sum_{i \in S} B_i'' - C_i''} < 0$ with $\frac{\sum_{i \in S} B_i''}{\sum_{i \in S} B_i'' - C_i''} > -1$. If and only if $B_i'' = 0$, then replacement and reaction functions have a slope of zero and are orthogonal but the second stage equilibrium is still unique.

Consider a coalition S with $Q^*(S)$ and a coalition $S \backslash \{i\}$ with $Q^*(S \backslash \{i\})$. We want to prove $Q^*(S) > Q^*(S \backslash \{i\})$. Suppose the opposite, namely, $Q^*(S) \leq Q^*(S \backslash \{i\})$ was true. Then,

$$\forall\, k \notin S : C_k'(q_k(S \backslash \{i\})) = B_k'(Q^*(S \backslash \{i\})) \leq B_k'(Q^*(S)) = C_k'(q_k(S))$$

$$\forall\, i \in S : C_i'(q_i(S \backslash \{i\})) = B_i'(Q^*(S \backslash \{i\})) < \sum_{i \in S} B_i'(Q^*(S \backslash \{i\})) < \sum_{i \in S} B_i'(Q^*(S)) = C_k'(q_k(S))$$

implying $q_k(S \backslash \{i\}) \leq q_k(S)$ and $q_i(S \backslash \{i\}) < q_i(S)$ which contradicts the initial assumption. Hence, the game is strictly effective. (Because $Q^*(S) > Q^*(S \backslash \{i\})$ for all $S \subseteq N$, even for symmetric players the equilibrium provision vectors are different for all possible coalitions.) Because $\frac{dq_k}{dQ} \leq 0$, the move from $S \backslash \{i\}$ to S implies equal or lower costs but

strictly higher benefits for all non-signatories and hence the positive externality property holds strictly.

A sufficient condition for superadditivity to hold is $B_i'' = 0$. Moving from $S\backslash\{i\}$ to S, $q_k(S\backslash\{i\}) = q_k(S) \ \forall \ k \notin S$ and $q_i(S\backslash\{i\}) < q_i(S) \ \forall \ i \in S$ where $q_i(S)$ follows $\max \sum_{i\in S} V_i(S)$. Hence, $\sum_{i\in S} V_i(S) > \sum_{j\in S\backslash\{i\}} V_j(S\backslash\{i\}) + V_i(S\backslash\{i\})$ must follow. For an example where superadditivity fails, consider the following payoff function: $V_i = b\left(aQ - \frac{1}{2}Q^2\right) - \frac{c}{2}q_i^2$ with a, b, and c positive parameters which are the same for all players. Let n be the total number of players and s the number of signatories. Then $Q^*(s) = \frac{ba(s^2-s+n)}{bs^2+bn-bs+c}$, $q_{i\notin S} = \frac{ba}{bs^2+bn-bs+c}$ and $q_{i\in S} = sq_{i\notin S}$. Computing $\Delta := V_{i\in S}(s=2) - V_{i\notin S}(s=1)$ (in which case superadditivity and internal stability are the same conditions) gives $\Delta = -\frac{b^2a^2c\Psi}{2(bn+2b+c)^2(bn+c)^2}$ with $\Psi = \gamma^2(3n^2-4n-4)+\gamma(2n-8)-1$ and $\gamma = b/c$. Now assume $n = 4$, then $\Psi = 28\gamma^2-1$ and hence $\Psi > 0$ and $\Delta < 0$ if γ is sufficiently large. It can be shown that if $\Delta < 0$, also no larger coalition is internally stable. For an example which shows that for no transfers, not all coalitions are Pareto-optimal and a stable non-trivial coalition may or may not exist, consider $V_i = b_iQ - \frac{c_i}{2}q_i^2$ with b_i, and c positive parameters. Note that for this payoff function superadditivity holds. Signatories' equilibrium provision level is given by $q_{i\in S} = \frac{\sum_{i\in S} b_i}{c}$ and non-signatories provision level by $q_{i\notin S} = \frac{b_i}{c}$. Assume for simplicity $n = 3$ and let $c = 1$ $\forall \ i \in N$. Example 1 assumes $b_1 = 1$, $b_2 = 2$ and $b_3 = 3$ and Example 2 assumes $b_1 = 1$, $b_2 = 1.1$ and $b_3 = 1.2$ with the results displayed in Table 1.

[**Table 1 about here**]

In Example 1, no non-trivial coalition is stable but all coalitions are Pareto-optimal. In Example 2, all two-player coalitions are stable, except a coalition of player 1 and 2, but only the grand coalition and the coalition of player 2 and 3 are Pareto-optimal.

B Weakest-Link Technology: Proofs

B.1 Lemma 4

The first order conditions in an interior equilibrium are $\hat{c}\frac{\partial C}{\partial q} = \hat{b}\frac{\partial B}{\partial q}$ where $\hat{b} = \sum_{i\in S} b_i$ and $\hat{c} = \sum_{i\in S} c_i$. Thus, we can define functions f and h as follows:

$$\begin{aligned} f(q) &= \frac{C'(q)}{B'(q)} = \frac{\hat{b}}{\hat{c}} = \frac{\bar{b}}{\bar{c}} = \theta_S \ , \\ q_S^A &= f^{-1}(\theta_S) = h(\theta_S) \ . \end{aligned}$$

To show that h is strictly concave and strictly increasing, we show that f is strictly increasing and strictly convex:

$$\frac{\partial f(q)}{\partial q} = \frac{B'(q)C''(q) - C'(q)B''(q)}{\left(B'(q)\right)^2} > 0 \ ,$$

$$\frac{\partial^2 f(q)}{\partial q \partial q} = \frac{C'''(q)\left(B'(q)\right)^2 + 2C'(q)\left(B''(q)\right)^2 - 2C''(q)B'(q)B''(q) - C'(q)B'(q)B'''(q)}{\left(B'(q)\right)^3} > 0$$

which is true due to the assumptions about the first and second derivatives summarized in Assumption 1 and the assumptions about the third derivatives mentioned in Section 6.2, namely, $C''' \geq 0$ and $B''' \leq 0$ (or if $B''' > 0$, $B'''(q) < -2B''(q)C''(q)/C'(q)$).

B.2 Proposition 6

Before proving the proposition itself, we proof a lemma that is useful for the subsequent analysis.

Lemma 5 *The function $k_i(\theta) = v_i(h(\theta))$ is strictly concave and increasing in θ, $\theta \in [0, \theta_i^A]$.*

Proof. $k_i'(\theta) = v_i'(h(\theta))h'(\theta)$ is increasing if $v_i'(h(\theta)) = \frac{\partial v(q)}{\partial q} > 0$, as we have shown in Lemma 4 that $h(\theta)$ is strictly concave and increasing and we thus have $h'(\theta) > 0$. Due to Assumption 1, v is a strictly concave function with respect to q_i with a maximum at $q_i^A = h(\theta_i^A)$, and it is therefore increasing for $q_i \in [0, q_i^A]$. As $h(\theta_i)$ is increasing everywhere we also know that $v_i(h(\theta_i))$ is increasing for $\theta_i \in [0, \theta_i^A]$ because for any $\theta_i \in [0, \theta_i^A]$ we know that $q_i = h(\theta_i) \leq q_i^A$. Thus, we have $v_i'(h(\theta)) = \frac{\partial v(q)}{\partial q} > 0$.
For k to be strictly concave, we need:

$$k''(\theta) = v''(h(\theta))\left(h'(\theta)\right)^2 + v'(h(\theta))h''(\theta) < 0. \tag{B.1}$$

We have just shown that $v_i'(h(\theta)) = \frac{\partial v(q)}{\partial q} > 0$ for $\theta_i \in [0, \theta_i^A]$, and by the strict concavity of v with respect to $q = h(\theta)$, due to Assumption 1, and that of h with respect to θ, shown in Lemma 4, we know $v''(h(\theta)) < 0$ and $h''(\theta) < 0$. Hence, $k''(\theta) < 0$ and $k(\theta)$ is strictly concave. ∎

Before proceeding, let us first write equation (4) in the text in a more disaggregated form:

$$\sigma_S(S, \Theta) = \sum_{i \in S} v_i(\theta_y) - \sum_{i \in S_1} v_i(\theta_i) - \sum_{i \in S_2} v_i(\theta_{S \setminus \{i\}}) - \sum_{i \in S_3} v_i(\theta_y) \geq 0. \tag{B.2}$$

After the marginal changes in the distribution mentioned in the Proposition, b_k becomes $b_k - \epsilon$ and b_l becomes $b_l + \epsilon$. These changes do neither affect θ_m nor $\theta_S = \sum_{i \in S} b_i / \sum_{i \in S} c_i$, because $c_i = c \ \forall \ i \in S$.

We now proof the three conditions of the θ-values listed in Proposition 6.

i) $\theta_l < \theta_k \leq \theta_y$. Consider first the case where both players are in S_1, i.e. $\theta_l < \theta_k < \theta_y$. We denote the new valuation function by $\tilde{v}$ and, slightly abusing notation, $\theta_{\{k-\epsilon\}}$ and $\theta_{\{l+\epsilon\}}$ the two values that have changed in $\tilde{\Theta}$. The third and fourth sum in condition (B.2) remain unchanged. In the first sum in (B.2) the value of θ_y is the same, but the valuation function has changed for players k and l. However, as $v_k(\theta_y) + v_l(\theta_y) = \tilde{v}_k(\theta_y) + \tilde{v}_l(\theta_y)$ still holds, the aggregate value of the sum does not change. Thus, only the second sum in condition (B.2) changes and in order for $\sigma_S(S, \Theta) < \sigma_S(S, \tilde{\Theta})$ to hold, we need:

$$v_k(\theta_k) + v_l(\theta_l) > \tilde{v}_k\left(\theta_{\{k-\epsilon\}}\right) + \tilde{v}_l\left(\theta_{\{l+\epsilon\}}\right)$$

or

$$\begin{aligned}
& \left[b_k B(h(\theta_k)) - cC(h(\theta_k))\right] - \left[b_k B(h(\theta_{\{k-\epsilon\}})) - cC(h(\theta_{\{k-\epsilon\}}))\right] \hspace{2em} \text{(B.3)} \\
& + \epsilon \left[B(h(\theta_{\{k-\epsilon\}})) - B(h(\theta_{\{l+\epsilon\}}))\right] \\
> \; & \left[b_l B(h(\theta_{\{l+\epsilon\}})) - cC(h(\theta_{\{l+\epsilon\}}))\right] - \left[b_l B(h(\theta_l)) - cC(h(\theta_l))\right] \; .
\end{aligned}$$

Recalling the definition of derivatives, dividing both sides by ϵ and taking the limit $\epsilon \to 0$, inequality (B.3) becomes:

$$\left[B(h(\theta_{\{k-\epsilon\}})) - B(h(\theta_{\{l+\epsilon\}}))\right] + v_k'(\theta)\big|_{\theta_{\{k-\epsilon\}}} \geq v_l'(\theta)\big|_{\theta_l} \; . \hspace{2em} \text{(B.4)}$$

For $\epsilon \to 0$, $\theta_k > \theta_l$ implies $\theta_{k-\epsilon} \geq \theta_{l+\epsilon}$ and therefore the first term on the LHS of inequality (B.4) is non-negative. Thus, inequality (B.4) always holds as we have $v_k'(\theta)\big|_{\theta_{\{k-\epsilon\}}} \gtrsim v_k'(\theta)\big|_{\theta_k} = 0$ and $v_l'(\theta)\big|_{\theta_l} = 0$, given that $\theta_{\{k-\epsilon\}} < \theta_k$, θ_k and θ_l maximize v_k and v_l, respectively, and $v_i(\theta)$ is an increasing and strictly concave function for $\theta \in [0, \theta_i^A]$ by Lemma 4. If player k is initially in S_3, i.e. if $\theta_k = \theta_y$, the last sum in (B.2) does change, but the relevant marginal changes are still summarized in condition (B.3) and the proof continues to hold.

(ii) $\theta_y < \theta_k \leq \theta_l$. Assume first $k, l \in S_2$. Following a similar argument as before, it is clear that only the third sum in condition (B.2) has changed, and for $\sigma_S(S, \Theta) < \sigma_S(S, \tilde{\Theta})$ to hold, we need:

$$v_k\left(\theta_{S \setminus k}\right) + v_l\left(\theta_{S \setminus l}\right) > \tilde{v}_k\left(\theta_{S \setminus \{k-\epsilon\}}\right) + \tilde{v}_l\left(\theta_{S \setminus \{l+\epsilon\}}\right) \hspace{2em} \text{(B.5)}$$

or

$$\left[b_l B(h(\theta_{S\setminus l})) - cC(h(\theta_{S\setminus l}))\right] - \left[b_l B(h(\theta_{S\setminus\{l+\epsilon\}})) - cC(h(\theta_{S\setminus\{l+\epsilon\}}))\right] \qquad \text{(B.6)}$$
$$+\epsilon \left[B(h(\theta_{S\setminus\{k-\epsilon\}})) - B(h(\theta_{S\setminus\{l+\epsilon\}}))\right]$$
$$> \left[b_k B(h(\theta_{S\setminus\{k-\epsilon\}})) - cC(h(\theta_{S\setminus\{k-\epsilon\}}))\right] - \left[b_k B(h(\theta_{S\setminus k})) - cC(h(\theta_{S\setminus k}))\right].$$

Noting that $\theta_{S\setminus\{l+\epsilon\}} < \theta_{S\setminus l}$ and $\theta_{S\setminus k} < \theta_{S\setminus\{k-\epsilon\}}$, we have that $\theta_{S\setminus\{l+\epsilon\}} < \theta_{S\setminus\{k-\epsilon\}}$ and the third term on the LHS of inequality (B.6) is positive. Thus, a sufficient condition is:

$$\left[b_l B(h(\theta_{S\setminus l})) - cC(h(\theta_{S\setminus l}))\right] - \left[b_l B(h(\theta_{S\setminus\{l+\epsilon\}})) - cC(h(\theta_{S\setminus\{l+\epsilon\}}))\right]$$
$$> \left[b_k B(h(\theta_{S\setminus\{k-\epsilon\}})) - cC(h(\theta_{S\setminus\{k-\epsilon\}}))\right] - \left[b_k B(h(\theta_{S\setminus k})) - cC(h(\theta_{S\setminus k}))\right]$$

and dividing both sides by ϵ and taking the limit $\epsilon \to 0$ this becomes:

$$v_l'\left(\theta\right)\big|_{\theta_{S\setminus\{l+\epsilon\}}} > v_k'\left(\theta\right)\big|_{\theta_{S\setminus k}}. \qquad \text{(B.7)}$$

Because we have

$$v_j'\left(\theta\right) = v_j'(h(\theta))h'(\theta) = \left[b_j B'(h(\theta)) - c_j C'(h(\theta))\right] h'(\theta) > 0$$

inequality (B.7) can be written as:

$$\left[b_l B'(h(\theta_{S\setminus\{l+\epsilon\}})) - cC'(h(\theta_{S\setminus\{l+\epsilon\}}))\right] h'(\theta_{S\setminus\{l+\epsilon\}}) > \left[b_k B'(h(\theta_{S\setminus k})) - cC'(h(\theta_{S\setminus k}))\right] h'(\theta_{S\setminus k}).$$

Because $\theta_l > \theta_k$, we also know that

$$\left[b_l B'(h(\theta_{S\setminus\{l+\epsilon\}})) - cC'(h(\theta_{S\setminus\{l+\epsilon\}}))\right] h'(\theta_{S\setminus\{l+\epsilon\}}) > \left[b_k B'(h(\theta_{S\setminus\{l+\epsilon\}})) - cC'(h(\theta_{S\setminus\{l+\epsilon\}}))\right] h'(\theta_{S\setminus\{l+\epsilon\}}).$$

Hence, a sufficient condition for inequality (B.7) to hold is

$$\left[b_k B'(h(\theta_{S\setminus\{l+\epsilon\}})) - cC'(h(\theta_{S\setminus\{l+\epsilon\}}))\right] h'(\theta_{S\setminus\{l+\epsilon\}}) > \left[b_k B'(h(\theta_{S\setminus k})) - cC'(h(\theta_{S\setminus k}))\right] h'(\theta_{S\setminus k})$$

or

$$v_k'\left(\theta\right)\big|_{\theta_{S\setminus\{l+\epsilon\}}} > v_k'\left(\theta\right)\big|_{\theta_{S\setminus k}}.$$

This holds for $\theta_{S\setminus\{l+\epsilon\}} < \theta_{S\setminus k} < \theta_k^A$, as $v_k(\theta)$ is an increasing and strictly concave function for $\theta \in [0, \theta_i^A]$ by Lemma 4.

Assume now $k, l \in S_3$. Then, equation (B.5) simplifies to

$$v_k\left(\theta_y\right) + v_l\left(\theta_y\right) = \tilde{v}_k\left(\theta_y\right) + \tilde{v}_l\left(\theta_y\right)$$

and $\sigma_S(S, \Theta) = \sigma_S(S, \tilde{\Theta})$.

If $k \in S_3$ and $l \in S_2$ (the opposite is not possible, see figure 2) then equation (B.5) simplifies to

$$v_k\left(\theta_y\right) + v_l\left(\theta_{S\backslash l}\right) > \tilde{v}_k\left(\theta_y\right) + \tilde{v}_l\left(\theta_{S\backslash\{l+\epsilon\}}\right)$$

or

$$\begin{aligned}
0 \quad < \quad & \left[b_l B(h(\theta_{S\backslash l})) - cC(h(\theta_{S\backslash l}))\right] - \left[b_l B(h(\theta_{S\backslash\{l+\epsilon\}})) - cC(h(\theta_{S\backslash\{l+\epsilon\}}))\right] \\
& + \epsilon\left[B(h(\theta_y)) - B(h(\theta_{S\backslash\{l+\epsilon\}}))\right]
\end{aligned}$$

which holds because $v_l'(\theta)|_{\theta_{S\backslash\{l+\epsilon\}}} > 0$ (for the first two terms) and $\theta_{S\backslash\{l+\epsilon\}} < \theta_y$ (for the last term).

Finally, in the "marginal case" where initially $\theta_y < \theta_k \le \theta_l$ but finally $\theta_{k-\epsilon} = \theta_y < \theta_{l+\epsilon}$, it is easy to check that the conclusions derived above hold. One just needs to note that in case $k, l \in S_2$, $\theta_{S\backslash\{k-\epsilon\}} = \theta_y$ holds.

(iii) $\theta_k \le \theta_y < \theta_l$ and $\theta_{\{k-\epsilon\}} \ge \theta_{S\backslash\{l+\epsilon\}}$. Assume first $k \in S_1$ and $l \in S_2$. Because nothing has changed for the remaining players, in order to have $\sigma_S(S, \Theta) < \sigma_S(S, \tilde{\Theta})$, we need:

$$\begin{aligned}
& \left[b_l B(h(\theta_{S\backslash l})) - cC(h(\theta_{S\backslash l}))\right] - \left[b_l B(h(\theta_{S\backslash\{l+\epsilon\}})) - cC(h(\theta_{S\backslash\{l+\epsilon\}}))\right] && \text{(B.8)} \\
& + \epsilon\left[B(h(\theta_{\{k-\epsilon\}})) - B(h(\theta_{S\backslash\{l+\epsilon\}}))\right] \\
> \quad & \left[b_k B(h(\theta_{\{k-\epsilon\}})) - cC(h(\theta_{\{k-\epsilon\}}))\right] - \left[b_k B(h(\theta_k)) - cC(h(\theta_k))\right] \ .
\end{aligned}$$

If $\theta_{\{k-\epsilon\}} > \theta_{S\backslash\{l+\epsilon\}}$, the third term on the LHS in inequality (B.8) is positive and a sufficient condition for (B.8) to hold is:

$$\begin{aligned}
& \left[b_l B(h(\theta_{S\backslash l})) - cC(h(\theta_{S\backslash l}))\right] - \left[b_l B(h(\theta_{S\backslash\{l+\epsilon\}})) - cC(h(\theta_{S\backslash\{l+\epsilon\}}))\right] \\
& + \left[b_k B(h(\theta_k)) - cC(h(\theta_k))\right] - \left[b_k B(h(\theta_{\{k-\epsilon\}})) - cC(h(\theta_{\{k-\epsilon\}}))\right] > 0 \ .
\end{aligned}$$

Noting that $\theta_{S\backslash\{l+\epsilon\}} < \theta_{S\backslash l}$, dividing both sides by ϵ and taking the limit $\epsilon \to 0$, this becomes:

$$v_l'(\theta)|_{\theta_{S\backslash\{l+\epsilon\}}} + v_k'(\theta)|_{\theta_{\{k-\epsilon\}}} > 0 \ .$$

This holds for $\theta_{S\backslash\{l+\epsilon\}} < \theta_l^A$ and $\theta_{\{k-\epsilon\}} < \theta_k^A$, as $v_i(\theta)$ is an increasing and strictly concave function for $\theta \in [0, \theta_i^A]$ by Lemma 5.

Consider now the case $k \in S_3$ and $l \in S_2$. This implies that we are considering the particular case where initially $\theta_k = \theta_y < \theta_l$ and finally $\theta_{k-\epsilon} < \theta_y < \theta_{l+\epsilon}$. Because now $\theta_{\{k-\epsilon\}} \ge \theta_{S\backslash\{l+\epsilon\}}$ always holds (because $\theta_{\{k-\epsilon\}}$ is infinitely close to θ_y) we have that $B(h(\theta_{\{k-\epsilon\}})) \ge B(h(\theta_{S\backslash\{l+\epsilon\}}))$. Thus, a sufficient condition for the equivalent to inequality

(B.8) to hold is:

$$\left[b_l B(h(\theta_{S\setminus l})) - cC(h(\theta_{S\setminus l})) \right] - \left[b_l B(h(\theta_{S\setminus\{l+\epsilon\}})) - cC(h(\theta_{S\setminus\{l+\epsilon\}})) \right] \qquad (B.9)$$
$$+ \left[b_k B(h(\theta_k)) - cC(h(\theta_k)) \right] - \left[b_k B(h(\theta_y)) - cC(h(\theta_y)) \right] > 0 \ .$$

We know that k was in S_3, i.e. $\theta_y \lesssim \theta_k$, but we also know that it only was in S_3 at the margin, as $(k - \epsilon) \in S_1$ and thus $\theta_{\{k-\epsilon\}} < \theta_y$. Hence $\theta_y = \theta_k$ or slightly above, i.e. $\theta_y \lesssim \theta_k$. Thus, either the second square bracket in inequality (B.9) is zero or it is equal to $v'_k(\theta)|_{\theta_y^A} > 0$. As the first square bracket is also positive (see above), the condition always holds.

For the case $k \in S_1$ and $l \in S_3$ we have that $\theta_{S\setminus l} > \theta_{S\setminus\{l+\epsilon\}} \geq \theta_y > \theta_k > \theta_{\{k-\epsilon\}}$ and hence the condition $\theta_{\{k-\epsilon\}} \geq \theta_{S\setminus\{l+\epsilon\}}$ fails. The same holds for the case $k, l \in S_3$, as in this case $\theta_{S\setminus l} > \theta_{S\setminus\{l+\epsilon\}} \geq \theta_y = \theta_k > \theta_{\{k-\epsilon\}}$.

B.3 Variance and skewness coefficient

We now define the conditions under which a marginal increase of stability (through the changes in Proposition 6) increases the variance and the skewness of the θ_i-distribution. Applying the standard definition of the variance (second moment) and the Fisher-Pearson coefficient of skewness to the distribution of θ_i-values (respectively b_i-values) we obtain the following definition:

Definition 6 *The skewness coefficient $g(\Theta)$ of the distribution Θ of θ_i-values within a coalition is:*

$$g(\Theta) = \frac{m_3(\Theta)}{(m_2(\Theta))^{3/2}}; \ m_2(\Theta) = \frac{1}{s} \sum_{i \in S} (\theta_i - \theta_S)^2; \ m_3(\Theta) = \frac{1}{s} \sum_{i \in S} (\theta_i - \theta_S)^3 \ ,$$

where $\theta_S = \frac{1}{s} \sum_{i \in S} \theta_i$ is the mean, $m_2(\Theta)$ the second moment (variance) and $m_3(\Theta)$ the third moment of the distribution Θ, respectively, and s is the number of coalition members.

Relating the distributions Θ and $\tilde{\Theta}$ defined in Proposition 6 to the variance and the skewness coefficient, we obtain the following proposition:[26]

Proposition 9 *Consider a coalition S determining the equilibrium provision level and two distributions Θ and $\tilde{\Theta}$ as defined in Proposition 6, then $m_2(\tilde{\Theta}) > m_2(\Theta)$ for cases (ii) and*

---

[26] If the assumption $c_i = c \ \forall \ i$ is substituted by the assumption $b_i = b \ \forall \ i$ Proposition 9 continues to hold. For the general case where players differ in their b_i's and their c_i's, the Proposition would continue to hold, but the coefficient $g(\Theta)$ would not any more be the standard skewness coefficient as $\theta_S = \frac{\sum_{i \in S} b_i}{\sum_{i \in S} c_i}$ is not anymore the average over all θ_i's (as it is when either $b_i = b \ \forall \ i$ or $c_i = c \ \forall \ i$).

(iii) in Proposition 6 and $g(\tilde{\Theta}) > g(\Theta)$ in all three cases if and only if

$$m_2(\Theta) > \frac{m_3(\Theta)}{\theta_k + \theta_l - 2\theta_S}. \tag{B.10}$$

Proof. All the marginal changes described in Proposition 6 imply that b_k becomes $b_k - \epsilon$ and b_l becomes $b_l + \epsilon$. Hence, $m_2(\tilde{\Theta}) > m_2(\Theta)$ implies

$$\frac{1}{n} \sum_{i \in S \setminus k, l} \left(\frac{b_i}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_k}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_l}{c} - \frac{b_S}{c} \right)^2$$

$$< \frac{1}{n} \sum_{i \in S \setminus k, l} \left(\frac{b_i}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_k - \epsilon}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_l + \epsilon}{c} - \frac{b_S}{c} \right)^2,$$

which simplifies to $b_k < \epsilon + b_l$. This holds for cases (ii) and (iii) in Proposition 6 but not for case (i). In addition, $g(\tilde{\Theta}) > g(\Theta)$ implies

$$\frac{\frac{1}{n} \left(\sum_{i \in S \setminus k, l} \left(\frac{b_i}{c} - \frac{b_S}{c} \right)^3 + \left(\frac{b_k}{c} - \frac{b_S}{c} \right)^3 + \left(\frac{b_l}{c} - \frac{b_S}{c} \right)^3 \right)}{\left(\frac{1}{n} \sum_{i \in S \setminus k, l} \left(\frac{b_i}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_k}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_l}{c} - \frac{b_S}{c} \right)^2 \right)^{3/2}}$$

$$< \frac{\frac{1}{n} \left(\sum_{i \in S \setminus k, l} \left(\frac{b_i}{c} - \frac{b_S}{c} \right)^3 + \left(\frac{b_k - \epsilon}{c} - \frac{b_S}{c} \right)^3 + \left(\frac{b_l + \epsilon}{c} - \frac{b_S}{c} \right)^3 \right)}{\left(\frac{1}{n} \sum_{i \in S \setminus k, l} \left(\frac{b_i}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_k - \epsilon}{c} - \frac{b_S}{c} \right)^2 + \frac{1}{n} \left(\frac{b_l + \epsilon}{c} - \frac{b_S}{c} \right)^2 \right)^{3/2}}.$$

This yields inequality (B.10) after tedious algebraic manipulations which are available from the authors upon request. ∎

Note that for case (i) in Proposition 6 we have that $\theta_k + \theta_l - 2\theta_S < 0$ and hence condition B.10 holds for any positively skewed distribution, where $m_3(\Theta) > 0$ (as $m_2(\Theta)$ is always positive), and for distributions that are not too negatively skewed (where the absolute value of $m_3(\Theta)$ is smaller than $m_2(\Theta)(\theta_k + \theta_l - 2\theta_S)$). For cases (ii) to (iii) in Proposition 6 we know that[27] $\theta_k + \theta_l - 2\theta_S > 0$, and thus condition B.10 holds for negatively skewed or not too positively skewed distributions. That is, the intuition that all marginal changes proposed in Proposition 6 increase skewness is correct for moderately skewed distributions (whether positively or negatively skewed). For "strongly" skewed distributions (where the absolute value of $m_3(\Theta)$ is larger than $m_2(\Theta)(\theta_k + \theta_l - 2\theta_S)$), there are exceptions, but one can always increase stability and skewness at the same time by selecting the appropriate changes in Proposition 6 (i.e. case (i) for "strongly" positively skewed distributions and cases (ii) to (iii) for "strongly" negatively skewed distributions).

---

[27] For case (iii) in Proposition 6 note that $\theta_k + \theta_l - 2\theta_S > 0$ if θ_l is further away from the average than θ_k, which holds if $\theta_{\{k-\epsilon\}} > \theta_{S \setminus \{l+\epsilon\}}$ as assumed in the proposition.

Proposition 9 has assumed that the coalition determines the equilibrium provision level. The reason is that if an outsider m determines the equilibrium, skewness needs to be replaced by an equivalent concept (coefficient) where the "moments" are defined around θ_m and not around the average θ_S. In other words, to extend Proposition 9 to any coalition not determining the equilibrium one needs to substitute $g(\Theta)$ by a similar coefficient, $\hat{g}(\Theta)$, defined using θ_m instead of the average of the distribution.

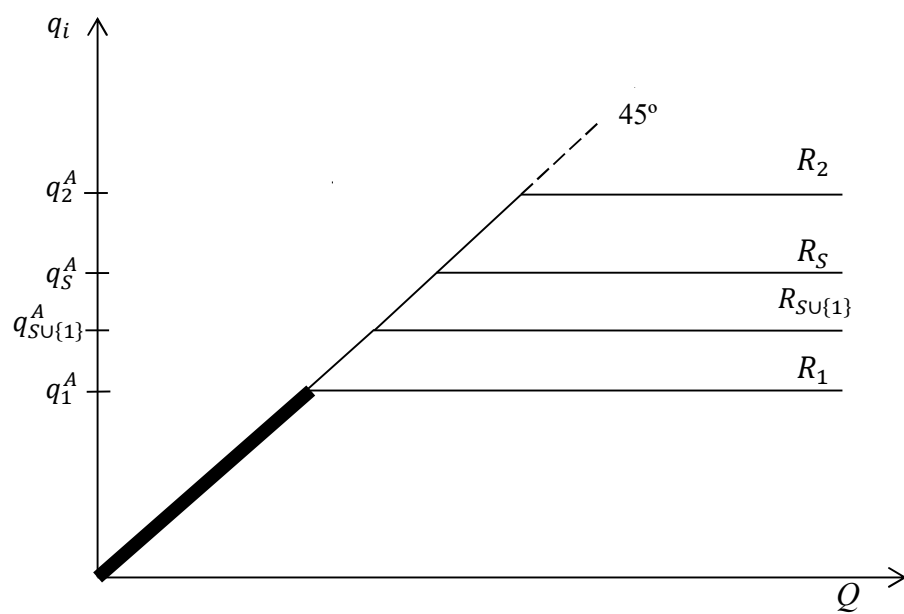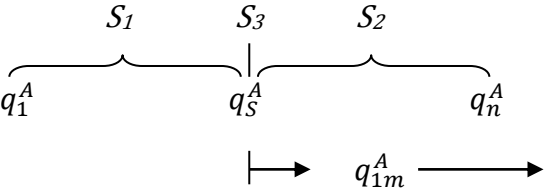**Figure 1: Replacement Functions and Equilibria for the Weakest-link Technology**
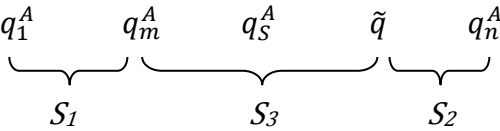
Figure 2: Weak, Strong, and Neutral Players
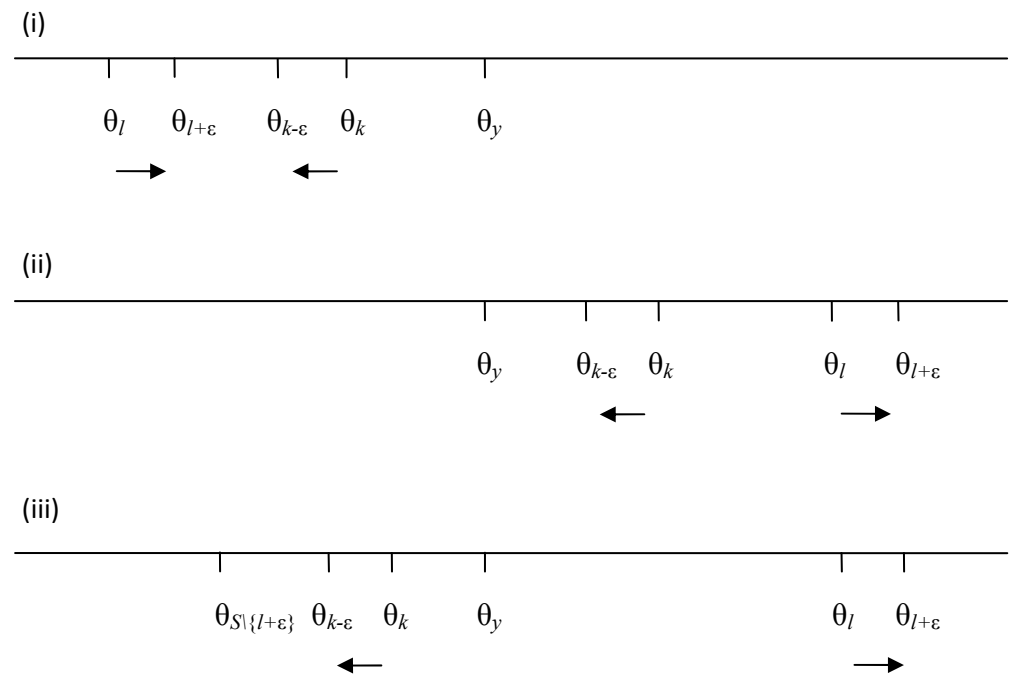
Case 1

S_1 S_3 S_2

q_1^A q_S^A q_n^A

q_{1m}^A

Case 2

q_1^A q_m^A q_S^A $\tilde{q}$ q_n^A

S_1 S_3 S_2

Figure 3: Illustration of Proposition 6

(i)

θ_l $\theta_{l+\varepsilon}$ $\theta_{k-\varepsilon}$ θ_k θ_y

(ii)

θ_y $\theta_{k-\varepsilon}$ θ_k θ_l $\theta_{l+\varepsilon}$

(iii)

$\theta_{S\setminus\{l+\varepsilon\}}$ $\theta_{k-\varepsilon}$ θ_k θ_y θ_l $\theta_{l+\varepsilon}$
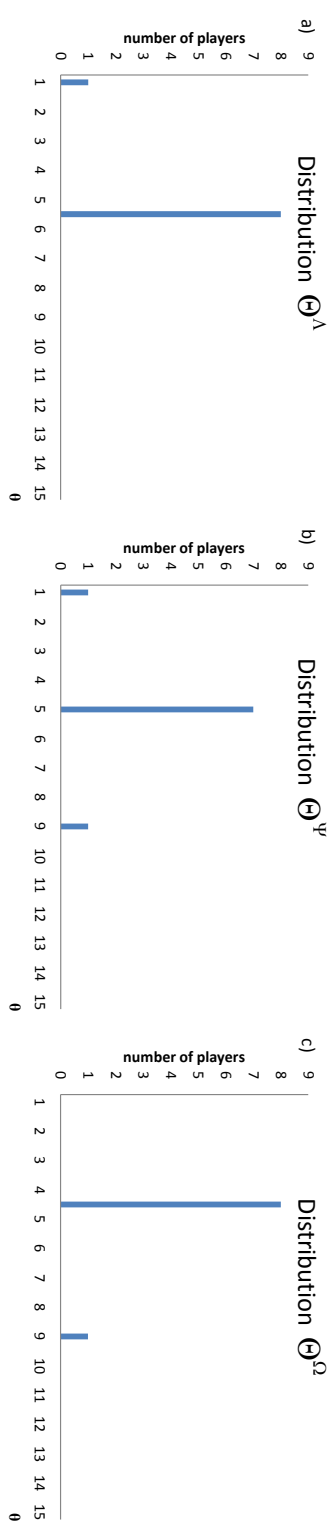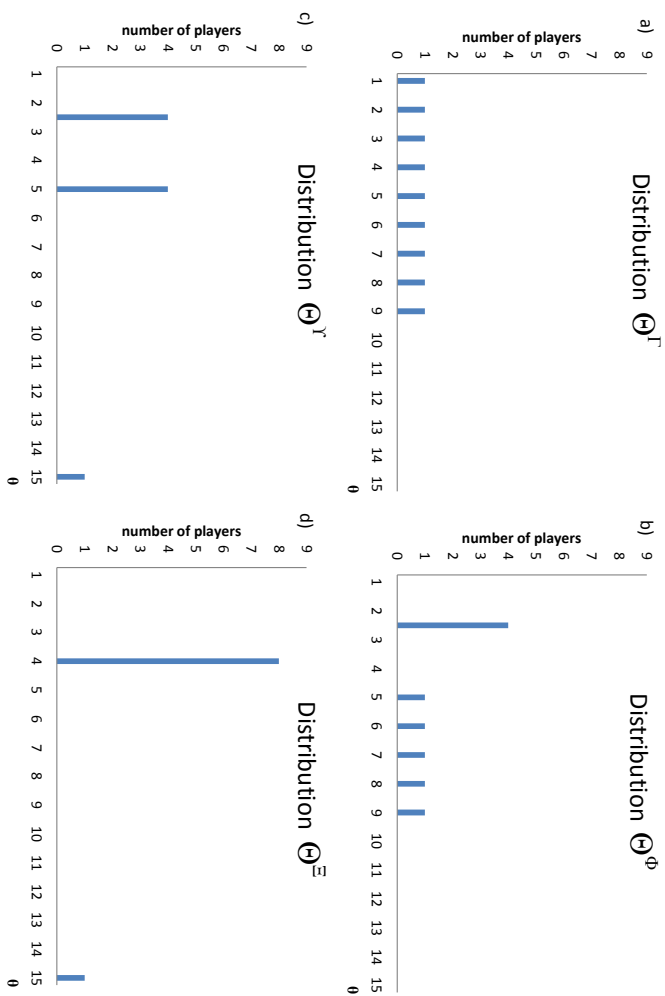
Figure 4: Illustration of Definition 5

Table 1: Coalition Structures and Payoffs

Coalition Structure	Example 1			Example 2		
	Player 1	Player 2	Player 3	Player 1	Player 2	Player 3
{{1},{2},{3}}	5.5	10	13.5	2.8	3.03	3.24
{{1,2},{3}}	4.5	13.5	22.5	3.2	3.74	5.76
{{1,3},{2}}	2	18	22	3.08	5.45	4.18
{{1},{2,3}}	10.5	9.5	20.5	5.1	3.52	4.08
{{1,2,3}}	0	18	36	4.46	5.45	6.44