PUBLIC GOOD AGREEMENTS UNDER THE WEAKEST-LINK TECHNOLOGY

ALEJANDRO CAPARROS
&
MICHAEL FINUS

CSIC

IPP

How to quote or cite this document:

Caparrós, A., & Finus, M. (2019). Public Good Agreements under the Weakest-link Technology. Instituto de Políticas y Bienes Públicos (IPP) CSIC, Working Paper. 2019-02

Available at: digital.csic.es

Public Good Agreements under the Weakest-link Technology

July 2019

Alejandro Caparrós
Institute for Public Goods and Policies (IPP)
Spanish National Research Council (CSIC)
Albasanz 26, 28037 Madrid, Spain
E-mail: alejandro.caparros@csic.es

Michael Finus
University of Graz (Austria)
Department of Economics
8010 Graz, Austria
&
University of Bath
Department of Economics
Bath, BA2 7AY, UK
E-mail: michael.finus@uni-graz.at & m.finus@bath.ac.uk

Abstract: We analyze the formation of public good agreements under the weakest-link technology. Whereas policy coordination is not necessary for symmetric players, it matters for asymmetric players; however, this fails in the absence of transfers. By contrast, with a transfer scheme, asymmetry may be an asset for cooperation. We characterize various types and degrees of asymmetry and relate them to the stability of self-enforcing agreements. Asymmetric distributions of autarky public good provision levels (also representing asymmetric interests in cooperation) that are positively skewed tend to be conducive to the stability of agreements. We show that under such conditions, even a coalition including all players can be stable. However, asymmetries that foster stability (instability) tend to be associated with low (high) gains from cooperation.

Key words: public goods, weakest-link technology, agreement formation
JEL classification: C7, D7, H4, H7

1   Introduction

A central aspect of the theory of public goods is the understanding of the incentive structure that typically leads to the underprovision of public goods as well as the possibilities of rectifying this. In this study, we answer the research question already posed by Cornes (1993): how do cooperative institutions develop under different aggregation technologies? Among the three typical technologies, namely, the summation, best-shot and weakest-link technology, we focus on the latter.

Under a weakest-link technology the benefits of public good provision depend on the smallest contribution.[1] Many policy issues are characterized by the weakest-link technology, including the classical example in Hirshleifer (1983) of building dykes to protect against flooding on an island: the lowest (rather than the average or the highest) dyke determines the level of protection. Other examples (e.g., Arce 2001, Barrett 2007 and Sandler 1998, 2004) include the spread of an epidemic or infectious disease such as small-pox and polio. Only global vaccination programs helped eradicate these diseases. If some countries had been omitted from this global effort, success would have been jeopardized. In addition, the success of the inspection of invasive species at ports, security measures at airports to protect against terrorist attacks, and protection of the integrity of computer networks through firewalls and anti-virus programs depend on the weakest-link effort. Moreover, the success of European Union (EU) measures to address illegal migration depends on the "weakest" member at the periphery of the EU area, irrespective of the measures taken by "stronger" members. Similarly, measures against money laundering and tax evasion will fail as long as some countries offer tax havens and have lax financial monitoring standards. Finally, the fight of a fire that threatens several communities will be as successful as the quality of the weakest fire brigade. Local fires will spread to other communities if not extinguished.

In most examples above, the benefits and costs of public good provision are expected to be unevenly distributed because of the range of preferences for public good supply as well as different technologies used to provide the public good. Therefore, asymmetry is a central focus of our study.

By cooperative institutions, we mean agreements among a group of players who depart from Nash equilibrium provision levels. However, we assume that players cannot write binding contracts; instead, agreements need to be "self-enforcing". That is, the benefits of an agreement must outweigh the gains from free-riding. The assumption of non-binding agreements is adequate in most of the examples cited above, especially for global public goods

---

[1]Measures to address climate change are typically related to the summation technology and missile protection is often viewed as a type of best-shot technology.

affecting several sovereign nations and jurisdictions, as no global entity has the ultimate enforcement power. However, coordination across several communities may also not be as straightforward as one may hope.

Based on the foregoing, we examine the conditions under which stable agreements emerge as well as how large they will be and under which conditions the associated gains from cooperation will be large.

For our analysis, we combine approaches from two strands of the literature that have developed independently: the literature on non-cooperative or privately provided public goods under the weakest-link technology and the literature on cooperatively provided public goods under the summation technology. We subsequently review these two strands of the literature in Section 2 and set out our two-stage coalition formation model in Section 3. We argue that coalition formation is only interesting in the context of asymmetric players and we therefore analyze the existence of stable agreements under this assumption in Section 4. We show that transfers are a necessary condition for the existence of effective agreements, namely those that improve upon the situation without agreement. Consequently, Section 5 provides a detailed analysis of the type and degree of asymmetry that fosters stability and explains how this relates to the welfare gains from cooperation in the presence of transfers. Section 6 concludes and discusses policy implications.

2   Related Literature

2.1   Non-cooperative Public Good Provision under the Weakest-Link Technology

Studies of non-cooperative public good provision have adopted three main approaches to understand the incentive structure under 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 2006 and Sandler and Arce 2002). Moreover, it is argued that either a third party or the best-off players should have an incentive to support the worst-off through monetary or in-kind transfers to increase the provision level. Cooperative agreements are not considered.

The second approach is formal (Cornes 1993, Cornes and Hartley 2007, Vicary 1990, and Vicary and Sandler 2002). It is shown that no unique Nash equilibrium exists, although Nash equilibria can be Pareto-ranked. Moreover, Nash equilibria are Pareto-inefficient except if players are symmetric. Improvements to inefficient outcomes are only considered in the

form of transfers in a Nash equilibrium. However, the departure from Nash equilibria in the form of policy coordination via agreements is not considered. Because monetary transfers change players' endowments, this may also change their Nash equilibrium strategies, as income neutrality no longer holds (as 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 for all players. In some models (e.g., Cornes and Hartley 2007 and Vicary and Sandler 2002) that allow for different prices across players (i.e., different marginal opportunity costs in the form of the foregone consumption of the private good), this is reinforced if recipients face a lower price than the donor. Vicary and Sandler (2002) also investigate how the Nash equilibrium provision level changes if monetary transfers are either substituted or complemented by in-kind transfers.

Finally, the third approach is also formal but specific. This approach considers various forms of 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 study differs from the literature by offering three new aspects. First, we combine a formal model of coalition formation with general payoff functions and continuous strategies. That is, we benefit from the motivation and intuition of the first and third approach as well as the formal concepts developed within the second approach. However, our analysis of cooperative institutions is not based on examples or matrix games, as is the third approach. Instead, we continue in the tradition of the second general and formal approach but consider not only Nash but also coalition equilibria. Second, we measure the degree of underprovision not only in physical terms (Cornes, 1993) but also in welfare terms. Admittedly, this is easier in our transferable utility framework, as equilibrium strategies are unaffected by monetary transfers. Third, our model considers not only different marginal costs but also non-constant marginal costs of public good provision, although we do not consider in-kind transfers.

2.2 Public Good Agreements under the Summation Technology

Within the literature on public good agreements under the summation technology, our approach applies a general theory of non-cooperative coalition formation in the presence of externalities, as summarized by Bloch (2003) and Yi (1997). First, agreements emerge from a non-cooperative game. Moreover, they are not stable per se (i.e., cooperation is not only about sharing a pie but also about enforcement), and, hence, full membership may not be obtained because of free-riding. A general conclusion emerging from this literature is that most economic problems can be broadly categorized into positive or negative externalities.

In positive (negative) externality games, players who are not involved in the enlargement of coalitions are better (worse) off through such a move. Hence, in positive externality games, only small coalitions are typically stable, as players have an incentive to stay outside coalitions. In negative externality games, by contrast, this is reversed and the grand coalition is usually stable. Therefore, it is more interesting to analyze positive externality games; however, in such games, making precise predictions about which coalitions will be stable is also more difficult.

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, players outside the cartel 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 who are not involved in the expansion of a coalition benefit from a higher total provision level but lower costs. We show that public good provision under the weakest-link technology also belongs to this group of externality games.

Until now, non-cooperative coalition theory has mainly assumed symmetric agents, as coalition formation complicates the analysis (see the surveys by Bloch 2003 and Yi 1997). In the context of global public goods, this literature can be traced to Barrett (1994) and Carraro and Siniscalco (1993); a collection of the most relevant contributions is provided by Finus and Caparrós (2015) and one of the most recent and advanced papers is Battaglini and Harstad (2016). In most cases, the analyses that consider asymmetric agents are either based on simulations or analytical results have been derived assuming particular functional forms and have captured asymmetry by assuming two types of players (e.g., Fuentes-Albero and Rubio 2010 and Pavlova and de Zeeuw 2013). Our study differs from these works in two fundamental respects. First, none of these papers has investigated the weakest-link technology. Second, we are able to characterize precisely the type and degree of asymmetry conducive to larger stable coalitions for general payoff functions, including the evaluation of the associated global welfare gains.

3 Model and Definitions

Following d'Aspremont et al. (1983), we consider a simple two-stage coalition formation game, also called cartel formation game. In the first stage, all players simultaneously choose their membership strategy. All players who choose to remain outside coalition S act as single players and are called non-signatories or non-members, whereas all players who choose to join coalition $S \subseteq N$ are called signatories or members, where N denotes the set of all players.

In the second stage, all non-signatories choose simultaneously their provision level and all signatories do the same. The game is solved by backward induction.

In the second stage, players choose their equilibrium strategies based on the following payoff function of player $k \in N$:

$$
\begin{aligned}
V_k(Q, q_k) &= B_k(Q) - C_k(q_k) \\
Q &= \min_{k \in N} \{q_k\}
\end{aligned}
\tag{1}
$$

where 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 k is q_k. Payoffs comprise benefits, $B_k(Q)$, and costs, $C_k(q_k)$. The externalities across players are captured by Q on the benefit side.

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

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

These assumptions ensure the strict concavity of all payoff functions and interior solutions. Together with the assumptions made below, they guarantee the existence of a unique Nash equilibrium in the second stage for every coalition $S \subseteq N$. Thus, instead of $V_k(Q^*(S), q_k^*(S))$, where the asterisks denote the Nash equilibrium provision levels in the second stage, we can simply write $V_k^*(S)$, which is the equilibrium payoff if coalition S forms.

In the first stage, given the payoffs obtained in the second stage, coalition S is called stable if

$$
\begin{aligned}
\text{Internal Stability:} \quad & V_i^*(S) \geq V_i^*(S \setminus \{i\}) \ \forall \ i \in S \\
\text{External Stability} \quad : \quad & V_j^*(S) \geq V_j^*(S \cup \{j\}) \ \forall \ j \notin S
\end{aligned}
\tag{2}
$$

hold simultaneously. That is, no signatory in stable coalition S, $i \in S$, has an incentive to leave (internal stability), as his or her payoff inside coalition S, $V_i^*(S)$, is larger than that when leaving, implying coalition $S \setminus \{i\}$ and payoff $V_i^*(S \setminus \{i\})$. In addition, no non-signatory $j \notin S$ with payoff $V_j^*(S)$ has an incentive to join coalition S (external stability) and receive payoff $V_j^*(S \cup \{j\})$. This definition corresponds to a Nash equilibrium in membership strategies. Given the equilibrium membership announcements of all players, no player has an incentive to change his or her strategy.

In the remainder of this section, we discuss some of the general features of the second stage. We define the term "autarky provision level" and consider two benchmarks: a) the empty or trivial coalition, denoted by $S = \varnothing$, which corresponds to the "classical"

Nash equilibrium without coalition formation, and b) the grand coalition, $S = N$, which corresponds to the social optimum. We discuss coalitions that entail partial cooperation, namely $S \neq \varnothing$ and $S \neq N$, in Section 4.

Following Vicary (1990), the "autarky" provision level is the provision level player k would choose if he or she were the only player in the game and would thus be able to determine the provision level. It can also be viewed as the "optimal" provision level of player i if he or she could also decide on the provision levels of all other players.

Definition 1 **Autarky Provision Level** *The (unique) autarky provision level of player k, q_k^A, follows from $\max_{q_k}. V_k(q_k) = B_k(q_k) - C_k(q_k)$, implying $B_k'(q_k^A) = C_k'(q_k^A)$ in an interior optimum.*

If all players are symmetric (i.e., they have the same payoff function), then all autarky levels will coincide. Otherwise, if players are asymmetric, autarky levels can be ranked: $q_1^A \leq q_2^A \leq ... \leq q_n^A$, where n is the cardinality of N. Without loss of generality, we denote each player by his or her position in the ranking of autarky levels. Henceforth, if we talk about asymmetric players, we assume that at least one inequality is strict.[2]

If no non-trivial coalition of at least two players has formed in the first stage of the game, $S = \varnothing$, all players act as single players. We also refer to this situation as "no agreement" below. It corresponds to the classical Nash equilibrium in the weakest-link game without coalition formation. As discussed by Hirshleifer (1983), Cornes (1993) or Cornes and Hartley (2007), among others, there are multiple Nash equilibrium provision levels given by $Q^{\varnothing} \in \left[0, \overline{Q}\right]$, $\overline{Q} = q_1^A$ and $Q^{\varnothing} = q_1^{\varnothing} = q_2^{\varnothing} = ... = q_n^{\varnothing}$. That is, all players choose the same provision level, which can be any level between zero and the smallest autarky provision level.

Given the weakest-link technology, choosing a higher provision level than the minimum of other players entails only costs but no benefits. Any provision below the own autarky provision level would entail a welfare loss, while any provision above it can de facto be vetoed. Hence, players will match any provision level from zero up to their autarky provision level. Any provision level above the smallest autarky provision will not be obtained, as the weakest player has veto power.

The fact that there is no unique Nash equilibrium implies that determining the provision level becomes a coordination game. However, the strict concavity of all payoff functions means that the smallest autarky provision level Pareto-dominates all smaller provision levels. Therefore, it seems natural to assume that players play the Pareto-optimal Nash equilibrium

---

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

if no coalition forms, an assumption that we make henceforth, i.e., $Q^{\varnothing *} = q_1^A$, $Q^{\varnothing *} = q_1^{\varnothing *} = q_2^{\varnothing *} = ... = q_n^{\varnothing *}$, where the asterisk denotes the equilibrium provision levels. We relax this assumption in Appendix C and discuss alternative equilibrium selection criteria. We find that most of our subsequent results hold under these alternative assumptions.

From a social planner's point of view, the social optimum follows from $\max_q \sum_{k \in N} V_k(q) \implies$ $\sum_{k \in N} B_k'(q_k^{SO}) = \sum_{k \in N} C_k'(q_k^{SO})$ with q a vector of all provision levels, and $Q^{SO} = q_1^{SO} = q_2^{SO} = q_k^{SO} = ... = q_n^{SO}$ in an interior equilibrium, which is ensured by Assumption 1. The social optimum is unique and all players choose the same provision level. Because the maximum of the sum of strictly concave payoff functions is located between the lowest and highest maximum of the individual payoff functions, we have $q_1^A \leq q_k^{SO} \leq q_n^A$. Consequently, as we have $Q^{\varnothing *} = q_1^A$ and $Q^{SO} = q_k^{SO}$, it is clear that $Q^{\varnothing *} \leq Q^{SO}$ must hold. That is, the socially optimal provision level is at least as large as in the Nash equilibrium that prevails if no coalition has formed. If we assume that coalitions choose their provision level by maximizing the aggregate payoff to the coalition (see Section 4 for details), then for $S = N$, and denoting the equilibrium provision if coalition S forms by $Q^*(S)$, we have $Q^*(N) = Q^{SO}$.

It is evident that if players are symmetric, the social optimum and the Nash equilibrium if no coalition forms coincide, and hence, there is no need for cooperation. That is, the analysis of coalition formation is only interesting for the setting with asymmetric players (with at least two players having a different autarky provision level). Henceforth, we assume this to be the case. Consequently, the strict inequality sign holds, i.e., $Q^{\phi *} < Q^{SO}$. In the following sections, we analyze the opportunities to form stable agreements.

4 Cooperation, Transfers, and Existence of Effective Stable Coalitions

We discuss the role of cooperative institutions, namely "self-enforcing" agreements (coalitions) among a group of players that aim to provide a higher provision level than that without agreement. Moreover, we introduce our transfer scheme and establish some initial results.

We call agreements with (strictly) higher provision levels than without agreement, effective agreements. Owing to the properties of our game (i.e., full cohesiveness, as discussed below), these agreements also generate a strictly higher total payoff than without agreement. In the context of the weakest-link technology, agreements can only be effective if the coalition includes player 1 with the lowest autarky provision level as player 1 can veto any provision level above q_1^A. Moreover, it requires at least one player j, with $q_j^A > q_1^A$, such that the coalition would have an incentive to implement a provision level above q_1^A. However, player 1 will

be worse off under such an agreement as $Q^{\varnothing *} = q_1^A$ is his or her preferred provision level. Thus, by leaving such an agreement, player 1 can obtain a higher payoff. This implies that in the absence of transfers, no effective agreement can be stable. As non-effective coalitions are not particularly interesting, we focus hereafter on transfers. In particular, we assume that coalition members negotiate an agreement that sets out the obligations to the parties (e.g., their provision levels and the associated transfers, which compensate for possible asymmetric gains (or losses) from cooperation).

The standard assumption in bargaining theory is that each player will only agree if he or she receives at least his or her disagreement payoff (i.e., the payoff obtained in the absence of an agreement) as well as a share of the surplus remaining after granting each player his or her disagreement payoff (Rubinstein 1982 and Muthoo 1999). Any agreement that fails to meet this criterion will not be honored.

In the context of the cartel formation game, as introduced in the previous section, the implicit assumption is that if one player leaves an agreement (coalition), the remaining players will continue cooperating, even though they reoptimize their strategies (d'Aspremont et al., 1983), reflecting the fact that the coalition comprises fewer one player.

Under this assumption, the deviating player can obtain the disagreement payoff by acting as a singleton. We call this the free-rider payoff $V_i^*(S\backslash\{i\})$. Consequently, in any equilibrium agreement with transfers, each player has to receive his or her free-rider payoff plus a portion of the surplus remaining after granting each player his or her free-rider payoff. As under the standard assumption in bargaining theory, any agreement that fails to meet this criterion will not be honored.

To determine the total surplus available to the coalition, we first need to specify the provision level implemented by the coalition. Consider first the grand coalition (and hence $S = N$). In autarky, our assumption is that before negotiating the distribution of the surplus, all coalition members agree to set the provision at the level that maximizes the coalition surplus. That is, signatories' autarky provision level q_N^A follows from $\max_{q_N} \sum_{i\in N} V_i(q_N) \implies \sum_{i\in N} B_i'(q_N^A) = \sum_{i\in N} C_i'(q_N^A)$ with q_N the vector of the provision levels in the grand coalition. All members will choose the same provision level q_N^A: it never pays coalition members to choose a provision vector with asymmetric entries, as this will only increase aggregate costs without any benefit. Hence, in autarky, for the grand coalition, we have $q_k^*(N) = q_N^A = Q^*(N)$ for all $k \in N$, and the total payoff available to the grand coalition is $\sum_{i\in N} V_i^*(N)$. For the grand coalition, the equilibrium provision level and total payoff thus correspond to the social optimum.

Consider now the case in which a coalition smaller than the grand coalition forms, namely some players are outside coalition S, $S \subset N$. Both the coalition and the non-signatories first

determine their autarky provision levels. The coalition can be viewed as one player in this coordination game apart from $n - s$ single players outside coalition S, where s and n denote the cardinality of S and N, respectively. The autarky level of coalition S, q_S^A, follows from $\max_{q_S} \sum_{i \in S} V_i(q_S)$ and those of single players j, q_j^A, from $\max_{q_j} V_j(q_j)$ as described above. Assuming, as mentioned in Section 2, that players coordinate on the Pareto-dominant equilibrium, the equilibrium in the second stage is given by $Q^*(S) = \min\{q_m^A, q_S^A\}$, with q_m^A, $m \notin S$, the minimum autarky level of all players $j \notin S$ such that $q_j^A \geq q_m^A$. All players $k \in N$ choose the same provision level $q_k^*(S) = Q^*(S)$ in equilibrium. Thus, the total payoff available to coalition S is $\sum_{i \in S} V_i^*(S)$. Typically, the total payoff in a coalition that is not the grand coalition is less than that in the social optimum.

Thus, regardless of whether the grand coalition or only a smaller coalition forms, we can define the surplus after granting $V_i^*(S \backslash \{i\})$ to each player inside coalition $S \subseteq N$ as follows:

$$\sigma(S) = \sum_{i \in S} V_i^*(S) - \sum_{i \in S} V_i^*(S \backslash \{i\}).$$

This surplus is the total payoff minus the sum of all free-rider payoffs. Clearly, an agreement is only stable if there is a sufficient surplus to grant each member $V_i^*(S \backslash \{i\})$, i.e., we must have $\sigma(S) \geq 0$. If $\sigma(S) \geq 0$, then any agreement where each member receives his or her free-rider payoff $V_i^*(S \backslash \{i\})$ plus an arbitrary share $\lambda_i > 0$ of $\sigma(S)$, with $i \in S$ and $\sum_{i \in S} \lambda_i = 1$, will be honored by all members (as the alternative would be to receive only $V_i^*(S \backslash \{i\})$). For future reference, we define the payoff obtained by member i after transfers as $V_i^{T*}(S) = V_i^*(S \backslash \{i\}) + \lambda_i \sigma(S)$. For non-members $j \notin S$, we simply have $V_j^{T*}(S) = V_j^*(S)$. We assume budget neutrality, $\sum_{i \in S} V_i^{T*}(S) = \sum_{i \in S} V_i^*(S)$, i.e., no resources are wasted and transfers are exclusively financed out of the gains from cooperation. Simple calculations show that the transfer received (paid if negative) by member $i \in S$ is given by

$$T_i = \lambda_i \left[\sum_{k \in S \backslash \{i\}} (V_k^*(S) - V_k^*(S \backslash \{k\})) \right] - (1 - \lambda_i) \left[V_i^*(S) - V_i^*(S \backslash \{i\}) \right] . \qquad (3)$$

The first term in equation (3) shows that member i receives a share λ_i of the net surplus obtained by the other members of the coalition and the second term captures the idea that member i has to pay a share $(1 - \lambda_i)$ of the net surplus that the coalition generates to member i. The properties of this class of transfers were analyzed by Eyckmans et al. (2012), who called it the "optimal transfer scheme", with similar notions used by Fuentes-Albero and Rubio (2010), McGinty (2007), Weikard (2009) as well as Caparrós and Péreau (2017). Below, we summarize the most important properties.

First, if the surplus is positive, $\sigma(S) \geq 0$, then the transfer scheme considered here makes S internally stable, which may not be the case for other transfer schemes. If the free-rider incentives are too large and/or the total payoff generated by coalition S is too small, then $\sigma(S) < 0$, and neither our nor any other transfer scheme will make coalition S internally stable.

Second, although the shares λ_i matter for individual payoffs, they do not affect internal stability, and, as we argue below, do not affect external stability and hence stability. Thus, the set of stable coalitions is robust against the particular assumption of weights as long as $\lambda_i > 0$ for all $i \in S$ and $\sum_{i \in S} \lambda_i = 1$.

Third, if coalition S is not externally stable against the accession of player j, then coalition $S \cup \{j\}$ is internally stable.

Fourth, this enlarged coalition $S \cup \{j\}$ implies a higher aggregate payoff, i.e., the sum of payoffs over all players. Hence, we are less concerned about external stability in this public good game. If coalition S is externally unstable, a larger coalition with a higher total payoff is stable (see also the fifth point below). The aggregate payoff increases with the enlargement of coalitions because of two properties: superadditivity and positive externalities. Superadditivity means that the aggregate payoff of those involved in enlarging a coalition increases, $\sum_{k \in S} V_k^*(S \cup \{j\}) \geq V_j^*(S) + \sum_{i \in S} V_i^*(S)$ for all $S \subset N$, and all $k \in S \cup \{j\}$. Positive externalities means that outsiders benefit from the enlargement of the coalition, $V_l^*(S \cup \{l\}) \geq V_l^*(S)$ for all $S \subset N$, and all $l \notin S \cup \{j\}$. Together, both properties imply that the aggregate payoff increases with the enlargement of coalitions (also called full cohesiveness). This also normatively motivates the search for large stable coalitions, with the grand coalition corresponding to the social optimum generating the largest global payoff compared with all other coalitions that do not imply full membership. In Appendix A.1, we show that the weakest-link coalition formation game exhibits all these properties.

Fifth, among those coalitions that can be internally stabilized using the transfer scheme defined above, the one with the highest aggregate payoff is also externally stable and hence, stable.

Taken together, the transfer scheme defined above maximizes what can be achieved given the constraint that players can leave a coalition. Subsequently, when we consider transfers, we mean the transfer scheme defined in equation (3) for any $\lambda_i > 0$, for all $i \in S$, and $\sum_{i \in S} \lambda_i = 1$. Proposition 1 summarizes the discussion above and provides an additional result.

Proposition 1 ***Existence of an Effective Stable Coalition*** *Assume payoff function (1) and asymmetric players.*

(i) Without transfers, no effective stable coalition exists.

(ii) For the transfer scheme defined in (3), a sufficient condition for an internally stable coalition is $\sigma(S) \geq 0$.

(iii) For the transfer scheme defined in (3), at least one stable effective coalition S exists.

Proof. Statements (i) and (ii) follow from the discussion above. For (iii), see Appendix A.1. ∎

Proposition 1 may be viewed as an "optimistic" result in that at least an effective agreement is stable with transfers. However, predicting which specific coalitions are stable, i.e., whether they are large or small, and whether this includes the grand coalition, is not straightforward at this level of generality. In particular, in the context of positive externality coalition games, it is well-known that only small coalitions may be stable, which achieve very little. In the next section, we analyze how the nature and degree of asymmetry affect stability and to what extent this improves the situation without agreement.

5 Characterizing Stable Agreements with Transfers

5.1 Stability

To render the analysis interesting, we focus on effective coalitions. Moreover, we focus on internal stability and hence on $\sigma(S) = \sum_{i \in S} \sigma_i(S)$. This is because if coalition S is internally stable with transfers, but not externally stable, then a larger coalition with a higher total payoff is stable (see the fourth and fifth points discussed in the previous section). In other words, external stability is of less concern. Analytically, we cannot simply consider the impact of different distributions of autarky levels on $\sigma(S)$, as they may be derived from different payoff functions. Therefore, we need to construct a framework that allows us to relate autarky levels to the parameters of the payoff functions. Hence, we consider a payoff function that has slightly more structure than our general payoff function (1), but that is still far more general than those that are typically considered in the literature on non-cooperative coalition formation in general and in the context of public good provision with a summation technology in particular.[3] We use the notation $v_i(Q, q_i)$ to indicate the difference from our

---

[3] All the specifications used in the context of the summation technology are a special case of payoff function (4) assuming $Q = \sum_{i \in N} q_i$ instead of $Q = \min_{i \in N} \{q_i\}$. See, for instance, Barrett (1994), Courtois and Haeringer (2012), Finus and Maus (2008), McGinty (2007) and Ray and Vohra (2001). This is also true for Rubio and Ulph (2006) and Dimantoudi and Sartzetakis (2006), although they analyze the dual problem of a public bad game.

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

$$
\begin{aligned}
v_i(Q, q_i) &= b_i B(Q) - c_i C(q_i) & (4) \\
Q &= \min_{i \in N} \{q_i\}
\end{aligned}
$$

where Assumption 1 summarizes the properties of functions B and C. 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, to simplify the subsequent analysis, we assume $C''' \geq 0$ and $B''' \leq 0$ (or, if $B''' > 0$, then B''' is sufficiently small).[4]

The following lemma shows the key advantage of payoff function (4): it allows us to characterize the autarky provision of any trivial or non-trivial coalition S based on a single parameter.

Lemma 1 *Autarky Provision Level and Benefit and Cost Parameters Consider payoff function (4). The autarky provision level of 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 A.2. ∎

That is, players can be ranked based on their parameter θ_i through function h. Players with higher parameters θ_i will have higher autarky levels. We say that player k is "bigger" than player l if $\theta_k > \theta_l$ and "smaller" if the opposite relation holds. For the subsequent analysis, we need to distinguish the two cases illustrated in Figure 1.

In case 1, coalition S determines the equilibrium provision level, i.e., $q_S^A = Q^*(S) \leq q_m^A$. Therefore, S may be a subcoalition or the grand coalition. In case 2, an outsider m determines the equilibrium provision, i.e., $q_S^A > q_m^A = Q^*(S)$, and $S \subset N$. Because S is assumed to be strictly effective, 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, apart from other players).

[**Figure 1 about here**]

In both cases (which are identical if $q_m^A = q_S^A$), we distinguish three groups of players in coalition S: (i) "small players" $i \in S_1$ for whom $q_i^A = Q^*(S \setminus \{i\}) < Q^*(S) < q_{S \setminus \{i\}}^A$ holds after they leave coalition S, (ii) "big players" $j \in S_2$ for whom $q_{S \setminus \{j\}}^A = Q^*(S \setminus \{j\}) < Q^*(S) < q_j^A$ is true, and (iii) "neutral" players $k \in S_3$ for whom $Q^*(S \setminus \{k\}) = Q^*(S) \leq q_k^A$ holds.[5] Small players have an autarky provision level below the equilibrium provision level when S forms,

---

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

[5]We use these terms for easy reference having in mind a small, big or neutral interest regarding a high level of public good provision. We could also use the terms "weak" and "strong" players; however, as a

and hence, gain from leaving coalition S, i.e., $\sigma_i(S) = V_i^*(S) - V_i^*(S \setminus \{i\}) < 0$. For big 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, i.e., $\sigma_j(S) = V_j^*(S) - V_j^*(S \setminus \{j\}) > 0$. 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. In case 2, their autarky provision level is larger than $q_S^A > Q^*(S) = q_m^A$, but insufficiently large to define the new equilibrium provision level if they leave ($q_k^A \leq \widetilde{q}$ in Figure 1) such that when player k leaves coalition S, we have $q_{S \setminus k}^A \geq Q^*(S) = Q^*(S \setminus \{k\}) = q_m^A$. That is, neutral players do not affect the equilibrium provision level after they leave. Clearly, $S = S_1 \cup S_2 \cup S_3$ noting that the sets of players in different groups do not coincide in cases 1 and 2, as is evident from Figure 1.

With reference to payoff function (4), in case 1, small players are those members for whom $\theta_i < \theta_S$ holds, big players are those members for whom $\theta_j > \theta_S$ holds and neutral players are those members for whom $\theta_l = \theta_S$ holds. In case 2, small players are coalition members for which $\theta_i < \theta_m$ holds, big 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.

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

Accordingly, the surplus of coalition S can be written as follows:

$$\sigma(S, \Theta) = \sum_{j \in S_2} \left[v_j(\theta_y) - v_j \left(\theta_{S \setminus \{j\}} \right) \right] - \sum_{i \in S_1} \left[v_i(\theta_i) - v_i(\theta_y) \right] \tag{5}$$

where $\theta_y = \theta_S$ in case 1 and $\theta_y = \theta_m$ in case 2, with $\sigma(S, \Theta)$ indicating that the internal stability of coalition S depends on the distribution of the θ_i-values of players in S, Θ.

We now examine which distributions of Θ favour stability. In particular, we are interested in whether it is more likely that a more asymmetric or symmetric distribution of the θ_i-values and hence autarky levels allows coalition S to be internally stable.

To compare different distributions, it is sensible to assume the same mean. Technically, we compare distributions by deriving a distribution $\widetilde{\Theta}$ from a distribution Θ as a sequence of small marginal changes ϵ that, if repeated, can be interpreted as large discrete changes Δ. In reality, parameters are given, and hence, we only compare the stability properties of different distributions of autarky levels.

To simplify the exposition, we focus on the case in which all players in S share a common $c_i = c$ and changes affect only the parameters b_i. However, all the results shown in this section hold if coalition members share a common $b_i = b$ and marginal changes affect the parameters c_i instead (in the opposite direction), and only minor adjustments are needed to

---

referee correctly pointed out, and as will become apparent below, weak (small) players are de facto strong, as they extract transfers from strong players. The opposite holds for strong (big) players, who are weak as they need to pay transfers to weak (small) players to make S stable.

accommodate the case in which players differ with respect to both parameters. We describe these adjustments in footnote 6.[6] We denote the modified b_i by $\overleftarrow{b_i}$ if the value has decreased and by $\overrightarrow{b_i}$ if it has increased. Correspondingly, we denote the modified θ as $\overleftarrow{\theta_i}$ or $\overrightarrow{\theta_i}$ depending on the direction of the change.

Proposition 2 *__Asymmetry and Stability__ Consider payoff function (4), with $c_i = c \ \forall$ $i \in S$, an effective coalition S 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 in two b_i-values of players in S, such that $\overleftarrow{b_k}$ $= b_k - \epsilon$ and $\overrightarrow{b_l} = b_l + \epsilon$, implying $\overleftarrow{\theta_k} < \theta_k$ and $\overrightarrow{\theta_l} > \theta_l$. Then, $\sigma(S, \tilde{\Theta}) \geq \sigma(S, \Theta)$ 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 $\overleftarrow{\theta_k} \geq \overleftarrow{\theta_{S \setminus l}}$.

Proof. See Appendix A.3. ∎

Figure 2 illustrates all three conditions, noting that θ_S remains the same through the marginal changes.

[**Figure 2 about here**]

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

In condition (ii), among the set of big players with θ-values above θ_y, the θ-value of a (weakly) bigger player l is increased at the expense of the θ-value of a (weakly) smaller player k. In case 1, in which the coalition determines the equilibrium provision level, the set of players involved in the marginal changes of the θ-values is the set of big players S_2 (see also Figure 1). In case 2, in which an outsider determines the equilibrium provision level, the set of involved players could be big players S_2 and/or neutral players S_3 (see also Figure 1).

In condition (iii), the marginal changes affect players from different sets, one player with a θ-value above and one below θ_y. The marginal changes involve an increase in the θ-value of a bigger player l at the expense of the θ-value of a smaller player k. The smaller player will typically belong to the group of small players S_1, but could also belong to the group of neutral players S_3 at the margin, namely in the special case of $\theta_k = \theta_y$. The bigger player l always belongs to the group of big players S_2 because otherwise, $\overleftarrow{\theta_k} \geq \overleftarrow{\theta_{S \setminus l}}$ would be violated

---

[6]Dropping the assumption $c_i = c \ \forall \ i \in S$, or $b_i = b \ \forall \ i \in S$, all the conditions in Proposition 2 remain unchanged except condition (ii) for which the additional condition $\left. \frac{\partial v_l(\theta)}{\partial \theta} \right|_{\theta_{S \setminus l}} \geq \left. \frac{\partial v_k(\theta)}{\partial \theta} \right|_{\theta_{S \setminus k}}$ would be needed. Propositions 3 and 4 would also continue to hold.

(see Appendix A.3 for details). The θ-value of one player involved in the change must be relatively big within its group because this ensures $\overleftarrow{\theta_k} > \overleftarrow{\theta_{S \setminus l}}$, where $\overleftarrow{\theta_{S \setminus l}}$ is the θ-value of coalition S if player l leaves coalition S under the distribution $\tilde{\Theta}$.

Note that the weak inequality sign in Proposition 2 in terms of stability only applies to the particular case in which 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 $\sigma(S, \tilde{\Theta}) > \sigma(S, \Theta)$ (see Appendix A.3. for details).

Given the abstract nature of Proposition 2, we consider two examples to illustrate how the marginal changes in Proposition 2 translate into discrete changes. Each example compares a sequence of distributions that can be generated by the marginal changes in Proposition 2. In Appendix A.4, we define the initial distributions and explain how the respective sequences are obtained, which are displayed in Figure 3. For simplicity, we assume that coalition S determines the equilibrium provision level (case 1). For these sequences, we find the following relations for the surplus of coalition S:

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

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

[**Figure 3 about here**]

In sequence 1, we move from a (very) negatively skewed distribution Θ^Λ to a symmetric distribution Θ^Ψ (using condition (ii) in Proposition 2), finally ending up in a (very) positively skewed distribution Θ^Ω (using condition (i) in Proposition 2). Along this sequence, the value of $\sigma(S)$ increases.

In sequence 2, we move from a symmetric, in fact, uniform distribution Θ^Γ, to a positively skewed distribution Θ^Φ (using condition (i) in Proposition 2). Imposing further changes (using conditions (i) and (ii) in Proposition 2), we generate the distributions Θ^Υ and Θ^Ξ, increasing $\sigma(S)$ along this sequence, noting that Θ^Ξ is a (very) positively skewed distribution. A similar sequence could have been generated starting from a normal distribution.

Whether $\sigma(S)$ is positive or negative along these sequences cannot be confirmed at this level of generality; we only know that finally $\sigma(S, \Theta^\Omega) > 0$ and $\sigma(S, \Theta^\Xi) > 0$. However, since the distribution Θ^Ω (like Θ^Ξ in sequence 2) can always be generated from any distribution, there always exists an asymmetric distribution for which coalition S is internally stable, including the grand coalition, according to Proposition 3.

Proposition 3 *Consider a strictly effective coalition $S \subseteq N$ with $\theta_S \leq \theta_m$ for all players $m \notin S$, and a distribution Θ^F with θ_S such that for all players in S, $\theta_k = \theta_S - \frac{\Delta}{s-1}$, except for one player for which $\theta_l = \theta_S + \Delta$, with s denoting the cardinality of S and Δ a positive*

number such that $\theta_S - \frac{\Delta}{s-1} > 0$. Then, $\sigma(S, \Theta^F) > 0$.

Proof. For the distribution Θ^F, regardless of which player leaves coalition S, the subsequent equilibrium provision will be that without agreement. Moreover, because there is a strictly positive aggregate gain for coalition members in S from moving from a situation without agreement to any non-trivial effective coalition $S \neq \varnothing$ (due to strict superadditivity), $\sigma(S, \Theta^F) > 0$ must be true. (The proof is similar to the proof in Appendix A.1.) ∎

Our analysis thus far suggests that asymmetric distributions of autarky levels, which are positively skewed, may be more conducive to the stability of coalitions than symmetric distributions if the asymmetric gains from cooperation can be balanced through any transfer scheme included in the class of transfer schemes defined in (3). 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 actually be 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 2). At the upper end, this is reversed. Instead of having many big players, it is better for stability to have one outlier at the top (condition (ii) in Proposition 2). If only one big player is left, he or she would pay transfers to all the other small players.

We close this subsection with two qualifying remarks. First, no unique measure exists to compare different distributions. The analysis above suggested that skewness could be a good measure. This is indeed the case for most distributions, although a one-to-one correspondence between the marginal changes listed in Proposition 2 and skewness may sometimes be lacking. As detailed in Appendix B, using the Fisher-Pearson coefficient of skewness, the intuition that all the marginal changes proposed in Proposition 2 increase skewness is correct for moderately skewed distributions (whether positively or negatively skewed). For "strongly" skewed distributions, there are exceptions; however, we can always increase stability and skewness at the same time by selecting one of the changes listed in Proposition 2.

Second, in contrast to the summation technology, when 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 2 highlights, stability only depends on the distribution of the autarky levels of the players in coalition S, namely, the θ_i-values. Adding a player l outside coalition S to S, a new distribution is generated for which $\sigma(S) < \sigma(S \cup \{l\})$ is possible.[7]

---

[7] 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 $\sigma_S(\{1, 2, 3\}, \{4\}) = -\frac{120}{2057} < 0$.

5.2 Welfare

From the previous subsection, we know how the distributions of the 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 $W(\Theta) = \sum_{i \in N} v_i(\Theta)$. We further define the gain from forming coalition $S \neq \varnothing$ by $\Delta W(S, \Theta) = W^S(\Theta) - W^\varnothing(\Theta)$, recalling that $\varnothing$ refers to a situation without agreement, i.e., all players play as singletons. (We would obtain the same conclusion shown below if we used a relative measure, say $\Delta W(S, \Theta) = \frac{W^S(\Theta)}{W^\varnothing(\Theta)}$). These definitions cover the case in which S is the grand coalition and therefore, the social optimum.

Proposition 4 *Asymmetry and Welfare Gains* *Consider payoff function (4), an effective coalition S with respect to no agreement, and two distributions Θ and $\tilde{\Theta}$ as defined in Proposition 2. 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. See Appendix A.5. ■

Thus, the smaller the smallest autarky level, the smaller is the provision level without agreement, and hence, the larger are the gains from cooperation, keeping constant the coalitional autarky level and equilibrium provision level if coalition S forms, as assumed by the marginal changes in Propositions 2 and 4. Thus, using the concept of a sequence of marginal changes in b_i-values (and/or c_i-values), as introduced in Proposition 2, the global payoff implications of different distributions can be compared. Again, using our two examples for illustrative purposes, with the two sequences of distributions displayed in Figure 3 and formally defined in Appendix A.4, we find that the relations are (almost) reversed.

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

That is, distributions that favour stability (instability) may be associated with a lower (higher) global gain from cooperation. 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 2 about stability and Proposition 4 about global payoffs reveals that the message is more nuanced. It is true for the marginal changes imposed in condition (i) in Proposition 2: the gains from cooperation decrease, whereas the stability value $\sigma(S)$ increases. It is not true, however, for the changes in conditions (ii) and (iii), which are payoff-neutral but increase the stability value $\sigma(S)$.

Again, a one-to-one correspondence between the marginal changes considered above and the skewness of the different θ_i-distributions may be lacking. However, as the two examples illustrate, we find the "paradox of cooperation" for larger discrete changes: stability increases when moving from a negatively skewed or symmetric distribution to a positively skewed distribution of the θ_i-values whereas the gains from cooperation decrease.

6   Conclusion and Discussion

Using a non-cooperative model of coalition formation, we analyzed the formation of institutions that coordinate the provision of a public good among agents. The public good is produced using a weakest-link aggregation technology. Institutions coordinate policies that depart from non-cooperative levels, may balance asymmetries via transfers, but cannot write binding agreements. Agents are assumed to behave opportunistically and leave an agreement if this pays, in which case the agreement is not self-enforcing.

It became apparent that there is no need for symmetric players to coordinate, as the Nash equilibrium without coalition formation is already Pareto-efficient, at least if we assume that players can "assure" each other to play the Pareto-efficient equilibrium. We assumed this to be the case and discuss alternative assumptions in Appendix C; showing that most of the results discussed in this article hold, with only minor changes. However, for the case of asymmetric players, cooperation can make a difference. We argued that in the absence of transfers, any form of cooperation either fails or is irrelevant. By contrast, we showed that with transfers stable agreements always exist. To understand whether these agreements enjoy small or large participation and how this relates to the welfare gains from stable agreements, we needed to characterize the type and degree of asymmetry. This translated into differences in autarky provision levels of players, which are related to the individual benefit-cost ratio of public good provision. It became apparent that asymmetry could actually be an asset for forming large stable agreements rather than an obstacle. Numerous players who prefer a similar provision below the average and one or a few players with a preference for a provision level well above the average favor stability. This ensures that there is no weakest-link outlier at the bottom and one or a few players with a very high individual benefit-cost ratio, well above those of all other signatories, who compensate the other signatories for their contributions to an efficient cooperative agreement. For an extremely positively skewed distribution of interests (i.e., autarky provision levels) we showed that even the grand coalition is stable. Unfortunately, the "paradox of cooperation" may also hold for the weakest-link technology: those asymmetries conducive to the stability of agreements may yield only low welfare gains from cooperation, whereas those asymmetries for which stability

is at risk would yield large welfare gains.

Although the bulk of the public goods literature on agreement formation has focused on the summation technology, many and important examples belong to the weakest-link technology (Barrett 2007 and Sandler 2004). Important regional or global public goods include the coordination of EU migration policies, compliance with minimum standards in marine law, protecting species whose habitat spreads across several countries, compliance with targets for fiscal convergence in a monetary union, fighting a fire that threatens several communities, and curbing the spread of a disease or epidemic such as small-pox and polio. For most of these examples, asymmetric interests seem to be likely, even though to the best of our knowledge, detailed individual cost-benefit data are lacking. Nevertheless, we briefly discuss how our results relate to some of these examples.

According to Barrett (2007), the eradication of small-pox and containment of polio can be judged as a success story. The interest in and resources used in fighting these diseases are unequally distributed globally, with industrialized countries having a much higher benefit-cost ratio than developing countries. Global action was coordinated by the World Health Organization. In terms of fighting small-pox, only eight countries donated funding to this effort between 1959 and 1966, which fell substantially short of what was needed (Barrett 2007, p.124). It took some time until sufficient funding could be raised, with the large majority coming from the United States and Sweden. In the case of polio, financing was very positively skewed, based on a UN assessment, with the EU required to pay 36.5% of total expenditure, the United States 22% and Japan 19.5%. See also Kaul et al. (2003, pp. 484) on global disease control.

Less successful seem to be the efforts of the EU in terms of fiscal convergence in the monetary union and the coordination of migration policies. Germany and some Nordic countries push heavily for budgetary discipline, whereas the bulk of EU-countries take a more relaxed view. This suggests that interests are positively skewed. Hence, according to our results, successful fiscal convergence, although possible, would need to be complemented by large transfers, something that Germany and the Nordic countries have thus far refused.

In addition, most countries at the periphery of the EU (Greece and Italy, as well as Spain and Portugal to a lesser extent), which serve as the gateways for most immigrants and asylum seekers, perceive the benefit-cost ratio of protection against illegal migration as very low. Those countries find it difficult to provide the infrastructure needed to process the large number of applications and benefit little in any case, as most applicants want to migrate to the richer countries in the EU. This suggests a very negatively skewed distribution of interests, which, according to our results, does not favor stable coordination, even though the welfare gains from coordination would be large.

Finally, let us briefly comment on the generality of our results. We focused on the most widespread coalition model and stability concept used in the literature on public good provision, but other concepts could also be considered (Bloch 2003, Caparrós et al. 2011, Finus and Rundshagen 2009 and Yi 1997). Internal and external stability implies that after a player leaves the coalition, the remaining members remain in the coalition. In a positive externality game, this is the weakest possible punishment after a deviation and hence implies the most pessimistic assumption about stability. This appears to be a good benchmark, especially because we showed that the grand coalition can be stable with transfers even under this assumption. Without transfers, other stability concepts would come to a similar negative conclusion, as individual rationality is a necessary condition for most equilibrium concepts, and without transfers, this condition is violated. Further, the assumption of open membership is not conducive to stability in positive externality games, as shown by Finus and Rundshagen (2009). In other words, those coalitions identified as being stable would also be stable under exclusive membership. Moreover, we restricted the analysis to a single coalition, but could have allowed for multiple coalitions (Bloch 2003, Finus and Rundshagen 2009 and Yi 1997). In the context of the weakest-link technology, this would not fundamentally change the incentive structure identified in the present study. An effective coalition structure must include the player with the lowest autarky level in one coalition. Only with transfers can effective coalition structures be stabilized. Moreover, within coalitions, autarky levels should be positively skewed for internal stability, with a few players with high autarky levels compensating the many players with low autarky levels. The sufficient conditions for the grand coalition being stable would remain the same. Future research could analyze the role of in-kind transfers, which was not considered in this paper.

References

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.

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

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

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

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

Barrett, S. (2007), Why Cooperate? The Incentive to Supply Global Public Goods. Oxford University Press.

Battaglini, M. and B. Harstad (2016), Participation and duration of environmental agreements. *Journal of Political Economy,* vol. 124(1), pp. 160-204.

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.

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.

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.

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

Caparrós, A. and J.C. Péreau (2017), Multilateral versus sequential negotiations over climate change. *Oxford Economic Papers,* vol. 69(2), pp. 365–387.

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

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

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

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

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.

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.

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

6-2012.

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

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

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.

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

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.

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

Kaul et al. (eds) (2003), Providing Global Public Goods. Oxford University Press.

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.

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

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

Muthoo, A. (1999), Bargaining Theory with Applications. Cambridge University Press.

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

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

Rubinstein, A. (1982), Perfect equilibrium in a bargaining model. *Econometrica,* vol 50, pp. 97-108.

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

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

Sandler T. (2004) Global Collective Action. Cambridge University Press.

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

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.

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

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.

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

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

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

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

A   Main Proofs

A.1   Proposition 1

For the subsequent proof, we need the following lemmas.

Lemma 2 *Enlarging a coalition (weakly) increases the public good provision level, i.e., $Q^*(S \cup \{i\}) \geq Q^*(S)$ for all $i \notin S$ and all $S \subset N$.*

Proof. First, if coalition S and player i merge, such that $S \cup \{i\}$ forms, then for the autarky level of the enlarged coalition, $q^A_{S \cup \{i\}}$, $\max\{q^A_S, q^A_i\} \geq q^A_{S \cup \{i\}} \geq \min\{q^A_S, q^A_i\}$ holds, with strict inequalities if $q^A_S \neq q^A_i$ due to the strict concavity of the payoff functions. Case 1: Suppose that $Q^*(S)$ is the autarky level of player j who does not belong to $S \cup \{i\}$. Hence, $q^A_j \leq q^A_i$, q^A_S and $q^A_j \leq q^A_{S \cup \{i\}}$, so that $Q^*(S) = Q^*(S \cup \{i\})$. Case 2: Suppose that $Q^*(S) = q^A_i$ initially, and hence, $q^A_j \geq Q^*(S)$ for all $j \notin S$. Moreover, $q^A_i \leq q^A_S$, and hence, $q^A_i \leq q^A_{S \cup \{i\}}$. Thus, regardless of whether $Q^*(S \cup \{i\})$ is equal to the autarky level of the enlarged coalition, $q^A_{S \cup \{i\}}$, or equal to the autarky level of another non-signatory j, $q^A_j \geq q^A_i$, $Q^*(S \cup \{i\}) \geq Q^*(S)$ must be true. Case 3: Suppose that $Q^*(S) = q^A_S$ (and hence $q^A_S \leq q^A_i$) before the enlargement, then the same argument applies as in Case 2. ∎

Lemma 3 *The public good coalition game with the weakest-link technology exhibits the properties of positive externality, superadditivity and full cohesiveness.*

1

Proof. *Positive Externality*: From Lemma 2, 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 or her autarky level if $q_j^A \leq Q^*(S \cup \{i\})$, and if $q_j^A > Q^*(S \cup \{i\})$, he or she 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 $S \cup \{i\}$ implies only $Q^*(S \cup \{i\}) = Q^*(S)$, weak superadditivity holds. If $Q^*(S \cup \{i\}) > Q^*(S)$, then either i or S must determine $Q^*(S)$ before the merger. Then, after the merger, $q_{S \cup \{i\}}^A > Q^*(S)$. Since the enlarged coalition $S \cup \{i\}$ can veto any provision level above $q_{S \cup \{i\}}^A$, moving from level $Q^*(S)$ toward $Q^*(S \cup \{i\}) \leq q_{S \cup \{i\}}^A$ must imply a move along the upward sloping part of the aggregate payoff 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. ∎

Because of asymmetry, we have $q_1^A < q_S^A < q_n^A$. Hence, an effective coalition S with $Q^*(S) > Q^{\varnothing *}$ exists, recalling that $q_1^A = Q^{\varnothing *}$. This effective coalition S must include all players i for whom $q_i^A = \min\{q_1^A, q_2^A, ..., q_n^A\}$ is true and player j with $q_j^A > q_i^A$ such that $q_i^A < q_S^A$. Because $Q^*(S) = \min\{q_k^A, q_S^A\}$, $q_k^A \geq Q^*(S) > q_i^A$, and all $k \notin S$ must be strictly better off in S than without an agreement (strict positive externality holds, and hence $V_k^*(S) > V_k^{\varnothing *}$). Let only one player j in S be different from i. Hence, regardless of which player leaves coalition S, $Q^*(S \setminus \{i\}) = Q^{\varnothing *}$ and so $V_j^*(S \setminus \{i\}) = V_j^{\varnothing *}$ for all i, $j \in S$. Therefore, because $q_S^A \geq Q^*(S) > Q^{\varnothing *}$, we must have $\sigma(S) = \sum_{j \in S}(V_j^*(S) - V_j^*(S \setminus \{i\})) > 0$ due to the strict concavity of the aggregate payoff function of S and hence $V_i^{T*}(S) = V_i^*(S \setminus \{i\}) + \lambda_i \sigma(S) > V_i^{\varnothing *}$. Hence, S constitutes a strict Pareto-improvement to all players in S (and, in fact, also to all players not in S) compared with the status quo 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 superadditivity (Lemma 3), the coalitions, which will be eventually stable, must (strictly) Pareto-dominate the situation without agreement.

A.2   Lemma 1

As discussed in the main text, the autarky provision level is the level that a coalition (including the trivial coalition formed by one player) would choose if it were the only player

in the game and would thus be able to define the provision level. In autarky, any coalition S maximizes $\sum_{i \in S} v_i(Q, q_i)$, with $v_i(Q, q_i)$, as defined in (4). The first order condition of any (trivial or non-trivial) coalition S to determine its autarky provision level (assuming an interior solution) is $\hat{c}\frac{\partial C}{\partial q} = \hat{b}\frac{\partial B}{\partial q}$ where $\hat{b} = \sum_{i \in S} b_i$ and $\hat{c} = \sum_{i \in S} c_i$. Thus, we can define function f as:

$$f(q) = \frac{C'(q)}{B'(q)} = \frac{\hat{b}}{\hat{c}},$$

and, with $\frac{\hat{b}}{\hat{c}} = \theta_S$, we can define function h as:

$$q_S^A = f^{-1}(\theta_S) := h(\theta_S).$$

Note that $\frac{\hat{b}}{\hat{c}} = \frac{\bar{b}}{\bar{c}}$, with $\bar{b} = \frac{\sum_{i \in S} b_i}{s}$ and $\bar{c} = \frac{\sum_{i \in S} c_i}{s}$. To show that h is strictly concave and strictly increasing, we show that f is strictly increasing and strictly convex:

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

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

which is true because of the assumptions about the first and second derivatives summarized in Assumption 1 and the assumptions about the third derivatives mentioned in Section 5, namely $C''' \geq 0$ and $B''' \leq 0$ (or if $B''' > 0$, then $B'''(q) < -2B''(q)C''(q)/C'(q)$).

A.3 Proposition 2

Before proving this proposition, we prove a lemma useful for the subsequent analysis. Along the proof, we denote $\frac{\partial v_i(h(\theta))}{\partial h} = \frac{\partial v_i(q)}{\partial q}$ by $v_i'(h(\theta))$ and $\frac{\partial v_i(\theta)}{\partial \theta} = \frac{\partial v_i(h(\theta))}{\partial \theta}$ by $v_i'(\theta)$. (This simple notation is consistent with viewing the payoffs $v(q)$ as a function of the provision level q, and the provision level as a function of the parameter θ, $q = h(\theta)$, without introducing a function $v = k(q)$.)

Lemma 4 *The function $v_i(\theta)$ is strictly concave and increasing in θ for $\theta \in [0, \theta_i]$.*

Proof. $v_i'(\theta) = v_i'(h(\theta))h'(\theta)$ is positive if $v_i'(h(\theta)) > 0$, as we have shown in Lemma 1, $h(\theta)$ is strictly concave and increasing and we thus have $h'(\theta) > 0$. Due to Assumption 1, v_i is a strictly concave function with respect to q_i with a maximum at $q_i^A = h(\theta_i)$, and it is therefore increasing for $q_i \in [0, q_i^A]$. As $h(\theta)$ is increasing everywhere, we also know that

3

$v_i(h(\theta))$ is increasing for $\theta \in [0, \theta_i]$ because for any $\theta \in [0, \theta_i]$, we know that $q_i = h(\theta) \leq q_i^A$. Thus, we have $v_i'(h(\theta)) > 0$.

For $v_i(\theta)$ to be strictly concave in θ, we need:

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

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

Before proceeding, let us write equation (5) in the text in a more disaggregated form:

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

using the sets described in Section 5.[8] 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 affect neither θ_m, for $m \notin S$, 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 prove the three conditions of the θ-values listed in Proposition 2.

(i) $\theta_l < \theta_k \leq \theta_y$. Consider first the case in which both players are in S_1, i.e., $\theta_l < \theta_k < \theta_y$. We denote the new valuation function by $\tilde{v}$ and the two values that have changed in $\tilde{\Theta}$ by $\overleftarrow{\theta_k}$ and $\overrightarrow{\theta_l}$. The third and fourth sums in condition (A.2) remain unchanged. In the first sum in (A.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 (A.2) changes and for $\sigma(S, \Theta) < \sigma(S, \tilde{\Theta})$ to hold, we need:

$$v_k(\theta_k) + v_l(\theta_l) > \tilde{v}_k\left(\overleftarrow{\theta_k}\right) + \tilde{v}_l\left(\overrightarrow{\theta_l}\right)$$

---

[8] These sets can be defined formally as follows: for 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 $S_3 = \left\{ k \in S \mid q_{S \setminus \{k\}}^A = q_k^A = q_S^A \right\}$; for 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 $S_3 = \left\{ k \in S \mid q_m^A \leq q_{S \setminus \{k\}}^A \right\}$. Note that the changes described in Proposition 2 modify neither q_m^A nor q_S^A.

or

$$\left[b_k B(h(\theta_k)) - cC(h(\theta_k))\right] - \left[b_k B(h(\overleftarrow{\theta_k})) - cC(h(\overleftarrow{\theta_k}))\right] \tag{A.3}$$
$$+\epsilon\left[B(h(\overleftarrow{\theta_k})) - B(h(\overrightarrow{\theta_l}))\right]$$
$$> \left[b_l B(h(\overrightarrow{\theta_l})) - cC(h(\overrightarrow{\theta_l}))\right] - \left[b_l B(h(\theta_l)) - cC(h(\theta_l))\right] .$$

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

$$\left[B(h(\overleftarrow{\theta_k})) - B(h(\overrightarrow{\theta_l}))\right] + v_k'(\theta)|_{\overleftarrow{\theta_k}} > v_l'(\theta)|_{\theta_l} . \tag{A.4}$$

For $\epsilon \to 0$, $\theta_k > \theta_l$ implies $\overleftarrow{\theta_k} \geq \overrightarrow{\theta_l}$ and therefore the first term on the LHS of inequality (A.4) is non-negative. Thus, inequality (A.4) always holds, as we have $v_k'(\theta)|_{\overleftarrow{\theta_k}} > v_k'(\theta)|_{\theta_k} = 0$ and $v_l'(\theta)|_{\theta_l} = 0$, given $\overleftarrow{\theta_k} < \theta_k$, θ_k and θ_l maximize v_k and v_l, respectively, and $v_i(\theta)$ is an increasing and strictly concave function for $\theta \in [0, \theta_i]$ by Lemma 4.

If player k is initially in S_3, i.e., if $\theta_k = \theta_y$, the last sum in (A.2) changes, but the relevant marginal changes are still summarized in condition (A.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 (A.2) has changed, and for $\sigma(S, \Theta) < \sigma(S, \tilde{\Theta})$ to hold, we need:

$$v_k\left(\theta_{S\setminus k}\right) + v_l\left(\theta_{S\setminus l}\right) > \tilde{v}_k\left(\overrightarrow{\theta_{S\setminus k}}\right) + \tilde{v}_l\left(\overleftarrow{\theta_{S\setminus l}}\right) \tag{A.5}$$

or

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

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

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

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

$$v_l'(\theta)|_{\overleftarrow{\theta_{S\setminus l}}} > v_k'(\theta)|_{\theta_{S\setminus k}} . \tag{A.7}$$

Because we have

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

inequality (A.7) can be written as:

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

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

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

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

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

or

$$v_k'(\theta)|_{\overleftarrow{\theta_{S\backslash l}}} > v_k'(\theta)|_{\theta_{S\backslash k}}.$$

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

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

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

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

If $k \in S_3$ and $l \in S_2$ (the opposite is not possible, see Figure 1), then equation (A.5) simplifies to

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

or

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

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

Finally, in the "marginal case" in which initially $\theta_y < \theta_k \leq \theta_l$ and finally $\overleftarrow{\theta_k} = \theta_y < \overrightarrow{\theta_l}$, it is easy to check that the conclusions derived above hold. We only need to note that in case $k, l \in S_2$, $\overrightarrow{\theta_{S\backslash k}} = \theta_y$ holds.

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

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

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

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

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

$$
v_l'(\theta)|_{\overleftarrow{\theta_{S\backslash l}}} + v_k'(\theta)|_{\overleftarrow{\theta_k}} > 0 \; .
$$

This holds for $\overleftarrow{\theta_{S\backslash l}} < \theta_l$ and $\overleftarrow{\theta_k} < \theta_k$, as $v_i(\theta)$ is an increasing and strictly concave function for $\theta \in [0, \theta_i]$ by Lemma 4.

Consider now the case $k \in S_3$ and $l \in S_2$. This implies that we consider the particular case in which initially $\theta_k = \theta_y < \theta_l$ and finally $\overleftarrow{\theta_k} < \theta_y < \overrightarrow{\theta_l}$. Because now $\overleftarrow{\theta_k} \geq \overleftarrow{\theta_{S\backslash l}}$ holds always (because $\overleftarrow{\theta_k}$ is infinitely close to θ_y), we have $B(h(\overleftarrow{\theta_k})) \geq B(h(\overleftarrow{\theta_{S\backslash l}}))$. Thus, a sufficient condition for the equivalent to inequality (A.8) to hold is:

$$
\begin{aligned}
&\left[b_l B(h(\theta_{S\backslash l})) - cC(h(\theta_{S\backslash l}))\right] - \left[b_l B(h(\overleftarrow{\theta_{S\backslash l}})) - cC(h(\overleftarrow{\theta_{S\backslash l}}))\right] \\
&+ \left[b_k B(h(\theta_k)) - cC(h(\theta_k))\right] - \left[b_k B(h(\theta_y)) - cC(h(\theta_y))\right] > 0 \; .
\end{aligned}
\tag{A.9}
$$

We know that k was in S_3, i.e., $\theta_y \lesssim \theta_k$, but we also know that it was only in S_3 at the margin, as $(k - \epsilon) \in S_1$, and thus, $\overleftarrow{\theta_k} < \theta_y$. Hence, $\theta_y = \theta_k$ or slightly above, i.e., $\theta_y \lesssim \theta_k$. Thus, the second square bracket in inequality (A.9) is either zero or equal to $v_k'(\theta)|_{\theta_y} > 0$. 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 $\theta_{S\backslash l} > \overleftarrow{\theta_{S\backslash l}} \geq \theta_y > \theta_k > \overleftarrow{\theta_k}$, and hence, the condition $\overleftarrow{\theta_k} \geq \overleftarrow{\theta_{S\backslash l}}$ fails. The same holds for the case $k, l \in S_3$, as in this case: $\theta_{S\backslash l} > \overleftarrow{\theta_{S\backslash l}} \geq \theta_y = \theta_k > \overleftarrow{\theta_k}$.

A.4 Definition of the Distributions Displayed in Figure 3

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 in the θ_i-values through marginal changes in b_i-values, as explained in Proposition 2, denoting the cardinality of S by s, with s being sufficiently large. Let $\Delta > 0$ in sequence 1 and $\hat{\Delta} > 0$ in sequence 2 be the result of a sequence of changes ϵ, as described in Proposition 2, conditional that $\theta_i > 0$ holds.

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 2 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 2 using condition (i).

Sequence 2:
(a) Uniform distribution Θ^Γ with $\theta_i = \theta_S - \left(\frac{(s-1)}{2} + 1 - i \right) \hat{\Delta}$, $\; i = 1, ..., s$ 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 $\theta_i = \theta_S - \frac{\hat{\Delta}}{4}(s+1)$ for all $i = 1, ..., \frac{(s-1)}{2}$, $\; \theta_k = \theta_S$ for player $k = \frac{(s-1)}{2} + 1$ and $\theta_j = \theta_S - \left(\frac{(s-1)}{2} + 1 - j \right) \hat{\Delta}$, $\; j = \frac{(s-1)}{2} + 2, ..., s$ for all $j \neq i, k$, generated from Θ^Γ by applying a sequence of changes in Proposition 2 using condition (i).
(c) Asymmetric distribution Θ^Υ with $\theta_i = \theta_S - \frac{\hat{\Delta}}{4}(s+1)$ for all $i = 1, ..., \frac{(s-1)}{2}$, $\; \theta_k = \theta_S$ for all players $k = \frac{(s-1)}{2} + 1, ..., s - 1$ and $\theta_j = \theta_S + \hat{\Delta}(\frac{s-1}{2}) + \hat{\Delta}\frac{(s-1)(s-3)}{8} = \theta_S + \hat{\Delta}\frac{(s^2-1)}{8}$ for player $j = s$, generated from Θ^Φ by applying a sequence of changes in Proposition 2 using condition (ii).
(d) Asymmetric distribution Θ^Ξ with $\theta_i = \theta_S - \frac{\hat{\Delta}}{8}(s+1)$ for all $i = 1, ..., s - 1$ and $\theta_j = \theta_S + \hat{\Delta}\frac{(s^2-1)}{8}$ for player $j = s$, generated from Θ^Υ by applying a sequence of changes in Proposition 2 using condition (i). Note that Θ^Ξ is equivalent to Θ^Ω if $\Delta = \hat{\Delta}\frac{(s^2-1)}{8}$.

A.5 Proposition 4

Using payoff function (4), $W = \sum_{i \in N} v_i = \sum_{i \in N} b_i B(q) - \sum_{i \in N} c_i C(q)$ from which it is evident that the marginal changes of b_i (or c_i-values) described in Proposition 2 do not change W. We know that W is strictly concave in q with $W'(q_S^A) = 0$ and $q_s^A = h(\theta_S)$ for all $S \subseteq N$ from Lemma 1. By construction, marginal changes do not affect θ_S and θ_m (or any $\theta_j, j \notin S$) but may affect the smallest autarky level $\theta_{\min}$. Therefore, $W^\varnothing(\Theta) < W^\varnothing(\tilde{\Theta})$ and

$$W^S(\Theta) = W^S(\tilde{\Theta}).$$

B Variance and Skewness Coefficient

We now define the conditions under which a marginal increase in stability (through the changes in Proposition 2) increases the variance and skewness of the θ_i-distribution. Applying the standard definition of 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 2 *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) of the distribution Θ and $m_3(\Theta)$ the third moment, and s the number of coalition members.

Relating the distributions Θ and $\tilde{\Theta}$ defined in Proposition 2 to the variance and skewness coefficient, we obtain the following proposition:[9]

Proposition 5 *Consider coalition S determining the equilibrium provision level and two distributions Θ and $\tilde{\Theta}$ as defined in Proposition 2, then $m_2(\tilde{\Theta}) > m_2(\Theta)$ for cases (ii) and (iii) in Proposition 2 and $g(\tilde{\Theta}) > g(\Theta)$ in all three cases if and only if*

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

Proof. All the marginal changes described in Proposition 2 imply that b_k becomes $b_k - \epsilon$ and b_l becomes $b_l + \epsilon$. Hence, $m_2(\tilde{\Theta}) > m_2(\Theta)$ implies

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

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

---

[9]If the assumption $c_i = c \; \forall \; i$ is substituted by the assumption $b_i = b \; \forall \; i$, Proposition 5 continues to hold. For the general case in which players differ in their b_i's and their c_i's, the proposition would continue to hold, but the coefficient $g(\Theta)$ would no longer be the standard skewness coefficient, as $\theta_S = \frac{\sum_{i \in S} b_i}{\sum_{i \in S} c_i}$ is no longer the average over all θ_i's (as this is the case if either $b_i = b \; \forall \; i$ or $c_i = c \; \forall \; i$).

which simplifies to $b_k < \epsilon + b_l$. This holds for cases (ii) and (iii) in Proposition 2 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 (A.10) after tedious algebraic manipulations, which are available from the authors upon request. ∎

Note that for case (i) in Proposition 2, we have $\theta_k + \theta_l - 2\theta_S < 0$ and hence condition (A.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 $m_2(\Theta)\left(\theta_k + \theta_l - 2\theta_S\right)$). For cases (ii) to (iii) in Proposition 2, we know that[10] $\theta_k + \theta_l - 2\theta_S > 0$, and thus, condition A.10 holds for negatively skewed or not too positively skewed distributions. As discussed in the main text, the intuition that all marginal changes proposed in Proposition 2 increase skewness is correct for moderately skewed distributions (whether positively or negatively skewed). For "strongly" skewed distributions (for which the absolute value of $m_3(\Theta)$ is larger than $m_2(\Theta)\left(\theta_k + \theta_l - 2\theta_S\right)$), there are exceptions: however, we can always increase stability and skewness at the same time by selecting one of the changes listed in Proposition 2 (i.e., case (i) for "strongly" positively skewed distributions and cases (ii) to (iii) for "strongly" negatively skewed distributions).

Proposition 5 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 rather than around the average θ_S. In other words, to extend Proposition 5 to any coalition not determining the equilibrium, we need to substitute $g(\Theta)$ by a similar coefficient, $\hat{g}(\Theta)$, defined using θ_m instead of the average of the distribution.

C   Equilibrium Selection

In the main text, we applied the criterion of Pareto-dominance to select the equilibrium provision level, in the Nash equilibrium, corresponding to the singleton coalition structure, but

---

[10]For case (iii) in Proposition 2 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 $\overleftarrow{\theta}_k \geq \overleftarrow{\theta}_{S\setminus l}$ as assumed in the proposition.

also if a non-trivial coalition S forms. Hence, generally, we assumed $Q^*(S) = \min\{q_m^A, ..., q_S^A\}$ for all $S \subseteq N$ with q_m^A being the smallest autarky level of players not belong to S. In contrast, experimental evidence suggests that efficient outcomes may be difficult to achieve when groups are large. However, as far as we are aware, like all theoretical contributions, all experiments assume that players play as single players, the possibility to form coalitions is not considered. For instance, 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 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 (2007) 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. 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, general payoff functions and coalition formation as in the present paper would be a paper in its own right. Therefore, we only take the main conclusions from these papers to motivate two simple alternative assumptions: a) $\widetilde{Q}^*(S) = \alpha(n)Q^*(S)$ and b) $\widetilde{\widetilde{Q}}^*(S) = \alpha(n,s)Q^*(S)$ for all $S \subseteq N$, with $0 \le \alpha(n) \le 1$ and $0 \le \alpha(n,s) \le 1$ 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 $Q^*(S)$, $Q^*(S) = \min\{q_j^A, q_m^A, ..., q_S^A\}$. We assume that $\alpha(n)$ and $\alpha(n,s)$

11

decrease in n and $\alpha(n,s)$ increases in s. The difference between both assumptions is how we count players; the second assumption treats the coalition as one player. Thus, the larger the coalition in relation to the total number of players, the smaller is the departure from the Pareto-optimal provision level. Hence, anything else being equal, coalition formation by itself leads to improved coordination for the second assumption.

The question we pose now is whether our results would still hold or, if not, what would change.

For symmetric players, we concluded that there is no need to analyze coalition formation because Nash equilibrium and social optimum coincide (See Section 3). If $\alpha < 1$, Nash equilibrium and social optimum will be different, the difference being inversely related to the value of α. For assumption a, coalition formation still cannot make a difference but for assumption b it can because $\widetilde{Q}^{*}(S)$ increases in $\alpha(n,s)$ which increases in s by assumption. Hence, the grand coalition is the unique stable coalition. The reason is simple. Payoffs increase in provision levels, which increase in s as long as $\alpha(n,s) < 1$. Hence, all coalitions S are internally stable as $\widetilde{V}_i^{*}(S) > \widetilde{V}_i^{*}(S\backslash\{i\})$ and therefore all coalitions $S \subset N$ are externally unstable, except the grand coalition, $S = N$. The larger the difference between $\alpha(n,n)$ and $\alpha(n,1)$, the larger will be the gain from cooperation in the grand coalition.
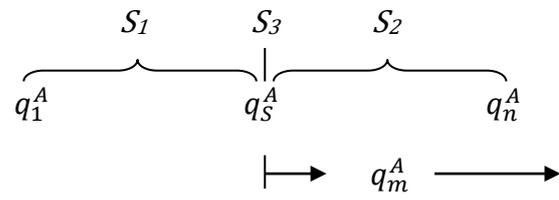
For asymmetric players and no transfers, we concluded that no agreement departing from the Nash equilibrium can be stable (Proposition 1). The reason was that an effective agreement $Q^{*}(S) > Q^{\varnothing} = q_1^A$ must include player 1 with the lowest autarky provision level but $V_1^{*}(S) < V_1^{*}(S\backslash\{1\}) = V_1^{\varnothing}$. Now, we can have $\widetilde{V}_1^{*}(S) \geq \widetilde{V}_1^{*}(S\backslash\{1\}) = \widetilde{V}_1^{\varnothing}$ ($\widetilde{\widetilde{V}}_1^{*}(S) \geq \widetilde{\widetilde{V}}_1^{*}(S\backslash\{1\}) = \widetilde{\widetilde{V}}_1^{\varnothing}$) which may include the grand coalition. (If for instance $\widetilde{Q}^{*}(S) = q_1^A$, respectively, $\widetilde{\widetilde{Q}}^{*}(S) = q_1^A$ this is obvious.) A sufficient (though not necessary) condition for internal stability of S is that $\alpha(n) \leq \underline{\alpha}(n) := \frac{q_1^A}{\widetilde{Q}^{*}(S)}$, respectively $\alpha(n,s) \leq \underline{\alpha}(n,s) := \frac{q_1^A}{\widetilde{\widetilde{Q}}^{*}(S)}$. That is, the equilibrium provision level if S forms is the below the autarky provision level of player 1. 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 stable coalitions. For a given coalition S, equilibrium provision levels will be below Pareto-optimal levels, but this could be compensated by larger coalitions being stable.

For asymmetric players and transfers, we argued in Section 4 that the transfer scheme which we considered is optimal if the coalition formation game is characterized by positive externalities and superadditivity because this gives us full cohesiveness. This was proved in Appendix A.1, see Lemma 3. It is easy to see that the lemma as well as the properties also hold for both alternative assumptions about α. Furthermore, in Proposition 1, we concluded

that an effective stable coalition exists. This was also proved in Appendix A.1. It is readily checked that the proof continues to hold with a slight modification, accounting for α. Also the further analysis in Section 5 (Propositions 2, 3 and 4) which characterized the nature and degree of asymmetry for stability and the associated global welfare gains continue to hold without any qualification. The only exception is condition (i) in Proposition 2. For condition (i), we used $v'_k\left(h(\theta)\right)\big|_{\theta_k} = v'_l\left(h(\theta)\right)\big|_{\theta_l} = 0$ in Appendix A.3 whereas now we need to assume for assumption a, $v'_k\left(\alpha(n)h(\theta)\right)\big|_{\theta_k} \geq v'_l\left(\alpha(n)h(\theta)\right)\big|_{\theta_l}$, and for assumption b, $v'_k\left(\alpha(n,s)h(\theta)\right)\big|_{\theta_k} \geq v'_l\left(\alpha(n,s)h(\theta)\right)\big|_{\theta_l}$.
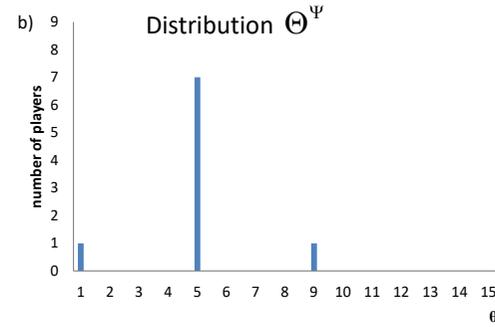
Figure 1: Weak, Strong, and Neutral Players

Case 1

$$S_1 \qquad S_3 \qquad S_2$$

$$q_1^A \qquad\qquad q_S^A \qquad\qquad q_n^A$$

$$\longmapsto \quad q_m^A \quad \longrightarrow$$

Case 2

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

$$S_1 \qquad\qquad S_3 \qquad\qquad S_2$$
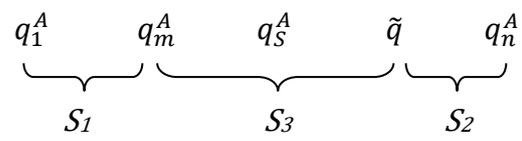
Figure 2: Illustration of Proposition 2

(i)

$\theta_l \quad \overrightarrow{\theta_l} \qquad \overleftarrow{\theta_k} \quad \theta_k \qquad \theta_y$

(ii)

$\theta_y \qquad \overleftarrow{\theta_k} \quad \theta_k \qquad \theta_l \quad \overrightarrow{\theta_l}$

(iii)

$\overleftarrow{\theta_{s\backslash l}} \qquad \overleftarrow{\theta_k} \quad \theta_k \quad \theta_y \qquad \theta_l \quad \overrightarrow{\theta_l}$

Sequence 1:

a) Distribution Θ^Λ

b) Distribution Θ^Ψ

c) Distribution Θ^Ω

Sequence 2:

a) Distribution Θ^Γ

b) Distribution Θ^Φ
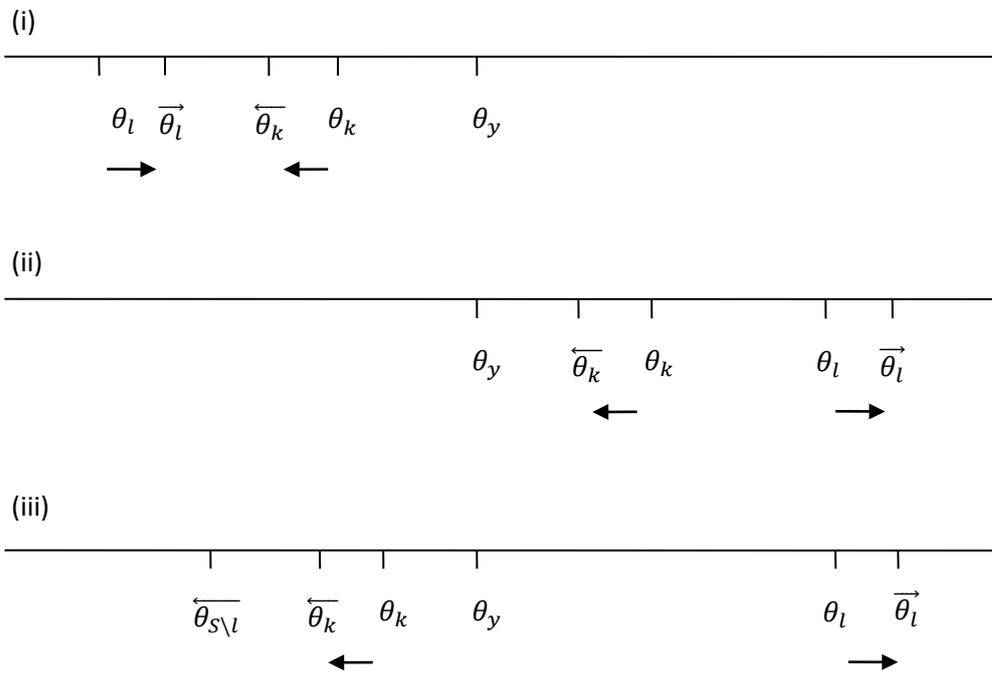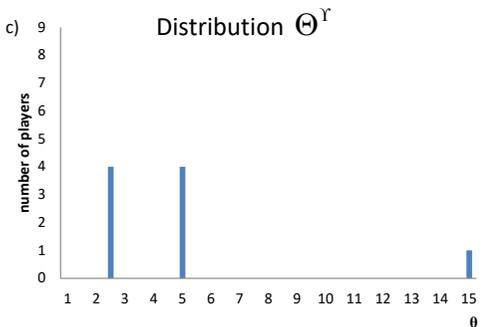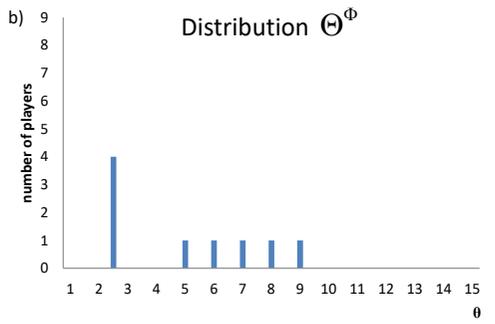
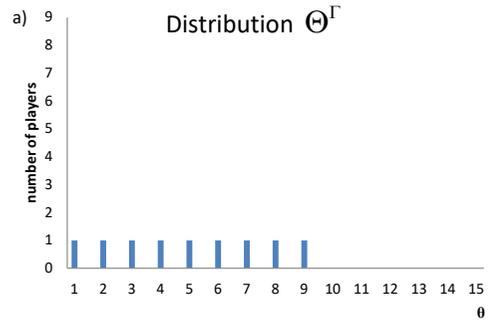c) Distribution Θ^Υ
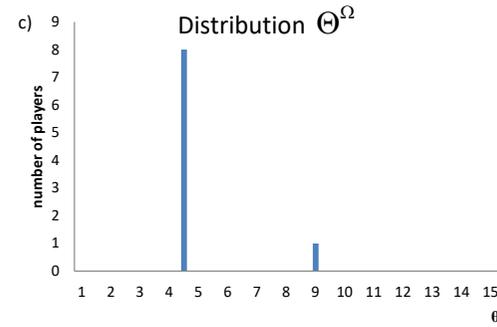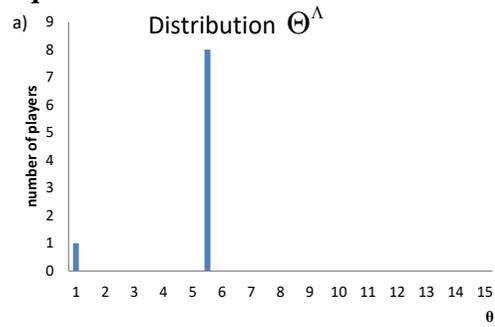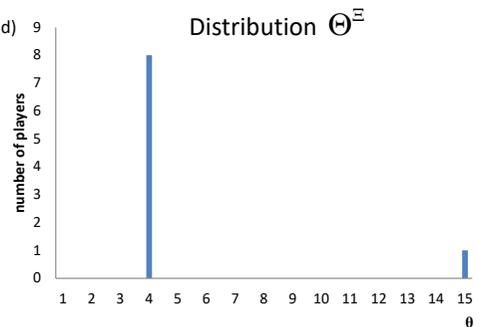
d) Distribution Θ^Ξ

Figure 3: Distributions of autarky public good provision levels