The Anti-paradox of Cooperation: Diversity Pays!

Michael Finus and Matthew McGinty

No. 40 /15

BATH ECONOMICS RESEARCH PAPERS

Department of Economics

The Anti-paradox of Cooperation: Diversity Pays!

Michael Finus and Matthew McGinty*

August 4th, 2015

Abstract

This paper considers the stability and success of a pubic good agreement. We allow for any type and degree of asymmetry regarding benefits and costs. We ask the question whether asymmetry and which type and degree of asymmetry is conducive to cooperation? We employ a simple non-cooperative game-theoretic model of coalition formation and derive analytical solutions for two scenarios: an agreement without and with optimal transfers. A central message of the paper is that asymmetry does not have to be an obstacle for successful cooperation but can be an asset. We qualify and reverse two central results in the literature. Firstly, the paradox of cooperation, known since Barrett (1994) and reiterated by many others afterwards, stating that under those conditions when cooperation would matter most, stable agreements achieve only little. Secondly, a kind of "coalition folk theorem", known (without proof) in the literature for a long time, stating that without transfers, stable coalitions will be smaller with asymmetric than symmetric players. We show that even without transfers the grand coalition can be stable if there is a negative covariance between benefit and cost parameters with massive gains from cooperation. Moreover, with transfers, many distributions of benefit and cost parameters lead to a stable grand coalition, again, some of them implying huge gains from cooperation. Stability and success greatly benefit from a very skewed asymmetric distribution of benefit and costs, i.e. diversity pays!

Keywords: public good provision, coalition formation, asymmetry, externalities, transfers

JEL classification: C7, D7, F5, H4

*Corresponding Author: Department of Economics, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI, 53201. Phone: 414-229-6146, email: mmcginty@uwm.edu.

1 Introduction

There are many cases of international and global public goods for which the decision in one country has consequences for other countries and which are not internalized via markets. Reducing global warming and the thinning of the ozone layer are examples in case. As Sandler (1998), p. 221, points out: "Technology continues to draw the nations of the world closer together and, in doing so, has created novel forms of public goods and bads that have diminished somewhat the relevancy of economic decisions at the nation-state level." The stabilization of financial markets, the fighting of contagious diseases and the efforts of non-proliferation of weapons of mass destruction have gained importance through globalization and the advancement of technologies. A central feature is the underprovision of most global public goods. Even in the absence of incomplete and asymmetric information, the lack of sufficient cooperation can be explained by the strategic behavior of governments. Differences across world regions and countries with respect to the benefits and costs of global public good provision add to the complication of signing meaningful treaties which depart from the non-cooperative statues quo. With a few exceptions (Arce and Sandler 2003, Ray and Vohra 2001 and Sandler 1999), the general literature on public goods focused on the voluntary provision in a Nash equilibrium (Bergstrom et al. 1986 and many others), but ignored the possibility of agreements, which is the focus of this paper. Only the literature with particular reference to international environmental agreements (IEAs), which has grown immensely in recent years (for a recent survey and a collection of some of the most influential papers, see Finus and Caparros 2015), predominantly focused on the formation of self-enforcing agreements.

Two main approaches have emerged (Finus 2001, 2003 and Tulkens 1998). The cooperative approach used mainly the stability concept of the core (e.g. Ambec and Sprumont 2002, Chander and Tulkens 1995, 1997, Eyckmans and Tulkens 2003). An imputation is said to be in the core if no single player or no subgroup of players has an incentive to deviate. An imputation is an allocation of the total worth of a coalition, the worth being the aggregate payoff to the coalition. Under the γ-core assumption (the most widespread assumption in the literature), a deviation by a set of players triggers the break-up of the remaining players in the coalition. The assumption is that the coalition members choose their economic strategies (e.g. emission or abatement levels) as to maximize the aggregate payoff to their coalition whereas singletons (if any) maximize their individual payoffs. In an externality game, it is easy to show that only an imputation in the grand coalition qualifies to be in the core. Chander and Tulkens analyzed this problem in various papers by considering particular imputations that are derived from a set of transfer rules, which were later termed the Chander-Tulkens transfer rules. Essentially, every player in the grand coalition receives his payoff in the Nash equilibrium, i.e. the payoff if all players played as singletons, plus a share of the surplus, which is defined as the difference between the worth in the social optimum (i.e. the grand coalition) and the sum of all Nash equilibrium payoffs. For the shares, they

considered different assumptions, but the most popular seems to be the ratio between the individual marginal damage from emissions and the sum of marginal damages in the social optimum. Essentially, this transfer rule is a Nash-bargaining solution in a TU-game with unequal weights. Under mild conditions, Chander and Tulkens (1995, 1997) show analytically that their imputation lies in the γ-core.[1] Clearly, the focus of the cooperative approach is on solving the asymmetry problem which may hamper the formation of an agreement. However, the cooperative approach is not very well suited to explain positive issues of agreement formation (Ray and Vohra 2001), like the lack of full participation in international treaties and inefficient provision levels.

In contrast, the non-cooperative approach, which we employ in this paper, predominately used the concept of internal & external stability (I&E-S) in a cartel formation game to test for stability of agreements.[2] Internal & external stability considers only deviations by a single player (either an insider leaving the coalition or an outsider joining the coalition) assuming that other players do not revise their membership strategy, though they revise their economic strategies. Economic strategies follow again from the assumption that the coalition maximizes the aggregate payoff whereas singletons maximize their individual payoffs. Hence, for instance in a public good provision game, leaving a coalition is more attractive under I&E-S than under the γ-core assumption and hence usually the grand coalition is not stable. The reason is simple. The public good provision game, like many others games, is a positive externality game.[3] That is, starting from a coalition structure where all players play as singletons, and hence there is no form of cooperation, gradually forming larger coalitions implies that the payoffs of outsiders not involved in the enlargement increase, each time one more player is added to the coalition. This means that for a player leaving the coalition, the weakest punishment is if all remaining coalition members remain in the coalition (I&E-S assumption) and the harshest punishment is if all remaining players break-up into singletons (γ-core assumption).

Many of the early papers (Barrett 1994 and Carraro and Siniscalco 1993) but also many later papers (e.g. Diamantoudi and Sartzetakis 2006 and Rubio and Ulph 2006) using the I&E-S concept assumed ex-ante symmetric players for simplicity.[4] Ex-ante means that all players have the same payoff function, though ex-post players may be different, depending whether they are coalition members or singletons.[5] The main conclusion from this literature

[1] Helm (2001) later generalised this result (without assuming a particular imputation) by showing that the γ-core is none empty, i.e. there exists at least one imputation which is immune to deviations by all possible groups of players. In later papers, e.g. Germain et al. 2003) extend the analysis to a difference game with a dynamic payoff structure and show that an imputation lies in the core along the entire time path.

[2] It has been shown later (Bloch 1997, Yi 1997) that this is a particular concept among a much broader class of coalitions games in partition function form. See Finus and Rundshagen (2009) for a theoretical exposition and Carraro and Marchiori (2003) and Finus and Rundshagen (2003) for applications to IEAs.

[3] Other games with positive externalities can be found in Bloch (1997) and Yi (1997).

[4] This assumption is not only widespread in the literature on IEAs but also in many other fields of coalition formation, using the partition function approach. See Bloch (1997) and Yi (1997) for surveys.

[5] In a more general setting, with multiple coalitions, payoffs depend on the size of the coalition to which a

under the standard assumptions is what Barrett (1994) called the "paradox of cooperation". That is, stable coalitions do not achieve a lot. Either stable coalitions are small or even if they are large, then the gap between the aggregate payoff in the grand coalition (social optimum or full cooperation) and the all singletons coalition structure (Nash equilibrium or no cooperation) is small and hence there is not really a need for cooperation. Starting from this pessimistic result, various extensions have been analyzed in the literature in order to find out whether they lead to more optimistic results.[6] In the context of this paper, the departure from ex-ante asymmetric players is the most interesting one. Under the assumption that coalition members choose their economic strategies by maximizing their aggregate welfare, this normally leads to an asymmetric distribution of the gains from cooperation with those receiving less than their fair share having an incentive to leave the coalition. In the absence of transfers, this implies smaller coalitions than under the symmetry assumption, at least this was the common view for a long time of most scholars working in this area, almost known like a "coalition folk-theorem".

One way to address the asymmetry problem in the absence of transfers was already suggested by Hoel (1992) who considered various second best-designs, which constitute a deviation from the assumption of joint welfare maximization. For instance, in Altamirano et al. (2008) and Finus and Rundshagen (1998) a bargaining process is considered about uniform emission reduction quotas or emission taxes where all coalition members make proposals and agree on the median (corresponding to majority voting) or smallest proposal (corresponding to unanimity voting). Essentially, there are two driving forces why a second-best design may lead to larger stable coalitions. First, the bargaining process implies that less ambitious emission reductions are implemented within the coalition, which reduces the free-rider incentive. Second, uniform emission reduction quotas, though not cost-effective, lead to a more symmetric distribution of the gains from cooperation than a cost-effective emission tax.

Another way to address this problem is to assume transfers. For a long time, concepts from cooperative game theory, like Nash bargaining solution, the Shapley value or the Chander-Tulkens transfer rule have been applied (Altamirano-Cabrera and Finus 2006, Barrett 2001 and Botteon and Carraro 1997, Eyckmans and Finus 2006 and Weikard et al. 2006). The problem is that different transfer rules lead to different coalition structures (sensitivity) and it is not clear whether other transfer rules would be even better suited to mitigate the free-rider incentive (optimality). In other words, similar to the Chander-Tulkens transfer rule in the context of the core, there was a necessity to develop an "optimal" transfer rule in the context of I&E-S. Such a transfer rule or sharing scheme, which Eyckmans and Finus (2004) call an "almost ideal sharing scheme", was independently developed by Eyckmans and Finus (2004), McGinty (2007) and Weikard (2009), illustrated with a calibrated climate model in Carraro,

player belongs.

[6] A comprehensive overview of alternative assumptions is presented in Finus (2008) and in the volume edited by Finus and Caparros (2015).

Eyckmans and Finus (2006), experimentally tested in McGinty et al. (2012) and further developed to capture the idea of trembles by players in McGinty (2011). Surprisingly, it turns out that the basic structure is very similar to the Chander-Tulkens transfer rule, which, as we have pointed out above, is a Nash bargaining solution in a TU-framework (Eyckmans, Finus and Mallozzi 2012). The only difference is that the threat point or disagreement point is not complete no cooperation but the payoff if one player leaves the coalition and all other players continue with cooperation. Accordingly, the surplus is defined as the sum of payoffs in a coalition minus the sum of free-rider payoffs. As Eyckmans and Finus (2004) show, weights do not matter for the set of stable coalitions, though equal sharing can be linked to some axiomatic properties, known from solution concepts or sharing rules of cooperative game theory as illustrated in Eyckmans, Finus and Mallozzi (2012). Moreover, among those coalitions, which can be potentially internally stable (i.e. the surplus is positive, which may not necessarily be true in the grand coalition), this almost ideal transfer scheme stabilizes the coalition with the highest aggregate welfare under rather general conditions.

Though all these general properties associated with the optimal transfer scheme are interesting, the papers above do not answer one fundamental question: How does asymmetry matter for coalition formation? From Finus and Pintassilgo (2013), McGinty (2007), Pintassilgo et al. (2010) and Weikard (2009) one derives two hints. First, with optimal transfers asymmetry may not necessarily lead to smaller coalitions but could also lead to larger stable coalitions than with symmetry. Second, even the grand coalition may be stable. A more systematic analysis is conducted Fuentes-Albero and Rubio (2010), Neitzel (2013) and Pavlova and de Zeeuw (2013), but, in the tradition of previous papers (e.g. Barrett 1997), they restrict the analysis to two or four types of players.[7] Consequently, modelling the type and degree of asymmetry among players is limited and some of our interesting results cannot be obtained. In this paper, we allow for any kind of asymmetry among players and analyze how asymmetry affects the size of stable coalitions, without transfers and with optimal transfers. We generalize a surprising result by Pavlova and de Zeeuw (2013), namely that even in the absence of transfers, "the right degree of asymmetry" among players allows to stabilize larger coalitions than under symmetry. Different from Pavlova and de Zeeuw (2013) we show that even in the absence of transfers, the grand coalition may be stable and, most importantly, the associated gains from cooperation may be huge. We show that with transfers, those gains may even be larger. The overall message is clear: asymmetry does not necessarily constitute an obstacle for successful cooperation but may in fact be an asset. We call this the anti-paradox of cooperation and characterize the type and degree of asymmetry which is conducive for large and successful stable coalitions, in the absence and presence of transfers.

The rest of the paper is organized as follows. Section 2 sets out our model and Section 3 derives some general properties useful to understand the incentive structure and the impli-

[7] Other papers in a similar spirit are Biancardi and Villani (2010) and Kolstad (2010).

cations of coalition formation. Section 4 characterizes stable coalitions without transfers and Section 5 does the same for transfers. Section 6 concludes and discusses the generality of our assumptions and directions for future research.

2 Model

2.1 Coalition Formation Game

Let the set of players be denoted by N with cardinality $n = |N|$ and consider the following simple two-stage coalition formation game due to d'Aspremont et al. (1983), which has been called cartel formation game or, more recently, referred to as open membership single coalition game (Yi 2003) in order to stress the institutional settings of this game. In the first stage, all players simultaneously choose whether they want to join coalition $S \subseteq N$ or remain a non-signatory, with cardinality $m = |S|$. This is essentially an announcement game with two possible strategies: all players who announce 1 are members of coalition S and all players who announce 0 are singletons. Given the simple structure of this "single coalition game", a coalition structure (i.e. the partition of players) is fully characterized by coalition S. In the second stage, players simultaneously choose their economic strategies. Coalition members, to whom we also refer as signatories, derive their strategies from maximizing the aggregate payoff to all signatories, which is the sum of all coalition members' payoffs. That is, the coalition acts like a meta player (Haeringer 2004), fully internalizing the externality among its members. In contrast, non-members, to whom we also refer as non-signatories, simply maximize their own payoff.

The game is solved by backwards induction. For any given coalition S, solving the maximization task of signatories and non-signatories in the second stage simultaneously, delivers a vector of equilibrium strategies $q^*(S) = (q_1^*(S), q_2^*(S), ..., q_n^*(S))$. Assuming a unique equilibrium for all possible coalitions S, equilibrium payoffs (also called valuations) follow simply from inserting strategies into the payoff functions of players, $\pi_i(q^*(S)) = \pi_i^*(S)$, and a vector of payoffs is derived: $\pi^*(S) = (\pi_1^*(S), \pi_2^*(S), ..., \pi_n^*(S))$. In the first stage, it is then tested which coalition(s) is (are) stable. Following d'Aspremont et al. (1983), we define a stable coalition as a coalition which is internally and externally stable:

$$\begin{aligned} &\text{internal stability:} && \pi_i^*(S) \geq \pi_i^*(S \setminus \{i\}) \ \forall i \in S \\ &\text{external stability:} && \pi_j^*(S) \geq \pi_j^*(S \cup \{j\}) \ \forall j \notin S. \end{aligned}$$

It is easy to see that internal & external stability is essentially a Nash equilibrium in membership strategies. A signatory has no incentive to leave coalition S, meaning that player i has no incentive to change announcement 1 to 0, given the announcement of all other players. By the same token, a non-signatory has no incentive to join coalition S, meaning

that player j has no incentive to announce 1 instead of 0.[8] Due to the definition of external stability, membership is open to all players, nobody can be precluded from joining coalition S.[9] Note that the "all singletons coalition structure" is generated by either $S = \{i\}$ or $S = \emptyset$ and, hence, strictly speaking, is always stable. If all players announce 0, and hence S is empty, a change of an individual player's membership strategy does change the coalition structure. Of course, subsequently, we are only interested in the stability of non-trivial coalitions, i.e. coalitions with $m > 1$. Moreover, in the case of multiple stable coalitions, we apply the Pareto-criterion and delete those stable coalitions from the set of stable coalitions which are Pareto-dominated by other stable coalitions. In our public good game, it turns out that the all singletons coalition structure is always Pareto-dominated by larger stable coalitions. This argument is briefly developed in Section 3. Finally note that we rule out knife-edge cases for simplicity by assuming henceforth that if a player is indifferent between remaining a non-signatory or joining coalition S, this player is assumed to join S.[10]

The definition of stability above assumes no transfers. Focusing on internal stability, it is evident that a necessary condition for internal stability is potential internal stability.

$$\begin{aligned} \text{potential internal stability:} \qquad & \sum_{i \in S} \pi_i^*(S) \geq \sum_{i \in S} \pi_i^*(S \setminus \{i\}) \\ \iff \qquad & \sigma(S) = \sum_{i \in S} \pi_i^*(S) - \sum_{i \in S} \pi_i^*(S \setminus \{i\}) \geq 0. \end{aligned}$$

That is, the surplus, $\sigma(S)$, defined as the difference between the total coalitional payoff and the sum of free-rider payoffs must be (weakly) positive. It is also clear that potential internal stability can be a sufficient condition for internal stability in the presence of transfers, provided that transfers are optimally designed. The optimal transfer scheme mentioned in the introduction does exactly this: no resources are wasted and every coalition member receives her free-rider payoff $\pi_i^*(S \setminus \{i\})$ plus a share $\lambda_i \geq 0$ of the surplus $\sigma(S)$, $\sum_{i \in S} \lambda_i = 1$.[11] Henceforth, if we talk about transfers, we mean the optimal transfer scheme.

Given the assumption about the second stage, clearly, if S is either empty or comprises only one player, which we may call "no cooperation", $q^*(S)$ is equivalent to a Nash equilibrium known from games without coalition formation. By the same token, if the grand coalition forms, i.e. $S = N$, which we may also call "full cooperation", this corresponds to the social optimum. Any coalition strictly larger than 1 but strictly smaller than n may be referred to as partial cooperation.

[8] Modeling the cartel formation game as an announcement game can be useful when comparing it with other games as for instance demonstrated in Finus and Rundshagen (2006) but would not add anything in the context of this paper.

[9] Exclusive membership games are described for instance in Bloch (1997), Finus and Rundshagen (2003) and Yi (1997).

[10] That is, henceforth, we replace the weak by a strong inequality sign in the external stability condition above as this is frequently done in the literature. This helps to reduce the number of stable equilibria.

[11] That is, payoffs after transfers, $\pi_i^{*T}(S)$, are given by $\pi_i^{*T}(S) = \pi_i^*(S \setminus \{i\}) + \lambda_i \, \sigma(S)$.

It is also obvious that the grand coalition must lead to an aggregate payoff which is at least as high than in any other coalition. In an externality game, this relation is strict, which is called strict cohesiveness. However, there are many interesting economic problems where the grand coalition is not stable, like in output or price cartels and public good games. Broadly speaking, this can be related to two reasons. Firstly, starting from the coalition structure with only single players and gradually enlarging the coalition by adding a player to S, superadditivity (see Definition 1 below) may not hold for all coalitions. That is, the aggregate payoff of those players involved in the enlargement of a coalition may not necessarily increase. For instance, in games in which strategies are strategic substitutes, superadditivity may fail for small coalitions. The joint efforts of the coalition members (e.g. reduction of output in a output cartel or increase of the provision level in a public good game) may be contradicted (through an increase of output and a decrease of contributions, respectively) by many free-riders However, even if superadditivity holds, it may still pay to stay outside a coalition in a game with positive externalities (see Definition 1 below). In an output cartel, output of signatories decrease with an enlargement of the coalition from which non-signatories benefit through a higher price. Similarly, in a public good game, the provision levels of signatories increase with the size of the coalition from which also non-signatories benefit as benefits are non-exclusive.

Definition 1: Superadditivity, Positive Externality and Cohesiveness

(i) A coalition game is (strictly) superadditive if for all $S \subseteq N$, $m > 1$, *and all* $i \in S$*:*

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

(ii) A coalition game exhibits a (strict) positive externality if for all $S \subseteq N$, $m > 1$, *and for all* $j \notin S$:

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

(iii) A game is (strictly) fully cohesive if for all $S \subseteq N$, $m > 1$*:*

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

It is obvious that a game which is superadditive and exhibits positive externalities is fully cohesive (and hence cohesive). Full cohesiveness is an important normative motivation to analyze the conditions under which large coalitions can be stable.[12]

[12] Note that in a (strictly) cohesive game we only know that the aggregate payoff in the grand coalition is (strictly) higher than in any other coalition. In a (strictly) fully cohesive game, we know that in each step of the enlargement coalition, the aggregate payoff (strictly) increases. Hence, even the grand coalition may not be stable, we know that "the larger the coalition, the larger will be global welfare".

2.2 Payoff Function

Consider the following pure public good game with individual contributions $q_i \geq 0$ and aggregate contribution $Q = \sum_{i \in N} q_i$ with payoff π_i defined as the difference between the benefit $B_i(Q)$ from the aggregate contribution and the cost from the individual contribution $C_i(q_i)$.

$$\pi_i = B_i(Q) - C_i(q_i) \tag{1}$$

$$\pi_i = \alpha_i b Q - \frac{\beta_i c}{2} (q_i)^2 \tag{2}$$

Payoff function (2) is probably the simplest representative of a strictly concave payoff function and has therefore been frequently considered in the literature (Barrett 1994, Breton et al. 2006, Finus and Pintassilgo 2013 and Ray and Vohra 2001). Though it would be possible to derive general properties regarding the second stage based on a general payoff function like the one given in (1), the analysis of stable coalitions, which is the central focus of this paper, necessitates the assumption of a specific payoff function, even for symmetric players as it is evident for instance from Bloch (1997), Ray and Vohra (2001) and Yi (1997).

In (2), $b > 0$ is a global benefit parameter; $\alpha_i > 0$ captures possible different benefits and can be interpreted as the share each player receives in which case $\sum_{i \in N} \alpha_i = 1$. Hence, one unit of public good provision generates $\alpha_i b$ marginal benefits to an individual player and b to all players. The global cost parameter is $c > 0$ and the individual cost parameter is $\beta_i > 0$. The individual marginal cost of a contribution is $\beta_i c q_i$ and hence the slope of the marginal cost curve is $\beta_i c$.

For instance, in the context of climate change, q_i can be interpreted as emission reduction or abatement. Nations such as the United States and China that use a relatively large proportion of coal have smaller $\beta_i > 0$ than nations such as Norway and France that use a relatively large amount of nuclear energy. Nations which are either exposed to large damages or perceive those damages to be high have a large share α_i and hence benefit more from emission reduction than other nations.

Note that we neither need to impose a normalization on the benefit parameter α_i nor the cost parameter β_i as all results will only depend on the ratio of these parameters. However, for ease of interpretation, we assume henceforth $\sum_{i \in N} \alpha_i = 1$. We note that assuming $\sum_{i \in N} \beta_i = n$ would allow to interpret the parameter $c > 0$ as the arithmetic mean of the marginal abatement cost slopes, though we do not use this normalization below.

For notational simplicity, we denote the set of players outside the coalition by T, where $S \cap T = \emptyset$ and $S \cup T = N$. Assuming an interior equilibrium, the first-order conditions of a non-signatory $j \in T$ read

$$\frac{\partial \pi_{j \in T}}{\partial q_j} = \alpha_j b - \beta_j c q_j = 0 \quad \Leftrightarrow \quad \alpha_j b = \beta_j c q_j$$

implying that individual marginal benefits, $\alpha_j b$, are set equal to individual marginal costs, $\beta_j c q_j$, from which the equilibrium provision level $q^*_{j\in T}$ follows:

$$q^*_{j\in T} = \frac{\alpha_j b}{\beta_j c}. \tag{3}$$

Note that in this example the equilibrium provision level of a non-signatory is independent of coalition S, a property which substantially eases computations. The first order condition of a signatory $i \in S$ is given by

$$\frac{\partial \sum_{k\in S} \pi_k}{\partial q_i} = \sum_{k\in S} \alpha_k b - \beta_i c q_i = 0 \quad \Leftrightarrow \quad \sum_{k\in S} \alpha_k b = \beta_i c q_i,$$

implying that the sum of marginal benefits of coalition S, $\sum_{k\in S} \alpha_k b$, is set equal to individual marginal cost, $\beta_i c q_i$, a kind of Samulson optimality condition for the coalition, from which the equilibrium provision level $q^*_{i\in S}$ follows:

$$q^*_{i\in S}(S) = \frac{b \sum_{k\in S} \alpha_k}{c\beta_i} \tag{4}$$

and hence the total contribution of signatories, $Q^*_S(S) = \sum_{i\in S} q^*_i(S)$, non-signatories, $Q^*_T(S) = \sum_{j\in T} q^*_j$, and over all players, $Q^*(S) = Q^*_S(S) + Q^*_T(S)$ is given by:

$$Q^*_S(S) = \frac{b}{c} \sum_{i\in S} \frac{1}{\beta_i} \sum_{i\in S} \alpha_i \tag{5}$$

$$Q^*_T(S) = \frac{b}{c} \sum_{j\in T} \frac{\alpha_j}{\beta_j} \tag{6}$$

$$Q^*(S) = \frac{b}{c} \left[\sum_{i\in S} \frac{1}{\beta_i} \sum_{i\in S} \alpha_i + \sum_{j\in T} \frac{\alpha_j}{\beta_j} \right]. \tag{7}$$

Among signatories, the externality is internalized and marginal contribution costs equalize, $\beta_i c q^*_{i\in S} = \beta_k c q^*_{k\in S}$ for all $i, k \in S$, $i \neq k$, meaning that $Q^*_S(S)$ is cost-effectively provided among signatories. The ratio of contributions is inverse to the individual cost parameters, i.e. $\frac{q^*_{i\in S}}{q^*_{k\in S}} = \frac{\beta_k}{\beta_i}$, implying that those with a flatter marginal cost curve should contribute more to the public good than those with a steep slope.

Inserting equilibrium provision levels into payoff functions, delivers signatories' payoffs $\pi^*_{i\in S}(S)$, the worth of the coalition, i.e. the sum of payoffs across all members, $\Pi^*_S(S) = \sum_{i\in S} \pi^*_{i\in s}(S)$, the payoffs to players outside the coalition, $\pi^*_{j\in T}(S)$, the aggregate payoff of those outside the coalition, $\Pi^*_T(S) = \sum_{j\in T} \pi^*_{j\in T}(S)$ and the global payoff for any given coalition S, $\Pi^*(S) = \Pi^*_S(S) + \Pi^*_T(S)$. The details are provided in Appendix 1.

3 General Properties

Before analyzing stability of coalitions, we look at some general properties of the public good coalition formation game. We assume throughout payoff function (2). The first Lemma proves useful on our way to derive general properties. An enlargement of a coalition is associated with an increase of total contributions. Thus from a normative point and measured in physical terms, coalition formation pays. The global provision level strictly increases each time an outsider joins coalition S.

Lemma 1

Consider any enlargement of a coalition from $S \setminus \{i\}$ to S, then total contributions over all players increase, i.e.

$$Q^*(S) > Q^*(S \setminus \{i\}) \text{ for all } S,\ m > 1. \tag{8}$$

Proof. From (3) and (4) we observe that $q^*_{k\in S}(S) > q^*_{k\in S}(S \setminus \{i\}) \geq q^*_{j\in T}$ and hence $Q^*(S) > Q^*(S \setminus \{i\})$ follows trivially. ■

An equivalent property holds for global welfare as Proposition 1 stresses, i.e. strict full cohesiveness holds. Moreover, any enlargement of the coalition pays at the aggregate for those players involved in the enlargement, i.e. superadditivity holds, but also outsiders benefit from this enlargement, i.e. the positive externality property holds.[13] The latter property implies that the grand coalition is not necessarily stable.[14]

Proposition 1 *The public good coalition formation game is strictly superadditive, exhibits strict positive externalities and hence is strictly fully cohesive.*

Proof. Superadditivity: Given that non-signatories have a dominant strategy, $\max \sum_{k\in s} \pi_k$ instead of $\max \sum_{k\in s\setminus\{i\}} \pi_k$ must imply $\sum_{k\in s} \pi^*_k(S) > \sum_{k\in s\setminus\{i\}} \pi^*_k(S \setminus \{i\}) + \pi^*_i(S \setminus \{i\})$. Positive Externality: Non-signatories' benefits increase in total contributions, and we have $Q^*(S) > Q^*(S \setminus \{i\})$ from Lemma 1, but cost remain the same because non-signatories' contribution levels do not change when moving from $S \setminus \{i\}$ to S. Full Cohesiveness: Superadditivity and positive externalities are sufficient for full cohesiveness. ■

Proposition 1 also provides the argument why any stable non-trivial coalition will Pareto-dominate the all singletons coalition structure in our public good coalition formation game. Due to the strict positive externality property, every non-signatory will be strictly better off in any non-trivial coalition than in the all singletons coalition structure and signatories must

[13] Note that the positive externality property holds generally for any payoff function of the form given in (1), assuming a concave benefit function and a strictly convex cost function. In contrast, superadditivity may fail to hold for some payoff functions. For instance, it can be shown that assuming a quadratic instead of a linear benefit function, a move from the singleton coalition structure to a two-player coaliton may not be superadditive.

[14] In Weikard (2009) it is shown that in the cartel formation game the grand coalition is the unique stable equilibrium if the game is superadditive and exihibts negative externalities.

be at least as well off otherwise a coalition S would not be internally stable. It is for this reason that we ignore stability of the all singleton coalition structure in the remainder of this paper and only report on non-trivial stable coalitions.

Lemma 1 and Proposition 1 are useful in comparing the welfare implications of transfers. Proposition 2 confirms our intuition that transfers weakly improve upon the outcome under no transfers, in physical and welfare terms. The interesting part is that this holds also in the strategic context of coalition formation.

Proposition 2 *Among the set of stable coalitions without transfers denote the coalition with the highest aggregate contribution (welfare) by* S^{**}*. In the case of transfers, denote the coalition with the highest aggregate contribution (welfare) by* S^{T**}*. Let* $Q(S^{**})$ *and* $Q(S^{T**})$ *denote the aggregate contribution over all players in these coalitions, and in terms of aggregate welfare* $\Pi(S^{**})$ *and* $\Pi(S^{T**})$*, then* $Q(S^{T**}) \geq Q(S^{**})$ *and* $\Pi(S^{T**}) \geq \Pi(S^{**})$*.*

Proof. We recall, every coalition S which is internally stable is potentially internally stable but not vice versa. Hence, every coalition which is internally stable without transfers will be internally stable with transfers. Suppose by contradiction $Q(S^{T**}) < Q(S^{**})$. Clearly, if S^{**} is also externally stable with transfers, this inequality cannot hold. However, suppose S^{**} is not externally stable with transfers, then a coalition $S^{**} \cup \{k\}$ is potentially internally stable but due to Lemma 1 $Q(S^{**}) < Q(S^{**} \cup \{k\})$. Either $S^{**} \cup \{k\}$ is externally stable or a larger coalition will eventually be externally stable, noting that the grand coalition is externally stable by definition. In any case, this contradicts $Q(S^{T**}) < Q(S^{**})$. For welfare, we proceed exactly along the same lines, making use of Proposition 1, noting $\Pi(S^{**}) < \Pi(S^{**} \cup \{k\})$ from Proposition 1. ■

It will be interesting to analyze below under which conditions transfers strictly improve upon no transfers and how big the difference will be.

In order to measure the paradox of cooperation (or to demonstrate the opposite), stable coalitions need to be benchmarked. Several measures have been proposed in the literature (e.g. Barrett 1994 and Eyckmans and Finus 2006). For our purpose, it is sufficient to measure the severeness of the externality as the difference between full cooperation and no cooperation. The larger this difference, the larger the need for cooperation. In Proposition 3 we measure the externality in terms of the total contribution and welfare.

As expected, the larger the global benefit parameter b and the smaller the global cost parameter c, the more pronounced the externality will be. If b is small and c is high, even under full cooperation it would not be rational to increase contribution levels substantially above the non-cooperative status quo.

A central focus of this paper is the impact of asymmetry on the outcome of cooperation. The problem is that there is no unique measure of asymmetry. Moreover, generally speaking, there can be asymmetry on the benefit and on the cost side. Players with high benefit shares

may also have high cost shares (positive covariance) or high benefit shares may go along with low cost shares (negative covariance). At the most general level, there is an infinite number of possible distributions of benefit shares α_i and cost shares β_i, and hence also the number of combinations is infinite. Therefore, in order to fix ideas, it is useful to start with the benchmark case of symmetric players. Then one may increase the α_i-values of some players at the expenses of some other players. In a sequential process, one can generate a skewed distribution by increasing the α_i-value of a player i with a larger or equal benefit share than a player j by decreasing her α_j-value by the same amount. Obviously, the same can be done for cost shares β_i. In Proposition 3, we cover the case of symmetric costs and changing benefit shares in the manner as just described, the case of symmetric benefits and changing cost shares proceeding in the same manner and the case of asymmetries on both sides by generating a distribution with a positive covariance between benefits and costs. According to Proposition 3, increasing the asymmetry in such a manner allows for the conclusion that the difference between full and no cooperation increases with asymmetry.

Proposition 3 *Measure the severeness of the externality as the difference between total contribution (total welfare) under full cooperation (FC) and no cooperation (NC), then the severeness of the externality is given by*

$$\begin{aligned} \Delta Q &: = Q_{FC} - Q_{NC} = \frac{b}{c}\left[\sum_{i\in N}\frac{1-\alpha_i}{\beta_i}\right] > 0 \\ \Delta \Pi &: = \Pi_{FC} - \Pi_{NC} = \frac{b^2}{2c}\left[\sum_{i\in N}\frac{(1-\alpha_i)^2}{\beta_i}\right] > 0 \end{aligned} \tag{9}$$

which increases in b and decreases in c. Moreover, consider a distribution of the benefit parameter Ψ^α with $\alpha_1 \leq \alpha_2 \leq ... \leq \alpha_n$ and a distribution of the cost parameter Ψ^β with $\beta_1 \leq \beta_2 \leq ... \leq \beta_n$ and let distributions $\widetilde{\Psi}^\alpha$ and $\widetilde{\Psi}^\beta$ be derived respectively by a marginal change ϵ of two α_i-values (β_i-values) such that $\alpha_k - \epsilon$ ($\beta_k - \epsilon$) and $\alpha_l + \epsilon$ ($\beta_l + \epsilon$), $\alpha_k \leq \alpha_l$ ($\beta_k \leq \beta_l$), then

$$\Delta Q(\widetilde{\Psi}^\alpha, \Psi^\beta) - \Delta Q(\Psi^\alpha, \Psi^\beta) \geq 0 \quad \textit{(increasing } \alpha\textit{-variance)} \tag{10}$$

$$\Delta Q(\Psi^\alpha, \widetilde{\Psi}^\beta) - \Delta Q(\Psi^\alpha, \Psi^\beta) > 0 \quad \textit{(increasing } \beta\textit{-variance)} \tag{11}$$

$$\Delta Q(\widetilde{\Psi}^\alpha, \widetilde{\Psi}^\beta) - \Delta Q(\Psi^\alpha, \Psi^\beta) > 0 \quad \textit{(increasing positive } \alpha\textit{-}\beta\textit{-covariance)} \tag{12}$$

$$\Delta \Pi(\widetilde{\Psi}^\alpha, \Psi^\beta) - \Delta \Pi(\Psi^\alpha, \Psi^\beta) > 0 \quad \textit{(increasing } \alpha\textit{-variance)} \tag{13}$$

$$\Delta \Pi(\Psi^\alpha, \widetilde{\Psi}^\beta) - \Delta \Pi(\Psi^\alpha, \Psi^\beta) > 0 \quad \textit{(increasing } \beta\textit{-variance)} \tag{14}$$

$$\Delta \Pi(\widetilde{\Psi}^\alpha, \widetilde{\Psi}^\beta) - \Delta \Pi(\Psi^\alpha, \Psi^\beta) > 0 \quad \textit{(increasing positive } \alpha\textit{-}\beta\textit{-covariance).} \tag{15}$$

Proof. ΔQ is computed by using (7) above and $\Delta \Pi$ by using Appendix 1, noticing that under full cooperation $S = N$ and under no cooperation $T = N$. Results regarding

distributions follow from using ΔQ and $\Delta\Pi$ above, considering two changes at the same time, which delivers the result after some basic calculations. ■

For a negative covariance between benefits and costs such a general conclusion cannot be derived; the gap between full and no cooperation can increase or decrease with a change of the degree of asymmetry.[15] Nevertheless, as will be apparent from Table 1 below and Section 4, also for a negatively correlated distribution the gap can be quite large.

By the nature of Proposition 3, which looks at marginal changes, not much can be concluded about absolute magnitudes. Table 1 illustrates those for a simple three player example. Scenario 1 assumes symmetry of benefit and cost shares. Then, in scenarios 2 to 4, the cost share distribution becomes more and more skewed, i.e. the β-variance increases. Going from scenario 4 to 5 generates additionally a very skewed benefit share distribution with a positive α-β-covariance. As Proposition 3 predicts along this sequence ΔQ and $\Delta\Pi$ increase and, as the example shows, the magnitudes become very large.

Table 1: Three player example

No.	α_1	α_2	α_3	β_1	β_2	β_3	$\frac{\Delta Q}{b/c}$	$\frac{\Delta\Pi}{b^2/2c}$
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	6	4
2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}-0.3$	$\frac{1}{3}$	$\frac{1}{3}+0.3$	23.08	15.37
3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}-0.3$	$\frac{1}{3}-0.3$	$\frac{1}{3}+0.6$	40.7	27.14
4	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}+0.64$	100.68	67.12
5	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}+0.64$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}+0.64$	148.03	146.23
6	$\frac{1}{3}+\frac{0.32}{2}$	$\frac{1}{3}+\frac{0.32}{2}$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}+0.64$	77.01	39.51
7	$\frac{1}{3}+0.64$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}+0.64$	77.01	74.07
8	$\frac{1}{3}-0.32$	$\frac{1}{3}-0.32$	$\frac{1}{3}+0.64$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	6	5.48

Scenarios 6 and 7 generates a negative α-β-covariance (not covered by Proposition 3) compared to scenario 4, which may increase or decrease ΔQ and $\Delta\Pi$. However, important compared to the symmetric case, scenario 1, ΔQ and $\Delta\Pi$ are still pretty large. It is also evident that starting from symmetry (scenario 1) and increasing only the asymmetry on the cost side gradually from scenario 2 to 4, increases ΔQ and $\Delta\Pi$ substantially, whereas increasing only the benefit asymmetry (going from scenario 1 to 8) has no implications for ΔQ and minor implications for $\Delta\Pi$. Hence, single asymmetry on the cost side can increase ΔQ and $\Delta\Pi$ by much, but single asymmetry on the benefit side has hardly any implication. The largest increases of ΔQ and $\Delta\Pi$ are generated by increasing the positive α-$\beta-$covariance, followed by an increase of the negative α-$\beta-$covariance and an increase of the $\beta-$variance; an increase of only the $\alpha-$variance for symmetric costs does not do much.

[15] In line with Proposition 3, increasing the degree of asymmetry of the benefit and cost parameter simultaneously, generating a negative covariance between benefit and costs, would mean $\alpha_k-\epsilon$, $\beta_k+\epsilon$, $\alpha_l+\epsilon$ and $\beta_l-\epsilon$, which does not allow for general conclusions regardless of the assumption about the relation between α_k and α_l as well as β_k and β_l.

4 Stable Coalitions without Transfers

Given the fact our public good coalition game is fully cohesive, it seems natural that one is more concerned about internal than external stability. Internal stability requires that $\sigma_i(S) = \pi_i^*(S) - \pi_i^*(S \setminus \{i\}) \geq 0$ holds for all $i \in S$. Using payoff function (2), and conducting some basic though cumbersome manipulations (see Appendix 2), leads to the following compact and closed form solution:

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j \neq i \in S} \frac{1}{\beta_j} - \frac{\left(\sum_{j \neq i \in S} \alpha_j\right)^2}{2\beta_i}\right] \tag{16}$$

which allows for the following statement.

Proposition 4 *In the absence of transfers, for any number of signatories, a necessary and sufficient condition for internal stability is:*

$$2\beta_i \sum_{j \neq i \in S} \frac{1}{\beta_j} \geq \left(\frac{\sum_{j \neq i \in S} \alpha_j}{\alpha_i}\right)^2 \text{ for all } i \in S. \tag{17}$$

Proof. See Appendix 2. ■

The internal stability condition (17) is remarkably simple compared to the complicated simulations found in the literature. The cost asymmetry is on the left-hand side and the benefit asymmetry is on the right-hand side. Only the symmetry or asymmetry of the m coalition members matter for internal stability, but not those of the $n-m$ outsiders. Moreover, the level of the global benefit and cost parameter b or c do not matter. Still, the interpretation of (17) may not be straightforward. Therefore, in a first step, we consider three benchmark cases: 1) full symmetry, 2) benefit symmetry but cost asymmetry and 3) benefit asymmetry but cost symmetry, before we focus in a second step on asymmetry on the benefit and cost side simultaneously. With asymmetry, we mean that at least two players in the population have a different benefit or cost share parameter.

Corollary 1 *Assume no transfers. Denote the size of an internally stable coalition by m^* and an internally and externally stable coalition by m^{**}. a) Full symmetry: $m^{**} = 3$. b) Benefit symmetry and cost asymmetry: i) $m^{**} < 3$ if at least two signatories have a different cost share and ii) $m^* = 3$ if and only if three coalition members have the same cost share which is externally stable if and only if no outsider $l \in T$ has a cost share β_l substantially higher than those of the three coalition members i, j and k, $\beta_i = \beta_j = \beta_k = \beta$, i.e. $\frac{\beta_l}{\beta} < \frac{3}{2}$. c) Benefit asymmetry and cost symmetry: i) $m^{**} < 3$ if at least two signatories have a different benefit share and ii) $m^* = 3$ if and only if three coalition members have the same benefit share which is externally stable if and only if no outsider $l \in T$ has a benefit share α_l substantially larger than those of the three coalition members i, j and k, $\alpha_i = \alpha_j = \alpha_k = \alpha$, i.e. $\frac{\alpha_l}{\alpha} < \frac{1}{2}\sqrt{6} \approx 1.22$.*

Proof. Define $\lambda_i :\equiv \left(\frac{\sum_{j\neq i\in S}\alpha_j}{\alpha_i}\right)^2$ and $\rho_i := 2\beta_i \sum_{j\neq i\in S}\frac{1}{\beta_j}$. Note that with identical benefit shares $\lambda_i = (m-1)^2$ and with identical cost shares $\rho_i = 2(m-1)$. Moreover, the mean of the benefit shares of a coalition of size m is given by $\bar{\alpha}_s \equiv \frac{\sum_{i\in S}\alpha_i}{m}$ and the mean of the cost shares is $\bar{\beta}_s \equiv \frac{\sum_{i\in S}\beta_i}{m}$. With single-sided asymmetry, we must have $\alpha_i < \bar{\alpha}_s < \alpha_j$ and $\beta_i < \bar{\beta}_s < \beta_j$. Therefore, $\lambda_i > (m-1)^2$, $\lambda_j < (m-1)^2$, $\rho_i < 2(m-1)$ and $\rho_j > 2(m-1)$. a) For symmetric players internally stability is given by $\rho_i := 2(m-1) \geq (m-1)^2 = \lambda_i$ which holds for $m \leq 3$ but is violated for $m > 3$. $m = 3$ is externally stable because $m = 4$ is not internally stable; $m = 2$ is not externally stable because outsiders are indifferent between staying outside and joining and hence we assume them to join. b) $\lambda_i = (m-1)^2$ and $\rho_i < 2(m-1)$ if two players in the coalition have different cost shares. Hence, $\lambda_i = (m-1)^2 \geq 2(m-1) > \rho_i$, and thus $\rho_i < \lambda_i$ for all $m \geq 3$, violating internal stability for player i. For three players with symmetric cost shares, $\lambda_i = (m-1)^2 = 2(m-1) = \rho_i$ for all $i \in S$ and hence $m^* = 3$. It is externally stable provided for $m = 4$, $\rho_l = 6\frac{\beta_l}{\beta} < 9 = \lambda_l$ or $\frac{\beta_l}{\beta} < \frac{3}{2}$. c) $\rho_i = 2(m-1)$ and $\lambda_i > (m-1)^2$ if two players in the coalition have different benefit shares. Hence, $\lambda_i > (m-1)^2 \geq 2(m-1) = \rho_i$ and thus $\rho_i < \lambda_i$ for all $m \geq 3$, violating internal stability for player i. For three players with symmetric benefit shares, $\lambda_i = (m-1)^2 = 2(m-1) = \rho_i$ for all $i \in S$ and hence $m^* = 3$. It is externally stable provided for $m = 4$, $\rho_l = 6 < \left(3\frac{\alpha}{\alpha_l}\right)^2 = \lambda_l$ or $\frac{\alpha_l}{\alpha} < \frac{1}{2}\sqrt{6} \approx 1.22$. ■

The result for symmetry is well-known in the literature and is a good benchmark: the largest stable coalition comprises three countries. With single asymmetry, either on the cost or benefit side, stable coalitions will be strictly smaller, except for the special case that three players are symmetric among a population of asymmetric players. Even in this special case, such a coalition of three symmetric players is only externally stable if the outsiders are not too different from the insiders, otherwise they would have an incentive to join the coalition. This result is very much in line with intuition and was known almost like a "folk-theorem" in coalition theory for a long time: departing from symmetry will lead to smaller stable coalitions in the absence of transfers. For our payoff function (2), internal stability holds at the margin for a coalition of three players. Any asymmetry will cause that some players get slightly more but others slightly less of the "cooperative cake", upsetting internal stability.

The next two corollaries look at asymmetry on the benefit and cost side. The first does not contradict the "coalition folk-theorem", assuming a positive covariance between benefit and cost shares, the second puts the folk theorem upside down, assuming a negative covariance.

Corollary 2 *In the absence of transfers, stable coalitions comprise strictly less than three players when there is a positive covariance between benefit and cost shares.*

Proof. Using the notation introduced in the proof of Corollary 1, a positive covariance between benefit and costs must imply for some coalition members i $\alpha_i < \bar{\alpha}_s$ and $\beta_i < \bar{\beta}_s$ simultaneously. Therefore $\lambda_i > (m-1)^2$ and $\rho_i < 2(m-1)$. Consequently, internal stability must be violated for coalitions with three or more members when there is a positive covariance

since $\lambda_i > (m-1)^2 \geq 2(m-1) > \rho_i$ for $m \geq 3$. ■

With a positive covariance between benefit and cost shares, stable coalitions are strictly smaller than 3, the benchmark size in the case of symmetric players. The intuition is clear. A member with below average cost share will contribute more than the average to the internalization of other members' benefits. However, with a positive covariance, this disadvantage is reinforced because the same member has also a below average benefit share. Hence, in order to compensate the lack of transfers, larger coalitions can only be obtained, if at all, if we have an asymmetry with a negative covariance. The next corollary sheds light on this conjecture.

Corollary 3 *In the absence of transfers, coalitions of size larger than 3 and even the grand coalition can be stable when there is a negative covariance between benefit and cost shares.*

Proof. Consider $m = n = 10$; let $\alpha_1 = 0.43$ and $\alpha_2 = 0.41$ and let the other 8 nations have the same benefit share $\alpha_3 = ...\alpha_{10} = 0.02$, so that $\sum_{i \in S} \alpha_i = 1$. Let $\beta_1 < \beta_2$ since $\alpha_1 > \alpha_2$ so that there is a negative covariance between the two nations. Let $\beta_1 = 0.0001$ and $\beta_2 = 0.00011$, and the other 8 nations have $\beta_3 = ... = \beta_{10} = 0.12497375$, so that $\sum_{i \in S} \beta_i = 1$, even though the absolute benefit and cost shares do not matter, but only their ratios. In Appendix 3, we show that the grand coalition is stable. ■

Clearly, the example chooses extreme values: player 1 and 2 have high benefit shares, which requires very low cost shares in order to not advantage them too much. For the other 8 players, this is just reversed; because of their low benefit shares they also cannot contribute too much to cooperation, otherwise it would be attractive to leave the grand coalition. Hence, they need high cost shares.

Viewing Corollary 1, 2 and 3 together, we can learn a couple of lessons.
First, even in the absence of transfers, asymmetry does not necessarily lead to worse outcomes than symmetry. Hence, when talking about asymmetry one needs to be precise about the nature of the asymmetry.
Second, Corollary 1 and 2 seem to confirm the paradox of cooperation. For symmetry, it is clear that a coalition of three players does not achieve a lot if the number of players n is large. For symmetry, regardless of the normalization of the benefit and cost shares, it is straightforward to show that $\partial(Q_{FC} - Q_{NC})/\partial n > 0$ and $\partial(\Pi_{FC} - \Pi_{NC})/\partial n > 0$. For asymmetry, Proposition 3 has argued that ΔQ and $\Delta\Pi := Q_{FC} - Q_{NC}$ and $\Delta\Pi := \Pi_{FC} - \Pi_{NC}$ increase in the degree of asymmetry if the asymmetry is increased only on the benefit share side (though ΔQ does not change for symmetric costs and the change of $\Delta\Pi$ will be small as Table 1 showed), only on the cost share side or on both sides if there is a positive α-β-covariance (with Table 1 illustrating that the change of ΔQ and $\Delta\Pi$ can be large in the latter two cases). Hence, the larger the degree of asymmetry for these types of asymmetry, the more pressing is the need for cooperation, but the size of stable coalitions falls short of that under symmetry.

Third, in contrast, Corollary 3 suggest that with the right type of asymmetry (negative α-β-covariance), at least in terms of coalition size, we can have even full cooperation. Certainly, this contradicts the coalition folk theorem. But can Corollary 3 be viewed as anti-paradox of cooperation? This depends whether ΔQ and $\Delta\Pi$ are large when the grand coalition is achieved. We know already from Table 1 that with a negative α-β-covariance ΔQ and $\Delta\Pi$ can be substantially larger than for symmetry. However, as we could not derive general results regarding these two measures for a negative α-β-covariance in Proposition 3, we compute ΔQ and $\Delta\Pi$ for the example in Proposition 4:

$$\begin{aligned}\Delta Q &: = Q_{FC} - Q_{NC} = \frac{b}{c}\left[\frac{1-0.43}{0.0001} + \frac{1-0.41}{0.00011} + \frac{8\,(1-.02)}{0.12497375}\right] = \frac{b}{c}\,[11,126] \\ \Delta\Pi &: = \Pi_{FC} - \Pi_{NC} = \frac{b^2}{2c}\left[\frac{(1-0.43)^2}{0.0001} + \frac{(1-0.41)^2}{0.00011} + \frac{8\,(1-.02)^2}{0.12497375}\right] = \frac{b^2}{2c}\,[6,475] \end{aligned} \quad (18)$$

noting that for symmetry we would have:

$$\begin{aligned}\Delta Q &: = Q_{FC} - Q_{NC} = \frac{b}{c}\left[10\frac{(1-0.1)}{0.1}\right] = \frac{b}{c}\,[90] \\ \Delta\Pi &: = \Pi_{FC} - \Pi_{NC} = \frac{b^2}{2c}\left[10\frac{(1-0.1)^2}{0.1}\right] = \frac{b^2}{2c}\,[81]\ . \end{aligned} \quad (19)$$

Clearly, those differences in (18) are large when compared with those that would emerge under symmetry in (19).[16] But also in relative terms, difference are huge. In the example, the total contribution under no cooperation is $Q_{NC} = \frac{b}{c}\,[8,028.6]$ and under full cooperation it is $Q_{FC} = \frac{b}{c}\,[19,154.9]$, and hence $\frac{Q_{FC}}{Q_{NC}} = 2.38$. Similarly for total payoffs we have: $\Pi_{NC} = \frac{b^2}{2c}\,[12,680]$ and $\Pi_{FC} = \frac{b^2}{2c}\,[19,155]$, and hence $\frac{\Pi_{FC}}{\Pi_{NC}} = 1.51$. Therefore, we have a stable grand coalition without transfers that achieves very meaningful gains relative to the non-cooperative outcome. This finding is in sharp contrast to Pavlova and de Zeuuw (2013). They were the first (and only to our knowledge) who showed that the coalitional folk theorem may break down, in that a coalition larger than three players may be stable without transfers. Without any doubt, this is an important result and full credit should be given to the authors for this finding. However, in their simulations the grand coalition does not emerge and their "large" stable coalitions do not achieve a lot. In fact, they are argue that smaller coalitions may be preferable. The crucial difference is that we allow for any asymmetry whereas they assume two types of players as many others do in the literature. By the nature of their assumption, this places an upper bound on the degree of asymmetry. What is needed for stability is a very skewed distribution on the benefit and cost side with a negative covariance between the benefit and cost parameters. Only this helps to equalize the gains from cooperation among

[16] Note the ratio between ΔQ ($\Delta\Pi$) with asymmetry and symmetry is independent of the normalization of the benefit and cost shares.

players compared to their free-rider payoffs in the absence of transfers and to stabilize large coalitions, including the grand coalition. As the example proves, this is exactly also the situation when the need for cooperation is large. We argue that this is an interesting version of an anti-paradox of cooperation.

5 Stable Coalitions with Transfers

Previous work has shown that with asymmetry an optimal transfer scheme can increase the size of stable coalitions compared to symmetry (Carraro et al. 2006, McGinty, 2007 and Weikard, 2009). However, the degree and type of asymmetry to generate this result has not been characterized. Again, we focus on internal stability and recall that potential internal stability requires that $\sigma(S) = \sum_{i\in S} \pi_i^*(S) - \sum_{i\in S} \pi_i^*(S \setminus \{i\}) \geq 0$ holds for all $i \in S$. Using payoff function (2), and conducting some basic manipulation (see Appendix 4), leads to the following compact and closed form solution:

$$\sigma(S) = \frac{b^2}{c}\left\{\sum_{i\in S}\frac{1}{\beta_i}\left[\sum_{i\in S}(\alpha_i)^2 - \frac{\left(\sum_{i\in S}\alpha_i\right)^2}{2}\right] + \sum_{i\in S}\alpha_i\sum_{i\in S}\frac{\alpha_i}{\beta_i} - \frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\right\} \geq 0 \qquad (20)$$

which allows for the following statement.

Proposition 5 *Under an optimal transfer scheme coalition S is internally stable if and only if*

$$\sum_{i\in S}\frac{1}{\beta_i}\left[\sum_{i\in S}(\alpha_i)^2 - \frac{\left(\sum_{i\in S}\alpha_i\right)^2}{2}\right] + \sum_{i\in S}\alpha_i\sum_{i\in S}\frac{\alpha_i}{\beta_i} - \frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i} \geq 0 \qquad (21)$$

Proof. See Appendix 4. ■

From (20) and (21) it is evident that neither the level of the global benefit and cost parameter nor their ratio does matter for internal stability, only benefit and cost shares matter. In order to draw further conclusions, it is helpful to write (20) in an alternative form

$$\sigma(S) = \frac{b^2}{2c}\sum_{i\in S}\frac{1}{\beta_i}\left(\sum_{j\neq i\in S}\alpha_j^2 - 2\sum_{j,k\neq i\in S}\alpha_j\alpha_k\right) \qquad (22)$$

which makes it clear that if the benefit shares are distributed such that the term in brackets is positive, cost shares do not matter for internal stability. This is also confirmed if we consider (20) for the grand coalition, noticing that $S = N$ and hence $\sum_{i\in N}\alpha_i = 1$:

$$\sigma(N) = \frac{b^2}{c}\left[\sum_{i\in N}\frac{1}{\beta_i}\left(\sum_{i\in N}(\alpha_i)^2 - \frac{1}{2}\right) + \frac{1}{2}\sum_{i\in N}\frac{\alpha_i(2-3\alpha_i)}{\beta_i}\right]. \qquad (23)$$

For instance, a sufficient condition for $\sigma(N) \geq 0$ is that $\sum_{i\in N}(\alpha_i)^2 - \frac{1}{2} \geq 0$ and $(2-3\alpha_i) \geq 0$ for all $i \in N$. $\sum_{i\in N}(\alpha_i)^2 - \frac{1}{2} \geq 0$ requires that the largest benefit share

is not smaller than $\frac{1}{2}$ and $(2-3\alpha_i) \geq 0$ requires that the largest benefit share is smaller than $\frac{2}{3}$. So if we have two players with a large benefit share, say for instance $\alpha_i = 0.65$ and $\alpha_j = 0.3$, both inequalities are satisfied and the grand coalition will be stable. Of course, there are many more possibilities to stabilize a large coalition or even the grand coalition. The example just illustrates that it is not that difficult to stabilize a large or even the grand coalition. Moreover, (22) and (23) make it very clear that what really matters for stability is the distribution of the benefit shares. Given Proposition 3 and the illustrations in Table 1 about the gains from cooperation when going from no to full cooperation, those gains can be really large. A skewed distribution of the benefit shares combined with a positive covariance between benefits and costs, for instance generated through a reversed very skewed distribution of the cost shares, implies a stable grand coalition with transfers and a large gain from cooperation, in physical and payoff terms. This would be a distribution of benefit and cost shares for which only a coalition smaller than three players was stable in the absence of transfers (see Corollary 2). However, as the example in Corollary 3 in Section 4 has shown, also for a negative covariance between benefits and costs, the absolute and relative gains from cooperation can be large, though in this example the grand coalition was already stable without transfers. The next two corollaries further clarify the nature of transfers and the type of asymmetry necessary for successful cooperation in the presence of transfers.

Corollary 4 *Under an optimal transfer scheme a stable coalition comprises at least three members.*

Proof. Using (20), we derive for a two player coalition $\sigma(S, m=2) = \frac{b^2}{2c}\sum_{i,j\in S}\frac{\alpha_i^2}{\beta_j}$ and for a three player coalition $\sigma(S, m=3) = \frac{b^2}{2c}\left[\sum_{i,j,k\in S}\frac{(\alpha_j-\alpha_k)^2}{2\beta_i}\right]$ which are obviously positive and $m=2$ is not externally stable because $m=3$ is internally stable. ■

Corollary 4 provides a good benchmark with the case of symmetric players. With transfers stable coalitions will comprise at least three and possibly more players. This confirms with intuition: with transfers and asymmetry we should be able to replicate at least the outcome under symmetry. However, as we know from above, we may be able to do much better. Hence, Corollary 4 may also be viewed as an upper bound for the inefficiency that can arise from free-riding in our public good game.

The next corollary confirms our conjecture from above that the distribution of the benefit shares is crucial for the stability of coalitions.

Corollary 5 *If benefit shares are symmetric, the stable coalition comprises three players. If cost shares are symmetric, a sufficient condition for an internally stable coalition is given by*

$$\begin{aligned} CV(\alpha_S) &\geq \sqrt{\frac{m^2-4m+3}{2m-3}} \text{ or} \\ HF(\alpha_S) &\geq \frac{(m-2)}{(2m-3)} \end{aligned} \tag{24}$$

where $CV(\alpha_S)$ is the coefficient of variation of the benefit shares α_i in coalition S, which is the standard deviation, $\sqrt{var(\alpha_S)}$, divided by the average benefit share $\overline{\alpha}_S = \frac{\sum_{i\in S}\alpha_i}{m}$, $\frac{\sqrt{var(\alpha_S)}}{\overline{\alpha}_S}$ where $var(\alpha_S)$ is the variance of the benefit shares in coalition S where $\sqrt{\frac{m^2-4m+3}{2m-3}}$ increases in m. Moreover, $HF(\alpha_S)$ is the (modified) Herfindahl index of benefit shares α_i in coalition S, $HF(\alpha_S) = \frac{\sum_{i\in S}(\alpha_i)^2}{\left(\sum_{i\in S}\alpha_i\right)^2}$ where $\frac{(m-2)}{(2m-3)}$ increases in m with $\lim_{m\to\infty}\frac{(m-2)}{(2m-3)} = \frac{1}{2}$.

Proof. See Appendix 5. ■

Corollary 5 clearly stresses once more that what matters is the type of asymmetry regarding the benefit shares. We provide two alternative measures of asymmetry. The first is the coefficient of variation of benefit shares which needs to be sufficiently large for a coalition S to be stable. The larger coalition S, i.e. the larger m, the larger this coefficient needs to be for stability. For instance, for $m = 10$ to be stable, $CV(\alpha_S) \geq 1.925$ and for $m = 20$, $CV(\alpha_S) \geq 2.955$. The second measure is the "modified" Herfindahl index which is frequently used to measure the concentration in markets in the US. In our context, the "modified" Herfindahl index of benefit shares of coalition members needs to be sufficiently high for stability. For instance, for $m = 10$ to be stable, $HF(\alpha_S) \geq 0.47$ and for $m = 20$, $HF(\alpha_S) \geq 0.49$ is required. For large m, the benchmark value is 0.5 which is satisfied for instance if one player has a benefit share larger than $\left(\frac{1}{2}\right)^{\frac{1}{2}} \approx 0.707$. In other words, all we need is a very skewed distribution of benefit shares such that the grand coalition is stable if costs are symmetric. We view this as another version of the anti-paradox of cooperation.

The interesting question is: what is the intuition that asymmetric benefit shares matter a lot but not asymmetric cost shares for stability? At first sight the result may appear to be counter-intuitive as asymmetric costs would suggest that the gains from cooperation should be large and hence the incentive for joining a coalition. This intuition is not completely wrong, but requires a slight twist of the argument. What matters for stability are not the absolute but the relative gains from cooperation compared to the free-rider gains. Joining a coalition has two implications. First, there is a gain from internalizing externalities among coalition members. This gain is non-exclusive and also accrues to outsiders. Second, the gain from cost-effective cooperation, which is exclusive to coalition members. Interesting, the relative exclusive gain depends crucially on the distribution of benefit shares. Inserting equilibrium contribution levels in the non-cooperative equilibrium $q_i^* = \frac{\alpha_i b}{\beta_i c}$, which may be viewed as the starting point before coalition formation takes place, into marginal contribution costs $\beta_i c q_i^*$, gives $\alpha_i b$. Hence, initially marginal costs are equalized across all players if and only if $\alpha_i = \alpha_j = ... = \alpha_n = \alpha$, irrespective of the individual cost shares, and the difference across players increases in the differences of benefit shares. Hence, the exclusive relative cost effectiveness effect from cooperation increases with the degree of asymmetry of benefit shares but not with the asymmetry of cost shares (which may sound like another paradox of cooperation). Nevertheless, as we have shown above, a skewed distribution of the cost shares matters for the absolute gains from cooperation.

6 Summary and Conclusions

We have analyzed a simple public good coalition formation game in which the enlargement of the agreement generates global welfare gains. However, joining an agreement is voluntary and there are free-rider incentives to stay outside. The free-rider incentive may be so strong that large coalitions may not be stable, letting alone the grand coalition. From the previous literature using a non-cooperative coalition theory approach, two central messages emerged. Firstly, whenever the gains from cooperation would be large, stable coalitions do not achieve a lot. This being the case because either coalitions are small or if they are large, the difference between full and no cooperation is small, both in contribution and welfare terms. This was called the paradox of cooperation by Barrett (1994). Secondly, the larger the asymmetry among players, the smaller will be stable coalitions in the absence of transfers. This conclusion was known as a kind of folk theorem for a long time. In this paper, we showed how the paradox can be transformed into an anti-paradox of cooperation and that the folk theorem does not always hold. Without and with transfers we need a strong asymmetry with skewed distributions of benefit and cost parameters. Without transfers, there must be a negative covariance of benefit and cost shares to generate large stable coalitions. This works like a compensation mechanism in the absence of transfers. Those players who contribute more than proportionally to cost-effective public good provision within the coalition need to be compensated with high benefit shares. Different from Pavlova and de Zeeuw (2013), we showed that even the grand coalition can be stable and most importantly that the gains from cooperation can be very large. Admittedly, stability without transfers requires a very skewed distribution on the cost and benefit side with a high negative covariance. However, with transfers, there are many benefit and cost share distributions which can lead to large stable coalitions. In fact, as we have shown in our model, a sufficient condition for the grand coalition being stable is that two players have a sufficiently large benefit share. If, additionally, there is a positive covariance of benefit and cost shares, then the gains from cooperation can be massive.

For instance, in climate change, we should expect a high positive covariance of benefit and cost shares. On the one hand, most industrialized countries face steep marginal abatement costs but also place a high emphasis on the benefits from emission reduction. On the other hand, most developing countries would have very flat marginal abatement cost slopes but view the climate problem as less important. Given the few big key players among industrialized countries in the climate change game suggests that those few have relatively large benefit shares - exactly those conditions which are conducive to cooperation and are associated with large global gains from cooperation. In the light of little progress in international climate change negotiations, this suggests that more emphasis should be placed on optimal compensation mechanisms in order to reap those gains in the future. If this is well understood, diversity is an asset and not an obstacle.

Finally, let us briefly address the question about the generality of our results and possible future research. Firstly, our results have been based on a simple payoff function with linear benefits from total contributions and quadratic costs from individual contributions. This allowed us to derive analytical solutions, which, admittedly, would most likely be impossible for more complicated payoff functions. Though quantitative results would differ for other functions, we strongly believe that all qualitative results (i.e. asymmetry can be an asset for successful cooperation) would carry over to other payoff functions. Secondly, we considered the simple open membership coalition game. On the one hand, in terms of stability, these are the most pessimistic assumptions. Players outside the coalition can join if they find this attractive due to open membership. Players leaving the coalition assume that the remaining coalition members continue to cooperate and only reoptimize their economic strategies which, in a game with positive externalities, is the weakest implicit punishment. Hence it appears that this is a sensible benchmark assumption in order to show that the "right degree of asymmetry" can overcome free-riding incentives. On the other hand, deviating from a single to a multiple coalition game would also not add much, given that we could show that the grand coalition is not an unlikely equilibrium.[17] Thirdly, one could depart from a public good setting and look at other economic problems with a similar incentive structure, like for instance coalition formation in a price and output oligopoly or trade agreements which exhibit positive externalities from coalition formation as considered in Bloch (1997) and Yi (1997). We expect that similar results could be shown. In the case of no transfers, the standard models would need to be extended in order to generate not only asymmetry on the cost side but also on the benefit side, such that our negative covariance result in Section 4 can be replicated. Fourthly, we believe a more interesting and qualitative different extension would look at the optimal allocation of contributions in the absence of transfers in a non-transferable utility framework. This implies a departure from the standard assumption which derives the equilibrium vector of coalitional contributions from the maximization of the aggregate payoff to coalition members. Instead, one would search for a contribution vector which maximizes a weighted sum of individual welfare, subject to the constraint of voluntary participation in coalitions and the possibility that players can free-ride.

[17] A systematic comparision of equilibrim coalition structures for different single and multiple coalition games in positive externality games is conducted in Finus and Rundshagen (2006, 2009).

References

[1] Altamirano-Cabrera, J.C. and M. Finus (2006), "Permit Trading and Stability of International Climate Agreements." *Journal of Applied Economics*, vol. 9(1), pp. 19-48.

[2] Altamirano-Cabrera, J.C., M. Finus and R. Dellink (2008), "Do Abatement Quotas Lead to More Successful Climate Coalitions?" *The Manchester School*, vol. 76(1), pp. 104-129.

[3] Ambec, S. and Y. Sprumont (2002), "Sharing a River." *Journal of Economic Theory*, vol. 107(2), pp. 453-462.

[4] Arce, D. and T. Sandler (2003), "Health-promoting Alliances." *European Journal of Political Economy*, vol. 19(2), pp. 355-375.

[5] 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.

[6] Barrett, S. (1994), "Self-enforcing International Environmental Agreements." *Oxford Economic Papers*, vol. 46, pp. 878-894.

[7] Barrett, S. (1997), "Heterogeneous International Environmental Agreements." In: Carraro, C. (ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 9-25.

[8] Barrett, S. (2001), "International Cooperation for Sale." *European Economic Review*, vol. 45(10), pp. 1835-1850.

[9] Bergstrom, T.C., L. Blume and H.R. Varian (1986), "On the Private Provision of Public Goods." *Journal of Public Economics*, vol. 29(1), pp. 25-49.

[10] Biancardi, M. and G. Villani (2010), "International Environmental Agreements with Asymmetric Countries." *Computational Economics*, vol. 36(1), pp. 69-92.

[11] Bloch, F. (1997), "Non-cooperative Models of Coalition Formation in Games with Spillovers." In: Carraro, C. and D. Siniscalco (eds.), New Directions in the Economic Theory of the Environment. Cambridge University Press, Cambridge, UK, ch. 10, pp. 311-352.

[12] Botteon, M. and C. Carraro (1997), "Burden-sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries." In: Carraro, C. (ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, UK et al., ch. 3, pp. 26-55.

[13] Breton, M., K. Fredj and G. Zaccour (2006), "International Cooperation, Coalitions Stability and Free Riding in a Game of Pollution Control." *The Manchester School*, vol. 74(1), pp. 103-122.

[14] Carraro, C., J. Eyckmans and M. Finus (2006), "Optimal Transfers and Participation Decisions in International Environmental Agreements." *Review of International Organizations*, vol. 1(4), pp. 379-396.

[15] Carraro, C. and C. Marchiori (2003), "Stable Coalitions." In: Carraro, C. (ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK et al., ch. 5, pp. 156-198.

[16] Carraro, C. and D. Siniscalco (1993), "Strategies for the International Protection of the Environment." *Journal of Public Economics*, vol. 52(3), pp. 309-328.

[17] Chander, P. and H. Tulkens (1995), "A Core-theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution." *International Tax and Public Finance*, vol. 2(2), pp. 279-293.

[18] Chander, P. and H. Tulkens (1997), "The Core of an Economy with Multilateral Environmental Externalities." *International Journal of Game Theory*, vol. 26(3), pp. 379-401.

[19] 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.

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

[21] Eyckmans, J. and M. Finus (2006), "Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Success of Environmental Treaty-making." *Natural Resource Modeling*, vol. 19(3), pp. 323-358.

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

[23] Eyckmans, J. and H. Tulkens (2003), "Simulating Coalitionally Stable Burden Sharing Agreements for the Climate Change Problem." *Resource and Energy Economics*, vol. 25(4), pp. 299-327.

[24] Finus, M. (2001), "Game Theory and International Environmental Cooperation." Edward Elgar, Cheltenham, UK et al.

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

[26] Finus, M. (2008), "Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks and Future Challenges." *International Review of Environmental and Resource Economics*, vol. 2(1), pp. 29-67.

[27] Finus, M. and A. Caparrós (2015), "Game Theory and International Environmental Cooperation." The International Library of Critical Writings in Economics. Edward Elgar, Cheltenham, UK.

[28] Finus, M. and P. Pintassilgo (2013), "The Role of Uncertainty and Learning for the Success of International Climate Agreements." *Journal of Public Economics*, vol. 103, pp. 29-43.

[29] Finus, M. and B. Rundshagen (1998), "Toward a Positive Theory of Coalition Formation and Endogenous Instrumental Choice in Global Pollution Control." *Public Choice*, vol. 96 (1-2), pp. 145-186.

[30] Finus, M. and B. Rundshagen (2003), "Endogenous Coalition Formation in Global Pollution Control: A Partition Function Approach." In: Carraro, C. (ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK et al., ch. 6, pp. 199-243.

[31] Finus, M. and B. Rundshagen (2006), "A Micro Foundation of Core Stability in Positive-Externality Coalition Games". *Journal of Institutional and Theoretical Economics*, vol. 162(2), pp. 329-346.

[32] 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.

[33] Fuentes-Albero, C. and S.J. Rubio (2010), "Can International Environmental Cooperation Be Bought?" *European Journal of Operational Research*, vol. 202, pp. 255-64.

[34] Germain, M., P.L. Toint, H. Tulkens and A. de Zeeuw (2003), "Transfers to Sustain Dynamic Core-theoretic Cooperation in International Stock Pollutant Control." *Journal of Economic Dynamics & Control*, vol. 28(1), pp. 79-99.

[35] Haeringer, G. (2004), "Equilibrium Binding Agreements: A Comment." *Journal of Economic Theory*, vol. 117(1), pp. 140-143.

[36] Helm, C. (2001), "On the Existence of a Cooperative Solution for a Coalitional Game with Externalities." *International Journal of Game Theory*, vol. 30(1), pp. 141-146.

[37] Hoel, M. (1992), "International Environment Conventions: The Case of Uniform Reductions of Emissions." *Environmental and Resource Economics*, vol. 2(2), pp. 141-159.

[38] Kolstad, C.D. (2010), "Equity, Heterogeneity and International Environmental Agreements." *The B.E. Journal of Economic Analysis and Policy*, vol. 10(2), Article 3.

[39] McGinty, M. (2007), "International Environmental Agreements among Asymmetric Nations." *Oxford Economic Papers*, vol. 59, pp. 45-62.

[40] McGinty, M. (2011), "A Risk-Dominant Allocation: Maximizing Coalition Stability." *Journal of Public Economic Theory*, vol. 13(2), pp. 311-325.

[41] McGinty, M., G. Milam and A. Gelves (2012), "Coalition Stability in Public Goods Provision: Testing an Optimal Allocation Rule." *Environmental and Resource Economics*, vol. 52, pp. 327–345.

[42] Neitzel, J. (2013), "The Stability of Coalitions when Countries are Heterogeneous." Working Paper, available at SSRN: http://ssrn.com/abstract=1997357 or http://dx.doi.org/10.2139/ ssrn.1997357.

[43] Pavlova, Y. and A. de Zeeuw (2013), "Asymmetries in International Environmental Agreements." *Environment and Development Economics*, vol. 18, pp. 51-68.

[44] Pintassilgo, P., M. Finus, M. Lindroos and G. Munro (2010), "Stability and Success of Regional Fisheries Management Organizations." *Environmental and Resource Economics*, vol. 46, pp. 377-402.

[45] Ray, D. and R. Vohra (2001), "Coalitional Power and Public Goods." *Journal of Political Economy*, vol. 109(6), pp. 1355-1384.

[46] Rubio, S.J. and A. Ulph (2006), "Self-enforcing International Environmental Agreements Revisited." *Oxford Economic Papers*, vol. 58(2), pp. 233-263.

[47] Sandler, T. (1998), "Global and Regional Public Goods: a Prognosis for Collective Action." *Fiscal Studies,* vol. 19, pp. 221-247.

[48] Sandler, T. (1999), "Alliance Formation, Alliance Expansion, and the Core." *Journal of Conflict Resolution*, vol. 43(6), pp. 727-747.

[49] Tulkens, H. (1998), "Cooperation versus Free-riding in International Environmental Affairs: Two Approaches." In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 30-44.

[50] Weikard, H.P. (2009), "Cartel Stability under Optimal Sharing Rule." *Manchester School*, vol. 77(5), pp. 575-93.

[51] Weikard, H.-P., M. Finus and J.C. Altamirano-Cabrera (2006), "The Impact of Surplus Sharing on the Stability of International Climate Agreements." *Oxford Economic Papers*, vol. 58(2), pp. 209-232.

[52] Yi, S.-S. (1997), "Stable Coalition Structures with Externalities." *Games and Economic Behavior*, vol. 20(2), pp. 201-237.

Appendix

Appendix 1

Inserting equilibrium provision levels (4) and (7) into payoff functions (2), respectively, delivers a signatory's payoff $\pi^*_{i\in S}$

$$\begin{aligned}
\pi^*_{i\in S} &= \alpha_i b Q^* - \frac{c\beta_i \left(q^*_{i\in S}\right)^2}{2} \\
\pi^*_{i\in S} &= \frac{\alpha_i b^2}{c}\left[\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right] - \frac{c\beta_i}{2}\left[\frac{b}{c\beta_i}\sum_{i\in S}\alpha_i\right]^2 \\
\pi^*_{i\in S} &= \frac{b^2}{c}\left[\alpha_i\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right) - \frac{1}{2\beta_i}\left(\sum_{i\in S}\alpha_i\right)^2\right]. \qquad \text{(A1)}
\end{aligned}$$

The worth of the coalition, the sum of payoffs across all members, $\Pi^*_S = \sum_{i\in S}\pi^*_{i\in s}$ is given by:

$$\begin{aligned}
\Pi^*_S &= \frac{b^2}{c}\left[\sum_{i\in S}\alpha_i\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right) - \frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\left(\sum_{i\in S}\alpha_i\right)^2\right] \\
\Pi^*_S &= \frac{b^2}{c}\left[\sum_{i\in S}\frac{1}{\beta_i}\left(\sum_{i\in S}\alpha_i\right)^2 + \sum_{i\in S}\alpha_i\sum_{j\in T}\frac{\alpha_j}{\beta_j} - \frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\left(\sum_{i\in S}\alpha_i\right)^2\right] \\
\Pi^*_S &= \frac{b^2}{c}\sum_{i\in S}\alpha_i\left[\frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right] \qquad \text{(A2)}
\end{aligned}$$

The payoff to each player outside the coalition, $\pi^*_{j\in T}$, using (3) and (7) is

$$\begin{aligned}
\pi^*_{j\in T} &= \alpha_j b Q^* - \frac{c\beta_j\left(q^*_{j\in T}\right)^2}{2} \\
\pi^*_{j\in T} &= \frac{\alpha_j b^2}{c}\left[\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right] - \frac{c\beta_j\left[\frac{b\alpha_j}{c\beta_j}\right]^2}{2} \\
\pi^*_{j\in T} &= \frac{b^2}{c}\left[\alpha_j\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right) - \frac{(\alpha_j)^2}{2\beta_j}\right]. \qquad \text{(A3)}
\end{aligned}$$

The aggregate payoff of those outside the coalition, $\Pi^*_T = \sum_{j\in T}\pi^*_{j\in T}$ is

$$\Pi^*_T = \frac{b^2}{c}\left[\sum_{j\in T}\alpha_j\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right) - \frac{1}{2}\sum_{j\in T}\frac{(\alpha_j)^2}{\beta_j}\right] \qquad \text{(A4)}$$

and hence global payoff, for any given coalition S, is $\Pi^* = \Pi^*_S + \Pi^*_T$.

$$\begin{aligned}
\Pi^* &= \frac{b^2}{c}\left[\sum_{i\in S}\alpha_i\left(\frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i+\sum_{j\in T}\frac{\alpha_j}{\beta_j}\right)+\sum_{j\in T}\alpha_j\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i+\sum_{j\in T}\frac{\alpha_j}{\beta_j}\right)-\frac{1}{2}\sum_{j\in T}\frac{(\alpha_j)^2}{\beta_j}\right]\\
\Pi^* &= \frac{b^2}{c}\left[\sum_{i\in S}\alpha_i\sum_{i\in S}\frac{1}{\beta_i}\left(\frac{1}{2}\sum_{i\in S}\alpha_i+\sum_{j\in T}\alpha_j\right)+\sum_{j\in T}\frac{\alpha_j}{\beta_j}\left(\sum_{i\in S}\alpha_i+\sum_{j\in T}\alpha_j\right)-\frac{1}{2}\sum_{j\in T}\frac{(\alpha_j)^2}{\beta_j}\right]\\
\Pi^* &= \frac{b^2}{c}\left[\sum_{i\in S}\alpha_i\sum_{i\in S}\frac{1}{\beta_i}\left(\frac{1}{2}\sum_{i\in S}\alpha_i+\left(1-\sum_{i\in S}\alpha_i\right)\right)+\sum_{j\in T}\frac{\alpha_j}{\beta_j}-\frac{1}{2}\sum_{j\in T}\frac{(\alpha_j)^2}{\beta_j}\right]\\
\Pi^* &= \frac{b^2}{c}\left[\sum_{i\in S}\alpha_i\sum_{i\in S}\frac{1}{\beta_i}\left(1-\frac{1}{2}\sum_{i\in S}\alpha_i\right)+\sum_{j\in T}\frac{\alpha_j}{\beta_j}-\frac{1}{2}\sum_{j\in T}\frac{(\alpha_j)^2}{\beta_j}\right]\\
\Pi^* &= \frac{b^2}{c}\left[\sum_{i\in S}\alpha_i\sum_{i\in S}\frac{1}{\beta_i}\left(1-\frac{1}{2}\sum_{i\in S}\alpha_i\right)+\frac{1}{2}\sum_{j\in T}\frac{\alpha_j\left(2-\alpha_j\right)}{\beta_j}\right] \qquad \text{(A5)}
\end{aligned}$$

Appendix 2

Consider a three member coalition $\{i,j,k\}$ and let non-signatories be denoted by l. The payoff of a signatory, using (A1) from Appendix 1 is given by

$$\pi^*_{i\in s}(S) = \frac{b^2}{c}\left[\alpha_i\left(\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right)(\alpha_i+\alpha_j+\alpha_k)+\alpha_i\sum_{l\in T}\frac{\alpha_l}{\beta l}-\frac{(\alpha_i+\alpha_j+\alpha_k)^2}{2\beta_i}\right]. \qquad \text{(A6)}$$

If player i leaves coalition S, then the payoff is given by

$$\pi^*_{i\in T}(S\setminus\{i\}) = \frac{b^2}{c}\left[\alpha_i\left(\frac{1}{\beta_j}+\frac{1}{\beta_k}\right)(\alpha_j+\alpha_k)+\alpha_i\sum_{l\notin T}\frac{\alpha_l}{\beta_l}+\frac{\alpha_i^2}{2\beta_i}\right]. \qquad \text{(A7)}$$

Let $\sigma_i(S) = \pi^*_{i\in s}(S) - \pi^*_{i\in T}(S\setminus\{i\})$. In the absence of side-payments, coalition S is internally stable for i if $\sigma_i(S)\geq 0$, or

$$
\begin{aligned}
\sigma_i(S) &= \frac{b^2}{c}\left[\alpha_i\left(\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right)(\alpha_i+\alpha_j+\alpha_k)+\alpha_i\sum_{l\in T}\frac{\alpha_l}{\beta_l}-\frac{(\alpha_i+\alpha_j+\alpha_k)^2}{2\beta_i}\right] \\
&\quad -\frac{b^2}{c}\left[\alpha_i\left(\frac{1}{\beta_j}+\frac{1}{\beta_k}\right)(\alpha_j+\alpha_k)+\alpha_i\sum_{l\in T}\frac{\alpha_l}{\beta_l}+\frac{\alpha_i^2}{2\beta_i}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\alpha_i^2\left(\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right)+\frac{\alpha_i(\alpha_j+\alpha_k)}{\beta_i}-\frac{(\alpha_i+\alpha_j+\alpha_k)^2}{2\beta_i}-\frac{\alpha_i^2}{2\beta_i}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_i^2}{\beta_j}+\frac{\alpha_i^2}{\beta_k}+\frac{\alpha_i(\alpha_j+\alpha_k)}{\beta_i}-\frac{(\alpha_i+\alpha_j+\alpha_k)^2}{2\beta_i}-\frac{\alpha_i^2}{2\beta_i}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\frac{2\alpha_i^2+2\alpha_i(\alpha_j+\alpha_k)-(\alpha_i+\alpha_j+\alpha_k)^2-\alpha_i^2}{2\beta_i}+\frac{\alpha_i^2}{\beta_j}+\frac{\alpha_i^2}{\beta_k}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\frac{\alpha_i^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k-(\alpha_i+\alpha_j+\alpha_k)^2}{2\beta_i}+\frac{\alpha_i^2}{\beta_j}+\frac{\alpha_i^2}{\beta_k}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\frac{\alpha_i^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k-\left(\alpha_i^2+\alpha_j^2+\alpha_k^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k+2\alpha_j\alpha_k\right)}{2\beta_i}+\frac{\alpha_i^2}{\beta_j}+\frac{\alpha_i^2}{\beta_k}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\frac{-\left(\alpha_j^2+\alpha_k^2+2\alpha_j\alpha_k\right)}{2\beta_i}+\frac{\alpha_i^2}{\beta_j}+\frac{\alpha_i^2}{\beta_k}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\frac{-(\alpha_j+\alpha_k)^2}{2\beta_i}+\frac{\alpha_i^2}{\beta_j}+\frac{\alpha_i^2}{\beta_k}\right] \\
\sigma_i(S) &= \frac{b^2}{c}\left[\alpha_i^2\sum_{j\neq i\in S}\frac{1}{\beta_j}-\frac{\left(\sum_{j\neq i\in S}\alpha_j\right)^2}{2\beta_i}\right] \qquad \text{(A8)}
\end{aligned}
$$

A generalization to more than three players is cumbersome but follows the same pattern and gives (20) in the text for which it is obvious that the term in brackets needs to be positive for internal stability to hold.

Appendix 3

Let there be $m = n = 10$ signatories. Using definitions $\lambda_i \equiv \left(\frac{\sum_{j\neq i\in S}\alpha_j}{\alpha_i}\right)^2$ and $\rho_i \equiv 2\beta_i\sum_{j\neq i\in S}\frac{1}{\beta_j}$, internal stability requires $\rho_i \geq \lambda_i$ for all $i \in N$. Let $\alpha_1 = 0.43$ and $\alpha_2 = 0.41$ and let the other 8 nations have the same benefit share $\alpha_3 = ...\alpha_{10} = 0.02$, so that $\sum_{i\in S}\alpha_i = 1$. With these benefit shares, we have in the grand coalition:

$$
\begin{aligned}
\lambda_1 &= \left(\frac{1-0.43}{0.43}\right)^2 = 1.76 \\
\lambda_2 &= \left(\frac{1-0.41}{0.41}\right)^2 = 2.07 \\
\lambda_3 &= .. = \lambda_{10} = \left(\frac{1-0.02}{0.02}\right)^2 = 2,401. \qquad \text{(A9)}
\end{aligned}
$$

Let $\beta_1 < \beta_2$ since $\alpha_1 > \alpha_2$ so there is a negative covariance between the two nations that matter. Let $\beta_1 = 0.0001$ and $\beta_2 = 0.00011$, and the other 8 nations have $\beta_3 = ... = \beta_{10} = 0.12497375$, so that $\sum_{i \in S} \beta_i = 1$, even though the absolute benefit and cost shares do not matter, but only their ratio. With this cost distribution at the grand coalition we have

$$\begin{aligned}
\rho_1 &= 2\left[\frac{0.0001}{0.00011} + \frac{8\,(0.0001)}{0.12497375}\right] = 1.83 \\
\rho_2 &= 2\left[\frac{0.00011}{0.0001} + \frac{8\,(0.00011)}{0.12497375}\right] = 2.21 \\
\rho_3 &= .. = \rho_{10} = 2\left[\frac{0.12497375}{0.0001} + \frac{0.12497375}{0.00011} + 7\right] = 4,785.7\,. \qquad \text{(A10)}
\end{aligned}$$

Hence, we have an internally stable grand coalition since $\lambda_i < \rho_i$ for all $i \in N$.

Appendix 4

The total contribution $Q(S,T)$ is given in (7) in the text which reads if a player k leaves coalition S:

$$Q(S\backslash\{k\}, T \cup \{k\}) = \frac{b}{c}\left[\left(\sum_{i \in S}\frac{1}{\beta_i} - \frac{1}{\beta_k}\right)\left(\sum_{i \in S}\alpha_i - \alpha_k\right) + \sum_{j \in T}\frac{\alpha_j}{\beta_j} + \frac{\alpha_k}{\beta_k}\right] \qquad \text{(A11)}$$

The payoff for player k leaving coalition S and choosing the dominant strategy $q_j^* = \frac{b\alpha_j}{c\beta_j}$ is then

$$\begin{aligned}
\pi_k &= b\alpha_k Q(S\backslash\{k\}, T \cup \{k\}) - \frac{b^2\,(\alpha_k)^2}{2c\beta_k} \\
\pi_k &= \frac{b^2\alpha_k}{c}\left[\left(\sum_{i \in S}\frac{1}{\beta_i} - \frac{1}{\beta_k}\right)\left(\sum_{i \in S}\alpha_i - \alpha_k\right) + \sum_{j \in T}\frac{\alpha_j}{\beta_j} + \frac{\alpha_k}{\beta_k} - \frac{\alpha_k}{2\beta_k}\right] \\
\pi_k &= \frac{b^2\alpha_k}{c}\left[\left(\sum_{i \in S}\frac{1}{\beta_i} - \frac{1}{\beta_k}\right)\left(\sum_{i \in S}\alpha_i - \alpha_k\right) + \sum_{j \in T}\frac{\alpha_j}{\beta_j} + \frac{\alpha_k}{2\beta_k}\right]. \qquad \text{(A12)}
\end{aligned}$$

Adding this across all coalition members $k \in S$, we get the aggregate payoff from leaving the coalition $\sum_{k \in S}\pi_k(S\backslash\{k\}, T \cup \{k\})$.

$$\begin{aligned}
&\sum_{k \in S}\pi_k(S\backslash\{k\}, T \cup \{k\}) \\
&= \frac{b^2}{c}\left[\begin{array}{c} \sum_{i \in S}\frac{1}{\beta_i}\left[\left(\sum_{i \in S}\alpha_i\right)^2 - \sum_{i \in S}(\alpha_i)^2\right] + \sum_{i \in S}\alpha_i\left[\sum_{j \in T}\frac{\alpha_j}{\beta_j} - \sum_{i \in S}\frac{\alpha_i}{\beta_i}\right] \\ +\frac{3}{2}\sum_{i \in S}\frac{(\alpha_i)^2}{\beta_i} \end{array}\right] \qquad \text{(A13)}
\end{aligned}$$

The coalition surplus, $\sigma(S) = \sum_{k \in S}\pi_k(S,T) - \sum_{k \in S}\pi_k(S\backslash\{k\}, T \cup \{k\})$, using (A2) and

(A13), is given by:

$$\begin{aligned}
\sigma(S) &= \frac{b^2}{c}\sum_{i\in S}\alpha_i\left[\frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i+\sum_{j\in T}\frac{\alpha_j}{\beta_j}\right] \\
&\quad -\left\{\frac{b^2}{c}\left[\sum_{i\in S}\frac{1}{\beta_i}\left[\left(\sum_{i\in S}\alpha_i\right)^2-\sum_{i\in S}(\alpha_i)^2\right]+\sum_{i\in S}\alpha_i\left[\sum_{j\in T}\frac{\alpha_j}{\beta_j}-\sum_{i\in S}\frac{\alpha_i}{\beta_i}\right]+\frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\right]\right\} \\
\sigma(S) &= \frac{b^2}{c}\left\{\begin{array}{c}\sum_{i\in S}\frac{1}{\beta_i}\left[\frac{\left(\sum_{i\in S}\alpha_i\right)^2}{2}-\left(\sum_{i\in S}\alpha_i\right)^2+\sum_{i\in S}(\alpha_i)^2\right] \\ +\sum_{i\in S}\alpha_i\left[\sum_{j\in T}\frac{\alpha_j}{\beta_j}+\sum_{i\in S}\frac{\alpha_i}{\beta_i}-\sum_{j\in T}\frac{\alpha_j}{\beta_j}\right]-\frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\end{array}\right\} \\
\sigma(S) &= \frac{b^2}{c}\left\{\sum_{i\in S}\frac{1}{\beta_i}\left[\sum_{i\in S}(\alpha_i)^2-\frac{\left(\sum_{i\in S}\alpha_i\right)^2}{2}\right]+\sum_{i\in S}\alpha_i\sum_{i\in S}\frac{\alpha_i}{\beta_i}-\frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\right\}. \qquad \text{(A14)}
\end{aligned}$$

Appendix 5

For symmetric benefit shares, using (20), $\sigma(S)=\frac{b^2}{2cn^2}\sum_{i\in S}\frac{1}{\beta_i}\left[4m-m^2-3\right]$ which is only positive for $m\leq 3$ and because $\sigma(S)=0$ for $m=3$, $m=2$ is not externally stable. For symmetric costs, the coalition surplus (20) reduces to

$$\begin{aligned}
\sigma(S) &= \frac{b^2 n}{2c}\left[(2m-3)\sum_{i\in S}(\alpha_i)^2+(2-m)\left(\sum_{i\in S}\alpha_i\right)^2\right] \\
\sigma(S) &= \frac{b^2 n}{2c}\left[(2m-3)m\left[var(\alpha_S)+\bar{\alpha}_S^2\right]+(2-m)m^2\bar{\alpha}_S^2)\right] \\
\sigma(S) &= \frac{b^2 nm}{2c}\left[(2m-3)var(\alpha_S)+(4m-m^2-3)\bar{\alpha}_S^2\right]. \qquad \text{(A15)}
\end{aligned}$$

noticing that $Var(\alpha_S)\equiv\frac{\sum_{i\in S}(\alpha_i)^2}{m}-(\bar{\alpha}_S)^2$. The surplus is non-negative if

$$\begin{aligned}
\frac{var(\alpha_S)}{\bar{\alpha}_S^2} &\geq \frac{m^2-4m+3}{2m-3} \\
CV(\alpha_S) &= \sqrt{\frac{var(\alpha_S)}{\bar{\alpha}_S^2}}=\frac{\sqrt{var(\alpha_S)}}{\bar{\alpha}_S}\geq\sqrt{\frac{m^2-4m+3}{2m-3}}. \qquad \text{(A16)}
\end{aligned}$$

Alternatively, from the first line in (A15) we have

$$\begin{aligned}
(2m-3)\sum_{i\in S}(\alpha_i)^2 &> (m-2)\left(\sum_{i\in S}\alpha_i\right)^2 \\
\frac{\sum_{i\in S}(\alpha_i)^2}{\left(\sum_{i\in S}\alpha_i\right)^2} &> \frac{(m-2)}{(2m-3)}. \qquad \text{(A17)}
\end{aligned}$$