Journal of Economic Behavior and Organization 157 (2019) 541–559

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

The anti-paradox of cooperation: Diversity may pay![☆]

Michael Finus[a,b,*], Matthew McGinty[c]

[a] Department of Economics, University of Graz, Universitäts-straße 15, 8010 Graz, Austria
[b] Department of Economics, University of Bath, 3 East, Bath, BA2 7AY, UK
[c] Department of Economics, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI 53201, USA

ARTICLE INFO

Article history:
Received 23 August 2017
Revised 25 July 2018
Accepted 23 October 2018
Available online 16 November 2018

JEL classification:
C7
D7
H4

Keywords:
International environmental agreements
Coalition formation
Asymmetry
Transfers

ABSTRACT

This paper considers the stability and success of a public 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 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 may pay!

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. Others include the stabilization of financial markets, reducing money laundering, the fighting of contagious diseases and the efforts of non-proliferation of weapons of mass destruction. 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

[☆] The authors would like to thank two anonymous reviewers and the editor Professor Hausser for helpful comments on an earlier draft. We also received helpful comments by the participants at the 2nd "Environmental Protection and Sustainability Forum" in Bath in 2015, as well as by John Heywood, Antonio Galvao and Suyong Song. Extremely helpful comments are acknowledged by Jean-Francois Jacques regarding the proof of Proposition 2. The second revision of this paper was done when Michael was a Visiting Professor at the University Autonoma, Madrid. He would like to thank the members of the Department of Economic Theory for their hospitality. Previous versions also benefitted from research visits at the University Dauphine, Paris in 2014 and 2016.

[*] Corresponding author.
E-mail addresses: michael.finus@uni-graz.at, m.finus@bath.ac.uk (M. Finus), mmcginty@uwm.edu (M. McGinty).

https://doi.org/10.1016/j.jebo.2018.10.015
0167-2681/© 2018 Elsevier B.V. All rights reserved.

be explained by the strategic behavior of governments, also referred to as free- or easy-riding. 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 status quo.

With a few exceptions (e.g., Ray and Vohra, 2001; Sandler, 1999), the general literature on public goods focused on the voluntary provision in a Nash equilibrium, 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 Caparrós, 2015), predominantly focused on the formation of self-enforcing agreements. The great importance of the topic is stressed by papers like (Harstad, 2012; 2016).

Two main approaches have emerged. The cooperative approach used mainly the stability concept of the core (e.g., Ambec and Sprumont, 2002; Chander and Tulkens, 1995). It is a normative approach, focusing how the gains from cooperation are shared in the grand coalition, based on some axiomatic properties. However, this 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. The main conclusion from this literature under the standard assumptions is what Barrett (1994) called the "paradox of cooperation" .[1] That is, stable coalitions do not achieve a lot. Either stable coalitions are small or 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. The reason is that the public good provision game, like many others games, is a positive externality game (Yi, 1997). That is, starting from a coalition structure where all players act as singletons, 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.

Many of the early papers using the non-cooperative approach (Barrett, 1994; Carraro and Siniscalco, 1993) but also many later papers (e.g., Diamantoudi and Sartzetakis, 2006; Rubio and Ulph, 2006) assumed ex-ante symmetric players for simplicity. 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. In this paper we give up this restrictive assumption and consider asymmetric players.[2] Under the standard 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 (e.g.,Fuentes-Albero and Rubio, 2010), at least this was the common view for a long time of most scholars working in this area, almost known as a "coalition folk-theorem."

In this paper, we question the general validity of this folk-theorem. We show that if benefit and cost parameters are negatively correlated, and their distributions are very skewed, large coalitions and even the grand coalition can be stable. Moreover, under those conditions we establish a kind of anti-paradox result as we can show that the gains from cooperation can be very large. Finally, we consider transfers, showing that also for a positive correlation between benefit and cost parameters large coalitions associated with large gains from cooperation can be stable, again, a kind of anti-paradox result. The overall message is clear: asymmetry does not necessarily constitute an obstacle for successful cooperation but may be an asset. We characterize those conditions in this paper.

In the context of no transfers, we generalize the results of Pavlova and de Zeeuw (2013). Though they conclude that even in the absence of transfers, larger coalitions than under symmetry may be stable, they can neither show the stability of the grand coalition nor that meaningful gains from cooperation can be obtained. To the contrary, they generally confirm the paradox of cooperation and argue that larger coalitions may be inferior to smaller coalitions from a global welfare point of view. A crucial difference of our approach is that we allow for any type and degree of asymmetry whereas they consider only two types of players.

In the context of transfers, we extend work by for instance Finus and Pintassilgo (2013), Fuentes-Albero and Rubio (2010), McGinty (2007), Pavlova and de Zeeuw (2013) and Weikard (2009), analyzing how asymmetry affects coalition formation, employing the "optimal transfer scheme" . Though it emerges from those papers that with transfers, larger coalitions can be obtained for asymmetric than symmetric players, they do not fully characterize how the type and degree of asymmetry affects the size and the success of stable coalitions. This is also due to the fact that those results have been obtained either based on calibrated simulations or under the restrictive assumption of two or four types of players. Hence, due to these restrictions, some of our interesting results could not have been obtained.

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

[1] An interesting departure from the standard assumptions is considered in Karp and Simon (2013) who show that for non-convex marginal abatement cost functions the paradox of cooperation may break down.

[2] A comprehensive overview of other departures from the standard assumptions is presented in Finus and Caparrós (2015).

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, 1997) in order to stress the institutional setting 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 $s = |S|$. 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 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 derive their strategies from maximizing the sum of all coalition members' payoffs. That is, the coalition acts as a meta player (Haeringer, 2004), fully internalizing the externality among its members. In contrast, non-members simply maximize their own payoff.

The game is solved by backwards induction. For any given coalition S, the second stage delivers a vector of equilibrium strategies $q^*(S) = (q_1^*(S), q_2^*(S), \ldots, q_n^*(S))$. Assuming a unique equilibrium, equilibrium payoffs follow from inserting strategies into the payoff functions of players, $\pi_i(q^*(S)) = \pi_i^*(S)$, and a vector of payoffs is derived: $\pi^*(S) = (\pi_1^*(S), \pi_2^*(S), \ldots, \pi_n^*(S))$. 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:

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

It is easy to see that internal & external stability is essentially a Nash equilibrium in membership strategies. No signatory has an incentive to leave coalition S, i.e., player i has no incentive to change announcement 1 to 0, given the announcement of all other players. By the same token, no non-signatory has an incentive to join coalition S, i.e., player j has no incentive to announce 1 instead of 0.[3] Due to the definition of external stability, membership is open to all players, nobody can be precluded from joining coalition S.[4] 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 $s > 1$. 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. Finally note that we rule out knife-edge cases 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.[5]

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.

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

That is, the surplus, $\sigma(S)$, defined as the difference between the total coalition payoff and the sum of free-rider payoffs must be (weakly) positive. It is clear that potential internal stability is 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 $\sigma(S)$, $\sum_{i \in S} \lambda_i = 1$.[6] Henceforth, if we talk about transfers, we mean the optimal transfer scheme. We note that with transfers if coalition S is internally stable, every coalition $S \setminus \{i\}$ is externally unstable.

Given the assumption about the second stage, clearly, if S is either empty or comprises only one player, $q^*(S)$ is equivalent to a Nash equilibrium known from games without coalition formation. By the same token, if the grand coalition forms, $q^*(S)$ corresponds to the social optimum.

It is obvious that the grand coalition must lead to an aggregate payoff which is at least as high as 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, as in output or price cartels and public good games. Broadly

[3] 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 to this paper.

[4] Exclusive membership games are described for instance in Finus and Rundshagen (2006) and Yi (1997).

[5] 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.

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

speaking, even though players may benefit from forming a coalition, it may be even more attractive to remain a singleton, enjoying the benefits from the actions of the coalition without sharing the costs.

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 from the aggregate contribution and the cost from the individual contribution $C_i(q_i)$.

$$\pi_i = \alpha_i bQ - \frac{\beta_i c}{2}(q_i)^2 \tag{1}$$

Payoff function (1) is probably the simplest representative of a strictly concave payoff function and has therefore been frequently considered in the literature (Barrett, 1994; Finus and Pintassilgo, 2013; Ray and Vohra, 2001). Though it would be possible to derive general properties regarding the second stage based on a general payoff function, 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 is evident for instance from Ray and Vohra (2001) and Yi (1997).[7]

In order to test whether the level or the distribution of benefits and costs matter for stability, we split the benefit and cost parameter into a global benefit parameter $b > 0$ and global cost parameter $c > 0$ and individual parameters, $\alpha_i > 0$ on the benefit and $\beta_i > 0$ on the cost side. This allows us to compare different mean preserving distributions of the benefit and cost parameter and to show that the level of benefits and costs do not matter for stability. Without loss of generality, we normalize the individual parameters such that $\sum_{i \in N} \alpha_i = 1$ and $\sum_{i \in N} \beta_i = 1$.[8]

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 condition of a non-signatory $j \in T$ reads

$$\frac{\partial \pi_{j \in T}}{\partial q_j} = \alpha_j b - \beta_j c q_j = 0 \quad \Leftrightarrow \quad \alpha_j b = \beta_j c q_j \quad \Leftrightarrow \quad q^*_{j \in T} = \frac{\alpha_j b}{\beta_j c} \tag{2}$$

implying that marginal benefits are set equal to marginal cost. Correspondingly, for a signatory $i \in S$, we have

$$\frac{\partial \sum_{k \in S} \pi_k}{\partial q_i} = \sum_{k \in S} \alpha_k b - \beta_i c q_i = 0 \quad \Leftrightarrow \quad \sum_{k \in S} \alpha_k b = \beta_i c q_i \quad \Leftrightarrow \quad q^*_{i \in S}(S) = \frac{\sum_{k \in S} \alpha_k b}{\beta_i c} \tag{3}$$

implying that the sum of marginal benefits of coalition S, is set equal to individual marginal cost, a kind of Samuelson optimality condition for the coalition. Among signatories, the externality is fully internalized and marginal contribution costs are equalized, $\beta_i c q^*_{i \in S}(S) = \beta_k c q^*_{k \in S}(S)$ for all $i, k \in S, i \neq k$, implying that the total contribution level of signatories is provided cost-effectively. The ratio of contributions is inverse to the individual cost parameters, i.e., $\frac{q^*_{i \in S}(S)}{q^*_{k \in S}(S)} = \frac{\beta_k}{\beta_i}$, implying that those with a flatter marginal cost curve should contribute more to the public good than those with a steep slope. Further details are provided in Appendix A.

3. General properties

Before analyzing stability of coalitions, we look at some general properties of the coalition formation game with payoff function (1). The first properties are discussed informally as proofs are provided in Finus and Pintassilgo (2013). From a normative point of view, and as pointed out above, in the public good provision game strict cohesiveness holds. However, even strict full cohesiveness holds, i.e., by starting from the all singleton coalition structure and gradually increasing the coalition by adding a player in every step until the grand coalition is reached, the aggregate payoff over all players strictly increases in each step, irrespective of the path of expansion. The same holds for the aggregate public good provision level. Hence from a social planner's point of view, larger coalitions are preferred to smaller coalitions. Coalition formation is also generally attractive to those involved in the enlargement of the coalition because their aggregate payoff increases (superadditivity). Of course, whether individual players benefit depends on how this gain is shared. Finally, all players not involved in the enlargement of the coalition also benefit (positive externality). As is evident from (3), the provision level of coalition members increase with the size of the coalition from which also non-signatories benefit as benefits are non-exclusive.

[7] Payoff function (1) allows us to derive all our results analytically. A quadratic benefit function would give similar qualitative results, though we have to rely on simulations. See Appendix E which confirms our main conclusions derived in Sections 4 and 5.

[8] That is, starting from a payoff function $\pi_i = b_i Q - \frac{c_i}{2}(q_i)^2$, we split the benefit parameter b_i into $\alpha_i b$ and the cost parameter c_i into $\beta_i c$. Note that any value and distribution of parameters b_i and c_i can be replicated by our formulation by choosing the appropriate values for the global and individual parameters. In terms of interpretation, we note that for instance low values of β_i do not necessarily imply very large contribution levels if c is large/or b is small. Our transformation is only for convenience, facilitates more elegant proofs, but does not affect results. It is important to point out that we do not pursue a traditional analysis by considering how changes of a single parameter (e.g., individual or global parameter) affects outcomes (e.g., equilibrium provision levels and stability). Instead, we compare vectors of benefit and cost parameters. The comparisons assume either the same variance but different arithmetic means or vice versa, same variance but different arithmetic means. It is also important to note that different individual benefit parameters do not affect the public good nature of the problem as benefits depend on the total contribution Q. Different individual benefit parameters only mean that players benefit to a different extent from the total contribution Q.

Hence, even if players are symmetric or even if the gains among coalition members are shared in an optimal way, if the positive externality effect is stronger than the superadditivity effect, large coalitions may not be stable. Obviously, without transfers, things do not become any easier. To the contrary, it is easy to show that the stable coalition with the highest aggregate payoff and contribution level with transfers is weakly larger and superior (in terms of the aggregate payoff and provision level) than without transfers. In other words, transfers weakly improve on outcomes. It is also easy to conclude that the all singleton coalition structure is Pareto-dominated by every stable non-trivial coalition. In a positive externality game, all non-signatories are obviously better off and all signatories must better off, otherwise internal stability would not hold.

In order to measure the paradox of cooperation (or to demonstrate the opposite), stable coalitions need to be benchmarked. For our purpose it is sufficient to measure the severity of the externality as the difference between full cooperation (social optimum) and no cooperation (Nash equilibrium). The larger this difference, the larger the need for cooperation will be. In Proposition 1 we measure the externality in terms of 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 non-cooperative levels. In terms of asymmetry, related to our individual parameters α_i and β_i, almost all distributions can be compared.

Proposition 1. *Measure the severity of the externality as the difference between total contribution (total welfare) under full cooperation (FC) and no cooperation (NC), then the severity of the externality is given by*

$$\Delta Q := Q_{FC} - Q_{NC} = \frac{b}{c}\left[\sum_{i\in N}\frac{1-\alpha_i}{\beta_i}\right] > 0$$

$$\Delta\Pi := \Pi_{FC} - \Pi_{NC} = \frac{b^2}{2c}\left[\sum_{i\in N}\frac{(1-\alpha_i)^2}{\beta_i}\right] > 0 \tag{4}$$

which increases in b and decreases in c. Moreover, consider a distribution of the individual benefit parameter Ψ^α with $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n$ and a distribution of the individual cost parameter Ψ^β with $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_n$ 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

$$\Delta Q(\widetilde{\Psi}^\alpha, \Psi^\beta) - \Delta Q(\Psi^\alpha, \Psi^\beta) \geq 0 \ \ (\text{increasing } \alpha\text{-variance}) \tag{5}$$

$$\Delta Q(\Psi^\alpha, \widetilde{\Psi}^\beta) - \Delta Q(\Psi^\alpha, \Psi^\beta) > 0 \ \ (\text{increasing } \beta\text{-variance}) \tag{6}$$

$$\Delta Q(\widetilde{\Psi}^\alpha, \widetilde{\Psi}^\beta) - \Delta Q(\Psi^\alpha, \Psi^\beta) > 0 \ \ (\text{increasing positive } \alpha\text{-}\beta\text{-covariance}) \tag{7}$$

$$\Delta\Pi(\widetilde{\Psi}^\alpha, \Psi^\beta) - \Delta\Pi(\Psi^\alpha, \Psi^\beta) > 0 \ \ (\text{increasing } \alpha\text{-variance}) \tag{8}$$

$$\Delta\Pi(\Psi^\alpha, \widetilde{\Psi}^\beta) - \Delta\Pi(\Psi^\alpha, \Psi^\beta) > 0 \ \ (\text{increasing } \beta\text{-variance}) \tag{9}$$

$$\Delta\Pi(\widetilde{\Psi}^\alpha, \widetilde{\Psi}^\beta) - \Delta\Pi(\Psi^\alpha, \Psi^\beta) > 0 \ \ (\text{increasing positive } \alpha\text{-}\beta\text{-covariance}). \tag{10}$$

Proof. ΔQ and $\Delta\Pi$ are computed using Appendix A, noticing that under full cooperation $S = N$ and under no cooperation $T = N$. Results regarding distributions follow from using ΔQ and $\Delta\Pi$ in (4) above, considering two marginal changes at the same time, which delivers the result after some basic calculations.[9] □

For the case "increasing the negative $\alpha - \beta$ -covariance" such a general conclusion cannot be derived; the gap between full and no cooperation can increase or decrease with marginal changes. Nevertheless, as will be apparent from Table 1 below and Section 4, also for negatively correlated distributions the gap can be quite large.

By the nature of Proposition 1, 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 2–4, the cost share asymmetry is increased, i.e., the β-variance increases. Going from scenario

[9] See footnote 8. The concept of marginal changes is a convenient way to compare two distributions, though in reality distributions are given. Moreover, it is important to note that, mathematically, we only operationalize the comparison of different vectors of parameters. The possible economic interpretation that this is at odds with the public good nature of our problem as increasing the individual benefit parameter of one player at the expenses of another player for instance would be misleading.

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	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}-0.3$	$\frac{1}{3}$	$\frac{1}{3}+0.3$	23.05	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

4 to 5 increases additionally the benefit share asymmetry, i.e. the α-variance increases, with a positive $\alpha - \beta$-covariance. As Proposition 1 predicts along this sequence ΔQ and $\Delta\Pi$ increase and, as the example shows, the magnitudes become very large. Proposition 1 also predicts that going from scenario 1 to 8, $\Delta\Pi$ increases. Other comparisons are not covered by Proposition 1.

Scenarios 6 and 7 generate a negative $\alpha - \beta$-covariance 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 $\Delta\Pi$. However, an increase of the asymmetry on the benefit side which is associated with an increase of the asymmetry on the cost side (going from scenario 1 to 4 and then 5), can make a big difference.

4. Stable coalitions without transfers

Given the fact our public good coalition game is fully cohesive, and the free-rider problem is mainly about leaving and not joining a coalition, it seems natural that one is more concerned about internal than external stability.[10] 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 (1), and conducting some basic though cumbersome manipulations (see Appendix B), leads to a compact and closed form solution which allows for the following statement.

Proposition 2. *In the absence of transfers, for any number of signatories, a necessary and sufficient condition for internal stability is:*

$$f(\theta_i) \equiv \frac{2\theta_i^2}{3\theta_i^2 - 2\theta_i + 1} \geq \psi_i \text{ for all } i \in S \tag{11}$$

where $\psi_i \equiv \frac{\frac{1}{\beta_i}}{\sum_{j\in S}\frac{1}{\beta_j}} \in (0, 1)$ *and* $\theta_i \equiv \frac{\alpha_i}{\sum_{j\in S}\alpha_j} \in (0, 1)$ *for* $|S| \geq 2$. *Hence, a necessary condition for any coalition* $|S| \geq 2$ *to be internally stable is that for one coalition member* $\theta_i \geq \frac{1}{3}$ *and* $\psi_i \geq \frac{1}{3}$.

Proof. See Appendix B. □

The internal stability condition (11) is remarkably simple compared to the complicated simulations found in the literature. The individual benefit parameters are on the left-hand side and the individual cost parameters are on the right-hand side, with θ_i the individual benefit parameter ratio and ψ_i the inverse individual cost parameter ratio. Only the individual parameters of the s coalition members matter for internal stability, but not those of the $n - s$ outsiders. Moreover, the level of the global benefit and cost parameter b or c do not matter.

We illustrate the internal stability condition (11) in Fig. 1. Note that $f(\theta_i) = \theta_i$ for $\theta_i \in \{0,\ 1/3,\ 1\}$ and $f(\theta_i) < \theta_i$ for all $\theta_i \in]0,\ 1/3[$ and $f(\theta_i) > \theta_i$ for all $\theta_i \in]1/3,\ 1[$. By definition $\sum_{j\in S}\theta_i = 1$ and $\sum_{j\in S}\psi_i = 1$. Hence, if for all $i \in S$, $\theta_i < 1/3$, $\theta_i > f(\theta_i)$ and hence $1 = \sum_{j\in S}\theta_i > \sum_{j\in S} f(\theta_i)$ holds. But a necessary condition for internal stability, $f(\theta_i) \geq \psi_i$ for all $i \in S$, is $\sum_{j\in S} f(\theta_i) \geq \sum_{j\in S}\psi_i = 1$ which cannot hold if $1 > \Sigma_{j\in S} f(\theta_i)$. Consequently, we need at least that $\theta_i \geq \frac{1}{3}$ holds for one coalition member. And for this member, the corresponding ψ_i must be larger than $\frac{1}{3}$. Obviously, there cannot be more than two members with $\theta_i \geq \frac{1}{3}$, except for symmetry, in which case there can be three members with $\theta_i = \frac{1}{3}$. With these considerations, it is straightforward to prove and understand the intuition of the following results.

[10] For completeness, Appendix D derives and discusses also the conditions for external stability for the case of no transfers, as discussed in this section, but also for the case of transfers, as discussed in the next section.

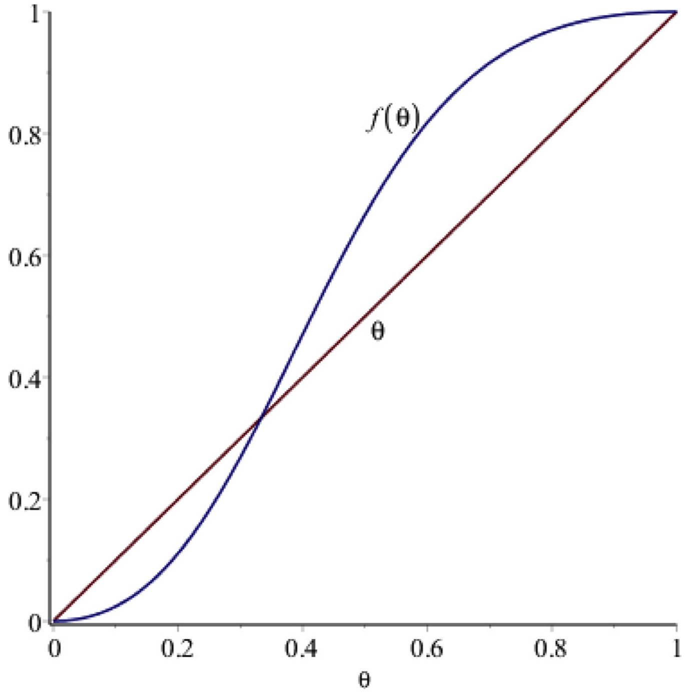

Fig. 1. θ and f(θ).

Corollary 1. *Assume no transfers. Denote the size of a stable coalition by s^*. (a) All coalition members have the same individual benefit and cost parameter: $s^* = 3$. (b) All coalition members have the same individual benefit parameter but at last two members have different individual cost parameters: $s^* < 3$. (c) All coalition members have the same individual cost parameter but at last two members have different individual benefit parameters: $s^* < 3$.*

Proof. See Appendix B. □

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. This result is 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 (1), internal stability holds at the margin for a coalition of three players. Any asymmetry will cause that some members to get slightly more but others slightly less of the “cooperative cake” upsetting internal stability.

The next corollary looks at simultaneous asymmetry on the benefit and cost side. Part (a) does not contradict the “coalition folk-theorem” ; part (b) puts it upside down.

Corollary 2. *In the absence of transfers, (a) if there is a positive covariance between the individual benefit and cost parameters of coalition members: $s^* < 3$ and (b) if there is a negative covariance: $s^* \geq 3$ is possible, including the grand coalition.*

Proof. See Appendix B. □

With a positive covariance between individual benefit and cost parameters, stable coalitions are strictly smaller than 3, the benchmark size in the case of symmetric players. A member with a below average cost parameter will contribute more than the average to the internalization of the externality among coalition members. With a positive covariance, this disadvantage is reinforced because the same member has also a below average benefit parameter. In contrast, with a negative covariance, this disadvantage can be balanced.

In Fig. 1, a positive $\alpha - \beta$-covariance shows up in a negative $\theta - \psi$-variance which is obviously not helpful for satisfying $f(\theta_i) \geq \psi_i$ for all $i \in S$. We just need the reverse. More specifically, for those one or two players with a large benefit parameter ratio $\theta_i \geq 1/3$, the inverse cost parameter ratio must be large, i.e., $\psi_i > \theta_i$ which means very low cost parameters β_i compared to the average of the coalition. Correspondingly, for all remaining players with a low benefit parameter ratio, i.e., $\theta_i < 1/3$, the inverse cost parameter ratio must be very small, i.e., $\psi_i < < \theta_i$, which means very high cost shares β_i compared to the average of the coalition. In other words, the negative $\alpha - \beta$-covariance needs to be very large, and we need a very skewed distribution on the benefit and cost side, with the individual benefit parameter distribution being extremely positively skewed and the individual cost parameter distribution being very negatively skewed.

For instance, consider the following three examples with $n = 10$ for which the grand coalition is stable. (This can be easily confirmed by plugging the numbers into (11) above.) Example 1: Let $\alpha_1 = 0.43$ and $\alpha_2 = 0.41$ and let the other 8 players have the same benefit parameter $\alpha_3 = \ldots \alpha_{10} = 0.02$, so that $\sum_{i \in S} \alpha_i = 1$. Further let $\beta_1 = 0.0001$ and $\beta_2 = 0.00011$, and the other 8 players have $\beta_3 = \ldots = \beta_{10} = 0.12497375$, so that $\sum_{i \in S} \beta_i = 1$. Example 1 can be modified such that $\alpha_1 = \alpha_2 = 0.42$ and $\beta_1 = \beta_2 = 0.00015$ and $\beta_3 = \ldots = \beta_{10} = 0.12496250$. Example 2: Let $\alpha_1 = \alpha_2 = 0.49$ and let the other 8 players have the same benefit parameter $\alpha_3 = \ldots \alpha_{10} = 0.0025$, so that $\sum_{i \in S} \alpha_i = 1$. Further let $\beta_1 = \beta_2 = 0.0000031408$, and the other 8 players have $\beta_3 = \ldots = \beta_{10} = 0.1249992148$, so that $\sum_{i \in S} \beta_i = 1$. Example 3: Let $\alpha_1 = 0.64$ and $\alpha_2 = 0.32$ and let the other 8 players have the same benefit parameter $\alpha_3 = \ldots \alpha_{10} = 0.005$, so that $\sum_{i \in S} \alpha_i = 1$. Further let $\beta_1 = 0.000007$ and $\beta_2 = 0.00004$, and the other 8 players have $\beta_3 = \ldots = \beta_{10} = 0.124994125$, so that $\sum_{i \in S} \beta_i = 1$. Example 1 and 2 assume two players with a relative benefit parameter above 1/3, whereas example 3 assumes only one player for whom this is true.

Viewing Corollary 1 and 2 together, we can learn a couple of lessons.

Firstly, 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 asymmetry.

Secondly, Corollary 1 and Part a of Corollary 2 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. Recalling our externality measures $\Delta Q := Q_{FC} - Q_{NC}$ and $\Delta\Pi := \Pi_{FC} - \Pi_{NC}$, it is easy to show that $\partial \Delta Q / \partial n > 0$ and $\partial \Delta\Pi / \partial n > 0$ for symmetry. Moreover, ΔQ and $\Delta\Pi$ increase in the level of the global benefit-cost ratio b/c. For asymmetry, Proposition 1 showed additionally that ΔQ and $\Delta\Pi$ increase in the degree of asymmetry if the asymmetry is increased only on the individual benefit parameter side (increasing the α-variance), only on the individual cost parameter side (increasing the β-variance) or on both sides if there is a positive $\alpha - \beta$-covariance. 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 even short of 3 players as observed under symmetry.

Thirdly, in contrast, Part b of Corollary 2 suggests that with the right type of asymmetry (negative $\alpha - \beta$-covariance), at least in terms of the coalition size, we can have even full cooperation. Certainly, this contradicts the coalition folk-theorem. But can this be viewed as anti-paradox of cooperation? This depends on whether ΔQ and $\Delta\Pi$ are large when for instance the grand coalition is stable. We know already from Table 1 that with a negative $\alpha - \beta$-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 $\alpha - \beta$-covariance in Proposition 1, we compute ΔQ and $\Delta\Pi$ for the three examples mentioned above. Example 1: $\Delta Q = 11,126\frac{b}{c}$ and $\Delta\Pi = 6,475\frac{b^2}{2c}$; Example 1 modified: $\Delta Q = 11,110\frac{b}{c}$ and $\Delta\Pi = 6,469\frac{b^2}{2c}$; Example 2: $\Delta Q = 324,822\frac{b}{c}$ and $\Delta\Pi = 165,690\frac{b^2}{2c}$; Example 3: $\Delta Q = 68,492\frac{b}{c}$ and $\Delta\Pi = 30,138\frac{b^2}{2c}$. In order to appreciate that these are very large numbers, it is useful to compare them with the case of symmetric benefit and cost parameters for which we find: $\Delta Q = 90\frac{b}{c}$ and $\Delta\Pi = 81\frac{b^2}{2c}$, even though this is not achieved because not the grand coalition but only a coalition of three players is stable. Therefore, we can 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 Zeeuw (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 argue that smaller coalitions may be preferable to larger coalitions. 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 individual benefit and cost parameters. Only this helps to equalize the gains from cooperation among players compared to their free-rider payoffs in the absence of transfers. As the examples proved, this is exactly the situation when the need for cooperation is large. We argue that this is an interesting version of an anti-paradox of cooperation.

In the context of climate change, it is typically argued that developed countries put more emphasis on the benefits of greenhouse gas reduction than developing countries but have steeper marginal abatement cost curves. That is, as a tendency, at the world scale, in climate change, we have a positive and not a negative covariance. The results of two empirically calibrated climate models, CLIMNEG (Carraro et al., 2006) and STACO (Nagashima et al., 2009) confirm this though payoff functions are different from ours. STACO has the same benefit function but a cubic cost function with twelve world regions. Without transfers, it emerges from a large set of sensitivity analyses that at best a coalition of only two regions can form a stable coalition, which does not depart much from non-cooperative behavior. CLIMNEG is a fully fledged general equilibrium model with 6 world regions. Without transfers, no non-trivial coalition is stable.

In contrast, for other problems, it is not unlikely that there is a negative covariance, though, unfortunately, we are not aware of empirical analyses which have tested stability of coalitions. In the context of the Montreal Protocol and follow-up protocols on the reduction of CFCs (causing the depletion of the ozone layer) it may be argued that success of this agreement is to a large extent due the fact of a negative covariance of benefits and costs, with extremely skewed distributions. Industrialized countries and in particular the US government pointed to the benefits of reduced damages (i.e., high individual benefit parameters) and at the same time had the technology to replace CFCs by substitutes (i.e., low individual cost parameters), which was not available to developing countries (i.e., high individual cost parameters), which also showed little interest in recognizing the environmental problem (i.e., low individual benefit parameters).

5. Stable coalitions with transfers

Whereas previous work has shown that with asymmetry an optimal transfer scheme can increase the size of stable coalitions compared to symmetry (e.g., Fuentes-Albero and Rubio, 2010; McGinty, 2007; Weikard, 2009), we are able to fully characterize which degree and type of asymmetry is necessary to generate this result. 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 coalitions S. Using payoff function (1), and conducting some basic manipulation (see Appendix C), a closed form solution for $\sigma(S)$ can be obtained which allows for the following statement.

Proposition 3. *Under an optimal transfer scheme coalition S is internally stable if and only if*

$$\left[\sum_{i\in S} (\theta_i)^2 - \frac{1}{2}\right] + \left[\frac{1}{2}\sum_{i\in S} \psi_i\theta_i(2-3\theta_i)\right] \geq 0 \tag{12}$$

holds for S with $\psi_i \equiv \frac{\frac{1}{\beta_i}}{\sum_{j\in S}\frac{1}{\beta_j}} \in (0,1)$ *and* $\theta_i \equiv \frac{\alpha_i}{\sum_{j\in S}\alpha_j} \in (0,1)$ *for* $|S| \geq 2$.

Proof. See Appendix C. □

From (12) it is evident that neither the level of the global benefit and cost parameter nor their ratio does matter for internal stability, only individual benefit and cost parameters matter, a result which we also found for no transfers. Nevertheless, it may not be straightforward to interpret (12). Hence, we proceed in three steps. Firstly, we establish the minimum size of stable coalitions with reference to symmetry for which $s^* = 3$. Secondly, we look at single asymmetry and thirdly we look at simultaneous asymmetry.

Corollary 3. *Under an optimal transfer scheme a stable coalition comprises at least three members, i.e.,* $s^* \geq 3$.

Proof. See Appendix C. □

Corollary 3 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. We now turn to single asymmetry.

Corollary 4. *If all coalition members have the same individual benefit parameter, the stable coalition comprises three players,* $s^* = 3$. *If all coalition members have the same individual cost parameter,* $s^* \geq 3$; *a sufficient condition for an internally stable coalition is given by*

$$HF(\theta_S) \equiv \sum_{i\in S} (\theta_i)^2 \geq \frac{(s-2)}{(2s-3)} \tag{13}$$

where $HF(\theta_S)$ *is the (modified) Herfindahl index of the individual benefit parameter ratio* θ_i *in coalition S where* $\frac{(s-2)}{(2s-3)}$ *increases in s with* $\lim_{s\to\infty} \frac{(s-2)}{(2s-3)} = \frac{1}{2}$.

Proof. See Appendix C □

Firstly, Corollary 4, when compared with Corollary 1, demonstrates that for single asymmetry transfers strictly improve upon no transfers. Stable coalitions will be strictly larger and hence also the total provision level and global welfare due to strict full cohesiveness. Secondly, Corollary 4 suggests that asymmetry of benefit parameters is crucial for internal stability of large coalitions. We measure asymmetry with the “modified” Herfindahl index (because we measure concentration among subgroup S and not the total population N) which is frequently used to measure the concentration in markets. In our context, the “modified” Herfindahl index of relative individual benefit parameters of coalition members needs to be sufficiently high for stability. For instance, for $s = 10$ to be stable, $HF(\theta_S) \geq 0.47$ and for $s = 20$, $HF(\theta_S) \geq 0.49$ is required. For large s, the benchmark value is 0.5 for which a sufficient condition is that one player has a benefit parameter ratio larger than $\left(\frac{1}{2}\right)^{\frac{1}{2}} \approx 0.707$. In other words, all we need is a very positively skewed distribution of individual benefit parameter if costs are symmetric in which case even the grand coalition can be stable. We may view this as another version of the anti-paradox of cooperation, even though Table 1 has shown that with cost symmetry the gains from cooperation may not be that large. This was different for cost asymmetry. For a negative covariance between benefits and costs, we have already shown for no transfers that large coalitions can be stable, including the grand coalition, and that the gains from cooperation can be very large. Since transfers weakly improve upon no transfers, the most interesting case is when there is a positive covariance between benefits and costs. In this case, stable coalitions are strictly smaller than 3 players without transfers (Corollary 2, result a) but the gains from cooperation would be very large according to Proposition 1 and as illustrated in Table 1. In order to develop this point briefly, let us start with some basic considerations related to internal stability condition (12) in Proposition 3, which allow us to state the following.

Corollary 5. *A sufficient condition for coalition S to be internally stable is $HF(\theta_S) \geq \frac{1}{2}$ but no individual benefit parameter ratio θ_i is larger than $\frac{2}{3}$.*

If the condition in Corollary 5 holds, both terms in (12) are positive. This reiterates what we found above, namely that the distribution of the individual benefit parameters are crucial for stability and that a very positively skewed distribution is conducive to cooperation. So if we have two players with a relatively large benefit parameter in the grand coalition, say for instance $\alpha_i = \theta_i = 0.65$ and $\alpha_j = \theta_j = 0.3$, then this coalition will be stable. Of course, many coalitions (including the grand coalition) may be stable even if the sufficient conditions do not hold. Suppose $\sum_{i \in S} (\theta_i)^2 - \frac{1}{2} \geq 0$ as before but that there is now one signatory with a really large individual benefit parameter ratio $\theta_i > \frac{2}{3}$ such that one element in the second term in (12) is negative, then for a sufficiently low value of ψ_i, (12) will still hold. That is, the high individual benefit parameters α_i should be positively correlated with a high individual cost parameter β_i and, again, the grand coalition could be stable. Thus, because not only large coalitions can be stable but also the welfare gains are larger for a positive covariance between benefits and costs (see Proposition 1), this may be viewed as another version of the anti-paradox of cooperation.[11]

In the context of the two empirically calibrated climate models, CLIMNEG and STACO, which we have discussed briefly above, it turns out that in STACO with optimal transfers coalitions of six out of twelve world regions can form a stable agreement which close the gap between no and full cooperation, which is very large, to a quite substantial amount. In CLIMNEG the same is true, with coalitions of four out of six world regions forming a stable coalition. Even though a direct comparison is not possible due to different payoff functions, this suggest that transfers could make quite some difference to the success of cooperation in addressing climate change, though the grand coalition may not be obtained as the distribution of the individual benefit parameters is not skewed enough.

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-riders incentive may be so strong so that large coalitions may not be stable, letting alone the grand coalition. From the previous literature, 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 stable coalitions are small or if they are large, the difference between full and no cooperation is small, both in terms of total contributions and welfare. This was called the paradox of cooperation by Barrett (1994). Secondly, the larger the asymmetry among players, the smaller stable coalitions will be in the absence of transfers. This conclusion was known as a kind of "coalition 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 parameters 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 parameters. 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 parameter 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 individual benefit parameter. If, additionally, there is a positive covariance of benefit and cost parameters, then the gains from cooperation can be massive.

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. 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. This is briefly demonstrated in Appendix E for a payoff function with quadratic benefits and quadratic costs, which has been frequently considered in the literature on international environmental agreements. Secondly, we considered the simple open membership coalition game. On the one hand, in terms of stability, this is the most pessimistic assumption. 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 re-optimize 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.[12] Thirdly, one could depart from a public good setting and look at other economic problems with a similar incentive structure, as for instance coalition formation in a price

[11] Clearly, also for less skewed distributions large coalitions can be stable, though conclusions are more specific and hence not discussed here.

[12] 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).

and output oligopoly or trade agreements which exhibit positive externalities from coalition formation as considered in 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.

Appendix A

Provision levels by non-signatories and signatories are given by (2) and (3) in the text, respectively, with the understanding that the total contribution of signatories is $Q_S^*(S) = \sum_{i\in S} q_i^*(S)$, that of non-signatories is $Q_T^*(S) = \sum_{j\in T} q_j^*$, and that of all players is $Q^*(S) = Q_S^*(S) + Q_T^*(S)$, which leads to $Q_S^*(S) = \frac{b}{c}\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i$, $Q_T^*(S) = \frac{b}{c}\sum_{j\in T}\frac{\alpha_j}{\beta_j}$ and hence $Q^*(S) = \frac{b}{c}\left[\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right]$. Inserting equilibrium provision levels into payoff function (1), delivers the payoff of a signatory $\pi_{i\in S}^*$:

$$\pi_{i\in S}^*(S) = \alpha_i b Q^*(S) - \frac{c\beta_i\left[q_{i\in S}^*(S)\right]^2}{2}$$

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

$$\pi_{i\in S}^*(S) = \frac{b^2}{c}\left[\alpha_i\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right) - \frac{1}{2\beta_i}\left(\sum_{i\in S}\alpha_i\right)^2\right]. \tag{A1}$$

The worth of the coalition, the sum of payoffs across all members, $\Pi_S^*(S) = \sum_{i\in S}\pi_{i\in S}^*(S)$, is given by:

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

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

$$\Pi_S^*(S) = \frac{b^2}{c}\sum_{i\in S}\alpha_i\left[\frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right]. \tag{A2}$$

The payoff to each player outside the coalition, $\pi_{j\in T}^*(S)$, is given by

$$\pi_{j\in T}^*(S) = \alpha_j b Q^*(S) - \frac{c\beta_j\left(q_{j\in T}^*\right)^2}{2}$$

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

$$\pi_{j\in T}^*(S) = \frac{b^2}{c}\left[\alpha_j\left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right) - \frac{\left(\alpha_j\right)^2}{2\beta_j}\right]. \tag{A3}$$

The aggregate payoff of those outside the coalition, $\Pi_T^*(S) = \sum_{j\in T}\pi_{j\in T}^*(S)$, is

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

and hence the global payoff, for any given coalition S, is $\Pi^*(S) = \Pi_S^*(S) + \Pi_T^*(S)$.

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

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

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

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

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

Appendix B

Consider a coalition S and let the set of non-signatories be denoted by T, where $S \cup T = N$ and $S \cap T = \emptyset$. The payoff of a signatory, using (A1) from Appendix A, is given by

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

If member i leaves coalition S, then the payoff using (A3) is given by

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

Let $\sigma_i(S) = \pi^*_{i\in s}(S) - \pi^*_{i\in T}(S\setminus\{i\})$. Using (A6) and (A7), we have

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

Combining terms $\alpha_i \sum_{l\in T}\frac{\alpha_l}{\beta^l} - \alpha_i \sum_{l\in T}\frac{\alpha_l}{\beta^l}$ results in

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i \sum_{j\in S}\frac{1}{\beta_j}\sum_{j\in S}\alpha_j - \frac{\left(\sum_{j\in S}\alpha_j\right)^2}{2\beta_i} - \alpha_i\sum_{j\neq i\in S}\frac{1}{\beta_j}\sum_{j\neq i\in S}\alpha_j - \frac{\alpha_i^2}{2\beta_i}\right]. \tag{A9}$$

Pulling apart the first term into two terms implies

$\alpha_i \sum_{j\in S}\frac{1}{\beta_j}\sum_{j\in S}\alpha_j = \alpha_i^2\sum_{j\in S}\frac{1}{\beta_j} + \alpha_i\sum_{j\in S}\frac{1}{\beta_j}\sum_{j\neq i\in S}\alpha_j$ and hence

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\in S}\frac{1}{\beta_j} + \alpha_i\sum_{j\in S}\frac{1}{\beta_j}\sum_{j\neq i\in S}\alpha_j - \frac{\left(\sum_{j\in S}\alpha_j\right)^2}{2\beta_i} - \alpha_i\sum_{j\neq i\in S}\frac{1}{\beta_j}\sum_{j\neq i\in S}\alpha_j - \frac{\alpha_i^2}{2\beta_i}\right]. \tag{A10}$$

Combining the second and fourth term implies

$\alpha_i \sum_{j\in S}\frac{1}{\beta_j}\sum_{j\neq i\in S}\alpha_j - \alpha_i\sum_{j\neq i\in S}\frac{1}{\beta_j}\sum_{j\neq i\in S}\alpha_j = \alpha_i\left(\frac{1}{\beta_i}\right)\sum_{j\neq i\in S}\alpha_j$ and therefore

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\in S}\frac{1}{\beta_j} + \alpha_i\left(\frac{1}{\beta_i}\right)\sum_{j\neq i\in S}\alpha_j - \frac{\left(\sum_{j\in S}\alpha_j\right)^2}{2\beta_i} - \frac{\alpha_i^2}{2\beta_i}\right]. \tag{A11}$$

Breaking the first sum up into two parts implies

$\alpha_i^2\sum_{j\in S}\frac{1}{\beta_j} = \alpha_i^2\sum_{j\neq i\in S}\frac{1}{\beta_j} + \frac{\alpha_i^2}{\beta_i}$, which gives

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} + \frac{\alpha_i^2}{\beta_i} + \alpha_i\left(\frac{1}{\beta_i}\right)\sum_{j\neq i\in S}\alpha_j - \frac{\left(\sum_{j\in S}\alpha_j\right)^2}{2\beta_i} - \frac{\alpha_i^2}{2\beta_i}\right]. \tag{A12}$$

Combining the second and fifth term $\frac{\alpha_i^2}{\beta_i} - \frac{\alpha_i^2}{2\beta_i} = \frac{\alpha_i^2}{2\beta_i}$, results in

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} + \alpha_i\left(\frac{1}{\beta_i}\right)\sum_{j\neq i\in S}\alpha_j - \frac{\left(\sum_{j\in S}\alpha_j\right)^2}{2\beta_i} + \frac{\alpha_i^2}{2\beta_i}\right]. \tag{A13}$$

Putting terms over a common denominator $2\beta_i$ results in

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} + \frac{2\alpha_i\sum_{j\neq i\in S}\alpha_j - \left(\sum_{j\in S}\alpha_j\right)^2 + \alpha_i^2}{2\beta_i}\right]. \tag{A14}$$

Breaking the second term in the numerator into two parts implies $-\left(\sum_{j\in S}\alpha_j\right)^2 = -\sum_{j\in S}\alpha_j^2 - 2\sum_{j,k\in S}\alpha_j\alpha_k$ such that

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} + \frac{2\alpha_i\sum_{j\neq i\in S}\alpha_j - \sum_{j\in S}\alpha_j^2 - 2\sum_{j,k\in S}\alpha_j\alpha_k + \alpha_i^2}{2\beta_i}\right]. \tag{A15}$$

Combining the second and fourth term in the numerator implies $-\sum_{j\in S}\alpha_j^2 + \alpha_i^2 = -\sum_{j\neq i\in S}\alpha_j^2$ and hence

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} + \frac{2\alpha_i\sum_{j\neq i\in S}\alpha_j - \sum_{j\neq i\in S}\alpha_j^2 - 2\sum_{j,k\in S}\alpha_j\alpha_k}{2\beta_i}\right]. \tag{A16}$$

Combining the first and third term in the numerator implies $2\alpha_i\sum_{j\neq i\in S}\alpha_j - 2\sum_{j,k\in S}\alpha_j\alpha_k = -2\sum_{j\neq i,k\neq i\in S}\alpha_j\alpha_k$ such that

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} + \frac{-\sum_{j\neq i\in S}\alpha_j^2 - 2\sum_{j\neq i,k\neq i\in S}\alpha_j\alpha_k}{2\beta_i}\right]. \tag{A17}$$

Combining the two terms in the numerator implies $-\sum_{j\neq i\in S}\alpha_j^2 - 2\sum_{j\neq i,k\neq i\in S}\alpha_j\alpha_k = -\left(\sum_{j\neq i\in S}\alpha_j\right)^2$ which leads to

$$\sigma_i(S) = \frac{b^2}{c}\left[\alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} - \frac{\left(\sum_{j\neq i\in S}\alpha_j\right)^2}{2\beta_i}\right]. \tag{A18}$$

In the absence of side-payments, coalition S is internally stable for i if $\sigma_i(S) \geq 0$, or since $b > 0$, $c > 0$, and $\alpha_i > 0$ and $\beta_i > 0$ for all $i \in N$:

$$\sigma_i(S) \geq 0 \Leftrightarrow \alpha_i^2 \sum_{j\neq i\in S}\frac{1}{\beta_j} - \frac{\left(\sum_{j\neq i\in S}\alpha_j\right)^2}{2\beta_i} \geq 0 \Leftrightarrow 2\beta_i \sum_{j\neq i\in S}\frac{1}{\beta_j} \geq \left(\frac{\sum_{j\neq i\in S}\alpha_j}{\alpha_i}\right)^2 \tag{A19}$$

Using $\gamma_i \equiv \frac{1}{\beta_i}$, $\psi_i \equiv \frac{\gamma_i}{\sum_{j\in S}\gamma_j}$ and $\theta_i \equiv \frac{\alpha_i}{\sum_{j\in S}\alpha_j}$, (A19) becomes

$$\begin{aligned}
\frac{2}{\gamma_i}\left(\sum_{j\in S}\gamma_j - \gamma_i\right) &\geq \left(\frac{\sum_{j\in S}\alpha_j - \alpha_i}{\alpha_i}\right)^2 \\
\frac{2\sum_{j\in S}\gamma_j}{\gamma_i}\left(1 - \frac{\gamma_i}{\sum_{j\in S}\gamma_j}\right) &\geq \left(\frac{1 - \frac{\alpha_i}{\sum_{j\in S}\alpha_j}}{\frac{\alpha_i}{\sum_{j\in S}\alpha_j}}\right)^2 \\
\frac{2}{\psi_i}(1 - \psi_i) &\geq \left(\frac{1-\theta_i}{\theta_i}\right)^2 \\
\frac{2\theta_i^2}{(1-\theta_i)^2}(1 - \psi_i) &\geq \psi_i \\
\frac{2\theta_i^2}{(1-\theta_i)^2} &\geq \psi_i\left[1 + \frac{2\theta_i^2}{(1-\theta_i)^2}\right] \\
2\theta_i^2 &\geq \psi_i\left[(1-\theta_i)^2 + 2\theta_i^2\right] \\
f(\theta_i) \equiv \frac{2\theta_i^2}{3\theta_i^2 - 2\theta_i + 1} &\geq \psi_i
\end{aligned} \tag{A20}$$

which is condition (11) in Proposition 2.

For Corollary 1 results (a)–(c) are immediately evident from Fig. 1. For symmetric benefits, the necessary condition $\theta_i \geq \frac{1}{3}$ can only hold for a coalition of up to three players. For symmetric costs, $\theta_i = \psi_i$ and hence $f(\theta_i) \geq \psi_i$ holds for 2 and 3 players but only the larger coalition is externally stable and hence $s^* = 3$. For three players with symmetric benefits $f(\theta_i) = \theta_i = \frac{1}{3}$ and therefore $f(\theta_i) \geq \psi_i$ for all $i \in S$ can only hold if and only if $\psi_i = \frac{1}{3}$ for all $i \in S$. Hence, for symmetric benefits but asymmetric costs $s^* < 3$. Finally, for symmetric costs, the necessary condition $\psi_i \geq \frac{1}{3}$ can only hold for a coalition of up to three players, with $\psi_i = \frac{1}{3}$ for all $i \in S$. But this will violate $f(\theta_i) \geq \psi_i$ for all $i \in S$ except if all benefits are symmetric, i.e., $\theta_i = \frac{1}{3}$. Hence for asymmetric benefits but symmetric costs, $s^* < 3$.

For Corollary 2 a covariance different from zero requires at least two θ_i-values and ψ_i-values to be different. Consequently, the necessary condition for internal stability $\theta_i \geq \frac{1}{3}$ and $\psi_i \geq \frac{1}{3}$ for at least one player i from Proposition 2, requires $\alpha_i > \bar{\alpha}_S$ and $\beta_i < \bar{\beta}_S$ (with $\bar{\alpha}_S$ and $\bar{\beta}_S$ referring to coalition averages) for any coalition of size $s \geq 3$, i.e., a negative covariance.

Appendix C

The total contribution $Q^*(S)$ is given at the beginning of Appendix A, which reads if a player k leaves coalition S:

$$Q^*(S\backslash\{k\}) = \frac{b}{c}\left[\left(\sum_{i\in S}\frac{1}{\beta_i} - \frac{1}{\beta_k}\right)\left(\sum_{i\in S}\alpha_i - \alpha_k\right) + \sum_{j\in T}\frac{\alpha_j}{\beta_j} + \frac{\alpha_k}{\beta_k}\right]. \tag{A21}$$

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

$$\pi_k^*(S\backslash\{k\}) = b\alpha_k Q^*(S\backslash\{k\}) - \frac{b^2(\alpha_k)^2}{2c\beta_k}$$

$$\pi_k^*(S\backslash\{k\}) = \frac{b^2\alpha_k}{c}\left[\left(\sum_{i\in S}\frac{1}{\beta_i} - \frac{1}{\beta_k}\right)\left(\sum_{i\in S}\alpha_i - \alpha_k\right) + \sum_{j\in T}\frac{\alpha_j}{\beta_j} + \frac{\alpha_k}{\beta_k} - \frac{\alpha_k}{2\beta_k}\right]$$

$$\pi_k^*(S\backslash\{k\}) = \frac{b^2\alpha_k}{c}\left[\left(\sum_{i\in S}\frac{1}{\beta_i} - \frac{1}{\beta_k}\right)\left(\sum_{i\in S}\alpha_i - \alpha_k\right) + \sum_{j\in T}\frac{\alpha_j}{\beta_j} + \frac{\alpha_k}{2\beta_k}\right]. \tag{A22}$$

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

$$\sum_{k\in S}\pi_k^*(S\backslash\{k\}) = \frac{b^2}{c}\left[\begin{array}{c}\sum_{i\in S}\frac{1}{\beta_i}\left[(\sum_{i\in S}\alpha_i)^2 - \sum_{i\in S}(\alpha_i)^2\right]\\ +\sum_{i\in S}\alpha_i\left[\sum_{j\in T}\frac{\alpha_j}{\beta_j} - \sum_{i\in S}\frac{\alpha_i}{\beta_i}\right] + \frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\end{array}\right] \tag{A23}$$

Hence, the coalition surplus, $\sigma(S) = \sum_{k\in S}\pi_k^*(S) - \sum_{k\in S}\pi_k^*(S\backslash\{k\})$, using (A2) and (A23), is given by:

$$\sigma(S) = \frac{b^2}{c}\sum_{i\in S}\alpha_i\left[\frac{1}{2}\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{j\in T}\frac{\alpha_j}{\beta_j}\right]$$

$$-\left\{\frac{b^2}{c}\left[\begin{array}{c}\sum_{i\in S}\frac{1}{\beta_i}\left[(\sum_{i\in S}\alpha_i)^2 - \sum_{i\in S}(\alpha_i)^2\right]\\ +\sum_{i\in S}\alpha_i\left[\sum_{j\in T}\frac{\alpha_j}{\beta_j} - \sum_{i\in S}\frac{\alpha_i}{\beta_i}\right] + \frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\end{array}\right]\right\}$$

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

$$\sigma(S) = \frac{b^2}{c}\left\{\sum_{i\in S}\frac{1}{\beta_i}\left[\sum_{i\in S}(\alpha_i)^2 - \frac{(\sum_{i\in S}\alpha_i)^2}{2}\right] + \sum_{i\in S}\alpha_i\sum_{i\in S}\frac{\alpha_i}{\beta_i} - \frac{3}{2}\sum_{i\in S}\frac{(\alpha_i)^2}{\beta_i}\right\}. \tag{A24}$$

Since $\frac{b^2}{c} > 0$, the stability condition depends on the term in brackets. Using $\gamma_i \equiv \frac{1}{\beta_i}$, this becomes

$$\sum_{i\in S}\gamma_i\left[\sum_{i\in S}(\alpha_i)^2 - \frac{(\sum_{i\in S}\alpha_i)^2}{2}\right] + \sum_{i\in S}\alpha_i\sum_{i\in S}\gamma_i\alpha_i - \frac{3}{2}\sum_{i\in S}\gamma_i(\alpha_i)^2 \geq 0. \tag{A25}$$

Then, dividing both sides by $(\Sigma_{i\in S}\alpha_i)^2$, we obtain

$$\sum_{i\in S}\gamma_i\left[\frac{\sum_{i\in S}(\alpha_i)^2}{(\sum_{i\in S}\alpha_i)^2} - \frac{1}{2}\right] + \frac{\sum_{i\in S}\alpha_i\sum_{i\in S}\gamma_i\alpha_i}{(\sum_{i\in S}\alpha_i)^2} - \frac{3}{2}\frac{\sum_{i\in S}\gamma_i(\alpha_i)^2}{(\sum_{i\in S}\alpha_i)^2} \geq 0. \tag{A26}$$

The middle term reduces to

$$\sum_{i\in S}\gamma_i\left[\frac{\sum_{i\in S}(\alpha_i)^2}{(\sum_{i\in S}\alpha_i)^2} - \frac{1}{2}\right] + \frac{\sum_{i\in S}\gamma_i\alpha_i}{\sum_{i\in S}\alpha_i} - \frac{3}{2}\frac{\sum_{i\in S}\gamma_i(\alpha_i)^2}{(\sum_{i\in S}\alpha_i)^2} \geq 0 \tag{A27}$$

and, given $\theta_i \equiv \frac{\alpha_i}{\sum_{j\in S}\alpha_j}$, the middle term simplifies to

$$\sum_{i\in S}\gamma_i\left[\frac{\sum_{i\in S}(\alpha_i)^2}{\left(\sum_{i\in S}\alpha_i\right)^2}-\frac{1}{2}\right]+\sum_{i\in S}\gamma_i\theta_i-\frac{3}{2}\frac{\sum_{i\in S}\gamma_i(\alpha_i)^2}{\left(\sum_{i\in S}\alpha_i\right)^2}\geq 0. \tag{A28}$$

We can square both sides and note that $\theta_i^2=\frac{(\alpha_i)^2}{\left(\sum_{j\in S}\alpha_j\right)^2}$. Then summing across all $i\in S$, we have $\sum_{i\in S}\theta_i^2=\frac{\sum_{i\in S}(\alpha_i)^2}{\left(\sum_{i\in S}\alpha_i\right)^2}$. Using this for the first and third term, the condition becomes

$$\sum_{i\in S}\gamma_i\left[\sum_{i\in S}\theta_i^2-\frac{1}{2}\right]+\sum_{i\in S}\gamma_i\theta_i-\frac{3}{2}\sum_{i\in S}\gamma_i\theta_i^2\geq 0. \tag{A29}$$

Now if we divide this condition by $\Sigma_{i\in S}\gamma_i>0$ for $s\geq 1$

$$\sum_{i\in S}\theta_i^2-\frac{1}{2}+\frac{\sum_{i\in S}\gamma_i\theta_i}{\sum_{i\in S}\gamma_i}-\frac{3}{2}\frac{\sum_{i\in S}\gamma_i\theta_i^2}{\sum_{i\in S}\gamma_i}\geq 0 \tag{A30}$$

and, using $\psi_i\equiv\frac{\gamma_i}{\sum_{i\in S}\gamma_i}=\frac{\frac{1}{\beta_i}}{\sum_{j\in S}\frac{1}{\beta_j}}$, this becomes

$$\sum_{i\in S}\theta_i^2-\frac{1}{2}+\sum_{i\in S}\psi_i\theta_i-\frac{3}{2}\sum_{i\in S}\psi_i\theta_i^2\geq 0. \tag{A31}$$

Then combining the last two sums, we obtain the stability condition (12) in Proposition 3.

For Corollary 3, we use (A24) from Appendix C to show that for a two player coalition with transfers the surplus is strictly positive:

$$\sigma(S,m=2)=\frac{b^2}{c}\left\{\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}\right]\left[\alpha_i^2+\alpha_j^2-\frac{(\alpha_i+\alpha_j)^2}{2}\right]+\left[\alpha_i+\alpha_j\right]\left[\frac{\alpha_i}{\beta_i}+\frac{\alpha_j}{\beta_j}\right]-\frac{3}{2}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}\right]\right\} \tag{A32}$$

$$\begin{aligned}
&=\frac{b^2}{c}\left\{\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}\right]\left[\frac{2\alpha_i^2+2\alpha_j^2-\alpha_i^2-\alpha_j^2-2\alpha_i\alpha_j}{2}\right]+\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_i\alpha_j}{\beta_i}+\frac{\alpha_i\alpha_j}{\beta_j}+\frac{\alpha_j^2}{\beta_j}\right]-\frac{3}{2}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}\right]\right\}\\
&=\frac{b^2}{c}\left\{\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}\right]\left[\frac{\alpha_i^2+\alpha_j^2-2\alpha_i\alpha_j}{2}\right]+\left[\frac{\alpha_i\alpha_j}{\beta_i}+\frac{\alpha_i\alpha_j}{\beta_j}\right]-\frac{1}{2}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}\right]\right\}\\
&=\frac{b^2}{2c}\left\{\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}\right]\left[\alpha_i^2+\alpha_j^2-2\alpha_i\alpha_j\right]+\left[\frac{2\alpha_i\alpha_j}{\beta_i}+\frac{2\alpha_i\alpha_j}{\beta_j}\right]-\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}\right]\right\}\\
&=\frac{b^2}{2c}\left\{\frac{\alpha_i^2+\alpha_j^2-2\alpha_i\alpha_j+2\alpha_i\alpha_j-\alpha_i^2}{\beta_i}+\frac{\alpha_i^2+\alpha_j^2-2\alpha_i\alpha_j+2\alpha_i\alpha_j-\alpha_j^2}{\beta_j}\right\}\\
&=\frac{b^2}{2c}\left\{\frac{\alpha_j^2}{\beta_i}+\frac{\alpha_i^2}{\beta_j}\right\}\\
&=\frac{b^2}{2c}\sum_{i,j\in S}\frac{\alpha_i^2}{\beta_j}>0\,.
\end{aligned} \tag{A33}$$

For a three player coalition with transfers the surplus is strictly positive:

$$\begin{aligned}
\sigma(S,m=3)&=\frac{b^2}{c}\left\{\begin{array}{l}\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right]\left[\alpha_i^2+\alpha_j^2+\alpha_k^2-\frac{(\alpha_i+\alpha_j+\alpha_k)^2}{2}\right]+\left[\alpha_i+\alpha_j+\alpha_k\right]\left[\frac{\alpha_i}{\beta_i}+\frac{\alpha_j}{\beta_j}+\frac{\alpha_k}{\beta_k}\right]\\ \quad-\frac{3}{2}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}+\frac{\alpha_k^2}{\beta_k}\right]\end{array}\right\}\\
&=\frac{b^2}{c}\left\{\begin{array}{l}\frac{1}{2}\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right]\left[2\alpha_i^2+2\alpha_j^2+2\alpha_k^2-\left(\alpha_i+\alpha_j+\alpha_k\right)^2\right]\\ \quad+\frac{2\alpha_i\left(\alpha_i+\alpha_j+\alpha_k\right)}{2\beta_i}+\frac{2\alpha_j\left(\alpha_i+\alpha_j+\alpha_k\right)}{2\beta_j}+\frac{2\alpha_k\left(\alpha_i+\alpha_j+\alpha_k\right)}{2\beta_k}-\frac{3}{2}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}+\frac{\alpha_k^2}{\beta_k}\right]\end{array}\right\}\\
&=\frac{b^2}{c}\left\{\begin{array}{l}\frac{1}{2}\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right]\left[\begin{array}{c}2\alpha_i^2+2\alpha_j^2+2\alpha_k^2\\-\left(\alpha_i^2+\alpha_j^2+\alpha_k^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k+2\alpha_j\alpha_k\right)\end{array}\right]\\ +\frac{2\alpha_i^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k}{2\beta_i}+\frac{2\alpha_j^2+2\alpha_i\alpha_j+2\alpha_j\alpha_k}{2\beta_j}+\frac{2\alpha_k^2+2\alpha_i\alpha_k+2\alpha_j\alpha_k}{2\beta_k}-\frac{3}{2}\left[\frac{\alpha_i^2}{\beta_i}+\frac{\alpha_j^2}{\beta_j}+\frac{\alpha_k^2}{\beta_k}\right]\end{array}\right\}
\end{aligned}$$

$$= \frac{b^2}{2c}\left\{\begin{array}{c}\left[\frac{1}{\beta_i}+\frac{1}{\beta_j}+\frac{1}{\beta_k}\right]\left[\alpha_i^2+\alpha_j^2+\alpha_k^2-\left(2\alpha_i\alpha_j+2\alpha_i\alpha_k+2\alpha_j\alpha_k\right)\right]\\ +\frac{-\alpha_i^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k}{\beta_i}+\frac{-\alpha_j^2+2\alpha_i\alpha_j+2\alpha_j\alpha_k}{\beta_j}+\frac{-\alpha_k^2+2\alpha_i\alpha_k+2\alpha_j\alpha_k}{\beta_k}\end{array}\right\}$$

$$= \frac{b^2}{2c}\left\{\begin{array}{c}\frac{\alpha_i^2+\alpha_j^2+\alpha_k^2-\left(2\alpha_i\alpha_j+2\alpha_i\alpha_k+2\alpha_j\alpha_k\right)-\alpha_i^2+2\alpha_i\alpha_j+2\alpha_i\alpha_k}{\beta_i}\\ +\frac{\alpha_i^2+\alpha_j^2+\alpha_k^2-\left(2\alpha_i\alpha_j+2\alpha_i\alpha_k+2\alpha_j\alpha_k\right)-\alpha_j^2+2\alpha_i\alpha_j+2\alpha_j\alpha_k}{\beta_j}\\ +\frac{\alpha_i^2+\alpha_j^2+\alpha_k^2-\left(2\alpha_i\alpha_j+2\alpha_i\alpha_k+2\alpha_j\alpha_k\right)-\alpha_k^2+2\alpha_i\alpha_k+2\alpha_j\alpha_k}{\beta_k}\end{array}\right\}$$

$$= \frac{b^2}{2c}\left[\frac{\alpha_j^2+\alpha_k^2-2\alpha_j\alpha_k}{\beta_i}+\frac{\alpha_i^2+\alpha_k^2-2\alpha_i\alpha_k}{\beta_j}+\frac{\alpha_i^2+\alpha_j^2-2\alpha_i\alpha_j}{\beta_k}\right]$$

$$= \frac{b^2}{2c}\left[\sum_{i,j,k\in S}\frac{\left(\alpha_j-\alpha_k\right)^2}{2\beta_i}\right] > 0. \tag{A34}$$

Note that $m = 2$ is not externally stable because $m = 3$ is internally stable.

For Corollary 4 with respect to benefit symmetry, note that $\theta_i = \frac{1}{s}$, $\theta_i^2 = \frac{1}{s^2}$ and $\sum_{i\in S}(\theta_i)^2 = \frac{1}{s}$ and hence (12) in the text becomes

$$\frac{1}{s}-\frac{1}{2}+\frac{1}{2s}\sum_{i\in S}\psi_i\left(2-\frac{3}{s}\right)\geq 0. \tag{A35}$$

Since $\sum_{j\in S}\psi_i = 1$, this becomes

$$\frac{1}{s}-\frac{1}{2}+\frac{2s-3}{2s^2}\geq 0. \tag{A36}$$

Then, multiplying by $2s^2$, we have

$$2s-s^2+2s-3\geq 0 \tag{A37}$$

or

$$\begin{aligned}&-s^2+4s-3\geq 0\\ &-(s-1)(s-3)\geq 0.\end{aligned} \tag{A38}$$

Clearly, (A38) holds for $s = 1$, $s = 2$, and $s = 3$. For $s \geq 4$, the condition does not hold and the coalition is not stable.

Regarding the second statement in Corollary 4 with respect to cost symmetry note that $\psi_i = \frac{1}{s}$ for all $i \in S$ and hence (12) in the text becomes

$$\sum_{i\in S}\theta_i^2-\frac{1}{2}+\frac{1}{2s}\sum_{i\in S}\theta_i(2-3\theta_i)\geq 0. \tag{A39}$$

Breaking up the second sum, we have

$$\sum_{i\in S}\theta_i^2-\frac{1}{2}+\frac{1}{s}\sum_{i\in S}\theta_i-\frac{3}{2s}\sum_{i\in S}\theta_i^2\geq 0. \tag{A40}$$

Since $\sum_{i\in S}\theta_i = 1$ by definition, this becomes

$$\frac{1}{s}-\frac{1}{2}+\sum_{i\in S}\theta_i^2\left(1-\frac{3}{2s}\right)\geq 0 \tag{A41}$$

which leads directly to (13) in Corollary 4.

Appendix D

We first note that the grand coalition is externally stable by definition. The total contribution $Q^*(S)$ is given at the beginning of Appendix 1. Take a non-signatory $j \in T$ who joins coalition S. Abatement after j joins S is

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

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

$$Q^*(S \cup \{j\}) = \frac{b}{c}\left[\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \frac{1}{\beta_j}\sum_{i\in S}\alpha_i + \alpha_j\sum_{i\in S}\frac{1}{\beta_i} + \sum_{k\in T}\frac{\alpha_k}{\beta_k}\right]. \tag{A42}$$

The change in benefit for j from joining is

$$\begin{aligned}
\Delta B_j &= \alpha_j b[Q^*(S\cup\{j\}) - Q^*(S)]\\
\Delta B_j &= \frac{\alpha_j b^2}{c}\left[\textstyle\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \frac{1}{\beta_j}\sum_{i\in S}\alpha_i + \alpha_j\sum_{i\in S}\frac{1}{\beta_i} + \sum_{k\in T}\frac{\alpha_k}{\beta_k} - \left(\sum_{i\in S}\frac{1}{\beta_i}\sum_{i\in S}\alpha_i + \sum_{k\in T}\frac{\alpha_k}{\beta_k}\right)\right]\\
\Delta B_j &= \frac{\alpha_j b^2}{c}\left[\frac{1}{\beta_j}\sum_{i\in S}\alpha_i + \alpha_j\sum_{i\in S}\frac{1}{\beta_i}\right]\\
\Delta B_j &= \frac{\alpha_j b^2}{c\beta_j}\left[\sum_{i\in S}\alpha_i + \alpha_j\beta_j\sum_{i\in S}\frac{1}{\beta_i}\right].
\end{aligned} \tag{A43}$$

When j joins S, j has abatement level

$$q^*_{j\in S\cup\{j\}}(S\cup\{j\}) = \frac{b}{c\beta_j}\left(\sum_{i\in S}\alpha_i + \alpha_j\right). \tag{A44}$$

The change in abatement cost for j from joining is

$$\begin{aligned}
\Delta C_j &= \frac{c\beta_j\left(q^*_{j\in S\cup\{j\}}(S+j)\right)^2}{2} - \frac{c\beta_j\left(q^*_{j\in T}(S)\right)^2}{2}\\
\Delta C_j &= \frac{c\beta_j}{2}\left[\left(\frac{b}{c\beta_j}\left(\sum_{i\in S}\alpha_i + \alpha_j\right)\right)^2 - \left(\frac{b\alpha_j}{c\beta_j}\right)^2\right]\\
\Delta C_j &= \frac{b^2}{2c\beta_j}\left[\left(\sum_{i\in S}\alpha_i + \alpha_j\right)^2 - \left(\alpha_j\right)^2\right]\\
\Delta C_j &= \frac{b^2}{2c\beta_j}\left[\left(\sum_{i\in S}\alpha_i\right)^2 + 2\alpha_j\sum_{i\in S}\alpha_i\right].
\end{aligned} \tag{A45}$$

The external stability condition is

$$\Delta\pi_i = \Delta B_i - \Delta C_i \le 0 \tag{A46}$$

which must hold for all $j \in T$. Using (A43) and (A45), this becomes

$$\begin{aligned}
\Delta B_i - \Delta C_i &\le 0\\
\frac{\alpha_j b^2}{c\beta_j}\left[\sum_{i\in S}\alpha_i + \alpha_j\beta_j\sum_{i\in S}\frac{1}{\beta_i}\right] &\le \frac{b^2}{2c\beta_j}\left[\left(\sum_{i\in S}\alpha_i\right)^2 + 2\alpha_j\sum_{i\in S}\alpha_i\right]\\
2\alpha_j\left[\sum_{i\in S}\alpha_i + \alpha_j\beta_j\sum_{i\in S}\frac{1}{\beta_i}\right] &\le \left(\sum_{i\in S}\alpha_i\right)^2 + 2\alpha_j\sum_{i\in S}\alpha_i\\
2\left(\alpha_j\right)^2\left[\beta_j\sum_{i\in S}\frac{1}{\beta_i}\right] &\le \left(\sum_{i\in S}\alpha_i\right)^2.
\end{aligned} \tag{A47}$$

Using $\theta_i \equiv \frac{\alpha_i}{\sum_{j\in S}\alpha_j}$, $\gamma_i \equiv \frac{1}{\beta_i}$ and $\psi_i \equiv \frac{\gamma_i}{\sum_{k\in S}\gamma_k}$, we have:

$$2\frac{(\alpha_j)^2}{(\sum_{i\in S}\alpha_i)^2} \le \frac{\frac{1}{\beta_j}}{\frac{1}{\sum_{i\in S}\frac{1}{\beta_i}}} \tag{A48}$$

or

$$2\frac{\frac{(\alpha_j)^2}{\left(\sum_{k\in N}\alpha_k\right)^2}}{\frac{\left(\sum_{i\in S}\alpha_i\right)^2}{\left(\sum_{k\in N}\alpha_k\right)^2}} \le \frac{\frac{\frac{1}{\beta_j}}{\frac{1}{\sum_{k\in N}\frac{1}{\beta_k}}}}{\frac{\frac{1}{\sum_{i\in S}\frac{1}{\beta_i}}}{\frac{1}{\sum_{k\in N}\frac{1}{\beta_k}}}} \tag{A49}$$

or

$$2\frac{{\theta_j}^2}{\psi_j} \le \frac{\theta_S^2}{\psi_S}. \tag{A50}$$

Hence, a coalition is externally stable if player *j's* benefit to inverse cost ratio is sufficiently small compared to the "average" of the current coalition members in *S*. In other words, player *j* would join coalition *S* if his individual benefit parameter and his individual cost parameter was sufficiently large compared to those of the current members in *S*, as he would contribute below and benefit above average.

For transfers, it is easy to show that if coalition *S* is not externally stable regarding accession of player *j*, then $S\cup\{j\}$ is potentially internally stable, and, due to strict full cohesiveness, total payoffs are higher. Axiomatically, among the set of potentially internally stable coalitions, the one with the highest aggregate payoff is externally stable. Thus, external stability is less of concern with transfers.

Appendix E

We claim that our results extend to other payoff functions, at least in qualitative terms, even though analytical results are difficult to obtain. Hence, we consider a payoff function with a quadratic benefit and quadratic cost function, which has been frequently considered in the IEA-literature, and consider 5 different parameter constellations, which we call Examples 1–5.

$$\pi_i = \alpha_i\left(bQ - \frac{a}{2}Q^2\right) - \beta_i\frac{c}{2}q_i^2 \tag{A51}$$

This implies the following first order conditions in the grand coalition, $S = N$:

$$\begin{aligned} b - aQ(N) &= c\beta_i q_i \\ \frac{1}{c\beta_i}(b - aQ(N)) &= q_i. \end{aligned} \tag{A52}$$

Summing over all $i \in N$, we have:

$$\begin{aligned} \sum_{i\in N}\frac{1}{\beta_i}\frac{1}{c}(b - aQ(N)) &= Q(N) \\ X(b - aQ(N)) &= Q(N) \\ \frac{Xb}{1 + aX} &= Q(N) \end{aligned} \tag{A53}$$

with $X = \sum_{i\in N}\frac{1}{\beta_i}\frac{1}{c}$. Substituting *Q* back into q_i above, gives individual contribution levels.

Now consider if one player *j* leaves the grand coalition and we have coalition *S* with $\sum_{i\in S}\alpha_i = \alpha_S$. Then we have:

$$\begin{aligned} \alpha_S(b - aQ(S)) &= c\beta_i q_i \\ \alpha_j(b - aQ(S)) &= c\beta_j q_j \end{aligned} \tag{A54}$$

or

$$\begin{aligned} \frac{1}{c\beta_i}\alpha_S(b - aQ(S)) &= q_i \\ \frac{1}{c\beta_j}\alpha_j(b - aQ(S)) &= q_j \end{aligned} \tag{A55}$$

Now summing over all signatories we have:

$$\sum_{i\in N}\frac{1}{\beta_i}\frac{1}{c}\alpha_S(b - aQ(S)) = Q_S \tag{A56}$$

together with

$$\frac{1}{c\beta_j}\alpha_j(b - aQ(S)) = q_j \tag{A57}$$

and hence if we let $X_{-j} = \sum_{i\in N}\frac{1}{\beta_i}\frac{1}{c}\alpha_S$ and $X_j = \frac{1}{c\beta_j}\alpha_j$, and sum over all first order conditions, we have:

$$\begin{aligned}\left(X_{-j}+X_j\right)(b - aQ(S)) &= Q(S)\\ \frac{\left(X_{-j}+X_j\right)b}{1+\left(X_{-j}+X_j\right)a} &= Q(S).\end{aligned} \tag{A58}$$

and, again, we can determine individual contribution levels by substituting $Q(S)$ back above. If we insert contribution levels in the payoff function, we can determine internal stability and potential internal stability. We have conducted this exercise with the software programme Maple for $n = 10$ players; detailed results are available upon request. It turns out that as for the simple payoff function considered in the text, stability depends on the distribution of the individual parameters and is independent of the value of the global benefit parameter b. Different from the simple payoff function, stability depends on the parameter c, and the new parameter a, though it turns out that only the ratio $\frac{a}{c}$ matters, which needs to be sufficiently small. In the following we choose $a = 6\cdot 10^{-9}$ and $c = 100$ which imply stability of the grand coalition in all five examples ($s^* = 10$), though for examples 4 and 5 transfers are required.

As in the paper in Section 4 (without transfers), Example 1 (modified) assumes $\alpha_1 = \alpha_2 = 0.42,\ \alpha_3 = ... = \alpha_{10} = 0.02$ and $\beta_1 = \beta_2 = 0.00015,\ \beta_3 = ... = \beta_{10} = 0.12496250$, Example 2 assumes $\alpha_1 = \alpha_2 = 0.49,\ \alpha_3 = ... = \alpha_{10} = 0.0025$ and $\beta_1 = \beta_2 = 0.0000031408,$ and $\beta_3 = ... = \beta_{10} = 0.1249992148$ and Example 3 assumes $\alpha_1 = 0.64$, $\alpha_2 = 0.32,\ \alpha_3 = ... = \alpha_{10} = 0.005$ and $\beta_1 = 0.000007,\ \beta_2 = 0.00004,\ \beta_3 = ... = \beta_{10} = 0.124994125$. In accordance with Section 5 (with transfers), we assume for Example 4 $\alpha_1 = 0.65$ and $\alpha_2 = 0.3,\ \alpha_3 = ... = \alpha_{10} = 0.00625$ and symmetric cost parameters and hence $\beta_1 = \beta_2 = \ldots = \beta_{10} = 0.1$ and for Example 5, again, $\alpha_1 = 0.65$ and $\alpha_2 = 0.3,\ \alpha_3 = ... = \alpha_{10} = 0.00625$ but asymmetric cost parameters with a positive covariance with $\beta_1 = 0.65$ and $\beta_2 = 0.3$ and $\beta_3 = ... = \beta_{10} = 0.00625$.

Moreover, we find for Example 1: $\Delta Q = 77.96b$ and $\Delta\Pi = 22.73b^2$; Example 2: $\Delta Q = 3248.03b$ and $\Delta\Pi = 828.38b^2$; Example 3: $\Delta Q = 684.91b$ and $\Delta\Pi = 150.68b^2$; Example 4: $\Delta Q = 0.899b$ and $\Delta\Pi = 0.4256b^2$ and Example 5: $\Delta Q = 12.74b$ and $\Delta\Pi = 6.329b^2$. With symmetry, $\Delta Q = 0.89b$ and $\Delta\Pi = 0.4b^2$ but for the actual stable coalition with only $s^* = 2$ we have $\Delta\widetilde{Q} = 0.1169b$ and $\Delta\widetilde{\Pi} = 0.1199b^2$ ($\Delta\widetilde{Q} := Q(s^*) - Q_{NC}$ and $\Delta\widetilde{\Pi} := \Pi(s^*) - \Pi_{NC}$). Note that for this payoff function and symmetry, generally $s^* < 3$.

References

Ambec, S., Sprumont, Y., 2002. Sharing a river. J. Econ. Theory 107 (2), 453–462.

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxf. Econ. Pap 46, 878–894.

Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. J. Public Econ. 52 (3), 309–328.

Carraro, C., Eyckmans, J., Finus, M., 2006. Optimal transfers and participation decisions in international environmental agreements. Rev. Int. Organ. 1 (4), 379–396.

Chander, P., Tulkens, H., 1995. A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. Int. Tax Public Financ. 2 (2), 279–293.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J.A., 1983. On the stability of collusive price leadership. Can. J. Econ. 16 (1), 17–25.

Diamantoudi, E., Sartzetakis, E.S., 2006. Stable international environmental agreements: an analytical approach. J. Public Econ. Theory 8 (2), 247–263.

Finus, M., Rundshagen, B., 2006. A micro foundation of core stability in positive-externality coalition games. J. Inst. Theor. Econ. 162 (2), 329–346.

Finus, M., Caparrós, A., 2015. Game theory and international environmental cooperation. The International Library of Critical Writings in Economics. Edward Elgar, Cheltenham, UK.

Finus, M., Pintassilgo, P., 2013. The role of uncertainty and learning for the success of international climate agreements. J. Public Econ. 103, 29–43.

Fuentes-Albero, C., Rubio, S.J., 2010. Can international environmental cooperation be bought? Eur. J. Oper. Res. 202, 255–264.

Haeringer, G., 2004. Equilibrium binding agreements: a comment. J. Econ. Theory 117 (1), 140–143.

Harstad, B., 2012. Climate contracts: a game of emissions, investments, negotiations, and renegotiations. Rev. Econ. Stud. 79, 1527–1557.

Harstad, B., 2016. The dynamics of climate agreements. J. Eur. Econ. Assoc. 14, 719–752.

Karp, L., Simon, L., 2013. Participation games and international environmental agreements: a non-parametric model. J. Environ. Econ. Manage. 65, 326–344.

McGinty, M., 2007. International environmental agreements among asymmetric nations. Oxf. Econ. Pap. 59, 45–62.

Nagashima, M., Dellink, R., van Ierland, E., Weikard, H.P., 2009. Stability of international climate coalitions – a comparison of transfer schemes. Ecol. Econ. 68, 1476–1487.

Pavlova, Y., de Zeeuw, A., 2013. Asymmetries in international environmental agreements. Environ. Dev. Econ. 18, 51–68.

Ray, D., Vohra, R., 2001. Coalitional power and public goods. J. Polit. Econ. 109 (6), 1355–1384.

Rubio, S.J., Ulph, A., 2006. Self-enforcing international environmental agreements revisited. Oxf. Econ. Pap. 58 (2), 233–263.

Sandler, T., 1999. Alliance formation, alliance expansion, and the core. J. Confl. Resolut. 43 (6), 727–747.

Weikard, H.P., 2009. Cartel Stability Under Optimal Sharing Rule, 77. Manchester School, pp. 575–593.

Yi, S.S., 1997. Stable coalition structures with externalities. Games Econ. Behav. 20 (2), 201–223.