Negotiating International Environmental Agreements: Alone or in a Pool?

Noemí Navarro* and Jean-Christophe Pereau‡

Abstract

We analyze the stability and efficiency of international environmental agreements where negotiations take place bilaterally and simultaneously with one climate secretariat being the leader of the negotiations. Countries have the choice to bargain alone or in a pool with the leader. Our results show that depending the shape of the benefit and cost function and the type of beliefs in case of disagreement, several configurations in terms of abatement efforts and size of stable coalition can emerge. Our approach encompasses the standard IEA game.

1 Introduction

International environmental agreements (IEA) dealing with issues like the ozone layer, long-range pollutants or climate change are often the outcome of a negotiation process which is never, or rarely, made explicit (Finus and Caparrós, 2015). In the case of climate change negotiations, which is our leading reference example, one can observe a gap between the practical and legal setting of the negotiations and their associated theoretical modeling. On the one hand, the legal framework for negotiations refers to the United Nations Framework Convention on Climate Change, where countries or parties negotiate reduction agreements under the aegis of a secretariat responsible for organizing and progressing these multilateral negotiations. On the other hand the theoretical representation of these negotiations is based on a IEA game consisting in an emission reduction competition model between countries and not a negotiation model (Carraro and Siniscalco, 1993; Barrett, 1994). The main message of this IEA game which had a huge influence in climate negotiations is that only small coalitions (agreements) are stable and "self-enforcing". The aim of this article is to show how this message remains the same when the IEA game is reshaped within a negotiation theoretical framework.

*Univ Rennes, CNRS, CREM UMR-6211, F-35000 Rennes. navarro.prada@gmail.com

†University of Bordeaux, CNRS, INRAE (BSE – Bordeaux School of Economics, UMR 6060, UMR 1441), 33600 Pessac, France.

‡Corresponding author: jean-christophe.pereau@u-bordeaux.fr

The structure of Carraro and Siniscalco (1993) and Barrett (1994) models consists in a one-shot two-stage game. Countries decide in a first stage whether to join the agreement and in a second stage decide their abatement efforts jointly if they are signatories of the IEA and independently if not. This second stage refers to a Nash-Cournot scenario or a stackelberg one (Finus et al, 2021). The size of the coalition in the first stage is obtained by using the internal and external stability conditions derived from the cartel formation game of d'Aspremont et al (1983). These conditions state that members of the pool do not want to leave it (internal stability) while non-members of the pool do not want to join it (external stability). This article aims at introducing negotiation introduced into this framework[1] by replacing the Nash-Cournot assumption of the second stage of the IEA game[2] with the Nash-in-Nash protocol consisting in a Nash equilibrium of Nash Bargaining Solutions (Collard-Wexler et al. 2019)[3]. To be more precise, in the second stage countries negotiate with the climate secretariat emission reductions and associated transfers. This secretariat has no reduction target and is not able to enforce formal sanctions as fees in case of non compliance. As it sets the agenda and manages bilateral and simultaneous negotiations, we call it the leader. In the first stage countries have the choice between negotiating alone with the leader as singletons or to pool their meeting with the leader. The size of the pool (or the coalition[4]) is then obtained using the stability conditions as in the IEA game. However our vision of an agreement is somewhat broader than the existing IEA literature which restricts the agreement to the members of the stable coalition. In our negotiation framework, the multilateral IEA will concern both the negotiated agreement between the leader and the pool and also all bilateral negotiated agreements between the leader and the singletons countries.

Our negotiation approach involves two new features to the existing literature. Firstly, it highlights the role of the disagreement payoffs in the case of failure of negotiations. Even if negotiations always result in an agreement, their outcome depends on what might happen if they fail. Contrary to the IEA game, the disagreement payoffs may be different from the non-cooperative situation obtained in the Nash-Cournot case, where each country maximizes his own welfare. More precisely, in a negotiation setting the final outcome depends on the type of beliefs of the players about the behaviors of the other players in all the negotiations they are involved or not which can failed. In our framework we consider *passive beliefs* as defined by McAfee and Schwartz (1994). It implies that in case of disagreement in one negotiation, all countries belief that all other bargaining meetings reach the equilibrium agreement. Secondly, it provides a

[1] Explicit dynamic bargaining IEA game have been analysed by Caparrós and Pereau (2017), Okada (2022, 2023) to allow a gradual coalition formation process.

[2] Finus et al (2021) compare the Nash-Cournot scenario with the Stackelberg one.

[3] In footnote 6, Barrett (1994) refers to the Nash bargaining solution as the solution of a cooperative game to justify that if all signatories are identical, each country will undertake the same level of abatement. But the NBS is not used as we do in this paper.

[4] We use these two terms as synonyms.

complete characterization of utility transfers between countries at the equilibrium. By negotiating reduction agreements, countries create a surplus and the transfer payments implemented by the leader show how countries share the benefits of their cooperation. This model will be solved assuming general concave benefit and convex cost functions. However in order to characterize explicit solutions concerning the abatement made by countries inside the pool or outside as singletons and their associated transfers, we will consider two general polar cases and intermediate ones. The first case refers to a constant marginal benefit and a homogeneous cost function of degree greater than one while the second consists in a constant marginal cost and a homogeneous benefit function of degree lower than one. Intermediate cases refers to some specific values for the degree of homogeneity of the cost and benefit functions. In all cases we are able to determine what will be the abatement countries implement and the size of the pool. Our benchmark model will be subject to two extensions. First we will consider another type of beliefs with *symmetric beliefs* when all countries belief that the other meetings result in a disagreement as well. Secondly we consider Stackelberg bargaining configurations where the leader can decide to bargain either first with the coalition and second with the singletons or vice versa.

This article aims at answering the following questions:

- How the shape of the benefit and cost functions impact the size of a stable coalition from the empty to the grand coalition?
- What are the different strategies of the players in terms of abatements including the possibility of doing nothing in a negotiation framework with different beliefs?
- What can be the role of the leader as an agenda setter?

The organization of this paper is structured as follows. Section 2 introduces the IEA bargaining model with passive beliefs. Section 3 provides the results assuming different shapes of the benefit and cost functions. Section 4 considers two extensions of the model with symmetric beliefs and Stackelberg configurations. Finally, Section 5 offers concluding remarks. All proofs are displayed in appendix.

2 The IEA bargaining model

Assume $n+1$ players with a leader 0 plus n identical players indexed by $i = 1, ..., n$. Each player $i = 1, ..., n$ contributes an emission reduction q_i, bearing a private cost $C(q_i)$ and enjoys a benefit from the total emission reduction (enjoyed by all players including the leader) given by $B(Q)$ with $Q = \sum_{i=1}^{n} q_i$. We assume $B(0) = 0$ and $C(0) = 0$. The benefit function is assumed to be increasing in total abatement at a decreasing rate ($B'(.) > 0$ and $B''(.) < 0$). Costs are assumed to be a strictly convex function of individual abatement ($C'(.) > 0$ and $C''(.) > 0$). We also assumed $B'(0) > C'(0)$ stating that the first unit of abatement is profitable (Muthoo, 1999).

Player 0 meets with each of the other players $i = 1, ..., n$ to agree on the emission reduction q_i and a transfer t_i (from the leader to i). The leader hence pays a total transfer $T = \sum_{i=1}^{n} t_i$ to the other players in exchange of emission reductions. All the negotiations are assumed to be bilateral and simultaneous with singleton players or with a coalition over two variables, the amount of abatement and the transfers. As explained in the introduction, the leader has no emission reduction target and is the agenda setter. Payoffs to players are $\pi_0 = B(Q) - T$ for the leader and $\pi_i = B(Q) - C(q_i) + t_i$ for i in case of an agreement consisting on emission reductions q_i for all i and transfers t_i from the leader to each of the other players i.

Players i can organize into a coalition of size J to bargain with the leader 0. We consider the case when there is at most one coalition of size $J > 1$. Players i not in the coalition bargain with the leader on their own.

We consider a two-stage game. In stage 1 countries decide whether they prefer to negotiate with the leader alone or in a pool using the internal and external stability conditions. In stage 2, the abatement levels are solution of the bargaining Nash-in-Nash protocol.

2.1 The abatement stage

Two configurations are examined, depending on whether the leader is negotiating facing singletons or a coalition.

2.1.1 The leader against singletons

If there is no coalition of size $J > 1$ then all players are bargaining with the leader on their own. At each bilateral negotiation between i and 0 the Nash-in-Nash solves

$$\max_{(q_i \geq 0, t_i \geq 0)} NP_{0,i} = \left(\pi_0 - \pi_{0,-i}^d\right)\left(\pi_i - \pi_{i,-i}^d\right), \tag{1}$$

where $\pi_{0,-i}^d$ and $\pi_{i,-i}^d$ denote, respectively, the leader's payoff and the other player i's payoff at disagreement. These disagreement payoffs depend on players' beliefs about what will happen if a disagreement is reached. We consider that the leader 0 and the other player i have passive beliefs (McAfee and Schwartz, 1994), meaning that in case of disagreement between 0 and i, those players believe that all other bargaining meetings reach the equilibrium agreement. The Nash Bargaining Solution (NBS) is given by the following proposition.

Proposition 1 *The NBS of the maximization problem (1) between the leader and the singletons give:*

$$t_i = \frac{1}{2}\left(C(q_i) - C(q_i^{BR}(Q_{-i}))\right), \tag{2}$$

and

$$C'(q_i) = 2B'(Q), \tag{3}$$

where $q_i^{BR}(Q_{-i})$ stands for the best reply emission reduction by player i in case of disagreement, given by

$$B'(q_i^{BR}(Q_{-i}) + Q_{-i}) = C'(q_i^{BR}(Q_{-i})). \tag{4}$$

The proof of Proposition 1 can be found in the appendix.

2.1.2 A coalition forms to bargain with the leader

We consider now the case of a coalition of size J, with $1 < J < n$ pooling their meeting with the leader and the rest of non leader players $n - J$ bargaining with the leader on their own. We note the set of players in the coalition $\mathcal{J}$ and the set of players outside the coalition as $N \backslash \mathcal{J}$. Note that $|\mathcal{J}| = J$. At the bilateral negotiation between the coalition $\mathcal{J}$ and 0 the Nash-in-Nash solves

$$\max_{(q_j \geq 0, t_j \geq 0)} NP_{0,\mathcal{J}} = \left(\pi_0 - \pi_{0,-\mathcal{J}}^d\right) \prod_{j \in \mathcal{J}} \left(\pi_j - \pi_{j,-\mathcal{J}}^d\right), \tag{5}$$

where $\pi_{0,-\mathcal{J}}^d$ and $\pi_{j,-\mathcal{J}}^d$ denote, respectively, the leader's payoff and player j's payoff, $j \in \mathcal{J}$, at disagreement.

At each bilateral negotiation between players k outside the coalition $\mathcal{J}$ and 0 the Nash-in-Nash solves

$$\max_{(q_k \geq 0, t_k \geq 0)} NP_{0,k} = \left(\pi_0 - \pi_{0,-k}^d\right) \left(\pi_k - \pi_{k,-k}^d\right), \tag{6}$$

where $\pi_{0,-k}^d$ and $\pi_{k,-k}^d$ denote, respectively, the leader's payoff and player k's payoff, $k \notin \mathcal{J}$, at disagreement.

Note that in case of disagreement when the coalition bargains with the leader all players in the coalition disagree (the coalition must reach unanimity). As argued before, the disagreement payoffs depend on players' beliefs about what will happen if a disagreement is reached. The leader 0 and the other players $j \in \mathcal{J}$ having passive beliefs means that in case of disagreement between 0 and the coalition $\mathcal{J}$, those players believe that all other bargaining meetings (in this case all 0 with k, $k \notin \mathcal{J}$) reach the equilibrium agreement. Furthermore, we assume that in case of disagreement, players $j \in \mathcal{J}$ simultaneously choose their emission q_j taking the decision of the other players in $\mathcal{J}$ as given (as in a Cournot competition). Given that cost functions are identical, the resulting equilibrium emission in case of disagreement is the same for each player $j \in \mathcal{J}$ and will be denoted $q_j^{BR}(Q_{-\mathcal{J}})$. Concerning the bargaining problem between the leader 0 and each of the other players $k \notin \mathcal{J}$, we consider under passive beliefs that players 0 and k believe that all other bargaining meetings reach the equilibrium agreement. We obtain the following proposition.

Proposition 2 *When a coalition forms to bargain with the leader:*

1. The NBS of the maximization problem (5) between the leader and the coalition gives:

$$t_j(J) = \frac{1}{J+1}\left(C(q_j) - C(q_j^{BR}(Q_{-\mathcal{J}}))\right), \tag{7}$$

and

$$C'(q_j) = (J+1)B'(Q), \tag{8}$$

where $q_j^{BR}(Q_{-\mathcal{J}})$ stands for the emission reduction by player j in case of disagreement, given by

$$B'(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) = C'(q_j^{BR}(Q_{-\mathcal{J}})). \tag{9}$$

2. The NBS of the maximization problem (6) between the leader and the singletons give:

$$t_k = \frac{1}{2}\left(C(q_k) - C(q_k^{BR}(Q_{-k}))\right), \tag{10}$$

and

$$C'(q_k) = 2B'(Q), \tag{11}$$

where $q_j^{BR}(Q_{-\mathcal{J}})$ stands for the emission reduction by player j in case of disagreement, given by

$$B'(q_k^{BR}(Q_{-k}) + Q_{-k}) = C'(q_k^{BR}(Q_{-k})). \tag{12}$$

The proof of Proposition 2 can be found in the appendix.

At the end of this bargaining protocol the leader pays a total transfer $\sum_{j\in\mathcal{J}} t_j + \sum_{k\notin\mathcal{J}} t_k$ and obtains

$$\pi_0 = B(Jq_j+(n-J)q_k) - \frac{J}{J+1}\left(C(q_j) - C(q_j^{BR}(Q_{-\mathcal{J}}))\right) - \frac{n-J}{2}\left(C(q_k) - C(q_k^{BR}(Q_{-k}))\right).$$

Each agent $j \in \mathcal{J}$ and each agent $k \notin \mathcal{J}$ respectively obtain a payoff that depends on the size of the coalition J. Because benefit and cost functions are identical across players, we have

$$\begin{aligned} \pi_j(J) &= B(Jq_j + (n-J)q_k) - \frac{1}{J+1}\left(JC(q_j) + C(q_j^{BR}(Q_{-\mathcal{J}}))\right), && (13) \\ \pi_k(J) &= B(Jq_j + (n-J)q_k) - \frac{1}{2}\left(C(q_k) + C(q_k^{BR}(Q_{-k}))\right). && (14) \end{aligned}$$

2.2 The participation stage

After obtaining the abatement levels for players in the second stage of the game and their associated profits, the size of the stable coalition is determined in the first stage by using the internal and external conditions. Denoting $\pi_j(J)$ the payoff of a player belonging to a pool of size J and $\pi_k(J)$ the payoff of a player negotiating alone for an existing pool of size J, a stable coalition J^* must satisfy the two following internal and external stability conditions

$$\begin{aligned} \pi_j(J^*) - \pi_k(J^*-1) &\geq 0, && (15) \\ \pi_k(J^*) - \pi_j(J^*+1) &\geq 0. && (16) \end{aligned}$$

Condition (15) states that a member of the pool of size J^* has no incentive to leave the pool, which will be reduced by one unit, J^*-1, to negotiate alone with the leader. Condition (16) states that a player not belonging to the pool of size J^* has no incentive to join it and move up to size J^*+1.

3 The results

We consider cost and benefit functions such that $C(q) = q^{\alpha_C} C(1)$ homogeneous of degree $\alpha_C > 1$ and $B(Q) = Q^{\alpha_B} B(1)$ homogeneous of degree $\alpha_B < 1$ where $C(1)$ is the abatement cost of one unit and $B(1)$ the abatement benefit of one unit. We first focus on two polar cases a constant marginal benefit equal to b with a general cost function and a constant marginal cost equal to c with a general benefit function. Then we consider an intermediate case where $\alpha_C + \alpha_B = 2$.[5]

3.1 Constant marginal benefit

We consider first the case of a homogeneous cost function $C(q) = q^{\alpha_C} C(1)$ of degree $\alpha_C > 1$ and a benefit function $B(Q) = bQ$ (homogeneous of degree 1). When the players meet with the leader individually (or as singletons) we obtain for each i[6]

$$q_i = 2^{\frac{1}{\alpha_C - 1}} q_i^{BR},\ t_i = \frac{1}{2}\left(2^{\frac{\alpha_C}{\alpha_C - 1}} - 1\right) C(q_i^{BR}),$$

with

$$q_i^{BR}(Q_{-i}) = q^{BR} = \left(\frac{b}{\alpha_C C(1)}\right)^{\frac{1}{\alpha_C - 1}}.$$

As a result, the total emission reduction is equal to $Q = 2^{\frac{1}{\alpha_C - 1}} n q^{BR}$.

When a coalition of size J forms to bargain with the leader, we obtain for each k not in the coalition:

$$q_k = 2^{\frac{1}{\alpha_C - 1}} q^{BR},\ q_k^{BR}(Q_{-k}) = q^{BR},\ t_k = \frac{1}{2}\left(2^{\frac{\alpha_C}{\alpha_C - 1}} - 1\right) C(q^{BR}),$$

while for each j inside the coalition:

$$q_j(J) = (J+1)^{\frac{1}{\alpha_C - 1}} q^{BR},\ q_j^{BR}(Q_{-\mathcal{J}}) = q^{BR},\ t_j(J) = \frac{1}{J+1}\left((J+1)^{\frac{\alpha_C}{\alpha_C - 1}} - 1\right) C(q^{BR}).$$

As a result, the total emission reduction is equal to $Q(J) = Jq_j(J) + (n-J)q_k = \left(J(J+1)^{\frac{1}{\alpha_C - 1}} + (n-J)2^{\frac{1}{\alpha_C - 1}}\right) q^{BR}$.

With these values, we obtain the following proposition.

Proposition 3 *Assume that individual cost functions are homogeneous of degree $\alpha_C > 1$ and the benefit function is homogeneous of degree 1. Then:*

1. *When players bargain individually with the leader (singletons), the payoff of the leader is lower than the other players' payoff when $n \geq 3$ or when $n = 2$ and $\alpha_C < \frac{\ln 3}{\ln 3 - \ln 2}$.*

[5] With both linear cost and benefit functions, there is no interior solution and the problem consists in a threshold public game.

[6] The detailed computations can be found in the appendix.

2. *When a coalition of size J forms to bargain with the leader, the payoff of the singletons is higher than that of the leader and of the players inside the coalition. The payoff of the leader is lower than that of the players inside the coalition when $n - J \geq 2$ or when $J = n - 1$ and $\alpha_C < \frac{\ln 3}{\ln 3 - \ln 2}$. If the grand coalition forms, i.e., $J = n$ then the payoff of the leader is higher than that of all the other players.*

3. *The leader's preferred coalition size is $J = n$ (the grand coalition) while he always prefers a negotiation with a pool of size $J \geq 2$ than a negotiation with singletons.*

4. *The stable size is $J = 2$ for reasonable values of α_C, namely $\alpha_C \in [1.215, 4.059]$. The leader's preferred coalition size (i.e., the grand coalition) can be stable for α_C sufficiently close to 1. For $\alpha_C > 4.059$ no coalition size is internally stable and all non-leader players prefer to negotiate as singletons with the leader.*

The proof of Proposition 3 can be found in the appendix.

With constant marginal benefit, when the leader negotiates against a coalition of size $J \geq 2$ and against the remaining players as singletons, Proposition 3 implies that even though the members of the pool receive more transfers, their payoff is lower because they abate more than the singletons. Furthermore, the leader being the agenda setter always prefers to negotiate with players having the possibility to form a pool, and his preferred pool size is n. The size of the stable coalition depends on the degree of convexity of the cost function. With a quadratic cost function which is usually assumed in the literature, it implies that the stable coalition will be of size 2, as in Barrett's canonical model. However a higher coalition size being stable requires a degree of homogeneity closer to 1 (i.e., a less convex cost function).

Our negotiation model recovers this result of partial cooperation. The members of the pool abate more and even if they received more transfers, their payoffs are lower than the singletons. However they are all better off with respect to the case in which the leader negotiates separately and simultaneously with all the players.

It is interesting to compare our results with Barret (1994). The Barrett (1994) model with n symmetric players having a linear benefit function $B(Q) = bQ$ and a quadratic cost function $C(q) = \frac{c}{2}q^2$ corresponds in our case to a value $\alpha_C = 2$ and $C(1) = c/2$. When the leader bargains against singletons, we obtain $q^{BR} = \frac{b}{c}$, $C(q^{BR}) = \frac{1}{2}\frac{b^2}{c}$ and $t_i = \frac{3}{4}\frac{b^2}{c}$. Payoffs are $\pi_i = \frac{1}{4}(8n-5)\frac{b^2}{c}$ and $\pi_0 = \frac{1}{4}(5n)\frac{b^2}{c}$. Total abatement is $Q = 2n\frac{b}{c}$ and total welfare is $W = 2n^2\frac{b^2}{c}$. The leader obtains a lower payoff than the players as far as $n > 5/3$ (true for $n > 2$). With respect to the non-cooperative case defined as if each country maximizes its own welfare, we obtain $q_i^{nc} = \frac{b}{c}$ and $\pi^{nc} = \left(n - \frac{1}{2}\right)\frac{b^2}{c}$. It can be shown that $\pi_0 > \pi^{nc}$ for $n > 0$ and $\pi_i > \pi^{nc}$ for $n > 1$. When a pool forms to bargain with the leader , the size of the stable pool can be either $J^* = 2$ or $J^* = 3$. For $J^* = 2$, the abatement of an insider is $q_i = 3\frac{b}{c}$ and it is $q_k = 2\frac{b}{c}$ for an outsider.

The global amount of reduction is equal to $Q = 2\,(n+1)\,\frac{b}{c}$. We can determine the amount of utility transfers always for $J^* = 2$ with $t_j = \frac{4}{3}\frac{b^2}{c}$ and $t_k = \frac{3}{4}\frac{b^2}{c}$ implying $t_j > t_k$. The total amount of transfer is $T(J^* = 2) = \frac{1}{12}\frac{b^2}{c}\,(9n + 14)$. In terms of payoffs we obtain $\pi_k(J^* = 2) > \pi_j(J^* = 2) > \pi_0(J^* = 2)$ with $\pi_j(J^* = 2) = \frac{1}{6}\,(12n - 7)\,\frac{b^2}{c}$, $\pi_k(J^* = 2) = \frac{1}{4}\,(8n + 3)\,\frac{b^2}{c}$ and $\pi_0(J^* = 2) = \frac{5}{12}\,(3n + 2)\,\frac{b^2}{c}$. Welfare is $W(J^* = 2) = \left(2n^2 + 2n - 3\right)\frac{b^2}{c}$. We can check that this situation is profitable for all players with respect the situation where all players are singletons, $\pi_j(J^* = 2) - \pi_i = \frac{1}{12}\frac{b^2}{c}$, $\pi_k(J^* = 2) - \pi_i = 2\frac{b^2}{c}$ and for the leader $\pi_0(J^* = 2) - \pi_0^s = \frac{5}{6}\frac{b^2}{c}$. We have that $\pi(J^* = 2) > \pi$ for all players. Total Abatement and welfare are greater with $Q(J^* = 2) - Q = 2\frac{b}{c}$ and $W(J^* = 2) - W = (2n - 3)\,\frac{b^2}{c}$.

3.2 Constant marginal cost

We consider now the case of a homogeneous benefit function $B(Q) = Q^{\alpha_B} B(1)$ of degree $\alpha_B < 1$ and the cost function $C(q) = cq$ (homogeneous of degree 1). When the players meet with the leader individually (or as singletons) we obtain for each i[7]

$$q_i = \frac{1}{n}\left(\frac{2B(1)\alpha_B}{c}\right)^{\frac{1}{1-\alpha_B}} = q^s,\ t_i = \frac{c}{2}q^s,$$

with

$$q_i^{BR}(Q_{-i}) = 0.$$

As a result, the total emission reduction is equal to $Q^s = nq^s$.

When a coalition of size J forms to bargain with the leader, we obtain for each k not in the coalition:

$$q_k = q_k^{BR} = 0,\ t_k = 0,$$

while for each j inside the coalition:

$$\begin{aligned} q_j(J) &= \frac{n}{J}\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} q^s,\ q_j^{BR}(J) = \frac{n}{J}\left(\frac{1}{2}\right)^{\frac{1}{1-\alpha_B}} q^s, \\ t_j(J) &= \frac{c\ n}{J(J+1)} q^s \frac{(J+1)^{\frac{1}{1-\alpha_B}} - 1}{2^{\frac{1}{1-\alpha_B}}}. \end{aligned}$$

As a result, the total emission reduction is equal to $Q(J) = nq^s\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}$.

With these values, we obtain the following proposition.

Proposition 4 *Assume that the benefit function is homogeneous of degree $\alpha_B <$ 1 and that the cost function is homogeneous of degree 1. Then:*

1. *When players bargain individually with the leader (singletons), the payoff of the leader is lower than the other players' payoff .*

[7] The detailed computations can be found in the appendix.

2. *When a coalition of size J forms to bargain with the leader, the payoff of the singletons is higher than that of the leader and of the players inside the coalition. The payoff of the leader is higher than that of the players inside the coalition.*

3. *The leader's preferred coalition size is $J = n$ (the grand coalition) while he always prefers a negotiation with a pool of size $J \geq 2$ than a negotiation with singletons.*

4. *The stable size is $J = 2$ for $\alpha_B \in [0.711563, 0.880902]$. The leader's preferred coalition size (i.e., the grand coalition) can be stable for α_B sufficiently close to 1. For $\alpha_B < 0.711563$ no coalition size is internally stable and all non-leader players prefer to negotiate as singletons with the leader.*

The proof of Proposition 4 can be found in the appendix.

With constant marginal cost, when the leader negotiates against a coalition of size $J \geq 2$ and against the remaining players as singletons, only the members of the pool abate while the singletons fully free-ride and get higher payoffs. Note that the leader being the agenda setter will always prefer to negotiate with players having the possibility to form a pool, as was the case with the constant marginal benefit.

All coalition sizes are stable, from an empty coalition to a grand coalition, depending on the degree of concavity of the benefit function. A higher coalition size being stable requires a degree of homogeneity closer to 1 (i.e., a less concave benefit function). However if we assume a square-root benefit function $\alpha_B = 0.5$, ie $B(Q) = bQ^{0.5}$ with $B(1) = b$, there will be no stable coalition meaning all players will prefer to bargain individually with the leader to avoid the strong free-riding outside the coalition. In that case the amount of reduction of the singletons will be $q_i = \frac{1}{n}\left(\frac{b}{c}\right)^2$ and $q_0^s = 0$ for the leader. The transfer obtained by a singleton is $t_i = \frac{1}{2n}\frac{b^2}{c}$ leading to the payoffs $\pi_i = \frac{2n-1}{2n}\frac{b^2}{c}$ and $\pi_0 = \frac{1}{2}\frac{b^2}{c}$.

Now assume a higher value for α_B with $\alpha_B = \frac{4}{5}$, ie $B(Q) = bQ^{\frac{4}{5}}$ which ensures that a stable coalition exists with $J^* = 2$. In that case only the insiders reduce their emissions by an amount $q_j = \frac{1}{2}\left(\frac{12}{5}\right)^5\left(\frac{b}{c}\right)^5$ while the outsiders are doing nothing. The transfer received by the insiders is $t_j = \frac{121}{3}\left(\frac{4}{5}\right)^5\frac{b^5}{c^4}$ implying the following payoffs $\pi_j(J^* = 2) = \frac{241}{15}\left(\frac{4}{5}\right)^4\frac{b^5}{c^4}$, $\pi_k(J^* = 2) = 3^4\left(\frac{4}{5}\right)^4\frac{b^5}{c^4}$ and $\pi_0(J^* = 2) = \frac{976}{15}\left(\frac{4}{5}\right)^4\frac{b^5}{c^4}$ with $\pi_k > \pi_0 > \pi_j$. With respect to the singleton case, we obtain $q_i = \frac{1}{n}\left(\frac{8}{5}\right)^5\left(\frac{b}{c}\right)^5$, $t_i = \frac{1}{2}\frac{1}{n}\left(\frac{8}{5}\right)^5\frac{b^5}{c^4}$, $\pi_i = \frac{1}{n}\frac{b^5}{c^4}2^4\left(\frac{4}{5}\right)^4\left(n - \frac{4}{5}\right)$ and $\pi_0 = \frac{1}{5}\frac{b^5}{c^4}2^4\left(\frac{4}{5}\right)^4$. It can be show that $\pi_0(J^* = 2) > \pi_0$.

3.3 Intermediate case

We consider now the case of a homogeneous benefit function $B(Q) = Q^{\alpha_B}B(1)$ of degree $\alpha_B < 1$ and a homogeneous cost function $C(q) = q^{\alpha_C}B(1)$ of degree $\alpha_C = 2 - \alpha_B > 1$ to keep the computation tractable.

When the players meet with the leader individually, we obtain for each i

$$q_i = \left(\frac{2\alpha_B}{2-\alpha_B}\frac{B(1)}{C(1)}\right)^{\frac{1}{2(1-\alpha_B)}} n^{-\frac{1}{2}},\ Q = nq_i.$$

The best reply q_i^{BR} is given by

$$q_i^{BR} = \frac{1}{2}\left(\sqrt{\left(\left(\frac{n-1}{\sqrt{n}}\right)^2 + 2^{\frac{2\alpha_B-1}{\alpha_B-1}}\right)}n - (n-1)\right) q_i < q_i.$$

When a coalition of size J forms to bargain with the leader, we obtain for each j inside the coalition and each k not in the coalition

$$q_j = \left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} q_k,\ q_k = \left(\frac{n}{J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J}\right)^{\frac{1}{2}} q_i$$

The associated best-replies are

$$q_j^{BR} = \frac{\sqrt{(n-J)^2 + 2^{\frac{2\alpha_B-1}{\alpha_B-1}} J\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J\right)} - (n-J)}{2^{-\frac{\alpha_B}{1-\alpha_B}} J\left(J+1\right)^{\frac{1}{1-\alpha_B}}} q_j$$

$$q_k^{BR} = \frac{1}{2}\left(\begin{array}{c}\sqrt{\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J - 1\right)^2 + 2^{\frac{2\alpha_B-1}{\alpha_B-1}}\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J\right)} \\ -\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J - 1\right)\end{array}\right) q_k$$

It can be shown that $0 < q_j^{BR} < q_j$ and $0 < q_k^{BR} < q_k$.

After computing the associated payoffs and the internal stability condition $\pi_j(J) - \pi_k(J-1)$, we obtain the following proposition

Proposition 5 *Assume that the benefit function is homogeneous of degree $\alpha_B < 1$ and that the cost function is homogeneous of degree $\alpha_C = 2 - \alpha_B > 1$. Then:*

1. *For $\alpha_B = 1/2$ and $\alpha_C = 3/2$, the coalition is always empty and the leader negotiates with only singletons*

2. *For $\alpha_B = 3/4$ and $\alpha_C = 5/4$, only the grand coalition is stable*

The proposition shows that high values of α_C and low values of α_B imply an empty coalition. Conversely, low values of α_C and high values of α_B lead to the grand coalition. Under the constraint $\alpha_C + \alpha_B = 2$ necessary for an explicit calculation of the best replies, we were unable to obtain intermediate cases of partial coalition as in the two extreme cases.

To compare our results with Barrett (1994), we have considered the following benefit and cost functions $C(q) = \frac{c}{2}(q)^2$ and $B(Q) = \frac{b}{n}\left(aQ - \frac{1}{2}Q^2\right)$. The amount of abatement of an insider and an outsider are $q_j = \frac{(J+1)a}{J^2-J+2n+n\gamma}$ and $q_k = \frac{2a}{J^2-J+2n+n\gamma}$ with $\gamma = c/b$. Their best replies are $q_j^{BR} = \frac{J^2+J+n\gamma}{(J+n\gamma)(J+1)} q_j$ and $q_k^{BR} = \frac{(2+n\gamma)}{2(n\gamma+1)} q_k$. After computing the payoffs $\pi_j(J)$ and $\pi_k(J)$, it can be shown that the function $\pi_j(J) - \pi_k(J-1)$ is decreasing in J meaning that the he coalition is always empty and the leader negotiates with only singletons.

4 Extensions

4.1 The IEA bargaining model with symmetric beliefs

The assumption of symmetric beliefs states that in case of disagreement in one negotiation, players belief that all other bargaining meetings result in a disagreement as well and not reach the equilibrium agreements as it was the case with passive beliefs. More precisely, given the emission reduction by the other players in case of disagreement, denoted by $Q_{-i}^d = \sum_{j \neq i} q_j^d$, player i optimally chooses q_i. Player i thinks that all the other players $j \neq i$, will disagree and anticipates that they will choose optimally and simultaneously their emission reduction q_j^d. Hence q_i^d is given by the system of equations

$$q_i^d(Q_{-i}^d) = arg \max_{q_i \geq 0} B(q_i + Q_{-i}^d) - C(q_i),$$

for each i. The first-order condition of this maximization problem (interior solution) is given by

$$B'(q_i^d(Q_{-i}^d) + Q_{-i}^d) = C'(q_i^d(Q_{-i}^d)),$$

while the second-order condition is satisfied by concavity of B and convexity of C. Given that players have identical benefit and cost functions, $q_i^d = q_j^d$ whenever i and j are non leaders. The first-order condition stated just above can be rewritten as

$$B'(nq_i^d(Q_{-i}^d)) = C'(q_i^d(Q_{-i}^d)). \tag{17}$$

Propositions 1 and 2 remain the same but with a different disagreement payoff $q_i^d(Q_{-i}^d)$ instead of $q_i^{BR}(Q_{-i})$.

It gives the following proposition.

Proposition 6 *Assume that the benefit function is homogeneous of degree $\alpha_B \leq 1$ and that the cost function is homogeneous of $\alpha_C \geq 1$, we obtain that*

1. *When $\alpha_B = 1$ the results with symmetric beliefs are the same than with passive beliefs*

2. *When $\alpha_C = 1$ an internally stable coalition with passive beliefs implies internally stable with symmetric beliefs while an externally stable coalition with symmetric beliefs implies externally stable with passive beliefs*

3. *Depending the values of $\alpha_B < 1$ and $\alpha_C > 1$, the coalition can be empty or the grand coalition can be achieved.*

The proof of Proposition 6 can be found in the appendix.

The first part of the proposition shows that when we assume that a constant marginal benefit, our results remain the same independently of the beliefs in case of disagreement since marginal benefit does not depend on what other players are deciding. However in the other cases, results are different from the passive beliefs assumption. In comparison with the case of a constant marginal cost, we can show that when the leader bargains with singletons his payoff is higher with symmetric beliefs than passive beliefs since he has to pay a higher amount of transfers to the singletons in the passive beliefs case while the total abatement is the same in both cases. This situation is the opposite for the singletons. The interesting feature of the constant marginal cost case is that only the members of the coalition reduce their emission while the singletons not in the pool fully free ride. In that case we can show that the range of the coalition sizes can vary from an empty coalition, to a partial coalition up to the grand coalition according to the values of α_B. The second part of the proposition also show that if $J^* = n$ for passive beliefs, then $J^* = n$ for symmetric beliefs. If no coalition of at least two countries is stable with symmetric beliefs, then the same is true with passive beliefs. Moreover in the general case with $\alpha_B < 1$ and $\alpha_C > 1$, the grand coalition can be achieved for $\alpha_C = 3/2$, $\alpha_B = 1/2$, $n = 200$ contrary to the passive beliefs case and also for $\alpha_C = 5/4$, $\alpha_B = 3/4$, $n = 200$. For $\alpha_C = 1.9$ and $\alpha_B = 0.8$ the coalition is empty and all players prefer to negociate alone with the leader.

4.2 A Stackelberg bargaining setting

We have to analysed two configurations depending the leader bargains first or second with the coalition. We consider only passive beliefs.

Proposition 7 *When the leader bargains first with the coalition and second with the singletons,*

1. *The NBS between the leader and the singleton give*

$$t_k = \frac{1}{2}\left(C(q_k) - C(q_k^{BR}(Q_{\mathcal{J}}))\right), \tag{18}$$

$$C'(q_k(Q_{\mathcal{J}})) = 2B'(Kq_k(Q_{\mathcal{J}}) + Q_{\mathcal{J}}), \tag{19}$$

with

$$B'(q_k^{BR}(Q_{\mathcal{J}}) + (K-1)q_k(Q_{\mathcal{J}}) + Q_{\mathcal{J}}) = C'(q_k^{BR}(Q_{\mathcal{J}})). \tag{20}$$

2. *The NBS between the leader and the coalition give*

$$t_j = \frac{1}{J+1}\left(C(q_j) - C(q_j^{BR})\right), \tag{21}$$

$$C'(q_j) = (J+1)B'(Kq_k(Q_{\mathcal{J}}) + Q_{\mathcal{J}})\left(1 + K\frac{d\ q_k}{d\ Q_{\mathcal{J}}}\right), \tag{22}$$

where

$$\frac{dq_k}{dQ_{\mathcal{J}}} = \frac{2B''(Kq_k(Q_{\mathcal{J}}) + Q_{\mathcal{J}})}{C''(q_k(Q_{\mathcal{J}})) - 2KB''(Kq_k(Q_{\mathcal{J}}) + Q_{\mathcal{J}})} \leq 0,$$

and

$$B'(q_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}} + Kq_k(q_j + Q_{-\mathcal{J}})) = C'(q_j^{BR}(Q_{-\mathcal{J}})). \tag{23}$$

The proof of Proposition 6 can be found in the appendix.

At stage 1 the leader bargains with the coalition. At stage 2 the leader bargains with the singletons.

Proposition 8 *When the leader bargains first with the singletons and second with the coalition,*

1. *The NBS between the leader and the coalition give*

$$t_j = \frac{1}{J+1}\left(C(q_j) - C(q_j^{BR}(Q_{\mathcal{K}}))\right), \tag{24}$$

$$C'(q_j(Q_{\mathcal{K}})) = (J+1)B'(Jq_j(Q_{\mathcal{K}}) + Q_{\mathcal{K}}), \tag{25}$$

with

$$B'(Jq_j^{BR}(Q_{\mathcal{K}}) + Q_{\mathcal{K}}) = C'(q_j^{BR}(Q_{\mathcal{K}})). \tag{26}$$

2. *The NBS between the leader and the singletons give*

$$t_k = \frac{1}{2}\left(C(q_k) - C(q_k^{BR})\right), \tag{27}$$

$$C'(q_k) = 2B'(Kq_k + Jq_j(Kq_k))\left(1 + J\frac{dq_j}{dQ_{\mathcal{K}}}\right), \tag{28}$$

where

$$\frac{dq_j}{dQ_{\mathcal{K}}} \leq 0,$$

and

$$B'(q_k^{BR} + \sum_{h\in\mathcal{K}, h\neq k} q_h + \sum_{j\in\mathcal{J}} q_j\left(q_k + \sum_{h\in\mathcal{K}, h\neq k} q_h\right)) = C'(q_k^{BR}).$$

The proof of Proposition 7 can be found in the appendix.

If we consider the two polar cases according the shape of the benefit and cost function, it is immediate to see that given that marginal benefit does not depend on what other players are deciding, all quantities are like in the simultaneous case and all partial derivatives are zero. Hence we obtain on the other polac case with the following Proposition

Proposition 9 *Assume that the benefit function is homogeneous of degree $\alpha_B < 1$ and that the cost function is homogeneous of degree 1. Then:*

1. *The players involved in the first stage of the Stackelberg protocol contributes nothing but when the coalition bargains in the second stage his contribution will be larger than the one made by singletons also in the second stage*

2. *With passive beliefs the leader is indifferent between the simultaneous protocol and the stackelberg protocol when coalition bargains in the second stage*

3. *With passive beliefs the stability conditions are the same in the simultaneous protocol and the stackelberg protocol when coalition bargains in the second stage*

The proof of Proposition 8 can be found in the appendix.

5 Conclusion

The aim of this article was to reshape the IEA game and reinterpret its results on the optimal size of stable coalition. While the structure of the IEA is based on an abatement stage and a participation stage, we have replaced the first stage by an explicit negociation game à la Nash-in-Nash. Countries are supposed to negotiate simultaneously and bilaterally with a leader who is represented by the secretariat responsible for redistributing the gains from cooperation through appropriate transfers in exchange of emission reductions. This leader has no coercive power. Countries have the choice to negotiate alone or in a pool with the leader. As this is a negotiation game, it's important to consider how the players will behave in the event of a negotiation failure.

Our contribution shows more and new configurations in terms of size of coalition and abatement effort strategies. The size of the coalition can be from 0 to the grand coalition depending the values of the degree of homogeneity concerning the benefit and cost functions. With constant benefit and a quadratic cost function, the size of a stable coalition is equal to 2 as the well-known result in the IEA literature. However the size of the coalition can increase when the degree of homogeneity of the cost function is closed to one and/or the degree of homogeneity of the benefit function is also closed to one. Depending the respective values of these parameters several size of inetrmediate coalition can emerge. The bargaining approach also shows that the configuration of a constant marginal cost and a general benefit function creates the highest free riding behavior from the singletons which are doing nothing in terms of abatement. Only the members of the coalition reduce their emissions. This is a key difference with the IEA game à la Cournot-Nash because by definition all players are realising a positive effort.

Our results are robust to the type of player beliefs in the event of failed negotiations. Symmetrical beliefs imply that if one bilateral negotiation fails, all others will also fail, whereas in the case of passive beliefs, a failure does not impede all other negotiations from reaching an agreement. We have also assumed the case where the leader can decide to negotiate first or second with the pool, à la Stackelberg. Our results show that it is always in the leader's interest to negotiate simultaneously.

6 References

d'Aspremont, C. A., Jacquemin, J., Gabszeweiz, J. and Weymark, J. A. (1983) On the Stability of Collusive Price Leadership, *Canadian Journal of Economics* 16: 17-25.

Barrett, S., (1994) Self-Enforcing international environmental agreements, *Oxford Economic Papers*, 46:878–894.

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

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

Collard-Wexler, A., G. Gowrisankaran, and R. S. Lee (2019) 'Nash-in-Nash' bargaining: A microfoundation for applied work, *Journal of Political Economy*, 27(1):163–195

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

Finus, M., F. Furini and A.V. Rohrer (2021) The efficacy of international environmental agreements when adaptation matters: Nash-Cournot vs Stackelberg leadership, *Journal of Environmental Economics and Management*, 109:102461

McAfee, P., Schwartz, M., (1994). Opportunism in multilateral vertical contracting: nondiscrimination, exclusivity, and uniformity. *American Economic Review*, 84:210–230.

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

Okada, A. (2022) Dynamic bargaining with voluntary participation and externality, *Economic Theory*, 75:427-452

Okada, A. (2023) A Dynamic Climate Negotiation Game Achieving Full Cooperation, *Journal of Economic Behavior and Organization*, 214:657-669

Appendix

Proof of Proposition 1

In case of of disagreement between 0 and i, we assume that player i optimally chooses q_i, given the emission reduction by the other players in equilibrium, denoted by $Q_{-i} = \sum_{j \neq i} q_j$. We denote by $q_i^{BR}(Q_{-i})$ the best reply emission reduction by player i. By definition,

$$q_i^{BR}(Q_{-i}) = arg \max_{q_i \geq 0} B(q_i + Q_{-i}) - C(q_i).$$

The first-order condition of this maximization problem, assuming $q_i^{BR}(Q_{-i}) > 0$, i.e., strictly positive emission reductions in case of disagreement, is given by

$$B'(q_i^{BR}(Q_{-i}) + Q_{-i}) = C'(q_i^{BR}(Q_{-i})), \tag{29}$$

while the second-order condition is satisfied by concavity of B and convexity of C. Furthermore, in case of disagreement $t_i = 0$, so we have

$$\begin{aligned} \pi_{0,-i}^d &= B(q_i^{BR}(Q_{-i}) + Q_{-i}) - T_{-i}, \\ \pi_{i,-i}^d &= B(q_i^{BR}(Q_{-i}) + Q_{-i}) - C(q_i^{BR}(Q_{-i})). \end{aligned}$$

The leader keeps on paying a transfer to the other players, but not to the one with whom negotiations failed. With this in mind, net payoffs are

$$\begin{aligned} \pi_0 - \pi_{0,-i}^d &= B(Q) - B(q_i^{BR}(Q_{-i}) + Q_{-i}) - t_i, \\ \pi_i - \pi_{i,-i}^d &= B(Q) - B(q_i^{BR}(Q_{-i}) + Q_{-i}) + t_i - C(q_i) + C(q_i^{BR}(Q_{-i})). \end{aligned}$$

Recall that the Nash bargaining solution is found as the solution to the following maximization problem:

$$\max_{(q_i \geq 0, t_i \geq 0)} NP_{0,i} = \left(\pi_0 - \pi_{0,-i}^d\right)\left(\pi_i - \pi_{i,-i}^d\right).$$

The FOCs of this maximization problem are for each $i \in N$:

$$\begin{aligned} \tfrac{\partial NP_{i,0}}{\partial q_i} &\leq 0 \text{ with } q_i \tfrac{\partial NP_{i,0}}{\partial q_i} = 0, \\ \tfrac{\partial NP_{i,0}}{\partial t_i} &\leq 0 \text{ with } t_i \tfrac{\partial NP_{i,0}}{\partial t_i} = 0. \end{aligned}$$

The SOC of this maximization problem is satisfied by concavity of B and convexity of C. Assuming $q_i > 0$ and $t_i > 0$ at the solution, we must have $\frac{\partial NP_{i,0}}{\partial q_i} = 0$ and $\frac{\partial NP_{i,0}}{\partial t_i} = 0$. Note that

$$\frac{\partial NP_{i,0}}{\partial q_i} = 0 \leftrightarrow NP_{i,0}\left(\frac{B'(Q)}{\pi_0 - \pi_{0,-i}^d} + \frac{B'(Q) - C'(q_i)}{\pi_i - \pi_{i,-i}^d}\right) = 0,$$

and

$$\frac{\partial NP_{i,0}}{\partial t_i} = 0 \leftrightarrow NP_{i,0}\left(\frac{-1}{\pi_0 - \pi_{0,-i}^d} + \frac{1}{\pi_i - \pi_{i,-i}^d}\right) = 0.$$

Recall that $NP_{i,0} > 0$ because by assumption $\pi_0 - \pi_{0,-i}^d > 0$ and $\pi_i - \pi_{i,-i}^d > 0$. The latter condition, i.e., $\frac{\partial NP_{i,0}}{\partial t_i} = 0$ holds if and only if $\pi_0 - \pi_{0,-i}^d = \pi_i - \pi_{i,-i}^d$. This equality has two implications. First,

$$B(Q) - B(q_i^{BR}(Q_{-i}) + Q_{-i}) - t_i = B(Q) - B(q_i^{BR}(Q_{-i}) + Q_{-i}) + t_i - C(q_i) + C(q_i^{BR}(Q_{-i})),$$

if and only if $t_i = \frac{1}{2}\left(C(q_i) - C(q_i^{BR}(Q_{-i}))\right)$. Second, $\frac{\partial NP_{i,0}}{\partial q_i} = 0$ if and only if $2B'(Q) - C'(q_i) = 0$. ■

Proof of Proposition 2

Given the emission reduction by the players not in $\mathcal{J}$, denoted by $Q_{-\mathcal{J}} = \sum_{k \notin \mathcal{J}} q_k$, each player j optimally (and simultaneously) chooses q_j. For each possible decision of the other players $i \neq j, i \in \mathcal{J}$, player j has a best-response in terms of emission reductions in case of disagreement, that we denote by $q_j^{BR}(\sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}})$. By definition,

$$q_j^{BR}(\sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}}) = arg \max_{q_j \geq 0} B(q_j + \sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}}) - C(q_j). \tag{30}$$

The FOC of this maximization problem assuming that $q_j^{BR}(\sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}}) > 0$ (interior solution) is given by

$$B'(q_j^{BR}(\sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}}) + \sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}}) = C'(q_j^{BR}(\sum_{i \neq j, i \in \mathcal{J}} q_i + Q_{-\mathcal{J}})), \tag{31}$$

while the SOC is satisfied by concavity of B and convexity of C. Given that players different from the leader have identical benefit and cost functions the equilibrium emission at disagreement, denoted $q_j^{BR}(Q_{-\mathcal{J}})$, satisfies that $q_j^{BR}(Q_{-\mathcal{J}}) = q_i^{BR}(Q_{-\mathcal{J}})$ whenever i and j belong to the coalition $\mathcal{J}$. Furthermore, at equilibrium in case of disagreement all players $j \in \mathcal{J}$ choose an optimal emission reduction. Hence, from the first-order condition stated just above (31) and the symmetry condition just discussed we obtain:

$$B'(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) - C'(q_j^{BR}(Q_{-\mathcal{J}})) = 0. \tag{32}$$

Recall that in case of disagreement $t_j = 0$ for all $j \in \mathcal{J}$. We have

$$\begin{aligned} \pi_{0,-\mathcal{J}}^d &= B(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) - T_{-\mathcal{J}}, \\ \pi_{j,-\mathcal{J}}^d &= B(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) - C(q_j^{BR}(Q_{-\mathcal{J}})). \end{aligned}$$

With this in mind, net payoffs are

$$\begin{aligned} \pi_0 - \pi_{0,-\mathcal{J}}^d &= B(Q) - B(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) - \sum_{i \in \mathcal{J}} t_i, \\ \pi_j - \pi_{j,-\mathcal{J}}^d &= B(Q) - B(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) + t_j - C(q_j) + C(q_j^{BR}(Q_{-\mathcal{J}})). \end{aligned}$$

Recall that the Nash bargaining solution for the meeting between the leader and the coalition is found as the solution to the following maximization problem:

$$\max_{(q_j \geq 0, t_j \geq 0)} NP_{0,\mathcal{J}} = \left(\pi_0 - \pi_{0,-\mathcal{J}}^d\right) \prod_{j \in \mathcal{J}} \left(\pi_j - \pi_{j,-\mathcal{J}}^d\right),$$

The FOCs of this maximization problem are for each $j \in \mathcal{J}$:

$$\begin{aligned} \tfrac{\partial NP_{0,\mathcal{J}}}{\partial q_j} &\leq 0 \text{ with } q_j \tfrac{\partial NP_{0,\mathcal{J}}}{\partial q_j} = 0, \\ \tfrac{\partial NP_{0,\mathcal{J}}}{\partial t_j} &\leq 0 \text{ with } t_j \tfrac{\partial NP_{0,\mathcal{J}}}{\partial t_j} = 0. \end{aligned}$$

The SOC of this maximization problem is satisfied by concavity of B and convexity of C. Assuming $q_j > 0$ and $t_j > 0$ at the solution, we must have $\frac{\partial NP_{0,\mathcal{J}}}{\partial q_j} = 0$ and $\frac{\partial NP_{0,\mathcal{J}}}{\partial t_j} = 0$. Note that

$$\frac{\partial NP_{0,\mathcal{J}}}{\partial q_j} = 0 \leftrightarrow NP_{0,\mathcal{J}} \left(B'(Q) \left\{ \frac{1}{\pi_0 - \pi^d_{0,-\mathcal{J}}} + \sum_{i \neq j, i \in \mathcal{J}} \frac{1}{\pi_i - \pi^d_{i,-\mathcal{J}}} \right\} + \frac{B'(Q) - C'(q_j)}{\pi_j - \pi^d_{j,-\mathcal{J}}} \right) = 0,$$

and

$$\frac{\partial NP_{0,\mathcal{J}}}{\partial t_j} = 0 \leftrightarrow NP_{0,\mathcal{J}} \left(\frac{-1}{\pi_0 - \pi^d_{0,-\mathcal{J}}} + \frac{1}{\pi_j - \pi^d_{j,-\mathcal{J}}} \right) = 0.$$

Given that players all have the same benefit and cost functions, we concentrate on the symmetric solution where $q_j = q_i$ and $t_j = t_i$ for any i and j in $\mathcal{J}$. Again recall that $NP_{0,\mathcal{J}} > 0$ because by assumption $\pi_0 - \pi^d_{0,-\mathcal{J}} > 0$ and $\pi_j - \pi^d_{j,-\mathcal{J}} > 0$ for every $j \in \mathcal{J}$. The latter condition, i.e., $\frac{\partial NP_{0,\mathcal{J}}}{\partial t_j} = 0$ holds if and only if $\pi_0 - \pi^d_{0,-\mathcal{J}} = \pi_j - \pi^d_{j,-\mathcal{J}}$. Given that such condition must hold for all $j \in \mathcal{J}$ we find two implications. First,

$$\begin{aligned} B(Q) - B(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) - Jt_j &= B(Q) - B(Jq_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}}) \\ &\quad + t_j - C(q_j) + C(q_j^{BR}(Q_{-\mathcal{J}})), \end{aligned}$$

if and only if $t_j = \frac{1}{J+1}\left(C(q_j) - C(q_j^{BR}(Q_{-\mathcal{J}}))\right)$. Second, $\frac{\partial NP_{0,\mathcal{J}}}{\partial q_j} = 0$ if and only if $(J+1)B'(Q) - C'(q_j) = 0$.

Fix $k \notin \mathcal{J}$. Given the emission reduction by the other players in equilibrium (i.e., the ones inside the coalition $\mathcal{J}$ and the ones outside the coalition except k), denoted by $Q_{-k} = \sum_{i \neq k} q_i$, player k optimally chooses q_k. We denote by $q_k^{BR}(Q_{-k})$ the emission reduction by player k in case of disagreement, given by

$$q_k^{BR}(Q_{-k}) = arg \max_{q_k \geq 0} B(q_k + Q_{-k}) - C(q_k). \tag{33}$$

The FOC of this maximization problem assuming that $q_k^{BR}(Q_{-k}) > 0$ (interior solution) is given by

$$B'(q_k^{BR}(Q_{-k}) + Q_{-k}) - C'(q_k^{BR}(Q_{-k})) = 0. \tag{34}$$

while the SOC is satisfied by concavity of B and convexity of C. Again, in case of disagreement $t_k = 0$. With this in mind, net payoffs are

$$\begin{aligned} \pi_0 - \pi^d_{0,-k} &= B(Q) - B(q_k^{BR}(Q_{-k}) + Q_{-k}) - t_k, \\ \pi_k - \pi^d_{k,-k} &= B(Q) - B(q_k^{BR}(Q_{-k}) + Q_{-k}) + t_k - C(q_k) + C(q_k^{BR}(Q_{-k})). \end{aligned}$$

Recall that the Nash bargaining solution for the meeting between the leader and each of the singletons is found as the solution to the following maximization problem:

$$\max_{(q_k \geq 0, t_k \geq 0)} NP_{0,k} = \left(\pi_0 - \pi^d_{0,-k}\right)\left(\pi_k - \pi^d_{k,-k}\right),$$

The FOCs of this maximization problem are:

$$\frac{\partial NP_{0,k}}{\partial q_k} \leq 0 \text{ with } q_k \frac{\partial NP_{0,k}}{\partial q_k} = 0,$$
$$\frac{\partial NP_{0,k}}{\partial t_k} \leq 0 \text{ with } t_k \frac{\partial NP_{0,k}}{\partial t_k} = 0.$$

The SOC of this maximization problem is satisfied by concavity of B and convexity of C. Assuming $q_k > 0$ and $t_k > 0$ at the solution, we must have $\frac{\partial NP_{0,k}}{\partial q_k} = 0$ and $\frac{\partial NP_{0,k}}{\partial t_k} = 0$. Note that

$$\frac{\partial NP_{0,k}}{\partial q_k} = 0 \leftrightarrow NP_{0,k} \left(\frac{B'(Q)}{\pi_0 - \pi^d_{0,-k}} + \frac{B'(Q) - C'(q_k)}{\pi_k - \pi^d_{k,-k}} \right) = 0,$$

and

$$\frac{\partial NP_{0,k}}{\partial t_k} = 0 \leftrightarrow NP_{0,k} \left(\frac{-1}{\pi_0 - \pi^d_{0,-k}} + \frac{1}{\pi_k - \pi^d_{k,-k}} \right) = 0.$$

Again, $NP_{0,k} > 0$ because by assumption $\pi_0 - \pi^d_{0,-k} > 0$ and $\pi_k - \pi^d_{k,-k} > 0$. The latter condition, i.e., $\frac{\partial NP_{0,k}}{\partial t_k} = 0$ holds if and only if $\pi_0 - \pi^d_{0,-k} = \pi_k - \pi^d_{k,-k}$. This condition has two implications. First,

$$B(Q) - B(q_k^{BR}(Q_{-k}) + Q_{-k}) - t_k = B(Q) - B(q_k^{BR}(Q_{-k}) + Q_{-k}) + t_k - C(q_k) + C(q_k^{BR}(Q_{-k}))$$

if and only if $t_k = \frac{1}{2}\left(C(q_k) - C(q_k^{BR}(Q_{-k}))\right)$. Second, $\frac{\partial NP_{0,k}}{\partial q_k} = 0$ if and only if $2B'(Q) - C'(q_k) = 0$. ■

Proof of Proposition 3

We start by proving part 1 of Proposition 3. From Proposition 1 and the FOC given by (29) we have

$$b = C'(q_i^{BR}(Q_{-i})).$$

From the cost function $C(q) = q^{\alpha_C} C(1)$, it is easy to see using $C'(q) = \alpha_C C(q)/q$ that for any i

$$q_i^{BR}(Q_{-i}) = \left(\frac{b}{\alpha_C C(1)} \right)^{\frac{1}{\alpha_C - 1}}.$$

Let us fix the notation as in the main text $q^{BR} = \left(\frac{b}{\alpha_C C(1)} \right)^{\frac{1}{\alpha_C - 1}}$ because the best response does not depend on the decision of the other players when marginal benefit is constant. When the leader bargains with each of all the other players we obtain for each i

$$q_i = \left(\frac{2b}{\alpha_C C(1)} \right)^{\frac{1}{\alpha_C - 1}} = 2^{\frac{1}{\alpha_C - 1}} q^{BR}.$$

From (2), the transfer paid by the leader to each player is

$$t_i = \frac{1}{2} \left(C\left(2^{\frac{1}{\alpha_C - 1}} q^{BR}\right) - C(q^{BR}) \right),$$

and as $C\left(2^{\frac{1}{\alpha_C-1}}q^{BR}\right) = 2^{\frac{\alpha_C}{\alpha_C-1}}C(q^{BR})$ we have

$$t_i = \frac{1}{2}\left(2^{\frac{\alpha_C}{\alpha_C-1}} - 1\right) C(q^{BR}).$$

Payoffs are

$$\begin{aligned}
\pi_i &= \left(2^{\frac{1}{\alpha_C-1}}(n\alpha_C - 1) - \frac{1}{2}\right) C(q^{BR}), \\
\pi_0 &= \left(2^{\frac{1}{\alpha_C-1}}(n\alpha_C - n) + \frac{n}{2}\right) C(q^{BR}).
\end{aligned}$$

It is easy to see that $\pi_i > \pi_0$ if and only if $2^{\frac{1}{\alpha_C-1}}(n\alpha_C - 1) - \frac{1}{2} > 2^{\frac{1}{\alpha_C-1}}(n\alpha_C - n) + \frac{n}{2}$. After some rearrangements, this last condition gives $2^{\frac{\alpha_C}{\alpha_C-1}} > \frac{n+1}{n-1}$. This condition can be rewritten as $n > \frac{1+2^{\frac{\alpha_C}{\alpha_C-1}}}{2^{\frac{\alpha_C}{\alpha_C-1}}-1}$. Note that $\frac{1+2^{\frac{\alpha_C}{\alpha_C-1}}}{2^{\frac{\alpha_C}{\alpha_C-1}}-1} < 3$ because $2 < 2^{\frac{\alpha_C}{\alpha_C-1}}$. So if $n \geq 3$ then $n > \frac{1+2^{\frac{\alpha_C}{\alpha_C-1}}}{2^{\frac{\alpha_C}{\alpha_C-1}}-1}$. If $n = 2$ then the inequality becomes $2 > \frac{1+2^{\frac{\alpha_C}{\alpha_C-1}}}{2^{\frac{\alpha_C}{\alpha_C-1}}-1}$, which is satisfied for $1 < \alpha_C < \frac{\ln 3}{\ln 3 - \ln 2}$.

To prove part 2 of Proposition 3, we show the computations now for a size of the coalition equal to J. The marginal benefit being constant makes the result in case of disagreement to be as in the case when there is no coalition. For each player $j \in \mathcal{J}$ (with $|\mathcal{J}| = J$) the bargaining with the leader results in

$$\begin{aligned}
q_j(J) &= (J+1)^{\frac{1}{\alpha_C-1}} q^{BR}, \\
t_j(J) &= \left((J+1)^{\frac{1}{\alpha_C-1}} - \frac{1}{J+1}\right) C(q^{BR}).
\end{aligned}$$

For each player $k \notin \mathcal{J}$ the bargaining with the leader results in

$$q_k = 2^{\frac{1}{\alpha_C-1}} q^{BR},\ t_k = \left(2^{\frac{1}{\alpha_C-1}} - \frac{1}{2}\right) C(q^{BR}).$$

Comparison in abatement and transfers follows from the discussion above. Payoffs are

$$\begin{aligned}
\pi_j(J) &= bQ(J) - \left(J(J+1)^{\frac{1}{\alpha_C-1}} + \frac{1}{J+1}\right) C(q^{BR}), \\
\pi_k(J) &= bQ(J) - \left(2^{\frac{1}{\alpha_C-1}} + \frac{1}{2}\right) C(q^{BR}),
\end{aligned}$$

with $Q(J) = Jq_j(J) + (n-J)q_k$. Condition $\pi_k(J) > \pi_j(J)$ implies $J(J+1)^{\frac{1}{\alpha_C-1}} + \frac{1}{J+1} > 2^{\frac{1}{\alpha_C-1}} + \frac{1}{2}$ which holds when $J \geq 2$. The sum of the transfer paid by the leader is $T = Jt_j + (N-J)t_k$ and it gives

$$\pi_0(J) = bQ(J) - \left(J\left((J+1)^{\frac{1}{\alpha_C-1}} - \frac{1}{J+1}\right) + (n-J)\left(2^{\frac{1}{\alpha_C-1}} - \frac{1}{2}\right)\right) C(q^{BR}).$$

Condition $\pi_j(J) > \pi_0(J)$ can be rewritten as $(n-J)\left(2^{\frac{1}{\alpha_C-1}} - \frac{1}{2}\right) > 1$. Note that $2^{\frac{1}{\alpha_C-1}} - \frac{1}{2} > \frac{1}{2}$ because $2^{\frac{1}{\alpha_C-1}} > 1$. As a result, if $n-J \geq 2$ then $2^{\frac{1}{\alpha_C-1}} - \frac{1}{2} > \frac{1}{n-J}$. If $n-J=1$ then the condition becomes $2^{\frac{1}{\alpha_C-1}} > \frac{3}{2}$, which holds when $\alpha_C < \frac{\ln 3}{\ln 3 - \ln 2}$. Finally, $\pi_j(J) < \pi_0(J)$ when $n = J$, i.e, when the grand coalition forms to bargain against the leader.

Condition $\pi_k(J) > \pi_0(J)$ can be rewritten as $(n-J-1)\left(2^{\frac{1}{\alpha_C-1}} - \frac{1}{2}\right) + J\left((J+1)^{\frac{1}{\alpha_C-1}} - \frac{1}{J+1}\right) > 1$. As argued before, $2^{\frac{1}{\alpha_C-1}} - \frac{1}{2} > \frac{1}{2}$. Similarly, $(J+1)^{\frac{1}{\alpha_C-1}} - \frac{1}{J+1} > \frac{J}{J+1}$. We check that $(n-J-1)\frac{1}{2} + J\frac{J}{J+1} > 1$ whenever $n \geq 3$, which would imply that $\pi_k(J) > \pi_0(J)$. Note that n must be at least 3 if $J \geq 2$ because when $n = J$ there is no $k \notin \mathcal{J}$.

Indeed, $(n-J-1)\frac{1}{2} + J\frac{J}{J+1} > 1$ if and only if $J^2 + (n-4)J + n - 3 > 0$. It is easy to see that the inequality holds for $n \geq 4$. When $n = 3$ the inequality becomes $J(J-1) > 0$, which is true given that $J \geq 2$.

We prove part 3 of Proposition 3. In order to see that the leader's preferred coalition size is $J = n$ we check that $\pi_0(J)$ is increasing in J. Recall that

$$\pi_0(J) = bQ(J) - \left(J\left((J+1)^{\frac{1}{\alpha_C-1}} - \frac{1}{J+1}\right) + (n-J)\left(2^{\frac{1}{\alpha_C-1}} - \frac{1}{2}\right)\right)C(q^{BR}),$$

with $Q(J) = \left(J(J+1)^{\frac{1}{\alpha_C-1}} + (n-J)2^{\frac{1}{\alpha_C-1}}\right)q^{BR}$. Given that $bq^{BR} = \alpha_C C(q^{BR})$, we have that

$$\pi_0(J) = \left((\alpha_C - 1)\left(J\left((J+1)^{\frac{1}{\alpha_C-1}} + (n-J)2^{\frac{1}{\alpha_C-1}}\right)\right) + \frac{J}{J+1} + \frac{n-J}{2}\right)C(q^{BR}).$$

By checking that its derivative with respect to J is positive, we know that $\pi_0(J)$ is an increasing function in J. Note that

$$\pi_0'(J) = \left((\alpha_C - 1)\left\{(J+1)^{\frac{1}{\alpha_C-1}} - 2^{\frac{1}{\alpha_C-1}}\right\} + \frac{2J(J+1)^{\frac{\alpha_C}{\alpha_C-1}} - J(J+2) + 1}{2(J+1)^2}\right)C(q^{BR}),$$

which is positive given that $2(J+1)^{\frac{\alpha_C}{\alpha_C-1}} > 2(J+1) > (J+2)$.

In order to check that a pool of size $J \geq 2$ is preferred to singletons by the leader, it suffices to note that $\pi_0(2) > \pi_0$ because, as just shown, π_0 is an increasing function of J. Indeed,

$$(\alpha_C - 1)\left(2 \times 3^{\frac{1}{\alpha_C-1}} - (n-2)2^{\frac{1}{\alpha_C-1}}\right) + \frac{2}{3} + \frac{n-2}{2} \geq 2^{\frac{1}{\alpha_C-1}}(n\alpha_C - n) + \frac{n}{2},$$

if and only if

$$2(\alpha_C - 1)\left(\times 3^{\frac{1}{\alpha_C-1}} - 2^{\frac{1}{\alpha_C-1}}\right) - \frac{1}{3} \geq 0.$$

The left-hand-side of this inequality is a decreasing function of α_C for $\alpha_C > 1$. Its limit as α_C approaches 1 from the right is ∞ while its limit as α_C approaches

∞ is $-\frac{1}{3}+2\left(\ln 3-\ln 2\right)>0$. Hence, the inequality is always true for any value of $\alpha_C>1$.

We finally prove the last part of Proposition 3. From the internal (15) and external (16) stability conditions, the external stability condition can be rewritten as $\pi_j(J^*+1)-\pi_k(J^*)\leq 0$. Hence, a stable coalition J^* must satisfy the two internal and external stability conditions. We concentrate then on the analysis of the function $\Pi(J)=\pi_j(J)-\pi_k(J-1)$. Note that the internal stability condition for J is satisfied when $\Pi(J)\geq 0$ and the external stability condition for J is satisfied when $\Pi(J+1)\leq 0$. Recall that

$$\begin{aligned}\pi_j(J)-\pi_k(J-1) &= b\left(J(J+1)^{\frac{1}{\alpha_C-1}}-(J-1)J^{\frac{1}{\alpha_C-1}}-2^{\frac{1}{\alpha_C-1}}\right)q^{BR}- \\ &- \left(J(J+1)^{\frac{1}{\alpha_C-1}}+\frac{1}{J+1}-2^{\frac{1}{\alpha_C-1}}-\frac{1}{2}\right)C(q^{BR}).\end{aligned}$$

Given that $bq^{BR}=\alpha_C C(q^{BR})$,we have $\Pi(J)=F(J)C(q^{BR})$ with

$$F(J)=(\alpha_C-1)(J(J+1)^{\frac{1}{\alpha_C-1}}-2^{\frac{1}{\alpha_C-1}})-\alpha_C(J-1)J^{\frac{1}{\alpha_C-1}}-\frac{1}{J+1}+\frac{1}{2}.$$

This means that $\Pi(J)$ has the same sign as $F(J)$ and a stable coalition of size $J^*\geq 2$ must verify $F(J^*)\geq 0$ and $F(J^*+1)\leq 0$ for the internal and external coalitions, respectively.

We start by checking the conditions for $J^*=2$. Note that

$$F(2)\geq 0\leftrightarrow\alpha_C\leq 4.05937 \text{ and } F(3)\leq 0\leftrightarrow\alpha_C\geq 1.21459.$$

Hence, $J^*=2$ whenever $\alpha_C\in[1.21459,4.05937]$. This means that for a quadratic cost function $\alpha_C=2$ the stable coalition has a size of 2. We can also identify the conditions for $J^*=3$. Indeed,

$$F(3)\geq 0\leftrightarrow\alpha_C\leq 1.21459. \text{ and } F(4)\leq 0\leftrightarrow\alpha_C\geq 1.11045.$$

Hence, $J^*=3$ whenever $\alpha_C\in[1.11045,1.21459]$. In a similar way, we obtain $J^*=4$ whenever $\alpha_C\in[1.075,1.11045]$ and $J^*=5$ whenever $\alpha_C\in[1.056,1.075]$.

In general, note that when a coalition of size J is not internally stable then $F(J)<0$. This condition is as well telling us that $J-1$ is externally stable. Following a similar argument, if J is not externally stable then $F(J+1)>0$, which means that $J+1$ is internally stable. This explains the overlapping in the values of α_C for the stability of each consecutive coalition size.

Note that the function $F(J)$ for each J tends to $+\infty$ as α_C approaches 1 from the right (taking larger values). This means that $F(n)$ would also have such behavior. Interestingly, there is no external stability condition for the grand coalition to check because there are no further players k outside the coalition. If $F(n)$ is positive for α_C sufficiently close to 1, this means that the grand coalition is internally stable, and therefore stable, for α_C sufficiently close to 1.

Furthermore, as α_C goes to infinity $(J+1)^{\frac{1}{\alpha-1}}$, $J^{\frac{1}{\alpha-1}}$, $(J-1)^{\frac{1}{\alpha-1}}$ and $2^{\frac{1}{\alpha-1}}$ all go to 1 (faster than the multiplicative term). Hence

$$\lim_{\alpha_C \to \infty} F(J) \sim \frac{J-1}{2(J+1)}(-J-3) < 0.$$

Hence, for α_C sufficiently large $F(J) \leq 0$ for all J and all players inside the pool prefer leaving the coalition whatever the size of J. This completes the proof of Proposition 3. ■

Proof of Proposition 4

We start by proving part 1 of Proposition 4.

If we take the conditions for an interior solution given by (3) and (29) we must have

$$B'(q_i^{BR}(Q_{-i}) + Q_{-i}) = c = 2B'(q_i + Q_{-i}) = B'\left(2^{-\frac{1}{1-\alpha_B}} q_i + 2^{-\frac{1}{1-\alpha_B}} Q_{-i}\right),$$

because B' is a homogeneous function of degree $\alpha_B - 1$. Hence,

$$q_i + Q_{-i} = 2^{\frac{1}{1-\alpha_B}} q_i^{BR}(Q_{-i}) + 2^{\frac{1}{1-\alpha_B}} Q_{-i}.$$

But if $Q_{-i} = (n-1)q_i$ (by symmetry of cost and benefit functions) then

$$nq_i = 2^{\frac{1}{1-\alpha_B}} q_i^{BR}(Q_{-i}) + 2^{\frac{1}{1-\alpha_B}} (n-1)q_i,$$

which is impossible because $n < 2^{\frac{1}{1-\alpha_B}}(n-1)$ (note that $2^{\frac{1}{1-\alpha_B}} > 2 > \frac{n+1}{n}$) and $q_i^{BR}(Q_{-i}) \geq 0$. The Nash-in-Nash bargaining solution must yield a corner solution.

We cannot have $q_i = 0$ and $q_i^{BR} > 0$ as the Nash-in-Nash bargaining result. If that would be the case, then $Q_{-i} = 0$ with $2B'(0) \leq c = B'(q_i^{BR}(0))$. Again, by homogeneity of B' we would have $B'(2^{\frac{-1}{1-\alpha_B}} 0) = B'(0) \leq B'(q_i^{BR}(0))$. By strict concavity of B ($\alpha_B < 1$), B' is strictly decreasing and hence $0 \geq q_i^{BR}(0)$, a contradiction with $q_i = 0$ and $q_i^{BR} > 0$. The only possibility is then $q_i^{BR}(Q_{-i}) = 0$ and $q_i > 0$ solves

$$2B'(nq_i) = c. \tag{35}$$

We obtain then

$$q_i = \frac{1}{n}\left(\frac{2B(1)\alpha_B}{c}\right)^{\frac{1}{1-\alpha_B}} = q^s,$$

with $t_i = \frac{c}{2}q^s$ from (2). The total level of abatement is then $Q = nq^s$.

Note that $c\, nq^s = 2\alpha_B B(nq^s)$ from the FOC in equation (35) and the benefit function being homogeneous of degree α_B, the payoffs of singleton players and for the leader are

$$\begin{aligned}
\pi_i &= B(nq^s) - \frac{c}{2}q^s = B(nq^s)\left(\frac{n-\alpha_B}{n}\right), \\
\pi_0 &= B(nq^s) - n\frac{c}{2}q^s = B(nq^s)\left(1-\alpha_B\right).
\end{aligned}$$

It shows that $\pi_i > \pi_0$ for $n > 1$.

To prove part 2 of the Proposition, we proceed with the computations for a coalition of size J. We assume throughout that $J > 2$ because $J = 1$ is equivalent to the leader bargaining with singletons. In this case we cannot have $q_j > 0$ for $j \in \mathcal{J}$ and $q_k > 0$ for $k \notin \mathcal{J}$ because that would mean $c = 2B'(Q) = (J+1)B'(Q)$. Given that $2 < (J+1)$ the bargaining results in $q_k = 0$ for $k \notin \mathcal{J}$. We also must have $q_k^{BR}(Jq_J) = 0$ because $B'(Q) = B'(Jq_J) = \frac{c}{J+1} < c$.

We obtain

$$
\begin{aligned}
q_j^{BR} &= \frac{1}{J}\left(\frac{\alpha_B B(1)}{c}\right)^{\frac{1}{1-\alpha_B}} = \frac{n}{J}\left(\frac{1}{2}\right)^{\frac{1}{1-\alpha_B}} q^s, \\
q_j(J) &= \frac{1}{J}\left(\frac{(J+1)\alpha_B B(1)}{c}\right)^{\frac{1}{1-\alpha_B}} = \frac{n}{J}\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} q^s.
\end{aligned}
$$

After some rearrangement, we obtain the payofffs:

$$
\begin{aligned}
\pi_j(J) &= 2^{-\frac{\alpha_B}{1-\alpha_B}}\left((J+1)^{\frac{\alpha_B}{1-\alpha_B}}(1-\alpha_B) - \frac{\alpha_B}{J\,(J+1)}\right) B(nq^s), \\
\pi_k(J) &= 2^{-\frac{\alpha_B}{1-\alpha_B}}(J+1)^{\frac{\alpha_B}{1-\alpha_B}} B(nq^s), \\
\pi_0(J) &= 2^{-\frac{\alpha_B}{1-\alpha_B}}\left((J+1)^{\frac{\alpha_B}{1-\alpha_B}}(1-\alpha_B) + \frac{\alpha_B}{J+1}\right) B(nq^s).
\end{aligned}
$$

It shows that $\pi_k(J) > \pi_0(J) > \pi_j(J)$.

We prove part 3 of Proposition 4. In order to see that the leader's preferred coalition size is $J = n$ we check that $\pi_0(J)$ is increasing in J. Recall that

$$
\pi_0(J) = \frac{B(nq^s)}{2^{\frac{\alpha_B}{1-\alpha_B}}}\left((J+1)^{\frac{\alpha_B}{1-\alpha_B}}(1-\alpha_B) + \frac{\alpha_B}{J+1}\right).
$$

By checking that its derivative with respect to J is positive, we know that $\pi_0(J)$ is an increasing function in J. Note that

$$
\pi_0'(J) = \frac{B(nq^s)}{2^{\frac{\alpha_B}{1-\alpha_B}}}\frac{\alpha_B}{(J+1)^2}\left((J+1)^{\frac{1}{1-\alpha_B}} - \alpha_B\right),
$$

which is positive given that $(J+1)^{\frac{1}{1-\alpha_B}} > 1 > \alpha_B$.

In order to check that a pool of size $J \geq 2$ is preferred to singletons by the leader, it suffices to note that $\pi_0(2) > \pi_0$ because, as just shown, π_0 is an increasing function of J. Indeed,

$$
\frac{B(nq^s)}{2^{\frac{\alpha_B}{1-\alpha_B}}}\left((J+1)^{\frac{\alpha_B}{1-\alpha_B}}(1-\alpha_B) + \frac{\alpha_B}{J+1}\right) \geq B(nq^s)(1-\alpha_B),
$$

if and only if

$$
(J+1)^{\frac{\alpha_B}{1-\alpha_B}}(1-\alpha_B) + \frac{\alpha_B}{J+1} \geq 2^{\frac{\alpha_B}{1-\alpha_B}}(1-\alpha_B).
$$

This inequality is true because $J \geq 2$.

We finally prove the last part of Proposition. A stable coalition J^* must satisfy the two internal and external stability conditions. We concentrate then on the analysis of the function $\Pi(J) = \pi_j(J) - \pi_k(J-1)$. Note that the internal stability condition for J is satisfied when $\Pi(J) \geq 0$ and the external stability condition for J is satisfied when $\Pi(J+1) \leq 0$. Recall that

$$\pi_j(J) - \pi_k(J-1) = 2^{-\frac{\alpha_B}{1-\alpha_B}} \left((J+1)^{\frac{\alpha_B}{1-\alpha_B}} (1-\alpha_B) - \frac{\alpha_B}{J(J+1)} - J^{\frac{\alpha_B}{1-\alpha_B}} \right) B(nq^s),$$

for $J \geq 3$. However the case $J = 1$ refers to the situation of the leader facing singletons (with $q_k \neq 0$). In that case the payoff of a singleton is

$$\pi_k(1) = \left(\frac{n - \alpha_B}{n} \right) B(nq^s).$$

Since $\pi_j(2) = 2^{-\frac{\alpha_B}{1-\alpha_B}} \left(3^{\frac{\alpha_B}{1-\alpha_B}} (1-\alpha_B) - \frac{\alpha_B}{6} \right) B(nq^s)$ it implies that

$$\pi_j(2) - \pi_k(1) = 2^{-\frac{\alpha_B}{1-\alpha_B}} \left((1-\alpha_B) 3^{\frac{\alpha_B}{1-\alpha_B}} - 2^{\frac{\alpha_B}{1-\alpha_B}} \frac{n-\alpha_B}{n} - \frac{\alpha_B}{6} \right) B(nq^s).$$

Let us define $H(J) = (1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - J^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{J(J+1)}$, for $J \geq 3$ and $H(2) = (1-\alpha_B)3^{\frac{\alpha_B}{1-\alpha_B}} - 2^{\frac{\alpha_B}{1-\alpha_B}} \frac{n-\alpha_B}{n} - \frac{\alpha_B}{6}$. A stable coalition of size $J^* \geq 2$ must verify $\pi_j(J^*) - \pi_k(J^*-1) \geq 0$ and $\pi_j(J^*+1) - \pi_k(J^*) \leq 0$. These two conditions hold if and only if $H(J^*) \geq 0$ and $H(J^*+1) \leq 0$, respectively. We start by checking the conditions for $J^* = 2$. Note that $H(2)$ depends on n while $H(3) \leq 0 \leftrightarrow \alpha_B \leq 0.880902$. More precisely, for $n = 3$ $H(2) \leq 0 \leftrightarrow \alpha_B \geq 0.711563$, which means that $J^* = 2$ whenever $\alpha_B \in [0.711563, 0.880902]$. For $n = 4$ $H(2) \leq 0 \leftrightarrow \alpha_B \geq 0.743462$, which means that $J^* = 2$ whenever $\alpha_B \in [0.743462, 0.880902]$. For $n = 200$ $H(2) \leq 0 \leftrightarrow \alpha_B \geq 0.798448$, which means that $J^* = 2$ whenever $\alpha_B \in [0.798448, 0.880902]$. As n tends to infinity, $H(2)$ tends to $H^{\infty}(2) = (1-\alpha_B)3^{\frac{\alpha_B}{1-\alpha_B}} - 2^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{6}$. Given that $H^{\infty}(2) \leq 0 \leftrightarrow \alpha_B \geq 0.799213$, which means that $J^* = 2$ whenever $\alpha_B \in [0.799213, 0.880902]$ for any value of n. This also means that for a square-root benefit function $\alpha_B = 0.5$ there is not stable coalition because $H(J) \leq 0$ for all $J > 2$, and all players prefer bargaining individually with the leader to avoid the strong free-riding outside the coalition. We can more easily identify the conditions for $J^* = 3$. Indeed,

$$H(3) \geq 0 \leftrightarrow \alpha_B \geq 0.880902 \text{ and } H(4) \leq 0 \leftrightarrow \alpha_B \leq 0.918139.$$

Hence, $J^* = 3$ whenever $\alpha_B \in [0.880902, 0.918139]$. In a similar way, we obtain $J^* = 4$ whenever $\alpha_B \in [0.918139, 0.938701]$.

In general, note that when a coalition of size J is not internally stable then $H(J) < 0$. This condition is as well telling us that $J-1$ is externally stable.

Following a similar argument, if J is not externally stable then $H(J+1) > 0$, which means that $J+1$ is internally stable. This explains the overlapping in the values of α_B for the stability of each consecutive coalition size. Note that the function $H(J)$ for each J tends to $+\infty$ as α_B approaches 1 from the left (taking smaller values). This means that $H(n)$ would also have such behavior. Interestingly, there is no external stability condition for the grand coalition to check because there are no further players k outside the coalition. If $H(n)$ is positive for α_B sufficiently close to 1, this means that the grand coalition is internally stable, and therefore stable, for α_B sufficiently close to 1. Furthermore, for any coalition size J

$$\lim_{\alpha_B \to 0} H(J) = 0, \quad \lim_{\alpha_B \to 1^+} H(J) = \infty,$$

with

$$\left.\frac{\partial H(J)}{\partial J}\right|_{\alpha_B=0} < 0,$$

and a unique local minimum and no local maximum between 0 and 1. Hence, for sufficiently low α_B we have that $H(J) \leq 0$ for all J and all players inside the pool prefer leaving the coalition whatever the size of J. This completes the proof of Proposition 4. ∎

6.1 Proof of Proposition 5

We assume $B(Q) = Q^{\alpha_B} B(1)$ and $C(q) = q^{2-\alpha_B} C(1)$. From the FOC $C'(q_i) = 2B'(Q)$ with $Q = nq_i$ we obtain $q_i = \left(\frac{2\alpha_B}{2-\alpha_B}\frac{B(1)}{C(1)}\right)^{\frac{1}{2(1-\alpha_B)}} n^{-\frac{1}{2}}$. Using the FOC $B'(q_i^{BR} + Q_{-i}) = C'(q_i^{BR})$, the best reply is solution of the second order condition $\left(q_i^{BR}\right)^2 + q_i^{BR}(Q - q_i) - \left(\frac{\alpha_B B(1)}{(2-\alpha_B)C(1)}\right)^{\frac{1}{1-\alpha_B}} = 0$ which gives $q_i^{BR} = \frac{1}{2}\left(\sqrt{2^{-\frac{1}{1-\alpha_B}}\left(\left(\frac{n-1}{\sqrt{n}}\right)^2 2^{\frac{1}{1-\alpha_B}} + 4\right)n} - (n-1)\right) q_i$. When a coalition forms, the ratio of the FOC gives $q_j = \left(\frac{J+1}{2}\right)^{1-\alpha_B} q_k$ and $Q = \left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J\right) q_k$ with $q_k = \left(\frac{2\alpha_B}{2-\alpha_B}\frac{B(1)}{C(1)}\right)^{\frac{1}{2(1-\alpha_B)}} \left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J\right)^{-\frac{1}{2}}$. Concerning the best reply q_j^{BR} is solution of the second order equation

$$J\left(q_j^{BR}\right)^2 + (n-J)q_k q_j^{BR} - \left(\frac{\alpha_B}{2-\alpha_B}\frac{B(1)}{C(1)}\right)^{\frac{1}{1-\alpha_B}} = 0$$

The determinant is $\Delta = \left((n-J)^2 + 2^{\frac{2\alpha_B - 1}{\alpha_B - 1}} J\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J\right)\right) q_k^2$ and it gives

$$q_j^{BR} = \frac{\sqrt{(n-J)^2 + 2^{\frac{2\alpha_B-1}{\alpha_B-1}} J\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} + n - J\right)} - (n-J)}{2^{-\frac{\alpha_B}{1-\alpha_B}} J (J+1)^{\frac{1}{1-\alpha_B}}} q_j$$

The best remply q_k^{BR} is solution of the second order equation

$$\left(q_k^{BR}\right)^2+\left(\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}+n-J-1\right)q_k\right)q_k^{BR}-\left(\frac{\alpha_B}{2-\alpha_B}\frac{B(1)}{C(1)}\right)^{\frac{1}{1-\alpha_B}}=0$$

The determinant is $\Delta=\left(\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}+n-J-1\right)^2+2^{\frac{2\alpha_B-1}{\alpha_B-1}}\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}+n-J\right)\right)q_k^2$ and it gives

$$q_k^{BR}=\frac{1}{2}\left(\begin{array}{c}\sqrt{\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}+n-J-1\right)^2+2^{\frac{(2\alpha_B-1)}{\alpha_B-1}}\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}+n-J\right)}\\-\left(J\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}+n-J-1\right)\end{array}\right)q_k$$

For $\alpha_B=1/2$ and $\alpha_C=3/2$ we obtain

$$\begin{aligned}
q_j &= \left(\frac{J+1}{2}\right)^2 q_k,\ q_k=\frac{2b}{3c}\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)^{-\frac{1}{2}},\\
Q &= \left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)q_k,\\
q_j^{BR} &= \frac{2}{J(J+1)^2}\left(\sqrt{(n-J)^2+J\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)}-(n-J)\right)q_j,\\
q_k^{BR} &= \frac{1}{2}\left(\begin{array}{c}\sqrt{\left(J\left(\frac{J+1}{2}\right)^2+n-J-1\right)^2+J\left(\frac{J+1}{2}\right)^2+(n-J)}\\-\left(J\left(\frac{J+1}{2}\right)^2+n-J-1\right)\end{array}\right)q_k.
\end{aligned}$$

It can be shown that $0<q_j^{BR}<q_j$ and $0<q_k^{BR}<q_k$.
Payoffs are

$$\pi_j(J)=\frac{b^{\frac{3}{2}}}{3^{\frac{1}{2}}c^{\frac{1}{2}}}\frac{2^{\frac{1}{2}}\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)-\frac{1}{J+1}\frac{1}{3}\left(J(J+1)^3 2^{-\frac{3}{2}}+\left(\frac{1}{J}\right)^{\frac{3}{2}}\left(\sqrt{(n-J)^2+J\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)}-(n-J)\right)^{\frac{3}{2}}\right)}{\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)^{\frac{3}{4}}}$$

$$\pi_k(J)=\frac{b^{\frac{3}{2}}}{3^{\frac{1}{2}}c^{\frac{1}{2}}}\frac{2^{\frac{1}{2}}\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)-\frac{1}{6}\left(2^{\frac{3}{2}}+\left(\sqrt{\left(J\left(\frac{J+1}{2}\right)^2+n-J-1\right)^2+J\left(\frac{J+1}{2}\right)^2+(n-J)}-\left(J\left(\frac{J+1}{2}\right)^2+n-J-1\right)\right)^{\frac{3}{2}}\right)}{\left(J\left(\frac{J+1}{2}\right)^2+(n-J)\right)^{\frac{3}{4}}}$$

It can be graphically shown that $F(J)=\pi_j(J)-\pi_k(J-1)$ is a decreasing function of J meaning that the only stable coalition is the empty coalition $J=\emptyset$ $\forall n$.

For $\alpha_B = 3/4$ and $\alpha_C = 5/4$, we obtain

$$
\begin{aligned}
q_j &= \left(\frac{J+1}{2}\right)^4 q_k,\, q_k = \left(\frac{6b}{5c}\right)^2 \left(\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right)\right)^{-\frac{1}{2}}, \\
Q &= \left(J\left(\frac{J+1}{2}\right)^4 + n - J\right) q_k, \\
q_j^{BR} &= \frac{4}{J(J+1)^4}\left(\sqrt{4(n-J)^2 + J\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right)} - 2(n-J)\right) q_j, \\
q_k^{BR} &= \frac{1}{4}\left(\begin{array}{c} \sqrt{4\left(J\left(\frac{J+1}{2}\right)^4 + n - J - 1\right)^2 + J\left(\frac{J+1}{2}\right)^4 + n - J} \\ -2\left(J\left(\frac{J+1}{2}\right)^4 + n - J - 1\right) \end{array}\right) q_k.
\end{aligned}
$$

It can be shown that $0 < q_j^{BR} < q_j$ and $0 < q_k^{BR} < q_k$. Payoffs are

$$
\pi_j(J) = \frac{3^{\frac{3}{2}} b^{\frac{5}{2}}}{5^{\frac{3}{2}} c^{\frac{3}{2}}} \frac{2^{\frac{3}{2}}\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right) - \frac{1}{J+1}\frac{3}{5}\left(2^{-\frac{5}{2}}(J+1)^5 + \left(\frac{1}{J}\right)^{\frac{5}{4}}\left(\sqrt{4(n-J)^2 + J\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right)} - 2(n-J)\right)^{\frac{5}{4}}\right)}{\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right)^{\frac{5}{8}}}
$$

$$
\pi_k(J) = \frac{3^{\frac{3}{2}} b^{\frac{5}{2}}}{5^{\frac{3}{2}} c^{\frac{3}{2}}} \frac{2^{\frac{3}{2}}\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right) - \frac{3}{10}\left(2^{\frac{5}{2}} + \left(\sqrt{4\left(J\left(\frac{J+1}{2}\right)^4 + n - J - 1\right)^2 + J\left(\frac{J+1}{2}\right)^4 + n - J} - 2\left(J\left(\frac{J+1}{2}\right)^4 + n - J - 1\right)\right)^{\frac{5}{4}}\right)}{\left(J\left(\frac{J+1}{2}\right)^4 + n - J\right)^{\frac{5}{8}}}
$$

It can be graphically shown that $F(J) = \pi_j(J) - \pi_k(J-1)$ is an increasing function of J meaning that the grand coalition $J = n$ is stable $\forall n$.

Proof of Proposition 6

We consider general functions $B(Q) = Q^{\alpha_B} B(1)$ and $C(q) = q^{\alpha_C} C(1)$. When the leader negotiates against singletons, the FOCs $B'(Q^d) = C'(q_i^d)$ given by (17) and $2B'(Q) = C'(q_i)$ given by (3) implies $q_i^d > 0$ and $q_i > 0$, with

$$
q_i^d = \left(\frac{\alpha_B B(1)}{\alpha_C C(1)}\right)^{\frac{1}{\alpha_C - \alpha_B}} n^{-\frac{1-\alpha_B}{\alpha_C - \alpha_B}},\; q_i = 2^{\frac{1}{\alpha_C - \alpha_B}} q_i^d,
$$

leading to the following payoff for the singleton

$$
\pi_i = \left(2^{\frac{\alpha_B}{\alpha_C - \alpha_B}}\left(\frac{\alpha_C}{\alpha_B} n - 1\right) - \frac{1}{2}\right) C(1)(q_i^d)^{\alpha_C}.
$$

From $B'(Q^d) = C'(q_i^d)$ we also obtain $C(q_i^d) = \frac{\alpha_B}{\alpha_C}\frac{B(Q^d)}{n}$ since $B'(Q^d) = \alpha_B \frac{B(Q^d)}{Q^d}$ and $C'(q_i^d) = \alpha_C \frac{C(q_i^d)}{q_i^d}$.

When a coalition forms to bargain with the leader, we assume throughout that $J > 1$ because $J = 1$ is equivalent the leader bargaining with singletons. From the two FOCs in Proposition 2 $C'(q_j) = (J+1)B'(Q)$ given by (8) and

$C'(q_k) = 2B'(Q)$ given by (11) we obtain the following relation between q_j and q_k: $q_j = \left(\frac{J+1)}{2}\right)^{\frac{1}{\alpha_C - 1}} q_k$. It helps to obtain a relation between the abatement realised by the singleton outside the pool and its abatement in case of disagreement

$$q_k = 2^{\frac{1}{\alpha_C - \alpha_B}} \left(\frac{n}{J\left(\frac{J+1)}{2}\right)^{\frac{1}{\alpha_C-1}} + n - J} \right)^{\frac{1-\alpha_B}{\alpha_C - \alpha_B}} q_k^d.$$

Depending on the different values of n, J, α_C and α_B, we may have $q_k > q_k^d$ or $q_k < q_k^d$. From the expression of the transfer $t_k = \frac{1}{2}\left(C(q_k) - C(q_k^d)\right)$ given by (10), a positive transfer implies the condition $q_k > q_k^d$. However for $q_k < q_k^d$ we fix $t_k = 0$. If we do not impose any condition on the sign of t_k, it is player k who transfers money to the leader, who in turn will transfer it to the coalition J, in exchange of free-riding: $t_k < 0$. But since the leader is not allowed to enforce formal sanction, he cannot collect fees from singletons. Such a configuration is not possible and we add the restriction $t_k \geq 0$.

Considering the case of a constant marginal benefit $\alpha_B = 1$, we obtain $q_k > q_k^d$ since $2^{\frac{1}{\alpha_C - \alpha_B}} > 1$. In that case payoffs are

$$\begin{aligned}
\pi_j(J) &= \frac{1}{n}\left(J\left(J+1\right)^{\frac{1}{\alpha_C-1}} \left(1 - \frac{1}{\alpha_C}\right) + 2^{\frac{1}{\alpha_C-1}}\left(n-J\right) - \frac{1}{J+1}\frac{1}{\alpha_C}\right) B\left(Q^d\right), \\
\pi_k(J) &= \frac{1}{n}\left(J\left(J+1\right)\right)^{\frac{1}{\alpha_C-1}} + 2^{\frac{1}{\alpha_C-1}}\left(n - J - \frac{\alpha_B}{\alpha_C}\right) - \frac{1}{2}\frac{1}{\alpha_C}\right) B(Q^d).
\end{aligned}$$

Define $\pi_j(J) - \pi_k(J-1) = G(J)C(Q^d)$ since $C(q_i^d) = \frac{\alpha_B}{\alpha_C}\frac{B(Q^d)}{n}$ with

$$G(J) = \left(\left(\alpha_C - 1\right)\left(J\left(J+1\right)^{\frac{1}{\alpha_C-1}} - 2^{\frac{1}{\alpha_C-1}} \right) - \alpha_C\left(J-1\right)J^{\frac{1}{\alpha_C-1}} - \frac{1}{J+1} + \frac{1}{2} \right),$$

which is the same expression denoted by $F(J)$ obtained in the case of passive beliefs. It proves part 1 of the proposition.

Considering the case of a constant marginal cost $\alpha_C = 1$. When the leader negotiates against singletons, we obtain $q_i^d > 0$ and $q_i > 0$, with

$$q_i^d = \frac{1}{n}\left(\frac{\alpha_B B(1)}{c}\right)^{\frac{1}{1-\alpha_B}}, \; q_i = 2^{\frac{1}{1-\alpha_B}} q_i^d.$$

With symmetric beliefs (with the uperscript SB), payoffs in the singleton case are

$$\begin{aligned}
\pi_i^{SB} &= B(Q) - cq_i + t_i = B(1)^{\frac{1}{1-\alpha_B}}\left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}}\left(2^{\frac{\alpha_B}{1-\alpha_B}}\left(1 - \frac{\alpha_B}{n}\right) - \frac{\alpha_B}{2n}\right), \\
\pi_0^{SB} &= B(Q) - n\frac{c}{2}(q_i - q_i^d) = B(1)^{\frac{1}{1-\alpha_B}}\left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}}\left(\left(1-\alpha_B\right)2^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{2}\right).
\end{aligned}$$

It shows that $\pi_i^{SB} > \pi_0^{SB}$ for $n > 1$. Moreover, comparing these payoffs with the ones obtained in the passive beliefs (with the upperscript PB)

$$\begin{aligned}
\pi_i^{PB} &= B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} 2^{\frac{\alpha_B}{1-\alpha_B}} \left(1 - \frac{\alpha_B}{n}\right), \\
\pi_0^{PB} &= B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} (1-\alpha_B)\, 2^{\frac{\alpha_B}{1-\alpha_B}}.
\end{aligned}$$

It can be shown that $\pi_i^{PB} > \pi_i^{SB}$. Total abatement is the same in both cases but the transfer paid by the leader to the singletons is higher in the passive beliefs case $t_i^{PB} > t_i^{SB}$ and this explains why singletons prefer this case and the leader the second case. This is the opposite for the leader $\pi_0^{PB} < \pi_0^{SB}$.

When a coalition forms, we cannot have $q_j > 0$ for $j \in \mathcal{J}$ and $q_k > 0$ for $k \notin \mathcal{J}$ because that would mean $c = 2B'(Q) = (J+1)B'(Q)$. Given that $2 < (J+1)$ the bargaining results in $q_k = 0$ for $k \notin \mathcal{J}$. Interestingly, we obtain

$$q_j^d = q_k^d = q^d = \frac{1}{n} \left(\frac{\alpha_B B(1)}{c}\right)^{\frac{1}{1-\alpha_B}},$$

and

$$q_j = \frac{n}{J} (J+1)^{\frac{1}{1-\alpha_B}} q^d > q^d$$

In this case $q_k^d > q_k = 0$ and $t_k = 0$. To prove that $t_k = 0$ consider the the maximization problem

$$\max_{(q_k, t_k)} NP_{0,k} = \left(\pi_0 - \pi_{0,-k}^d\right) \left(\pi_k - \pi_{k,-k}^d\right),$$

subject to $t_k \geq 0$. The first order condition implies that

$$\frac{\partial NP}{\partial t_k} = NP_{0,k} \left(\frac{-1}{\pi_0 - \pi_{0,-k}^d} + \frac{1}{\pi_k - \pi_{k,-k}^d} \right),$$

is negative at the (corner) solution, i.e., when $t_k = 0$, $q_k^d = q^d$, $q_k = 0$. This is so because $\pi_0 - \pi_{0,-k}^d < \pi_k - \pi_{k,-k}^d \Leftrightarrow \pi_0 - \pi_k = -Jt_j < \pi_{0,-k}^d - \pi_{k,-k}^d = cq^d$.

Payoffs are

$$\begin{aligned}
\pi_j(J) &= B(Jq_j) - cq_j + t_j = B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} \left((1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{n(J+1)} \right), \\
\pi_k(J) &= B(Jq_j) = \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}} (J+1)^{\frac{\alpha_B}{1-\alpha_B}}, \\
\pi_0(J) &= B(Jq_j) - Jt_j = B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} \left((1-\alpha_B J)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} + \frac{\alpha_B J}{n(J+1)} \right).
\end{aligned}$$

It is clear that $\pi_k(J) > \pi_j(J)$. From all the above,

$$\pi_j(J) - \pi_k(J-1) = B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} \left[(1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - J^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{n(J+1)} \right],$$

for $J \geq 3$.

However the case $J = 1$ refers to the situation of the leader facing singletons (with $q_k \neq 0$). In that case the payoff of a singleton is

$$\pi_k(1) = B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} \left(2^{\frac{\alpha_B}{1-\alpha_B}} \left(1 - \frac{\alpha_B}{n}\right) - \frac{\alpha_B}{2n}\right),$$

while

$$\pi_j(2) = B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} \left((1-\alpha_B)\, 3^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{3n}\right).$$

It implies that

$$\pi_j(2) - \pi_k(1) = B(1)^{\frac{1}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} \left((1-\alpha_B)\, 3^{\frac{\alpha_B}{1-\alpha_B}} - 2^{\frac{\alpha_B}{1-\alpha_B}} \left(1 - \frac{\alpha_B}{n}\right) + \frac{\alpha_B}{6n}\right).$$

Let us define $G(J) = (1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - J^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{n(J+1)}$, for $J \geq 3$ and $G(2) = (1-\alpha_B)3^{\frac{\alpha_B}{1-\alpha_B}} - 2^{\frac{\alpha_B}{1-\alpha_B}} \frac{n-\alpha_B}{n} + \frac{\alpha_B}{6n}$.

A stable coalition of size $J^* \geq 2$ must verify $\pi_j(J^*) - \pi_k(J^*-1) \geq 0$ and $\pi_j(J^*+1) - \pi_k(J^*) \leq 0$. These two conditions hold if and only if $G(J^*) \geq 0$ and $G(J^*+1) \leq 0$, respectively. In this case both inequalities depend on the value of n.

1. Fix first $n = 3$. We have $J^* = 2$ ($G(2) \geq 0$ and $G(3) \leq 0$) whenever $\alpha_B \in [0.79694, 0.880902]$ and $J^* = 3$ ($G(3) \geq 0$, no external stability because there are no further agents) whenever $\alpha_B \geq 0.880902$.

2. Consider $n = 4$. We have $J^* = 2$ whenever $\alpha_B \in [0.797005, 0.880902]$, $J^* = 3$ whenever $\alpha_B \in [0.880902, 0.918139]$, and $J^* = 4$ whenever $\alpha_B \geq 0.918139$.

If we consider higher numbers, for example $n = 200$ we obtain $J^* = 200$ whenever $\alpha_B \geq 0.999316$. Again, any coalition size can be stable for α_B taking values in an interval sufficiently close to 1 but for reasonable values of α_B, for example $\alpha_B \leq 0.7$, there is no stable coalition because all players prefer bargaining individually with the leader to avoid the strong free-riding outside the coalition.

For the sake of completeness, if we consider n going to infinity, we obtain that $G(J)$ tends to $G^\infty(J) = (1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - J^{\frac{\alpha_B}{1-\alpha_B}}$ for all $J \geq 2$. With these functions, $J^* = 2$ whenever $\alpha_B \in [0.797537, 0.88091]$, $J^* = 3$ whenever $\alpha_B \in [0.88091, 0.918139]$, $J^* = 4$ whenever $\alpha_B \in [0.918139, 0.938701]$, etc. These numbers can be used as an approximation.

Quantities in case of agreement do not depend on beliefs, it is the transfers (and hence final payoffs) that depend on beliefs. The conclusions are similar, the limits separating each stable coalition size slightly depend on n but stay close to the numbers for passive beliefs. The difference is that with symmetric beliefs the disagreement quantities are strictly positive. We have

$$q_k^{BR} = 0 < q_k^d \text{ and } q_j^{BR} = 0 > q_j^d.$$

In both cases $t_k = 0$ but $t_j^{PB} < t_j^{SB}$. This means that $\pi_k^{PB} = \pi_k^{SB}$ and $\pi_j^{PB} < \pi_j^{SB}$ for all J.

We have the following inequalities regarding the internal and stability conditions:

$$\pi_j^{PB}(J)-\pi_k^{PB}(J-1) < \pi_j^{SB}(J)-\pi_k^{SB}(J-1) \text{ and } \pi_j^{PB}(J+1)-\pi_k^{PB}(J) < \pi_j^{SB}(J+1)-\pi_k^{SB}(J).$$

This means that the internal stability condition with passive beliefs implies the internal stability condition with symmetric beliefs, and that the external stability condition with symmetric beliefs implies the external stability condition with passive beliefs. Another interpretation is that countries or agents gain less to add themselves to the coalition under passive beliefs than under symmetric beliefs, independently of the size of the coalition.

It completess the proof of part 2 of the proposition.

Now if we consider different values of α_C and α_B, we have to check whether $t_k = 0$ or $t_k > 0$ since it impacts the payoff of player k. We obtain

$$q_k > (<) q_k^d \Leftrightarrow t_k > (=) 0 \text{ for } 2^{\frac{1}{\alpha_C - \alpha_B}} \left(\frac{n}{J\left(\frac{J+1)}{2}\right)^{\frac{1}{\alpha_C - 1}} + n - J} \right)^{\frac{1-\alpha_B}{\alpha_C - \alpha_B}} > (<) 1$$

For $\alpha_C = \frac{3}{2}$, $\alpha_B = \frac{1}{2}$ and $n = 20$, we can show that from $J = 1$ to 5 the transfer is positive $t_k > 0$ but for $J \in [6, n]$ $t_k = 0$. The computation of the internal stability condition shows that the grand coalition is stable. Increasing the number of countries to $n = 200$ modifies the above condition since $t_k > 0$ for $J \in [1, 12]$ and $t_k = 0$ for $J \in [13, n]$ but the grand coalition remains stable. We obtain similar results when $\alpha_C = 5/4$, $\alpha_B = 3/4$, $n = 200$. However for $\alpha_C = 1.9$ and $\alpha_B = 0.8$ the coalition is empty and all players prefer to negociate alone with the leader.

Proof of Proposition 7

At stage 1 ce assume that a coalition forms to bargain with the leader. At stage 2 the leader bargains with the singletons. The game is solved backwards, starting at stage 2. We will denote by $Q_{\mathcal{J}}^h$ the total emission reduction by the players in the coalition. In history h there could have been an agreement or not, but at stage 2 it is *known* whether that is the case, and the corresponding quantity $Q_{\mathcal{J}}^h$.

We assume that all possible agreements are beneficial in equilibrium and will check the conditions that guarantee so. Each of the singletons bargain with the leader assuming that the others are agreeing as *in the equilibrium path starting at history* $Q_{\mathcal{J}}^h$.

$$\max_{(q_k, t_k)} NP_{0,k} = \left(\pi_0 - \pi_{0,-k}^d\right)\left(\pi_k - \pi_{k,-k}^d\right).$$

Under passive beliefs, players 0 and k believe that all other bargaining meetings reach the equilibrium agreement in stage 2 knowing $Q_{\mathcal{J}}^h$. We denote the equilibrium agreement emission reduction by $q_k(Q_{\mathcal{J}}^h)$, as it is a function of the known emission by players j in the coalition. By symmetry, the emission reduction by the other singleton players in equilibrium in stage 2, is equal to $(K-1)q_k(Q_{\mathcal{J}}^h)$. In case of disagreement with the leader, player k optimally chooses her disagreement emission reduction that we denote by $q_k^{BR}(Q_{\mathcal{J}}^h)$, given by

$$q_k^{BR}(Q_{\mathcal{J}}^h) = arg\max_{q_k \geq 0} B(q_k + (K-1)q_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h) - C(q_k),$$

where $(K-1)q_k(Q_{\mathcal{J}}^h)$ and $Q_{\mathcal{J}}^h$ are taken as given.

The FOC of this maximization problem (interior solution) is given by

$$B'(q_k^{BR}(Q_{\mathcal{J}}^h) + (K-1)q_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h) = C'(q_k^{BR}(Q_{\mathcal{J}}^h)).$$

By the implicit function theorem

$$\frac{dq_k^{BR}}{dQ_{\mathcal{J}}^h} = \frac{B''(q_k^{BR}(Q_{\mathcal{J}}^h) + (K-1)q_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h)}{C''(q_k^{BR}(Q_{\mathcal{J}}^h)) - B''(q_k^{BR}(Q_{\mathcal{J}}^h) + (K-1)q_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h)} \left((K-1)\frac{dq_k}{dQ_{\mathcal{J}}^h} + 1 \right).$$

Again, in case of disagreement $t_k = 0$. With this in mind,

$$\begin{aligned} \pi_0 - \pi_{0,-k}^d &= B(Q) - B(q_k^{BR}(Q_{\mathcal{J}}^h) + (K-1)q_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h) - t_k, \\ \pi_k - \pi_{k,-k}^d &= B(Q) - B(q_k^{BR}(Q_{\mathcal{J}}^h) + (K-1)q_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h) + t_k - C(q_k) + C(q_k^{BR}(Q_{\mathcal{J}}^h)). \end{aligned}$$

At each bilateral negotiation between 0 and k Nash-in-Nash solves $\max\limits_{(q_k \geq 0, t_k \geq 0)} NP_{0,k}$. The FOCs of this maximization problem give:

$$\begin{aligned} t_k &= \frac{1}{2}\left(C(q_k) - C(q_k^{BR}(Q_{\mathcal{J}}^h))\right), \\ C'(q_k(Q_{\mathcal{J}}^h)) &= 2B'(Q). \end{aligned}$$

with $Q = Kq_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h$.

By the implicit function theorem the reaction function $q_k(Q_{\mathcal{J}}^h)$ has the derivative given by:

$$\frac{dq_k}{dQ_{\mathcal{J}}^h} = \frac{2B''(Kq_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h)}{C''(q_k(Q_{\mathcal{J}}^h)) - 2KB''(Kq_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h)} \leq 0.$$

At stage 1 the leader negotiates with the coalition

$$\max_{(q_j, t_j)} NBP_{0,\mathcal{J}} = \left(\pi_0 - \pi_{0,-\mathcal{J}}^d\right) \prod_{j \in \mathcal{J}} \left(\pi_j - \pi_{j,-\mathcal{J}}^d\right). \tag{36}$$

Players know that in case of agreement the resulting $Jq_j = Q_{\mathcal{J}}$ in stage 1 becomes known in stage 2, following the rule $q_k(Q_{\mathcal{J}}^h)$, with $Q_{\mathcal{J}} = Q_{\mathcal{J}}^h$. This means what is decided today becomes history tomorrow. Players in the coalition anticipate this.

We consider first the case when all singleton players reach an agreement, even when the coalition does not agree. Each player j optimally (and simultaneously) chooses q_j, anticipating the behavior of singleton players in stage 2. We denote by q_j^{BR} the emission reduction by player j in case of disagreement, given by

$$q_j^{BR} = arg\max_{q_j \geq 0} B(q_j + Q_{-\mathcal{J}} + Kq_k(q_j + Q_{-\mathcal{J}})) - C(q_j).$$

with $Q_{-\mathcal{J}} = \sum_{i \neq j, i \in \mathcal{J}} q_i$.

The FOC of this maximization problem (interior solution) is given by

$$B'(q_j^{BR}(Q_{-\mathcal{J}}) + Q_{-\mathcal{J}} + Kq_k(q_j + Q_{-\mathcal{J}})) = C'(q_j^{BR}(Q_{-\mathcal{J}})).$$

Such a condition, together with $Q_{-\mathcal{J}} = (J-1)q_j^{BR}$, determines q_j^{BR}. With this in mind, net payoffs are

$$\begin{aligned}
\pi_0 - \pi_{0,-\mathcal{J}}^d &= B(Q) - B(Jq_j^{BR} + Kq_k(Jq_j^{BR})) - \sum_{j \in \mathcal{J}} t_j, \\
\pi_j - \pi_{j,-\mathcal{J}}^d &= B(Q) - B(Jq_j^{BR} + Kq_k(Jq_j^{BR})) + t_j - C(q_j) + C(q_j^{BR}).
\end{aligned}$$

Recall that $Q = \sum_{n \in \mathcal{J}} q_n + Kq_k(\sum_{n \in \mathcal{J}} q_n)$. At the bilateral negotiation between 0 and $\mathcal{J}$ Nash-in-Nash solves $\max_{(q_j \geq 0, t_j \geq 0)} NBP_{0,\mathcal{J}}$ given by (5). The FOCs of this maximization problem (interior solution) give:

$$\begin{aligned}
t_j(J) &= \frac{1}{J+1}\left(C(q_j) - C(q_j^{BR})\right), \\
C'(q_j) &= (J+1)B'(Q)\left(1 + K\frac{d\ q_k}{d\ Q_{\mathcal{J}}}\right) \leq (J+1)B'(Q).
\end{aligned}$$

Given that marginal cost is non-decreasing, we have that q_j is smaller in this stackelberg setting than in the static setting.

Proof of Proposition 8

At stage 1 we assume that the singletons bargain first with the leader. At stage 2 the leader bargains with the coalition. The game is solved backwards, starting at stage 2. We will denote by $Q_{\mathcal{K}}^h$ the total emission reduction by the players outside the coalition. In history h there could have been agreements or not, but at stage 2 it is *known* whether that is the case, and the corresponding quantity $Q_{\mathcal{K}}^h$. We assume that all possible agreements are beneficial in equilibrium and will check the conditions that guarantee so. We denote the equilibrium agreement emission reduction by $q_j(Q_{\mathcal{K}}^h)$, as it is a function of the known emission by players k not in the coalition. In case of disagreement with the leader, player $j \in \mathcal{J}$ optimally chooses her disagreement emission reduction that we denote by $q_j^{BR}(Q_{\mathcal{K}}^h)$, given by

$$q_j^{BR}(Q_{\mathcal{K}}^h) = arg\max_{q_j \geq 0} B(q_j + \sum_{h \in \mathcal{J}, h \neq j} q_h + Q_{\mathcal{K}}^h) - C(q_j).$$

where $\sum_{h\in\mathcal{J},h\neq j} q_h$ and $Q^h_{\mathcal{J}}$ are taken as given. The FOC of this maximization problem (interior solution) is given by

$$B'(q_j^{BR}(Q^h_{\mathcal{K}}) + \sum_{h\in\mathcal{J},h\neq j} q_h + Q^h_{\mathcal{K}}) = C'(q_j^{BR}(Q^h_{\mathcal{K}})).$$

By symmetry, $q_h = q_j^{BR}(Q^h_{\mathcal{K}})$ for $h \in \mathcal{J}$, and the FOC becomes

$$B'(Jq_j^{BR}(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}}) = C'(q_j^{BR}(Q^h_{\mathcal{K}})).$$

By the implicit function theorem

$$\frac{dq_j^{BR}}{dQ^h_{\mathcal{K}}} = \frac{B''(Jq_j^{BR}(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}})}{C''(q_j^{BR}(Q^h_{\mathcal{K}})) - JB''(Jq_j^{BR}(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}})} \leq 0.$$

In case of disagreement $t_j = 0$. With this in mind,

$$\begin{aligned}
\pi_0 - \pi^d_{0,-\mathcal{J}} &= B(Q) - B(Jq_j^{BR}(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}}) - Jt_j,\\
\pi_j - \pi^d_{j,-\mathcal{J}} &= B(Q) - B(Jq_j^{BR}(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}}) + t_j - C(q_j) + C(q_j^{BR}(Q^h_{\mathcal{K}})).
\end{aligned}$$

At the negotiation between 0 and $\mathcal{J}$ Nash-in-Nash solves $\max\limits_{(q_j\geq 0, t_j\geq 0)} NBP_{0,\mathcal{J}}$. The FOCs of this maximization problem give:

$$\begin{aligned}
t_j &= \frac{1}{J+1}\left(C(q_j) - C(q_j^{BR}(Q^h_{\mathcal{K}}))\right),\\
C'(q_j(Q^h_{\mathcal{K}})) &= (J+1)B'(Q),
\end{aligned}$$

with $Q = Jq_j(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}}$.

By the implicit function theorem the reaction function $q_k(Q^h_{\mathcal{J}})$ has the derivative given by:

$$\frac{dq_j}{dQ^h_{\mathcal{K}}} = \frac{(J+1)B''(Jq_j(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}})}{C''(q_j(Q^h_{\mathcal{K}})) - (J+1)JB''Jq_j(Q^h_{\mathcal{K}}) + Q^h_{\mathcal{K}})} \leq 0.$$

At stage 1, the leader bargains simultaneously with the singletons $k \in \mathcal{K}$. Players know that in case of agreement the resulting $\sum_{k\in\mathcal{K}} = Q_{\mathcal{K}}$ in stage 1 becomes known in stage 2, following the rule $q_j(Q^h_{\mathcal{K}})$, with $Q_{\mathcal{K}} = Q^h_{\mathcal{K}}$. This means what is decided today becomes history tomorrow. Singleton players anticipate this.

We consider first the case when the coalition reaches an agreement, even when a singleton does not agree. Each player k optimally (and simultaneously) chooses q_k, anticipating the behavior of coalition players in stage 2. We denote by q_k^{BR} the emission reduction by player k in case of disagreement, given by

$$q_k^{BR} = arg\max_{q_k\geq 0} B(q_k + \sum_{h\in\mathcal{K},h\neq k} q_h + \sum_{j\in\mathcal{J}} q_j\left(q_k + \sum_{h\in\mathcal{K},h\neq k} q_h\right)) - C(q_k),$$

where $\sum_{h\in\mathcal{K},h\neq k} q_h$ is taken as given and equal to the equilibrium emission reductions.

The FOC of this maximization problem (interior solution) is given by

$$B'(q_k^{BR} + \sum_{h\in\mathcal{K},h\neq k} q_h + \sum_{j\in\mathcal{J}} q_j \left(q_k + \sum_{h\in\mathcal{K},h\neq k} q_h\right)) = C'(q_k^{BR}).$$

Recall that in case of disagreement $t_k = 0$. With this in mind,

$$\begin{aligned}
\pi_0 - \pi_{0,-k}^d &= B(Q) - B(q_k^{BR} + \sum_{h\in\mathcal{K},h\neq k} q_h + \sum_{j\in\mathcal{J}} q_j \left(q_k^{BR} + \sum_{h\in\mathcal{K},h\neq k} q_h\right)) - t_k, \\
\pi_k - \pi_{k,-k}^d &= B(Q) - B(q_k^{BR} + \sum_{h\in\mathcal{K},h\neq k} q_h + \sum_{j\in\mathcal{J}} q_j \left(q_k^{BR} + \sum_{h\in\mathcal{K},h\neq k} q_h\right)) + t_k - C(q_k) + C(q_k^{BR}).
\end{aligned}$$

At each bilateral negotiation between 0 and k Nash-in-Nash solves $\max_{(q_k\geq 0, t_k\geq 0)} NP_{0,k}$. The FOCs of this maximization problem give:

$$\begin{aligned}
t_k &= \frac{1}{2}\left(C(q_k) - C(q_k^{BR})\right), \\
C'(q_k) &= 2B'(Q)\left(1 + J\frac{dq_j}{dQ_{\mathcal{K}}^h}\right) \leq 2B'(Q),
\end{aligned}$$

with $Q = q_k + \sum_{h\in\mathcal{K},h\neq k} q_h + \sum_{j\in\mathcal{J}} q_j \left(q_k + \sum_{h\in\mathcal{K},h\neq k} q_h\right)$. In equilibrium, $q_k = q_h$ for $h,k\in\mathcal{K}$ and Q becomes $Q = Kq_k + Jq_j\,(Kq_k)$

Given that marginal cost is non-decreasing, we have that q_k is smaller in this stackelberg setting than in the static setting.

Proof of Proposition 9

To prove part 1 of the proposition, consider stage 2 when the leader negotiates with the singletons: From $c = 2B'(Q_{\mathcal{J}}^h + Kq_k)$, we obtain

$$q_k(Q_{\mathcal{J}}^h) = \frac{1}{K}\left[\left(\frac{2B(1)\alpha_B}{c}\right)^{\frac{1}{1-\alpha_B}} - Q_{\mathcal{J}}^h\right].$$

We see that $\frac{d\ q_k}{d\ Q_{\mathcal{J}}} = -\frac{1}{K}$ and that $Kq_k(Q_{\mathcal{J}}^h) + Q_{\mathcal{J}}^h = nq^s$ is constant. From $B'(q_k^{BR} + Q_{\mathcal{J}}^h + (K-1)q_k) = c$, we obtain that

$$q_k^{BR}(Q_{\mathcal{J}}^h) = q_k - (2^{\frac{1}{1-\alpha_B}} - 1)nq^s.$$

Note that this value is negative because $(2^{\frac{1}{1-\alpha_B}} - 1)nq^s > nq^s > q_k$. Hence, $q_k^{BR}(Q_{\mathcal{J}}^h) = 0$. At stage 1 with the coalition, we cannot have

$$B'(nq^s) = c = (J+1)B'(nq^s),$$

so $q_j^{BR} = 0$. Given that $\frac{d\ q_k}{d\ Q_{\mathcal{J}}} = -\frac{1}{K}$ then $q_j = 0$ as far as $K > 1$. It shows that the coalition contributes nothing when they bargain first. The total amount of abatement is $Q = nq^s = \left(\frac{2B(1)\alpha_B}{c}\right)^{\frac{1}{1-\alpha_B}}$ and is only made by the singletons which receive $t_k = \frac{1}{2}C(q_k)$. It gives the payoffs

$$\begin{aligned}
\pi_0^{C,S} &= B(Q) - Kt_k = (1-\alpha_B)\, B(nq^S) = 2^{\frac{\alpha_B}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}} (1-\alpha_B)\,, \\
\pi_j^{C,S} &= B(Q) = B(nq^S) = 2^{\frac{\alpha_B}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}}\,, \\
\pi_k^{C,S} &= B(Q) + t_k - C(q_k) = \left(1 - \frac{\alpha_B}{K}\right) B(nq^S) = 2^{\frac{\alpha_B}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}} \left(1 - \frac{\alpha_B}{K}\right).
\end{aligned}$$

We now prove part 2 of the proposition.

Stage 2 with the coalition: From $c = (J+1)B'(Q_{\mathcal{K}}^h + Jq_j)$, we obtain

$$q_j(Q_{\mathcal{K}}^h) = \frac{1}{J}\left[\left(\frac{(J+1)B(1)\alpha_B}{c}\right)^{\frac{1}{1-\alpha_B}} - Q_{\mathcal{K}}^h\right].$$

We see that $\frac{d\ q_j}{d\ Q} = -\frac{1}{J}$ and that $Jq_j(Q_{\mathcal{K}}^h) + Q_{\mathcal{K}}^h) = nq^s\left(\frac{J+1}{S}\right)^{\frac{1}{1-\alpha_B}}$ is constant. From

$$B'(Jq_j^{BR} + Q_{\mathcal{K}}^h) = c,$$

we obtain that

$$q_j^{BR}(Q_{\mathcal{K}}^h) = \frac{1}{J}\left[\frac{B(1)\alpha_B}{c}^{\frac{1}{1-\alpha_B}} - Q_{\mathcal{K}}^h\right]$$

At stage 1 with the coalition, we cannot have

$$B'(nq^s\left(\frac{J+1}{S}\right)^{\frac{1}{1-\alpha_B}}) = c = 2B'(nq^s\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}),$$

so there is no interior solution for $q_k^{BR} > 0$ and $q_k > 0$. It implies that $q_k^{BR} = 0$. Futhermore we cannot have $c = 2B'(nq^s\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}})$ and $c = (J+1)B'(nq^s\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}})$, so we obtain $q_k = 0$.

The singletons didn't abate and the total amount of abatement made by the insiders is $Q = \left(\frac{(J+1)B(1)\alpha_B}{c}\right)^{\frac{1}{1-\alpha_B}} = nq^s\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}}$ and they receive $t_j = \frac{1}{J+1}\left(C(q_j) - C\left(q_j^{BR}(Q_{\mathcal{K}}^h)\right)\right)$.

If the coalition bargains at the end, the reduction will be larger than when the singletons bargain at the end since $nq^s\left(\frac{J+1}{2}\right)^{\frac{1}{1-\alpha_B}} > nq^s$.

It gives the payoffs

$$\begin{aligned}
\pi_0^{S,C}(J) &= B(Q) - Jt_j = \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}} \left((1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} + \frac{\alpha_B}{J+1}\right), \\
\pi_j^{S,C}(J) &= B(Q) + t_j - C(q_j) = \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}} \left((1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{J(J+1)}\right), \\
\pi_k^{S,C}(J) &= B(Q) = (J+1)^{\frac{\alpha_B}{1-\alpha_B}} \left(\frac{\alpha_B}{c}\right)^{\frac{\alpha_B}{1-\alpha_B}} B(1)^{\frac{1}{1-\alpha_B}}.
\end{aligned}$$

It can be shown that $\pi_0^{S,C}(J) > \pi_0^{C,S}$ since $(1-\alpha_B)\left((J+1)^{\frac{\alpha_B}{1-\alpha_B}} - 2^{\frac{\alpha_B}{1-\alpha_B}}\right) + \frac{\alpha_B}{J+1} > 0$. The leader prefers to negotiate first with the singletons and second with the coalition. It completes part one of the proposition.

To prove part 2 of the propositio, it straightforward to show that $\pi_0^{S,C}(J) = \pi_0(J)$ and $Q^{S,C} = Q$.The leader is indifferent in terms of payoffs between the stackelberg case when he negotiates with the coalition at the second stage and simultaneously with the coalition and the singletons. In both cases the total abatement is the same as well as the transfers because the singletons fully free ride.

Based on the internal and external stability conditions, we analyse the function $\Pi(J) = \pi_j(J) - \pi_k(J-1)$. Note that the internal stability condition for J is satisfied when $\Pi(J) \geq 0$ and the external stability condition for J is satisfied when $\Pi(J+1) \leq 0$.

Recall that

$$\pi_j(J) - \pi_k(J-1) = \frac{B(nq^s)}{2^{\frac{\alpha_B}{1-\alpha_B}}} \left((J+1)^{\frac{\alpha_B}{1-\alpha_B}} (1-\alpha_B) - \frac{\alpha_B}{J(J+1)} - J^{\frac{\alpha_B}{1-\alpha_B}}\right),$$

for $J \geq 3$. However the case $J = 1$ refers to the situation where in the first stage of the protocol $n-1$ agents negotiates withe the leader simultaneously and one agent negotiates in the second stage with the leader. In that case it means that all the abatement will be implemented by this last player. Consequently he will prefer to bargain simultaneously with the other sigletons in the first stage, meaning that $J = 1$ is not a stable coalition.

The rest of the proof is the same as the one in proposition 4 since tha analysis of the fonction $H(J) = (1-\alpha_B)(J+1)^{\frac{\alpha_B}{1-\alpha_B}} - J^{\frac{\alpha_B}{1-\alpha_B}} - \frac{\alpha_B}{J(J+1)}$, for $J \geq 3$ is the same.

This completes the proof of the proposition. ■