Environment and Development Economics 18: 51–68 © Cambridge University Press 2012
doi:10.1017/S1355770X12000289

Asymmetries in international environmental agreements

YULIA PAVLOVA
MTT Agrifood Research, Helsinki, Finland. Email: yulia.pavlova@mtt.fi

AART DE ZEEUW
Department of Economics, CentER and TSC, Tilburg University, P.O. Box 90153, Tilburg, The Netherlands. Email: a.j.dezeeuw@uvt.nl

Submitted September 7, 2011; revised January 7 and May 8, 2012; accepted June 12, 2012; first published online July 25, 2012

ABSTRACT. This paper considers self-enforcing international environmental agreements when countries are asymmetric with respect to emission-related benefits and environmental damage. Considering these asymmetries simultaneously yields large stable coalitions, also without the option of transfers between signatories. However, these large stable coalitions are only possible if they include countries that have relatively high marginal benefits and a relatively low marginal environmental damage. This type of countries hardly contributes to the common good and the gains of cooperation from including this type of countries in the stable coalition are small. This confirms a persistent result in this literature that large stable coalitions usually go hand in hand with low gains of cooperation. Without the option of transfers it is always better to have a small stable coalition with countries that matter than a large stable coalition with countries that do not matter. Only with transfers might a large stable coalition be able to perform better.

1. Introduction

Managing the global commons through institutions like international environmental agreements (e.g., the Kyoto Protocol for climate change) has proven to be very difficult. The standard theoretical model, introduced by Barrett (1994), shows that free-rider incentives are stronger than incentives to cooperate. This model is a two-stage non-cooperative game, developed by d'Aspremont *et al.* (1983) in cartel theory. In the first stage countries decide whether they want to be a member of the coalition or not, and in the second stage the coalition and the individual outsiders choose their production and emission levels. The subgame perfect Nash equilibrium of the resulting two-stage game is usually interpreted in terms of internal and external stability of the agreement: no signatory has an incentive to leave

This paper was presented in the 8th GTP Meeting on Global and International Issues at UC Riverside, USA, July 11–12, 2011. Comments from the participants and especially from two anonymous referees are gratefully acknowledged.

the agreement and no outsider has an incentive to join, which is the Nash equilibrium of the first stage given the equilibrium in the second stage. The basic model yields the pessimistic conclusion that the size of the stable coalition is small, regardless of the total number of countries. This model has been challenged by models with other behavioural assumptions (see, e.g., de Zeeuw, 2008, for a discussion) which can lead to more optimistic conclusions. These other models, however, assume that agreements fully collapse if a deviation from agreed behaviour occurs. Because these other behavioural assumptions are not considered to be very realistic in the international context, most of the literature continues to use the original basic model.

The model in this paper allows for asymmetries. This is realistic because countries are simply not identical in marginal benefits related to emissions and in vulnerability to environmental damage. For example, countries with an energy-intensive economic structure have higher costs in reducing greenhouse gas emissions, and countries that have already reduced emissions have higher costs, because the low-hanging fruit has been picked. On the other hand, countries below sea-level are more vulnerable to sea-level rise, and countries that rely strongly on agriculture are more vulnerable to temperature change. The literature on asymmetries has not yet been fully developed. Botteon and Carraro (1997) analyse an empirical game for a small set of regions in the world with calibrated costs and benefits and basically conclude again that the size of stable coalitions is small. Barrett (2001) changes the rules of the game somewhat and shows that if one type of country with a higher interest in abatement can bribe the other type to join the agreement, the final result can be improved. However, his result is based on a model in which countries either abate or pollute so that leaving the stable coalition implies that all countries will pollute. In the standard model, emission levels are adjusted but not all the way down to full pollution. It remains to be seen what happens in such a situation. McGinty (2007) gives Stackelberg leadership to the countries in the agreement and derives stability conditions for quadratic benefits and costs, allowing for transfers between the signatories. The model is too complicated to be able to obtain an analytical solution, but numerical exercises show that the size of the stable coalitions is still small in the asymmetric case but that the gains of cooperation can be substantial. Ruis and de Zeeuw (2010) analyse an empirical game for a small set of regions with calibrated costs and benefits, with and without transfers, and show that the largest stable coalition consists of countries with different characteristics in terms of costs of emission reductions and vulnerability to climate change. A number of other recent papers (Biancardi and Villani, 2010; Osmani and Tol, 2010; Pintassilgo *et al.*, 2010) consider asymmetries between countries but all rely on numerical exercises, so that there is still a need for analytical results in this area. The paper that comes closest to our paper is Fuentes-Albero and Rubio (2010) (with a correction by Glanemann, 2012) who derive stability conditions for the model with quadratic costs and linear damage. They consider both the cases with and without transfers, but they restrict themselves to either asymmetry in emission reduction costs or asymmetry in vulnerability. They conclude that the size of the stable coalition can only become large in case of transfers and asymmetry in vulnerability, and only by including the countries that are less vulnerable.

Our paper considers asymmetries in benefits and vulnerability simultaneously, for the model with quadratic benefits related to emissions and linear environmental damage. Costs of emission reductions are identified as lower benefits of production. Internal and external stability conditions are derived for the general case and then specified for two types of country. It will be shown that the size of the stable coalition can also become large in the case without transfers, but again only by including one of the two types, namely the countries that are less vulnerable and have higher marginal benefits. The flexibility of considering both asymmetries simultaneously allows for large stable coalitions in the case without transfers. The reason is that stability requirements can be satisfied for both types when there are two degrees of freedom, namely asymmetry in benefits and asymmetry in vulnerability. However, the asymmetry has to be rather strong to get the result. With transfers, the conditions can be relaxed but not by much, although the feasible set of parameter values becomes larger. It follows that the result is mainly driven by the degree of asymmetry, which is also concluded by Barrett (2001). The good news is that the size of the stable coalition can be large but the bad news is that the size can only be large by including a type of country that will not contribute much to the common good and will not trigger much more in contributions from the other coalition members either. This confirms a persistent result in this literature: large stable coalitions go hand-in-hand with small gains of cooperation. Indeed, it will be shown that the stable coalition of three countries of the type that matters performs better than the stable coalition of two countries of this type and all the countries of the other type. This always holds in the case without transfers. In the case with transfers the large stable coalition may perform better but only if the asymmetry is weaker in marginal benefits but very strong in marginal environmental damage. These requirements are relaxed when the number of countries that are less vulnerable and have higher marginal benefits increases.

The paper is organized as follows. Section 2 presents the basic model and section 3 derives conditions for internal and external stability. In section 4 these conditions are specified for two types of country and it is shown that the maximal size of the stable coalition consists of two countries of one type and all countries of the other type, provided that the asymmetry is strong enough. Section 5 considers the option of transfers and shows that strong asymmetry is still required to make the large stable coalition, but that the feasible set of parameter values is larger. In section 6 it is shown that total emission reductions are larger for the small stable coalition than for the large one in the case without transfers, but that this may turn around in the case with transfers. Section 7 concludes the paper.

2. The model

We consider N non-identical players, countries of the world, each of which emits a pollutant q_i that damages a shared environmental resource.

A one-to-one relationship is assumed between production and emission so that $q_i \geq 0$ denotes both production and emission. Environmental damage is a function of total emissions Q. Each player $i = 1, \ldots, N$ maximizes her objective function π_i, which depends on her own production q_i and on the total emissions Q produced by all players,

$$\max_{q_i} \pi_i = \alpha_i(\beta_i q_i - 0.5q_i^2) - \gamma_i Q, \tag{2.1}$$

where $Q = \sum_{j=1}^{N} q_j$. In the absence of environmental damage, production q_i will be equal to the business-as-usual level denoted by β_i, $\beta_i > 0$. The positive parameters α_i (shifting marginal benefits), β_i and γ_i (vulnerability to environmental damage) can be different for the countries $i = 1, \ldots, N$. Fuentes-Albero and Rubio (2010) use the model with total costs as objectives:

$$\min\{0.5c_i(\delta_i - x_i)^2 + m_i X\}, X = \sum_{j=1}^{N} x_j, i = 1, 2, \ldots, N,$$

where x denotes emission. Note, however, that this is the same model by multiplying with -1 and identifying $x = q$, $\delta = \beta$, $m = \gamma$ and $c = \alpha$.

The countries play a two-stage game. In the second stage, a coalition S of size k plays a non-cooperative game with the $n - k$ individual outsiders but cooperates on the joint objective of the coalition members. In the first stage, all the countries decide whether they want to join the coalition or not. In the second stage, the equilibrium production or emission levels are given by:

$$q_i = \beta_i - \frac{\sum_{j \in S} \gamma_j}{\alpha_i}, i \in S, \tag{2.2a}$$

$$q_i = \beta_i - \frac{\gamma_i}{\alpha_i}, i \notin S. \tag{2.2b}$$

The members of the coalition choose higher emission reductions than the outsiders because they take account of the damage to the other coalition members. In general, high values of γ_i and low values of α_i lead to high emission reductions. Note that high values of γ_i for coalition members trigger higher emission reductions from the other coalition members. Throughout the paper we assume that the business-as-usual levels β_i are high enough to assure non-negative production or emission levels q_i in (2.2a) and (2.2b).

3. Coalition stability

A Nash equilibrium in the first stage, where countries decide on membership of the coalition, implies that the members do not have an incentive to leave the coalition and the outsiders do not have an incentive to join the coalition, given the consequences in the second stage. This yields conditions for so-called internal and external stability, respectively (see, e.g., Barrett, 1994; Finus, 2003). We first derive general expressions for the internal and external stability conditions for the model in section 2.

If coalition member i deviates in the first stage, or leaves the coalition S of size k, the coalition changes to $S \setminus \{i\}$ of size $k-1$, the emission levels (2.2a)–(2.2b) of coalition members and outsiders in the second stage change, and therefore also total emissions Q change. Coalition member i will not deviate, i.e., will not leave the coalition S, if the resulting value of objective π_i in (2.1) for country i does not increase. This yields the internal stability condition. Similarly, if outsider i deviates in the first stage, i.e., joins the coalition S of size k, the coalition changes to $S \cup \{i\}$ of size $k+1$ and emissions change in the second stage. Outsider i will not deviate, or will not join the coalition S, if objective π_i in (2.1) for country i does not improve. This yields the external stability condition.

Proposition 1. *Without transfers, the internal and external stability conditions, respectively, can be written as:*

$$\sqrt{\sum_{j \in S\setminus\{i\}} \frac{1}{\alpha_j}} \left(\sqrt{2\alpha_i}\right) \gamma_i \geq \sum_{j \in S\setminus\{i\}} \gamma_j, \, i \in S, \tag{3.1}$$

$$\sqrt{\sum_{j \in S} \frac{1}{\alpha_j}} \left(\sqrt{2\alpha_i}\right) \gamma_i \leq \sum_{j \in S} \gamma_j, \, i \notin S. \tag{3.2}$$

Proof: See appendix. □

Note that the business-as-usual emission level β does not play a role in the internal and external stability conditions. Furthermore, if external stability does not hold for country i outside coalition S, it follows immediately that internal stability must hold for country i in coalition S extended with country i. Therefore, we will only focus on the internal stability conditions.

If we only have one-sided asymmetry in the sense that either all the net-benefit parameters α are the same or all the damage parameters γ are the same, it is easy to see that the size of the stable coalition cannot be larger than three, which confirms the result in Fuentes-Albero and Rubio (2010) and is the classical result in this type of model. This is shown in the following corollary.

Corollary. *If either $\alpha_i = 1$, $i = 1, 2, \ldots, N$, or $\gamma_i = 1$, $i = 1, 2, \ldots, N$, the size of the stable coalition cannot be larger than 3.*

Proof: If $\alpha_i = 1$, $i = 1, 2, \ldots, N$, the internal stability condition (3.1) becomes

$$\sqrt{2(k-1)}\gamma_i \geq \sum_{j \in S\setminus\{i\}} \gamma_j, \quad i \in S.$$

Summation over i leads to

$$\sqrt{2(k-1)}\sum_{i\in S}\gamma_i \geq (k-1)\sum_{i\in S}\gamma_i,$$

so that $k \leq 3$.

If $\gamma_i = 1, i = 1, 2, \ldots, N$, the internal stability condition (3.1) becomes

$$2\sum_{j\in S\backslash\{i\}}\frac{\alpha_i}{\alpha_j} \geq (k-1)^2, \quad i \in S.$$

The smallest α_i leads to

$$2(k-1) \geq 2\sum_{j\in S\backslash\{i\}}\frac{\alpha_i}{\alpha_j} \geq (k-1)^2,$$

so that $k \leq 3$. □

Therefore, we need two-sided asymmetry in both α and γ in order to be able to get larger stable coalitions. In the sequel we consider only two types of countries, in order to make the analysis tractable, but these countries can differ in both parameters.

4. Two-sided asymmetry and two types of countries

Suppose there are $N_1 \geq 2$ countries of type 1 with objective parameters (α_1, γ_1) and $N_2 \geq 2$ countries of type 2 with objective parameters (α_2, γ_2). Furthermore, the number of countries of type 1 in the coalition S is denoted by n_1 and the number of countries of type 2 in the coalition is denoted by n_2. The net benefits of a type 1 country in and out of the coalition are denoted by π_1^c and π_1^o, respectively, and the net benefits of a type 2 country in and out of the coalition are denoted by π_2^c and π_2^o, respectively. Define

$$\alpha := \frac{\alpha_1}{\alpha_2}, \quad \gamma := \frac{\gamma_1}{\gamma_2}.$$

Note that if a type 1 country leaves the coalition S, the coalition consists of $n_1 - 1$ type 1 countries and n_2 type 2 countries. Similarly, if a type 2 country leaves the coalition S, the coalition consists of n_1 type 1 countries and $n_2 - 1$ type 2 countries. By applying Proposition 1, the condition for internal stability (3.1) for a type 1 country becomes

$$2(n_1 + n_2\alpha - 1) \geq \left(n_1 + \frac{n_2}{\gamma} - 1\right)^2 \tag{4.1}$$

and the condition for internal stability (3.1) for a type 2 country becomes

$$2\left(\frac{n_1}{\alpha} + n_2 - 1\right) \geq (n_1\gamma + n_2 - 1)^2. \tag{4.2}$$

It is clear that the flexibility in both α and γ stands a better chance of satisfying these two internal stability conditions than flexibility in one of these

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

parameters only. A large γ makes it easier to satisfy the first condition and harder to satisfy the second condition but for α it is the other way around.

Note that, under symmetry where $\alpha = \gamma = 1$, it follows immediately that $n_1 + n_2 \leq 3$ which is of course not surprising. It is interesting, however, that the same result occurs if α and γ are both smaller than 1 and if α and γ are both larger than 1. In the first case, the first internal stability condition (4.1) yields

$$2(n_1 + n_2 - 1) \geq 2(n_1 + n_2\alpha - 1) \geq \left(n_1 + \frac{n_2}{\gamma} - 1\right)^2 \geq (n_1 + n_2 - 1)^2,$$

and in the second case, the second internal stability condition (4.2) yields

$$2(n_1 + n_2 - 1) \geq 2\left(\frac{n_1}{\alpha} + n_2 - 1\right) \geq (n_1\gamma + n_2 - 1)^2 \geq (n_1 + n_2 - 1)^2.$$

In both these cases it follows immediately that $n_1 + n_2 \leq 3$. Therefore, larger stable coalitions are only possible if either α is smaller than 1 and γ is larger than 1 or the other way around. Without loss of generality, we shall further assume $\gamma > 1$. It follows that $\alpha < 1$, i.e., countries of type 2 have a higher marginal benefit and a lower marginal damage. We can now show the following proposition.

Proposition 2. *The maximal size of the stable coalition consists of two countries of type 1 and N_2 countries of type 2, provided that γ is large enough and α is small enough.*

Proof: First note that, if there is only one type, the internal stability conditions (4.1)–(4.2) simplify to either $n_1 \leq 3$ (if $n_2 = 0$) or $n_2 \leq 3$ (if $n_1 = 0$), which is the standard result in this literature.

The internal stability conditions (4.1)–(4.2) can be rewritten as

$$\alpha \geq p := \frac{\left(n_1 + \frac{n_2}{\gamma} - 1\right)^2 - 2(n_1 - 1)}{2n_2}, \tag{4.3}$$

$$\alpha \leq r := \frac{2n_1}{(n_1\gamma + n_2 - 1)^2 - 2(n_2 - 1)}. \tag{4.4}$$

This implies that a stable coalition (n_1, n_2) requires that

$$p \leq \alpha \leq r, \quad \alpha < 1, \quad \gamma > 1. \tag{4.5}$$

Both p and r are decreasing in $\gamma > 1$ and therefore (taking $\gamma \to \infty$ to obtain lower bound p_l for p, and $\gamma = 1$ to obtain upper bound r_u for r)

$$p > p_l = \frac{(n_1 - 1)(n_1 - 3)}{2n_2}, \quad r < r_u = \frac{2n_1}{(n_1 + n_2 - 1)^2 - 2(n_2 - 1)}.$$

It is clear that $p_l > r_u$ for $n_1 = 4$. Furthermore,

$$\frac{\partial p_l}{\partial n_1} = \frac{n_1 - 1}{n_2}, \quad \frac{\partial r_u}{\partial n_1} = \frac{-2n_1^2 + 2(n_2-1)(n_2-3)}{\left[n_1^2 + 2(n_2-1)n_1 + (n_2-1)(n_2-3)\right]^2}.$$

If $n_1^2 > (n_2-1)(n_2-3)$, p_l increases and r_u decreases in n_1 for $n_1 \geq 4$. Otherwise, p_l increases faster than r_u in n_1 for $n_1 \geq 4$ because then $n_2 > 6$ and thus

$$\frac{n_1 - 2}{n_2} > \frac{2}{(n_2-1)(n_2-3) + n_1^2} > \frac{-2n_1^2 + 2(n_2-1)(n_2-3)}{\left[n_1^2 + 2(n_2-1)n_1 + (n_2-1)(n_2-3)\right]^2}.$$

It follows that $p > r$ for $n_1 \geq 4$ so that there cannot be a solution to (4.5) for $n_1 \geq 4$.

If $n_1 = 3$, then (4.3)–(4.4) yields

$$p = \frac{n_2 + 4\gamma}{2\gamma^2}, \quad r = \frac{6}{9\gamma^2 + 6(n_2-1)\gamma + (n_2-1)(n_2-3)}.$$

It follows that a solution to (4.5) requires that

$$36\gamma^3 + (33n_2 - 36)\gamma^2 + (n_2-1)(10n_2-12)\gamma + n_2(n_2-1)(n_2-3) \leq 0.$$

This inequality cannot hold for any $n_2 > 0$ and $\gamma > 1$.

If $n_1 = 1$, then (4.3)–(4.4) yields

$$p = \frac{n_2}{2\gamma^2}, \quad r = \frac{2}{(\gamma + n_2 - 1)^2 - 2(n_2-1)}.$$

It follows that a solution to (4.5) requires that

$$(n_2 - 4)\gamma^2 + 2n_2(n_2-1)\gamma + n_2(n_2-1)(n_2-3) \leq 0.$$

This inequality can only have a solution for $n_2 = 1, 2, 3$ so that the coalition $(1, 3)$ with at least $\gamma = 12$ (combined with $\alpha = 1/96$) is a candidate for the largest stable coalition.

If $n_1 = 2$, then (4.3)–(4.4) yields

$$p = \frac{\left(\frac{n_2}{\gamma} + 1\right)^2 - 2}{2n_2}, \quad r = \frac{4}{(2\gamma + n_2 - 1)^2 - 2(n_2-1)}. \tag{4.6}$$

If $n_2 = 1$, it follows that $p > r$ for $\gamma > 1$ so that (4.5) does not have a solution.

If $n_2 \geq 2$, it follows that $p > r$ for $\gamma = 1$ and $p < r = 0$ for $\gamma \to \infty$. Furthermore,

$$\frac{\partial p}{\partial \gamma} = -\frac{n_2 + \gamma}{\gamma^3}, \quad \frac{\partial r}{\partial \gamma} = -\frac{2\gamma + n_2 - 1}{[\gamma^2 + (n_2-1)\gamma + (n_2-1)(n_2-3)/4]^2}.$$

It is straightforward to show that the absolute value of the derivative of p with respect to γ is larger than the absolute value of the derivative of

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

Table 1. *Minimum values for the parameter γ, without and with transfers, and maximum value for the parameter α, without transfers, to get the stable coalition* $(2, N_2)$ *for* $n_1 = 2 \leq N_1$.

N_2	α^*	γ^*	$\gamma^{**}_{\min}$
2	0.046805	4.176012	4
3	0.017345	6.658624	6.464102
4	0.009018	9.101367	8.898979
5	0.005522	11.531267	11.324555
6	0.003727	13.955426	13.745966
7	0.002685	16.376513	16.165151
8	0.002026	18.795761	18.583005
9	0.001583	21.213822	21
10	0.001271	23.631072	23.416408
50	0.000048	120.22045	120
100	0.000012	240.93361	240.71247

r with respect to γ so that p decreases faster in γ than r. It follows that, for each $n_2 \geq 2$, a $\gamma^*(n_2)$ exists so that $p = r$ for $\gamma = \gamma^*(n_2)$ and $p < r$ for $\gamma > \gamma^*(n_2)$. This $\gamma^*(n_2) > 1$ is implicitly given by $p = r$ in (4.6). Note that $r < 1$ so that for $\gamma \geq \gamma^*(n_2)$ all the requirements of (4.5) are fulfilled. Finally, note that for $\gamma = 12$ the coalition $(2, 5)$ is internally stable and dominates the coalition $(1, 3)$. Therefore, the latter drops out as a candidate for the largest stable coalition.

Summarizing, the maximal size of the stable coalition consists of two countries of type 1 and N_2 countries of type 2, provided that $\gamma \geq \gamma^*(N_2)$ and $p(\gamma) \leq \alpha \leq r(\gamma)$. Table 1 gives $\gamma^*(N_2)$ with the corresponding maximal value α^* for α that allows for the stable coalition $(2, N_2)$, for values of N_2.

Doubling N_2 requires a bit more than doubling γ^* and lowering α^* to less than 25 percent. The asymmetry needs to be very strong to get a large stable coalition. □

This pattern was also found in Fuentes-Albero and Rubio (2010), but only in the case of transfers. The reason is that we allow for two-sided asymmetry and then this pattern also occurs without transfers. The result appears to be driven by asymmetry and not necessarily by transfers (see also Barrett, 2001). In the next section we consider the option of transfers.

5. Transfers

The option of transfers may allow the coalition members to allocate their net benefits in such a way that a larger number of countries will not have incentives to leave the coalition so that a larger coalition may satisfy the internal stability conditions (e.g., McGinty, 2007). This implies that the total net benefits of the coalition must be larger than the total of the outside net-benefit options of all the coalition members. It follows that the internal stability condition in the case of transfers is the sum of the internal stability conditions for the case without transfers, since the transfers add up to zero.

In the literature this is called potential internal stability (see, e.g., Eyckmans and Finus, 2004; Weikard, 2009; Pintassilgo *et al.*, 2010). If potential internal stability holds, an allocation rule exists that makes the coalition internally stable. This leads to the following internal stability condition (applying (A.1) from the appendix):

$$n_1\left[\left(\frac{n_1-1}{\alpha_1}+\frac{n_2}{\alpha_2}\right)\gamma_1^2-\frac{1}{2\alpha_1}((n_1-1)\gamma_1+n_2\gamma_2)^2\right]+$$

$$n_2\left[\left(\frac{n_1}{\alpha_1}+\frac{n_2-1}{\alpha_2}\right)\gamma_2^2-\frac{1}{2\alpha_2}(n_1\gamma_1+(n_2-1)\gamma_2)^2\right]\geq 0$$

or

$$\begin{aligned}&n_1[2(n_1-1+n_2\alpha)\gamma^2-((n_1-1)\gamma+n_2)^2]+\\&n_2[2(n_1+(n_2-1)\alpha)-\alpha(n_1\gamma+n_2-1)^2]\geq 0.\end{aligned}\tag{5.1}$$

If $\alpha=1$ this is the same condition as in Fuentes-Albero and Rubio (2010) for one-sided asymmetry in environmental damage, and if $\gamma=1$ it follows from (5.1) that

$$2(n_1+n_2\alpha)(n_1+n_2-1)\geq(n_1+n_2\alpha)(n_1+n_2-1)^2$$

so that $n_1+n_2\leq 3$ holds again.

As in the previous section we assume $\gamma>1$. In the case of two-sided asymmetry with transfers, the result resembles the one that was found without the option of transfers in the previous section, but it is also different: strong asymmetry is still required but the set of parameter values α and γ for which the internal stability condition holds is larger. This is shown in the following proposition.

Proposition 3. *With the option of transfers, the maximal size of the stable coalition consists again of two countries of type 1 and N_2 countries of type 2, provided that γ is large enough, i.e.,*

$$\gamma\geq\gamma^{**}(\alpha,N_2),$$

where

$$\begin{aligned}&\gamma^{**}(\alpha,N_2)=N_2((N_2-1)\alpha+1)+\\&\sqrt{N_2^2((N_2-1)\alpha+1)^2+N_2(N_2-1)(N_2-3)\alpha/2+N_2(N_2-2)}.\end{aligned}\tag{5.2}$$

Proof: The internal stability condition (5.1) can be rewritten as

$$\begin{aligned}&\alpha n_2[n_1(n_1-2)\gamma^2+2n_1(n_2-1)\gamma+(n_2-1)(n_2-3)]+\\&n_1[(n_1-1)(n_1-3)\gamma^2+2n_2(n_1-1)\gamma+n_2(n_2-2)]\leq 0.\end{aligned}\tag{5.3}$$

This inequality cannot hold for $n_1\geq 3$, with $n_2\geq 1$.

If $n_1=1$, inequality (5.3) becomes

$$\alpha[(\gamma^2-2(n_2-1)\gamma-(n_2-1)(n_2-3)]-(n_2-2)\geq 0.\tag{5.4}$$

This inequality always holds for $n_2=1,2$. Furthermore, for each $n_2\geq 3$ and for each $\alpha>0$, a $\gamma'(\alpha,n_2)>1$ exists so that inequality (5.4) holds for

$\gamma \geq \gamma'(\alpha, n_2)$. It follows that the option of transfers allows for the stable coalition $(1, N_2)$ but this situation can be improved, as is shown in the sequel.

If $n_1 = 2$, inequality (5.3) becomes

$$2\gamma^2 - 4n_2(\alpha n_2 - \alpha + 1)\gamma - \alpha n_2^3 + (4\alpha - 2)n_2^2 - (3\alpha - 4)n_2 \geq 0. \qquad (5.5)$$

If $n_2 = 1$, α drops out and inequality (5.5) holds for $\gamma > 1$.

Furthermore, the largest root of the left-hand side of (5.5) is equal to

$$\gamma^{**}(\alpha, n_2) = n_2((n_2 - 1)\alpha + 1) + \sqrt{n_2^2((n_2 - 1)\alpha + 1)^2 + n_2(n_2 - 1)(n_2 - 3)\alpha/2 + n_2(n_2 - 2)}.$$

If $\gamma = 1$, the left-hand side of inequality (5.5) is equal to

$$(\alpha n_2 + 2)(1 - n_2^2)$$

and is therefore negative.

It follows that for each $n_2 \geq 2$ and each $\alpha, \alpha > 0$, inequality (5.5) holds for $\gamma \geq \gamma^{**}(\alpha, n_2) > 1$. Consequently, the maximal size of the stable coalition consists of two countries of type 1 and N_2 countries of type 2, provided that $\gamma \geq \gamma^{**}(\alpha, N_2)$. □

It is clear that $\gamma^{**}(\alpha, N_2)$ is monotonically increasing in α (for $\alpha > 0$ and $N_2 \geq 2$). It follows that, when the asymmetry in α is relaxed, the asymmetry in γ has to be stronger. Moreover, since $\alpha > 0$ it follows that

$$\gamma^{**}(\alpha, N_2) > \gamma^{**}_{min}(N_2) = N_2 + \sqrt{2N_2(N_2 - 1)}.$$

The lower bounds of $\gamma^{**}(N_2)$ are presented in table 1.

The option of transfers allows the coalition $(2, N_2)$ to be stable for a larger set of the parameter values α and γ. In the case without transfers (see Proposition 2), N_2 determines a minimal value for γ and this, in turn, determines a maximal value for α. However, in the case with transfers, larger values of α can lead to the stable coalition $(2, N_2)$, provided that γ is large enough. In order to make the comparison precise, note that, in the case without transfers, the minimal value of γ and the maximal value of α are determined by $p = r$ in (4.6) with $\alpha^* = p = r$. The following lemma is useful:

Lemma. *It holds that*

$$\gamma^{**}(\alpha^*, N_2) = \gamma^*(N_2).$$

Proof: By substituting r for α from (4.6) into (5.5) it follows that $\gamma^{**}(\alpha^*, N_2)$ is implicitly determined by

$$-4\gamma^4 + 4(N_2 + 1)\gamma^3 + (11N_2^2 - 12N_2 - 3)\gamma^2 + N_2(N_2 - 1)^2\gamma + N_2^2(N_2 - 1)(N_2 - 3) = 0.$$

Furthermore, by taking $p = r$ in (4.6) it follows that $\gamma^*(N_2)$ is implicitly determined by the same equation. As was shown earlier, this equation has a unique solution larger than 1, so that the lemma holds. □

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

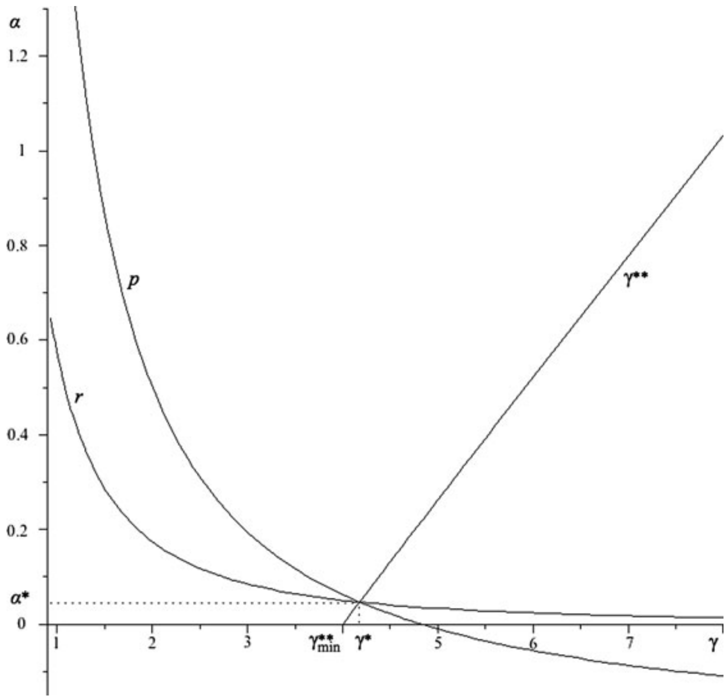

Figure 1. *Graphical representation of relationship between* p, r *and* γ^{**}

In the case without transfers, γ needs to be at least equal to $\gamma^*(N_2)$ in order to get the stable coalition $(2, N_2)$. If $\gamma = \gamma^*(N_2)$, then α must be equal to α^*. If $\gamma > \gamma^*(N_2)$, then α must be on the interval $[\max(0, p), r]$ and because p and r are decreasing in γ, $\alpha < \alpha^*$.

In the case with transfers, γ can be smaller but not much smaller: compare the minimal values $\gamma^{**}_{min}(N_2)$ with $\gamma^*(N_2)$ in table 1. However, if γ gets larger, more values of α yield the stable coalition $(2, N_2)$ in this case. The feasible set (γ, α) is determined by the curve that is given by (5.2). For example, if $N_2 = 2$ this curve and the curves for p and r in (4.6) are:

$$\gamma^{**}(\alpha, 2) = 2(\alpha + 1) + \sqrt{4\alpha^2 + 7\alpha + 4},$$

$$p = \frac{-\gamma^2 + 4\gamma + 4}{4\gamma^2}, \quad r = \frac{4}{4\gamma^2 + 4\gamma - 1}.$$

Figure 1 depicts this in the (γ, α)-plane.

We can conclude that in the case without transfers only the positive area between p and r to the right of $\gamma^*(N_2)$ yields the stable coalition $(2, N_2)$. In the case with transfers the positive area to the right of the curve $\gamma^{**}(\alpha, N_2)$ yields the stable coalition $(2, N_2)$.

Table 1 shows that the minimal value for γ in the case with transfers is hardly smaller than the minimal value for γ in the case without transfers. Moreover, the minimal value for γ in case with transfers $(\gamma^{**}_{min}(N_2))$ results

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

from $\alpha = 0$, implying a very high asymmetry on the side of α. We can conclude that the asymmetry requirements in the case with transfers are not much less demanding than in the case without transfers. With transfers, a larger set of parameter values allows for the large stable coalition but the asymmetry requirements essentially stay intact. The main difference is that it is possible to relax the demands on α, but γ has to increase when increasing α. In the next section we consider total emission reductions.

6. Total emission reductions

First we consider the case without transfers. The largest stable coalition consists of two countries of type 1 and N_2 countries of type 2. However, the countries of type 2 can only be included if γ is large and α is small, which means that type 2 countries have a low marginal damage and a high marginal benefit as compared to type 1 countries. Therefore, type 2 countries will not contribute much to the common good (compared to type 1 countries) and will not trigger much more in contributions from other coalition members when they join the coalition. Alternatively, a stable coalition of three countries of type 1 can be formed. It is not clear from the outset which of the stable coalitions, i.e., $(2, N_2)$ or $(3, 0)$, will perform better in terms of emission reductions.

With a coalition of size $n_1 + n_2$ (and therefore $(N_1 - n_1)$ and $(N_2 - n_2)$ outsiders of type 1 and type 2, respectively), total emission reductions with respect to the business-as-usual emission level $N_1\beta_1 + N_2\beta_2$ follow from (2.2a) and (2.2b) and are given by:

$$n_1 \frac{n_1\gamma_1 + n_2\gamma_2}{\alpha_1} + (N_1 - n_1)\frac{\gamma_1}{\alpha_1} + n_2 \frac{n_1\gamma_1 + n_2\gamma_2}{\alpha_2} + (N_2 - n_2)\frac{\gamma_2}{\alpha_2}$$

or

$$\frac{\gamma_1}{\alpha_1}\left[(n_1 + n_2\alpha)(n_1 + \frac{n_2}{\gamma}) + (N_1 - n_1) + (N_2 - n_2)\frac{\alpha}{\gamma}\right]. \tag{6.1}$$

For the coalitions $(3, 0)$ and $(2, N_2)$, (6.1) results in

$$\frac{\gamma_1}{\alpha_1}\left[N_1 + 6 + N_2\frac{\alpha}{\gamma}\right], \tag{6.2a}$$

$$\frac{\gamma_1}{\alpha_1}\left[N_2^2\frac{\alpha}{\gamma} + 2N_2\alpha + 2\frac{N_2}{\gamma} + N_1 + 2\right], \tag{6.2b}$$

respectively.

The question is whether emission reductions are higher with a coalition $(3, 0)$ or with a coalition $(2, N_2)$. The answer will be given in the following proposition.

Proposition 4. *Without the option of transfers, the total emission reductions are higher with the small stable coalition* $(3, 0)$ *than with the large stable coalition* $(2, N_2)$.

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

Proof: It has to be shown that the expression (6.2a) is larger than the expression (6.2b) or

$$N_2^2\frac{\alpha}{\gamma} + 2N_2\alpha + 2\frac{N_2}{\gamma} - N_2\frac{\alpha}{\gamma} < 4. \tag{6.3}$$

In section 4 it was shown that the coalition $(2, N_2)$ is stable for the minimal value γ^* and the maximal value α^* if $\gamma^*(N_2)$ is the unique solution larger than 1 of $p = r$ where, according to (4.6)

$$p = \frac{\left(\frac{N_2}{\gamma}+1\right)^2 - 2}{2N_2}, \quad r = \frac{4}{(2\gamma + N_2 - 1)^2 - 2(N_2 - 1)},$$

with

$$\alpha^* = p = r.$$

It is straightforward to show that $p > r$ for $\gamma = 2N_2$ and that $p < r$ for $\gamma = N_2/(\sqrt{2} - 1)$ so that

$$2N_2 < \gamma^*(N_2) < \frac{N_2}{\sqrt{2}-1}.$$

Using this and noticing that $\alpha^* = p$, it follows that

$$2\alpha^* N_2 = \left(\frac{N_2}{\gamma^*(N_2)} + 1\right)^2 - 2 < \left(\frac{1}{2} + 1\right)^2 - 2 = \frac{1}{4}.$$

These inequalities yield the higher bounds

$$\frac{N_2}{\gamma^*} < \frac{1}{2}, \quad \alpha^* N_2 < \frac{1}{8},$$

so that

$$N_2^2\frac{\alpha^*}{\gamma^*} + 2N_2\alpha^* + 2\frac{N_2}{\gamma^*} - N_2\frac{\alpha^*}{\gamma^*} < \frac{N_2}{\gamma^*}\alpha^* N_2 + 2\alpha^* N_2 + 2\frac{N_2}{\gamma^*} < \frac{21}{16} < 4.$$

□

We are back to the old grim story in this literature. It is possible to get a large stable coalition but only by including countries of type 2 that do not contribute much to the common good and do not trigger much more in contributions from the other coalition members. This can only be achieved at the expense of a country of type 1. It is better for total emission reductions to have a stable coalition of three countries of type 1 or, to put it differently, to have a small stable coalition of countries that matter.

With the option of transfers, the picture is different. In that case, parameter values of α and γ exist for which the large stable coalition performs better. The total emission reductions are higher with the large stable coalition $(2, N_2)$ than with the small stable coalition $(3, 0)$ if the opposite of (6.3) holds or

$$N_2^2\frac{\alpha}{\gamma} + 2N_2\alpha + 2\frac{N_2}{\gamma} - N_2\frac{\alpha}{\gamma} > 4.$$

A sufficient condition for this inequality to hold is that $\alpha > 2/N_2$. Moreover, in section 5 it was shown that the coalition $(2, N_2)$ is stable provided that $\gamma \geq \gamma^{**}(\alpha, N_2)$ where $\gamma^{**}(\alpha, N_2)$ is given by (5.2). Therefore, with the option of transfers it is possible that the large stable coalition $(2, N_2)$ performs better than the small stable coalition $(3, 0)$. This requires asymmetry to be weaker on the side of marginal benefits but stronger on the side of vulnerability. Note, however, that the asymmetry requirements on α and thus on γ are relaxed when the number N_2 of countries of type 2 increases. In general, we can conclude that the option of transfers is needed to allow for the possibility that the large stable coalition $(2, N_2)$ performs better than the small stable coalition $(3, 0)$.

7. Conclusion

This paper considers self-enforcing international environmental agreements among heterogeneous countries. In a model with individual quadratic benefits of production (or emissions) and linear damage of total emissions, the paper considers stability of coalitions when countries differ in both marginal benefits and marginal damage.

The results of this paper are two-fold. First, we show that large stable coalitions are possible without transfers if the asymmetries are sufficiently large. This extends the result in Fuentes-Albero and Rubio (2010) who only consider one-sided asymmetry in either marginal benefits or marginal damage and need transfers to get the result. It is confirmed that the large stable coalition mainly consists of countries that have a low marginal damage. In the case of two-sided asymmetry without transfers, these countries must have high marginal benefits of emissions as well. Second, we show that, in the absence of transfers, total emission reductions are lower with the large heterogeneous coalition than with the homogeneous coalition of size 3 that is composed of countries with lower marginal benefits of emissions and higher vulnerability to environmental damage. Only with transfers might the large heterogeneous coalition perform better because the asymmetry requirement regarding the marginal benefits can be relaxed.

References

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

Barrett, S. (2001), 'International cooperation for sale', *European Economic Review* **45**: 1835–1850.

Biancardi, M. and G. Villani (2010), 'International environmental agreements with asymmetric countries', *Computational Economics* **36**: 69–92.

Botteon, M. and C. Carraro (1997), 'Burden sharing and coalition stability in environmental negotiations with asymmetric countries', in C. Carraro (ed.), *International*

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

Environmental Negotiations, Strategic Policy Issues, Cheltenham: Edward Elgar, pp. 26–55.
d'Aspremont, C., A. Jacquemin, J. Gabszewicz, and J. Weymark (1983), 'On the stability of collusive price leadership', *Canadian Journal of Economics* **16**: 17–25.
de Zeeuw, A. (2008), 'Dynamic effects on the stability of international environmental agreements', *Journal of Environmental Economics and Management* **55**: 163–174.
Eyckmans, J. and M. Finus (2004), 'An almost ideal sharing scheme for coalition games with externalities', Working Paper Series 2004–14, Center for Economic Studies, K.U. Leuven.
Finus, M. (2003), 'Stability and design of international environmental agreements: the case of transboundary pollution', in H. Folmer and T. Tietenberg (eds), *International Yearbook of Environmental and Resource Economics, 2003/4*, Cheltenham: Edward Elgar, pp. 199–243.
Fuentes-Albero, C. and S.J. Rubio (2010), 'Can international environmental cooperation be bought?', *European Journal of Operational Research* **202**(1): 255–264.
Glanemann, N. (2012), 'Can international environmental cooperation be bought: Comment', *European Journal of Operational Research* **216**(3): 697–699.
McGinty, M. (2007), 'International environmental agreements among asymmetric nations', *Oxford Economic Papers* **59**(1): 45–62.
Osmani, D. and R.S.J. Tol (2010), 'The case of two self-enforcing international agreements for environmental protection with asymmetric countries', *Computational Economics* **36**: 93–119.
Pintissalgo, P., M. Finus, M. Lindroos, and G. Munro (2010), 'Stability and success of regional fisheries management organizations', *Environmental and Resource Economics* **46**(3): 377–402.
Ruis, A. and A. de Zeeuw (2010), 'International cooperation to combat climate change', *Public Finance and Management* **10**(2): 379–404.
Weikard, H.-P. (2009), 'Cartel stability under an optimal sharing rule', *The Manchester School* **77**: 599–616.

Appendix

Proof Proposition 1

Suppose that coalition member i considers deviating in the first stage, or leaving the coalition S of size k. The emission levels in the second stage would change from

$$q_j = \beta_j - \frac{\sum\limits_{l \in S \setminus \{i\}} \gamma_l}{\alpha_j} - \frac{\gamma_i}{\alpha_j},\ j \in S,$$
$$q_j = \beta_j - \frac{\gamma_j}{\alpha_j},\ j \notin S,$$

to

$$q_j = \beta_j - \frac{\sum\limits_{l \in S \setminus \{i\}} \gamma_l}{\alpha_j},\ j \in S \setminus \{i\},$$
$$q_j = \beta_j - \frac{\gamma_j}{\alpha_j},\ j \notin S,\ j = i.$$

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

The levels of total emissions Q before and after country i deviates are, respectively,

$$Q^k = \sum \beta_j - \sum_{j\in S} \frac{1}{\alpha_j} \sum_{j\in S\setminus\{i\}} \gamma_j - \sum_{j\in S} \frac{1}{\alpha_j}\gamma_i - \sum_{j\notin S} \frac{\gamma_j}{\alpha_j},$$

$$Q^{k-1} = \sum \beta_j - \sum_{j\in S\setminus\{i\}} \frac{1}{\alpha_j} \sum_{j\in S\setminus\{i\}} \gamma_j - \sum_{j\notin S} \frac{\gamma_j}{\alpha_j} - \frac{\gamma_i}{\alpha_i},$$

where k and $k-1$ denote the coalition size, so that the change in total emissions is given by

$$Q^{k-1} - Q^k = \sum_{j\in S\setminus\{i\}} \frac{1}{\alpha_j}\gamma_i + \frac{1}{\alpha_i} \sum_{j\in S\setminus\{i\}} \gamma_j.$$

It follows that the net benefits of country i in the second stage, respectively as a member of the coalition (π_i^c) and as an outsider (π_i^o), are

$$\pi_i^c = \alpha_i \left[\beta_i \left(\beta_i - \frac{1}{\alpha_i} \sum_{j\in S\setminus\{i\}} \gamma_j - \frac{1}{\alpha_i}\gamma_i \right) \right.$$
$$\left. -0.5 \left(\beta_i - \frac{1}{\alpha_i} \sum_{j\in S\setminus\{i\}} \gamma_j - \frac{1}{\alpha_i}\gamma_i \right)^2 \right] - \gamma_i Q^k,$$
$$\pi_i^o = \alpha_i \left[\beta_i (\beta_i - \frac{1}{\alpha_i}\gamma_i) - 0.5(\beta_i - \frac{1}{\alpha_i}\gamma_i)^2 \right] - \gamma_i Q^{k-1}.$$

A Nash equilibrium in the first stage or internal stability requires that for all i in S

$$\pi_i^c - \pi_i^o = \sum_{j\in S\setminus\{i\}} \frac{1}{\alpha_j}\gamma_i^2 - 0.5\frac{1}{\alpha_i} \left(\sum_{j\in S\setminus\{i\}} \gamma_j \right)^2 \geq 0 \tag{A.1}$$

or

$$\sqrt{\sum_{j\in S\setminus\{i\}} \frac{1}{\alpha_j}} \left(\sqrt{2\alpha_i}\right) \gamma_i \geq \sum_{j\in S\setminus\{i\}} \gamma_j, \, i \in S.$$

Suppose that outsider i considers deviating in the first stage, or joining the coalition S of size k. The first-order conditions in the second stage would

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press

change from

$$q_j = \beta_j - \frac{\sum\limits_{l \in S} \gamma_l}{\alpha_j}, \; j \in S,$$
$$q_j = \beta_j - \frac{\gamma_j}{\alpha_j}, \; j \notin S,$$

to

$$q_j = \beta_j - \frac{\sum_{l \in S} \gamma_l}{\alpha_j} - \frac{\gamma_i}{\alpha_j}, \; j \in S, \; j = i,$$
$$q_j = \beta_j - \frac{\gamma_j}{\alpha_j}, \; j \notin S, \; j \neq i.$$

The levels of total emissions Q before and after country i deviates are, respectively,

$$Q^k = \sum \beta_j - \sum_{j \in S} \frac{1}{\alpha_j} \sum_{j \in S} \gamma_j - \sum_{j \notin S} \frac{\gamma_j}{\alpha_j},$$
$$Q^{k+1} = \sum \beta_j - \left(\sum_{j \in S} \frac{1}{\alpha_j} + \frac{1}{\alpha_i} \right) \left(\sum_{j \in S} \gamma_j + \gamma_i \right) - \sum_{j \notin S} \frac{\gamma_j}{\alpha_j} + \frac{\gamma_i}{\alpha_i},$$

where $k+1$ and k denote the coalition size, so that the change in total emissions is given by

$$Q^k - Q^{k+1} = \sum_{j \in S} \frac{1}{\alpha_j} \gamma_i + \frac{1}{\alpha_i} \sum_{j \in S} \gamma_j.$$

It follows that the net benefits of country i in the second stage, respectively as an outsider (π_i^o) and as a member of the coalition (π_i^c), are

$$\pi_i^o = \alpha_i \left[\beta_i (\beta_i - \frac{1}{\alpha_i} \gamma_i) - 0.5(\beta_i - \frac{1}{\alpha_i} \gamma_i)^2 \right] - \gamma_i Q^k,$$
$$\pi_i^c = \alpha_i \left[\beta_i (\beta_i - \frac{1}{\alpha_i} \sum_{j \in S} \gamma_j - \frac{1}{\alpha_i} \gamma_i) - 0.5(\beta_i - \frac{1}{\alpha_i} \sum_{j \in S} \gamma_j - \frac{1}{\alpha_i} \gamma_i)^2 \right] - \gamma_i Q^{k+1}.$$

A Nash equilibrium in the first stage or external stability requires that for all i not in S

$$\pi_i^c - \pi_i^o = \sum_{j \in S} \frac{1}{\alpha_j} \gamma_i^2 - 0.5 \frac{1}{\alpha_i} \left(\sum_{j \in S} \gamma_j \right)^2 \leq 0,$$

or

$$\sqrt{\sum_{j \in S} \frac{1}{\alpha_j}} (\sqrt{2\alpha_i}) \gamma_i \leq \sum_{j \in S} \gamma_j, \; i \notin S.$$

□

https://doi.org/10.1017/S1355770X12000289 Published online by Cambridge University Press