The Role of Fixed Cost in International Environmental Negotiations

Basak BAYRAMOGLU* & Jean-François JACQUES^{†‡}

November 10, 2010

Abstract

We investigate the relative efficiency of an agreement based on a uniform standard without transfers and one based on differentiated standards with transfers when strictly identical countries deal with transboundary pollution. We especially ask what role fixed cost plays.

*INRA, UMR Economie Publique; AgroParisTech. Address: Avenue Lucien Brétignières, 78850 Thiverval Grignon, France. Tel: (0)1 30 81 45 35, E-mail: Basak.Bayramoglu@grignon.inra.fr.

[†]LEDa, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75016 Paris. Tel: (0)1 44 05 44 60, E-mail: jacques@dauphine.fr.

[‡]We gratefully acknowledge Bernard Caillaud and the three anonymous referees for their comments. We are particularly indebted to one referee who offered numerous, particularly constructive insights. Brian Copeland, Maia David and Stéphane de Cara also made helpful suggestions. We thank the participants of the European Summer School in Resource and Environmental Economics, Venice (July 2004) and the EUREQua-Université Paris 1 seminar, Paris (September 2004) for comments, and S. Tanis-Plant for fruitful discussions and editorial advice in English. This work was initiated during a research visit at the University British Columbia (Vancouver). We therefore also thank Ivar Ekeland for giving us that opportunity. Any errors are the responsibility of the authors.

Two approaches are examined: the Nash bargaining solution, implicating two countries, and the coalition formation framework, implicating numerous countries and emphasizing self-enforcing agreements. In the former, in terms of welfare, strictly identical countries may wish to reduce their emissions in a non-uniform way under the differentiated agreement. For this result to hold, the fixed cost of investment in abatement technology must be sufficiently high. The nature of the threat point of negotiations, however, also plays a crucial role. As concerns global abatement, the two countries abate more under the uniform agreement than under the differentiated one. In terms of coalition formation when numerous countries are involved, a grand coalition could emerge under a differentiated agreement.

Keywords: transboundary pollution, bargaining, standards, transfers, fixed cost, coalition stability.

JEL: Q50, C71.

Summary: *In this paper, a situation in which identical countries faced with a transboundary pollution problem is investigated. All countries are responsible for this pollution and seek to find an institutional arrangement in order to reduce it. We study the relative efficiency of an agreement based on a uniform standard without transfers and an agreement based on differentiated standards with transfers when two strictly identical countries are dealing with a problem of transboundary pollution. In the case of a uniform agreement, both countries abate and pay the fixed cost of abatement. In the case*

of a differentiated agreement, only one of the two countries would abate and pay for the fixed cost of investment. In return, its efforts would be compensated by monetary transfers. This analysis is carried out using the Nash bargaining solution (Nash (1950)) which makes it possible to determine the equilibrium of a negotiation game implicating two countries. Fixed cost, being part and parcel of abatement technology, plays a major role in determining the relative efficiency of these two types of agreement. We reveal that above a threshold level of fixed cost, the countries may prefer to sign the differentiated agreement, rather than the uniform one. The nature of the threat point of negotiations, however, also plays a crucial role. Furthermore, we show that the two countries abate more under the uniform agreement than under the differentiated one. When this problem is tackled in terms of self-enforcing agreements for numerous countries, a grand coalition could emerge in the case of a differentiated agreement.

1 Introduction

A situation in which similar countries facing the challenge of a transboundary pollution problem comes under study. Each country is both the cause and the victim of this pollution problem. The countries are identical in terms of abatement costs and their willingness to pay for cleaning up the environment they share. They are seeking an institutional arrangement in order to tackle this problem.

This institutional arrangement, in our case, is an International Environmental Agreement (hereafter referred to as an IEA). Such an agreement is based on an abatement standard: this is defined as the reduction of current emissions in order to reach a percentage of the emissions of a base year. In this case, the countries could choose one standard from a set of two: either a uniform standard or a differentiated standard. The uniform standard means that the above-mentioned percentage would be the same for any country signing an agreement. The Montreal Protocol on Substances that Deplete the Ozone Layer fell into this category of agreements. It included a provision which specified a reduction of emissions of CFCs (Chlorofluorocarbons) and halons by 20 percent based on 1986 emission levels, to be accomplished by 1998 (Finus (2001), chapter 11).¹ The differentiated standard means that the percentages would be different according to the country. The Kyoto Protocol on Climate Change (1997) and the Oslo Protocol on Further Re-

¹Another example is the Helsinki Protocol (1985) which suggested a reduction of sulphur dioxide from 1980 levels by 30 percent by 1993.

duction of Sulphur Emissions (1994) are both examples of agreements with differentiated standards.

Within the context a transboundary pollution problem across identical countries, we further define a uniform agreement to be the case where there is reciprocal action. This means that each country undertakes the same abatement effort, and any country involved pays for the fixed cost of investment in the abatement. This uniform abatement is a cost-effective solution given that the countries under study are identical. We call this undertaking an agreement based on a uniform standard without transfers or in short a uniform agreement. Since countries are identical, there is no need for a transfer payment scheme. We next define a differentiated agreement to be the case where there is unilateral action. This means that only some of the countries opt to make an abatement effort, and pay for the fixed cost, while any other countries compensate with transfers.² We refer to this undertaking as the agreement based on differentiated standards with transfers or in short the differentiated agreement.

In this paper, we emphasize the role of environmental protection fixed costs, which are part and parcel of abatement technology, for the outcome of international environmental negotiations. The recent environmental cooperation around the regional pollution problem in the Mediterranean Sea can

²Transfer payments are offered in order to increase participation in IEAs. Some examples of IEAs which include the possibility of transfers between countries are the Fur Seal Treaty (1911), the Montreal Protocol (1987), and the Stockholm Convention on Persistent Organic Pollutants (2001). See respectively Barrett (2003, p.34), Barrett (2003, p.346), and <http://chm.pops.int/default.aspx> for more details.

serve as an example. The Mediterranean Hot Spot Investment Programme (MeHSIP) and the Horizon 2020 initiative constitute the base of the European Union's cooperation with the southern and eastern Mediterranean countries. The objective of MeHSIP is to abate 80 % of Mediterranean pollution by 2020. To help countries to undertake investment projects, the Programme foresees several financing mechanisms, both bilateral and multilateral. The report provided by the Programme Horizon 2020³ states that *“Especially as concerns hazardous wastes, very little has been done so far in the MENA [Middle East and North Africa] countries to take care of this issue, the main reason for this being the high costs of the necessary investments e.g. incineration plants.”* (p.30). The high fixed installation cost of these plants represents some of the investment costs of the projects financed by MeHSIP. It is clear from this example that the abatement of hazardous wastes in these countries would not take place without MeHSIP's funding. As this example, among others, displays, the level of fixed costs in abatement technology can play a role in the outcome of international environmental negotiations.

In this paper, we ask what role fixed cost plays. Given the agreements (uniform agreement and differentiated agreement) as we have defined them, more specifically, we investigate their relative efficiency. To conduct this analysis, we examine two different approaches. Firstly, we adopt the Nash bargaining solution (Nash (1950)) as an equilibrium of a negotiation game implicating two countries. The threat point of negotiations is represented

³For more information: http://ec.europa.eu/environment/enlarg/med/pdf/mehsip_report.pdf

by Nash equilibrium. We will show that three types of Nash equilibria can emerge depending on the level of fixed cost. Secondly, we work within the coalition formation framework, implicating thereby numerous countries and emphasizing self-enforcing agreements. We extend the model provided by Barrett (1994) by including fixed costs in the abatement technology. Here, we assume that signatories simply maximize their joint welfare, whereas non-signatories individually play Nash equilibrium strategies. We then ask if the differentiated agreement could lead to a larger number of signatories than the uniform one.

A considerable amount of literature focuses on the possible explanations for why uniform standards prevail in the presence of asymmetric agents.⁴ In general, unless the countries are similar, uniform standards are less flexible, and therefore less efficient than differentiated standards (Harstad (2007), p.2).⁵ However, studies show that in the case of IEAs, uniform standards are frequently the case (Hoel (1991), p.64; Harstad (2007), p.2). The following arguments are put forward to explain this frequency: the stability of agreements argument (Finus and Rundshagen (1998)), the monetary transfers argument (Bayramoglu and Jacques (2005)) and the trade theory argument

⁴One can find in the literature several arguments explaining the use of uniform standards: the fairness argument (Welsch (1992)), the informational problems argument (Larson and Tobey (1994), Harstad (2007)), the “focal point” argument (Schelling (1960)), the agency problems argument (Boyer and Laffont (1999)).

⁵If the countries have different marginal implementation costs, then uniform standards will increase the total cost of attaining a given environmental objective (Hoel (1992), p.142). In an earlier work, Kolstad (1987) showed that the efficiency losses associated with uniform environmental regulations increase when marginal benefit and cost functions become more steeply sloped.

(Copeland and Taylor (2005)). Finus and Rundshagen (1998) show, in a coalition model, that for global pollutants, governments will decide on an agreement based on a uniform percentage reduction of emissions (known as the quota regime) rather than an agreement based on a uniform tax (known as the tax regime). In the tax regime, countries' net benefits are unevenly distributed. In contrast, the quota regime distributes net benefits more in line with the countries' characteristics, as abatement depends on initial emission levels. Bayramoglu and Jacques (2005) show, in a negotiation game implicating two countries, the possible Pareto superiority of an agreement based on a uniform standard with transfers compared to an agreement based on differentiated standards. The intuition in that case was that if it is less expensive to make transfers payments across the countries in order to incite the country, which benefits less from global abatement but has lower abatement costs to abate, then it is in the interest of both of the countries to sign the uniform agreement. The argument behind the Copeland and Taylor (2005) trade model is that trade in goods can act as a substitute for trade in emission permits. Therefore, uniform emission reductions in a world with freely traded goods can be efficient, even if trade in permits is banned.

Our paper differs from the literature in that it focuses on a distinct issue. Our interest lies in the analysis of whether or not differentiated standards can be optimal for perfectly symmetric countries. McAusland (2005) uses, with a similar aim in mind, a political economy model to highlight the possible inefficiency of uniform environmental regulations for identical countries.

McAusland shows that the harmonization of these regulations across jurisdictions could be negative for both the environment and global welfare, despite the countries being identical in all respects. This happens when politicians are captured by “dirty” industries and if the local effects of damages are sufficiently large. The former condition results in weak environmental policy. The latter condition makes the policy harmonization less effective in internalizing pollution, and also leads to a lower harmonized environmental standard. These two effects are detrimental for the environment and in turn for global welfare. By taking into account fixed cost in abatement technology, in the Nash-bargaining setting for two countries, we show that the uniform agreement is positive for the global environment, but it can be negative for the welfare of countries. In the coalition formation framework, implicating numerous countries, the uniform agreement can lead to a lower coalition size than the differentiated one.

The paper is organized as follows. Section 2 presents the negotiation model based on the Nash-bargaining approach implicating just two countries. The threat point of negotiations, the uniform agreement and the differentiated agreement are analyzed respectively, with the comparison of the individual welfare of each country across the agreements. Section 3 provides an extension of the model within the coalition formation framework, implicating thereby numerous countries and emphasizing self-enforcing agreements in the case of the two institutional arrangements described above. Finally, in Section 4 we discuss our findings in terms of the level of fixed cost, individual

welfare and self-enforcing agreement.

2 The Model

The utility function of country $i = 1, 2$ is written as follows:

$$NB_i = B(a_i + a_{-i}) - C(a_i) \quad (1)$$

where a_i is the individual abatement level of country $i = 1, 2$.

The benefits from global abatement are represented by the function $B(a_i + a_{-i})$, assumed to be increasing and concave. For simplicity, we assume that $B(0) = 0$.

The abatement costs are represented by the function $C(a_i) = c_o + c(a_i)$, when a_i is strictly positive. This function is composed of a fixed cost c_o and a variable cost $c(a_i)$. The variable cost function is assumed to be increasing and convex. We assume that the total cost of a country is zero when it does not abate, i.e., $C(a_i) = 0$ when $a_i = 0$ for $i = 1, 2$.

Throughout the paper, we will illustrate our theoretical results by an example with quadratic benefit and cost functions. The chosen functional forms are the following: $B(x) = \alpha x - \frac{\beta}{2}x^2$ with $x < \frac{\alpha}{\beta}$, and $c(x) = \frac{\gamma}{2}x^2$. Parameters α , β and γ are assumed to be strictly positive.

In this paper, we compare the abatement and welfare levels under the agreement based on a uniform standard without transfers, hereafter denoted

as U, and the agreement based on differentiated standards with transfers, hereafter denoted as DT. Agreement U implies the same level of abatement for the countries ($a_i = a_{-i}$), whereas agreement DT allows different levels of abatement ($a_i = 0$ and $a_{-i} \neq 0$ or, $a_i \neq 0$ and $a_{-i} = 0$). Moreover, under this agreement, the country which undertakes an abatement effort and pays for the fixed cost of investment receives transfer payments from the other country ($t > 0$). A question arises: is there a threshold level of fixed cost above which the DT agreement is better for each country than the U agreement? We shall show that it is not obviously so.

Before analyzing the outcome of the negotiations, we first study the non-cooperative equilibrium of the game, which constitutes the threat point in the negotiations.

2.1 Non-cooperative Equilibria

The non-cooperative game is represented here by Nash equilibrium. The objective of each country is to maximize its utility function by taking the abatement level of the other country as given:

$$\underset{a_i}{Max} NB_i = \underset{a_i}{Max} [B(a_i + a_{-i}) - c_o - c(a_i)] \quad (2)$$

We will show that two threshold levels of fixed cost c_{o1} and c_{o2} define three types of Nash equilibria:

1. For $c_o < c_{o1}$: each country abates (*Type 1 symmetric Nash equilibrium*)
2. For $c_{o1} < c_o < c_{o2}$: only one country (say country 1) abates (*Type 2 asymmetric Nash equilibrium*)
3. For $c_o > c_{o2}$: no country abates (*Type 3 symmetric Nash equilibrium*)

We now define the levels of abatement and the above mentioned threshold levels of fixed cost which limit the range of each type of Nash equilibrium.

Lemma 1: *Let $\hat{a}$ be the abatement of each country at Type 1 symmetric Nash equilibrium. It is characterized in the following way: $B'(2\hat{a}) = c'(\hat{a})$.*

Let $\hat{\hat{a}}$ be the abatement of country 1 at Type 2 asymmetric Nash equilibrium. It is characterized in the following way: $B'(\hat{\hat{a}}) = c'(\hat{\hat{a}})$.

Proof: see proof Proposition 1.

Lemma 2: *Let c_{o1} be the threshold level of fixed cost below which each country abates at Type 1 symmetric Nash equilibrium: c_{o1} is equal to $B(2\hat{a}) - c(\hat{a}) - B(\hat{\hat{a}})$.*

Let c_{o2} be the threshold level of fixed cost above which no country abates at Type 3 symmetric Nash equilibrium: c_{o2} is equal to $B(\hat{\hat{a}}) - c(\hat{\hat{a}})$.

Proof: see proof Proposition 1.

Lemma 3: $c_{o1} < c_{o2}$.

Proof. The function $2B(x) - c(x)$ is concave. It attains its maximum at point x^* . It is increasing for $x < x^*$ and decreasing for $x > x^*$. We know

that x^* is higher than $\widehat{a}$, because $2B'(\widehat{a}) - c'(\widehat{a}) = B'(\widehat{a}) > 0$. We also know that $\widehat{a}$ is higher than $\widehat{a}$, because $B'(\widehat{a}) > B'(2\widehat{a}) = c'(\widehat{a})$. This implies that $B'(\widehat{a}) - c'(\widehat{a}) > 0$. Consequently, $2B(x^*) - c(x^*) > 2B(\widehat{a}) - c(\widehat{a}) > 2B(\widehat{a}) - c(\widehat{a}) > B(2\widehat{a}) - c(\widehat{a})$. The final inequality is related to the concavity of the function $B(\cdot)$ which implies that $2B(x) > B(2x)$ (with $B(0) = 0$. If not, we would consider the function $(B(x) - B(0))$ and obtain an equivalent result). ■

Proposition 1 presents the general conditions under which the three Nash equilibria exist.

Proposition 1: *If $0 < c_o < c_{o1}$, then the equilibrium is a Type 1 symmetric Nash equilibrium with the following levels of utility for each country,*

$$NB_1^* = NB_2^* = B(2\widehat{a}) - c_o - c(\widehat{a}).$$

If $c_{o1} < c_o < c_{o2}$, then the equilibrium is a Type 2 asymmetric Nash equilibrium with the following levels of utility for the countries,
 $NB_1^* = B(\widehat{a}) - c_o - c(\widehat{a}); NB_2^* = B(\widehat{a}).$

If $c_o > c_{o2}$, then the equilibrium is a Type 3 symmetric Nash equilibrium with the following levels of utility for each country,
 $NB_1^* = NB_2^* = 0.$

Proof. At Type 1 equilibrium, the countries maximize their (identical) utility function: $Max_{a_i} [B(a_i + a_{-i}) - c_o - c(a_i)]$. We obtain $a_i = a_{-i} = \widehat{a}$ as the solution of these maximization programs. The payoffs of the countries at

the equilibrium are then equal to $B(2\hat{a}) - c_o - c(\hat{a})$.

At Type 2 equilibrium, the payoff of country i which abates is deduced from the following maximization program: $Max_{a_i}[B(a_i) - c_o - c(a_i)]$ because $a_{-i} = 0$. We obtain $a_i = \hat{a}$ as the solution of this program. At equilibrium, the payoff of the country which does not abate is equal to $B(\hat{a})$, and the payoff of the country which undertakes an abatement effort is equal to $B(\hat{a}) - c_o - c(\hat{a})$.

At Type 3 equilibrium, the payoffs of the countries which do not abate are null.

Let us first show that the payoffs of the countries at Type 1 equilibrium are deduced from Nash equilibrium given the assumptions $c_{o1} > 0$ and, $c_o < c_{o1}$. If country 1 unilaterally deviates (the same for country 2), i.e., if it does not abate it gets $B(\hat{a})$ which is lower than $[B(2\hat{a}) - c_o - c(\hat{a})]$ given the above assumptions, $c_{o1} > 0$ and $c_o < c_{o1}$.

Let us show now that the payoffs of the countries at Type 2 equilibrium are those of Nash equilibrium given the assumption $c_{o1} < c_o < c_{o2}$. Country 1 (which abates) has no interest in deviating if $[B(\hat{a}) - c_o - c(\hat{a})] > 0$, i.e., if $c_o < c_{o2}$. As concerns country 2, it has no incentive to deviate unilaterally if $B(\hat{a}) > [B(2\hat{a}) - c_o - c(\hat{a})]$, i.e., if $c_o > c_{o1}$.

Finally, if $c_o > c_{o2}$, country 1 deviates to Type 3 equilibrium. ■

Remark 1: If $c_{o1} > 0$ does not hold, then Type 1 symmetric Nash equilibrium would not exist. In this case, only Type 2 asymmetric Nash equilibrium and Type 3 symmetric Nash equilibrium would prevail.

We will illustrate the results of this section using the above-given quadratic example.

Example: We obtain: $\hat{a} = \frac{\alpha}{\gamma+2\beta}$; $\hat{\hat{a}} = \frac{\alpha}{\gamma+\beta}$; $c_{o1} = \frac{\alpha^2(4\beta+3\gamma)}{2(\gamma+2\beta)^2} - (\frac{\alpha^2(2\gamma+\beta)}{2(\gamma+\beta)^2})$; $c_{o2} = \frac{\alpha^2}{2(\gamma+\beta)}$; $B(\hat{a}) = \frac{\alpha^2(2\gamma+\beta)}{2(\gamma+\beta)^2}$; $B(2\hat{a}) - c(\hat{a}) = \frac{\alpha^2(4\beta+3\gamma)}{2(\gamma+2\beta)^2}$. It is clear that $c_{o1} < c_{o2}$ because $[4\gamma^2\beta + 15\gamma\beta^2 + 4\beta^3] > 0$. Assumption $c_{o1} > 0$ holds if and only if $\gamma(\gamma + \beta) > \beta^2$.

In the following section, we analyze the outcomes of negotiations on agreements U and DT. Since the countries are identical, we can assume that they have the same negotiation power. Therefore, our focus is on the simple Nash bargaining solution with identical negotiation powers.

2.2 Cooperation: the agreement on a uniform standard without transfers and the agreement on differentiated standards with transfers

Depending on the type of Nash equilibrium (1, 2 or 3), we determine the payoffs of the countries for each agreement: agreement U and agreement DT. We then show that for Type 1 and Type 2 Nash equilibria, at least one of the two agreements dominates the Nash equilibrium. Moreover, for each Nash equilibrium, we highlight the threshold level of fixed cost above which agreement DT outperforms, in terms of welfare, agreement U. It is worthwhile noting that these thresholds do not necessarily belong to the range of the related Nash equilibrium. This gives rise to different configurations

of equilibrium that we highlight below (Proposition 5). As concerns Type 3 symmetric Nash equilibrium, we show that there is a threshold level of fixed cost above which none of the two agreements leads to gains to cooperation. Finally, we compare the levels of global abatement across the agreements U and DT (Proposition 6).

Here, we define the two agreements (Definition 1), we characterize the associated levels of abatement (Lemma 4) and the thresholds levels of fixed cost (Propositions 2, 3 and 4). We analytically compare these levels of fixed costs (Lemma 5). Subsequently, we illustrate these results by an example with quadratic benefit and cost functions.

Definition 1: *The objective of the countries is to maximize the following Nash function with respect to a_1 , a_2 and t :*

$$N(a_1, a_2, t) = [B(a_1 + a_2) - c_o - c(a_1) + t - NB_1^*] \times [B(a_1 + a_2) - c_o - c(a_2) - t - NB_2^*]$$

where NB_1^ and NB_2^* are the respective payoffs of the countries at the threat point.*

The U agreement results from this maximization problem under the following constraints: $a_1 = a_2 = \bar{a}$ and $t = 0$.

The DT agreement results from the same maximization problem under the following constraints: $a_1 \neq 0$, $a_2 = 0$ and $t > 0$.

Remark 2: As concerns agreement DT, it is equivalent to define it in the following way: $a_1 = 0$, $a_2 \neq 0$ and $t < 0$. We arbitrarily assume here that the country which undertakes an abatement effort is country 1.

Lemma 4: *Let $\bar{a}$ be the uniform abatement of the countries under agreement U. It is characterized in the following way: $2B'(2\bar{a}) = c'(\bar{a})$.*

Let a_1 be the abatement of country 1 under agreement DT. It is characterized in the following way: $2B'(a_1) = c'(a_1)$.

As concerns agreement DT, the level of transfers from country 2 to country 1, given Types 1, 2 and 3 Nash equilibria are respectively equal to $\frac{(c(a_1)+c_o)}{2}$, $\frac{(c(a_1)-c(\hat{a}))}{2}$ and $\frac{(c(a_1)+c_o)}{2}$.

Proof: see Appendix.

Proposition 2: *If the threat point of negotiations is Type 1 Nash equilibrium, then agreement DT is better for each country, in terms of welfare, than agreement U, if the fixed cost is higher than $\bar{c}_o = 2(B(2\bar{a}) - c(\bar{a})) - (2B(a_1) - c(a_1))$.*

Agreement U always dominates, in a Pareto sense, Type 1 Nash equilibrium.

Proof: see Appendix.

The first part of this proposition is quite intuitive: when the fixed cost of investment in abatement technology is sufficiently high, it is better that only one of the countries abates. As we will see later on, the mere presence of a high level of fixed cost is, however, not enough for the superiority of agreement DT over agreement U.

Proposition 3: *If the threat point of negotiations is Type 2 Nash equilibrium, then agreement DT is better for each country, in terms of*

welfare, than agreement U , if the fixed cost is higher than $\overline{c}_o = B(2\bar{a}) - c(\bar{a}) - B(a_1) - \frac{[c(\hat{a}) - c(a_1)]}{2}$.

Agreement DT always dominates, in a Pareto sense, Type 2 Nash equilibrium.

Proof: see Appendix.

The first part of this proposition says the following: when the fixed cost of abatement technology is lower than $\overline{c}_o$, but higher than c_{o1} , then it is better to negotiate an agreement based on mutual abatement (U agreement), whereas only country 1 abates at the non-cooperation. This result is explained by the fact that, at the cooperation, global abatement benefits are sufficiently high that they counter balance the payment of abatement costs.

Proposition 4: *If the threat point of negotiations is Type 3 Nash equilibrium, then*

Agreement DT is better for each country, in terms of welfare, than agreement U , if the fixed cost is higher than $\overline{c}_o$.

Agreement U dominates, in a Pareto sense, Type 3 Nash equilibrium if and only if the fixed cost is lower than $c_{oU} = B(2\bar{a}) - c(\bar{a})$.

Agreement DT dominates, in a Pareto sense, Type 3 Nash equilibrium if and only if the fixed cost is lower than $c_{oDT} = 2B(a_1) - c(a_1)$.

Proof: see Appendix.

First, when the threat point of negotiations is represented by Type 1 Nash equilibrium, i.e., when the countries abate at the non-cooperative equilib-

rium, then agreement U always leads to gains to cooperation. Second, when the threat point of negotiations is represented by Type 2 Nash equilibrium, i.e., when only country 1 abates at the non-cooperative equilibrium, then agreement DT always improves upon the non-cooperative outcome. Hence, in these two cases there exist at least a cooperative agreement (agreement U or DT) which leads to gains to cooperation no matter the level of the fixed cost. However, when the threat point of negotiations is represented by Type 3 Nash equilibrium, i.e., when no country undertakes abatement activities at the non-cooperative equilibrium, the level of fixed cost comes into play. In this case, agreement U (*resp.* agreement DT) improves upon the non-cooperative outcome if $c_o < c_{oU}$ (*resp.* $c_o < c_{oDT}$). In the opposite case, there is no gain to cooperation.

Lemma 5: *The thresholds levels of fixed cost can be ranked in the following way:*

- 1) $c_{o2} < c_{oU} < c_{oDT}$.
- 2) $\overline{c_o} < c_{oU}$.
- 3) If $\overline{c_o} > 0$, then $\overline{\overline{c_o}} < \overline{c_o}$.

Proof: see Appendix.

Example: With the above example with quadratic benefit and cost functions, we obtain: $\overline{a} = \frac{2\alpha}{\gamma+4\beta}$; $a_1 = \frac{2\alpha}{\gamma+2\beta}$; $c_{oU} = \frac{2\alpha^2}{\gamma+4\beta}$; $c_{oDT} = \frac{2\alpha^2}{\gamma+2\beta}$; $\overline{c_o} = 2c_{oU} - c_{oDT} = \frac{4\alpha^2}{\gamma+4\beta} - \frac{2\alpha^2}{\gamma+2\beta}$; $\overline{\overline{c_o}} = \frac{\overline{c_o}}{2} - \frac{\gamma}{4} \left(\frac{\alpha}{\gamma+\beta} \right)^2 = \frac{2\alpha^2}{\gamma+4\beta} - \frac{\alpha^2}{\gamma+2\beta} - \frac{\gamma}{4} \left(\frac{\alpha}{\gamma+\beta} \right)^2$.

We then have:

$$1) c_{o2} = \frac{\alpha^2}{2(\gamma+\beta)} < c_{oU} = \frac{2\alpha^2}{\gamma+4\beta} < c_{oDT} = \frac{2\alpha^2}{\gamma+2\beta}.$$

2) $\overline{c_o} = \frac{4\alpha^2}{\gamma+4\beta} - \frac{2\alpha^2}{\gamma+2\beta} < c_{oU} = \frac{2\alpha^2}{\gamma+4\beta}$. Here, $\overline{c_o}$ is not negative no matter the value of the parameters.

3) $\overline{c_o} = \frac{4\alpha^2}{\gamma+4\beta} - \frac{2\alpha^2}{\gamma+2\beta} > 0$ and $c(\widehat{a}) = \frac{\gamma}{2}(\frac{\alpha}{\gamma+\beta})^2 > 0$. Recall that $\overline{\overline{c_o}} = \frac{1}{2}(\overline{c_o} - c(\widehat{a}))$, then $\overline{\overline{c_o}} < \overline{c_o}$.

The following proposition summarizes the three configurations in relation to the benefit function $B(\cdot)$ and to the variable abatement cost function $c(\cdot)$.⁶ Indeed, the threshold levels of fixed cost depend on functions $B(\cdot)$ and $c(\cdot)$.

Proposition 5: 1) If $\overline{\overline{c_o}} < c_{o1} < \overline{c_o} < c_{o2}$, as c_o increases, then the U agreement is first negotiated for a fixed cost lower than c_{o1} , afterwards the DT agreement is negotiated until c_{oDT} is reached. If the fixed cost exceeds c_{oDT} , Type 3 Nash equilibrium prevails (absence of abatement).

2) If $c_{o1} < \overline{\overline{c_o}} < c_{o2} < \overline{c_o} < c_{oU}$, as c_o increases, then the U agreement is first negotiated for a fixed cost lower than $\overline{\overline{c_o}}$, afterwards the DT agreement is negotiated until c_{o2} , the U agreement is again negotiated when $c_{o2} < c_o < \overline{c_o}$, the DT agreement again comes into force when $\overline{c_o} < c_o < c_{oDT}$. Finally, if the fixed cost exceeds c_{oDT} , Type 3 Nash equilibrium prevails (absence of abatement).

3) If $c_{o2} < \overline{\overline{c_o}} < \overline{c_o} < c_{oU}$, as c_o increases, then the U agreement is first negotiated for a fixed cost lower than $\overline{c_o}$, afterwards the DT agreement is negotiated until c_{oDT} is reached. If the fixed cost exceeds c_{oDT} , Type 3 Nash

⁶Two more cases can theoretically arise, but we have not found the numerical examples to illustrate them. They are qualitatively similar to Cases 1 and 3.

equilibrium prevails (absence of abatement).

Example: With the above example with quadratic benefit and cost functions, we illustrate the three cases of Proposition 5 in Tables 1 and 2 (note that the subscript Tj with $j = 1, 2, 3$ stands for Type 1, Type 2 and Type 3 Nash equilibrium).

Table 1: Illustration of Cases 1 and 2

Case 1: $\alpha = \beta = \gamma = 1$	Case 2: $\alpha = 1, \beta = 0.1, \gamma = 1$
$c_{o1} = 0.013; c_{o2} = 0.25$	$c_{o1} = 0.31; c_{o2} = 0.45$
$c_{oU} = 0.4; c_{oDT} = 0.66$	$c_{oU} = 1.42; c_{oDT} = 1.66$
$\bar{c}_o = 0.13; \bar{\bar{c}}_o = 0.004$	$\bar{c}_o = 1.19; \bar{\bar{c}}_o = 0.38$
$NB_{1T1}^* = NB_{2T1}^* = 0.38 - c_o$	$NB_{1T1}^* = NB_{2T1}^* = 1.18 - c_o$
$NB_{1T2}^* = 0.25 - c_o; NB_{2T2}^* = 0.37$	$NB_{1T2}^* = 0.45 - c_o; NB_{2T2}^* = 0.86$
$NB_{1T3}^* = NB_{2T3}^* = 0$	$NB_{1T3}^* = NB_{2T3}^* = 0$
$NB_{T1}^U = NB_{T2}^U = NB_{T3}^U = 0.4 - c_o$	$NB_{T1}^U = NB_{T2}^U = NB_{T3}^U = 1.42 - c_o$
$NB_{T1}^{DT} = NB_{T3}^{DT} = 0.33 - (c_o/2)$	$NB_{T1}^{DT} = NB_{T3}^{DT} = 0.83 - (c_o/2)$
$NB_{1T2}^{DT} = 0.27 - c_o; NB_{2T2}^{DT} = 0.39$	$NB_{1T2}^{DT} = 0.62 - c_o; NB_{2T2}^{DT} = 1.04$
$t_{T1} = t_{T3} = 0.11 + (c_o/2); t_{T2} = 0.04$	$t_{T1} = t_{T3} = 0.69 + (c_o/2); t_{T2} = 0.48$

Table 2: Illustration of Case 3

Case 3: $\alpha = 1, \beta = 0.1, \gamma = 2$
$c_{o1} = 0.19; c_{o2} = 0.23;$
$c_{oU} = 0.83; c_{oDT} = 0.90$
$\bar{c}_o = 0.75; \bar{\bar{c}}_o = 0.26$
$NB_{1T1}^* = NB_{2T1}^* = 0.66 - c_o$
$NB_{1T2}^* = 0.23 - c_o; NB_{2T2}^* = 0.46$
$NB_{1T3}^* = NB_{2T3}^* = 0$
$NB_{T1}^U = NB_{T2}^U = NB_{T3}^U = 0.83 - c_o$
$NB_{T1}^{DT} = NB_{T3}^{DT} = 0.45 - (c_o/2)$
$NB_{1T2}^{DT} = 0.34 - c_o; NB_{2T2}^{DT} = 0.56$
$t_{T1} = t_{T3} = 0.41 + (c_o/2); t_{T2} = 0.29$

Our findings as regards Proposition (5.1 and 5.3) can be summarized as follows. The U agreement is preferred by the countries to agreement DT for a sufficiently low level of fixed cost of investment. Conversely, when the fixed cost exceeds a threshold level ($c_o > c_{o1}$ in (1) and $c_o > \bar{c}_o$ in (3)), the countries are better off under agreement DT. These findings illustrate that strictly identical countries can have an interest in reducing their emissions differently, and not in a uniform way. This result can be explained by the assumption of fixed cost in the abatement technology which implies a local non-convexity of the abatement cost function. Identical countries could be better off by signing an agreement based on differentiated standards with transfers in order to take advantage of local increasing returns to scale in

abatement activities. In this case, one of the countries abates for both, and pays for the fixed cost of investment. In return, it is compensated by monetary transfers for this effort. We have shown that in such a case the level of fixed cost must be sufficiently high.

A new result emerges for a different configuration given by Proposition (5.2). In this case, the preference of countries over the two alternative agreements changes as the level of fixed cost increases. The optimality of the U agreement and the DT agreement are occurring by turns. The U agreement which is first negotiated for a low level of fixed cost ($c_o < \bar{c}_o$) could be also preferred by the countries for a higher level of fixed cost ($c_{o2} < c_o < \bar{c}_o$). The DT agreement which is first agreed on for a medium level of fixed cost ($\bar{\bar{c}}_o < c_o < c_{o2}$) could be also negotiated for a high level of fixed cost ($\bar{c}_o < c_o < c_{oDT}$). The explanation of this result is as follows. As the level of fixed cost increases, the threat point of negotiations changes. As we have already seen, for each threat point, there exists a threshold level of fixed cost under which it is better to negotiate a U agreement and above which it is better to negotiate a DT agreement. Consequently, as c_o increases, it is possible that the type of the best agreement changes because the threat point changes. If we do not observe this alternation (as in cases 1 and 3), it is because for a given threat point, this threshold level of fixed cost (under which it is better to negotiate the U agreement, and above which it is better to negotiate the DT agreement) does not belong to the range of that threat point. In this case, the intuitive result prevails: as c_o increases from 0, the

best agreement is first the U agreement, and above the threshold level of fixed cost, the best agreement is the DT agreement.

The difference in these three configurations is due to the forms of the benefit function $B(\cdot)$ and the variable abatement cost function $c(\cdot)$, which modify the ranking of the threshold levels of fixed cost. In all three configurations, the countries prefer to not cooperate for a very high level of fixed cost ($c_o > c_{oDT}$). In such a situation, none of the countries abates because it is too costly (Type 3 Nash equilibrium).

We can now compare the levels of abatement of different institutional arrangements.

Proposition 6: *Given the assumptions on the concavity of the benefit function from global abatement $B(\cdot)$ and the convexity of the variable abatement cost function $c(\cdot)$, we have the following ranking of the abatement levels: $2\hat{a} > \hat{\hat{a}}$ and $2\bar{a} > a_1$.*

Proof. $B'(2\hat{a}) = c'(\hat{a}) < c'(2\bar{a})$ and $B'(\hat{\hat{a}}) = c'(\hat{\hat{a}})$. The property that the function $B'(\cdot)$ is decreasing and the function $c'(\cdot)$ is increasing implies that $2\hat{a} > \hat{\hat{a}}$. A similar argument leads to $2\bar{a} > a_1$. ■

Proposition 6 states that the total abatement is more when both countries make an effort to abate (the case of a uniform standard without transfers) rather than when only one country abates and the other country compensates it for this effort (the case of differentiated standards with transfers). This also holds for the non-cooperative equilibria. To be more specific, the level of total

abatement is higher when the two countries abate (Type 1 Nash equilibrium) than when only one of the countries abates (Type 2 Nash equilibrium). This is due to the concavity of the benefit function and to the convexity of the abatement cost function. Nevertheless, when we take into account the level of individual welfare, which is influenced by the level of fixed cost, we have seen that it could be better that only one of the countries undertakes an abatement effort.

In the next section, we tackle the issue of fixed costs in terms of self-enforcing agreements for numerous countries. The theoretical framework put forward by Barrett (1994) is adopted.

3 Extension: Stability of Agreements

We posit the context of $i = 1, 2, \dots, N$ identical countries facing a trans-boundary pollution problem, as is the case when the issue is the climate change or the ozone layer. We will compare the number of countries that join a cooperative agreement for the given two types of agreements: the uniform (U) and the differentiated (DT) ones. We are guided by the literature on the internal and external stability of IEAs (Barrett (1994), Carraro and Siniscalco (1993)).⁷ The majority of the models adopting this approach come

⁷This concept of stability of coalitions is taken from the literature on cartel stability (d'Aspremont et al. (1983)). This concept has some drawbacks. First, it excludes group deviations. This issue is tackled in Finus and Rundshagen (2003). Secondly, it assumes that a given country believes that other countries do not react to a change in its behaviour. This last point was challenged by the concept of farsighted stability (see for a recent

to the conclusion that the size of the stable coalition is very small, unless punishment strategies in a repeated game framework are taken into account. As de Zeeuw (2008) stresses, trigger mechanisms highlighted in repeated games are similar to the γ -core concept used in cooperative game theory to analyze the stability of the grand coalition (Chander and Tulkens (1995)).

We assume that the coalition of signatory countries plays Nash equilibrium strategies with the individual outsiders (non-signatories). The outsiders are also assumed to have Nash equilibrium strategies between them (symmetric Nash equilibrium). We adopt one of the models in Barrett (1994) with linear benefit and quadratic cost functions, for which it is possible to obtain analytical results.⁸ We also assume that there is a fixed cost in abatement technology. When a country does not abate, its abatement cost is assumed to be null.

For agreement DT, we posit the existence of an arrangement between (p) signatory countries in the coalition which abate ($a_s > 0$) and pay a fixed cost c_o , and (m) signatory countries in the coalition which do not abate, but pay a transfer (t) to other coalition members. Then the number of non-signatories is $N - (p + m)$. Their individual abatement level is represented by (a_n) . The agreement U is the case with $m = 0$ and $t = 0$: all the coalition members abate and pay the fixed cost, so there is no need to a

application de Zeeuw (2008)).

⁸Barrett (1994) assumes that a Stackelberg game is played between the coalition of signatory countries, which moves first, and the outsiders (followers). In the case of a constant marginal benefit function, this assumption yields the same outcomes as those under the assumption of Nash equilibrium strategies (Rubio and Casino (2005), p.90, footnote 2).

transfer scheme. The calculations for agreement U are identical to the below calculations of agreement DT by putting $m = 0$.

Each non-signatory maximizes *ex-ante* its own utility by taking the abatement levels of all other countries as given. The program of a non-signatory is written as follows:

$$\max_{a_n} \left\{ NB_n = w(pa_s + \sum_{i=p+m+1}^{i=N} a_n^i) - c_o - \frac{ca_n^2}{2} \right\} \quad (3)$$

where w represents the slope of each country's damage curve and c represents the slope of each country's marginal abatement cost curve.⁹ We obtain $a_n^i = a_n = \frac{w}{c}$. We assume that an outsider country will abate if its utility when it abates is higher or equal to that when it does not abate. This condition reduces to the following: $w \geq \sqrt{2c_o c}$.¹⁰ This condition states that the outsider abates if the marginal benefit from abatement is sufficiently high, or, in other words, if the fixed cost of abatement is sufficiently low.

At equilibrium, the utility level of a coalition member which abates and receives a transfer payment (recipient) from other coalition members is written in the following way: $NB_s = w(pa_s + (N - (m + p))a_n) - c_o - \frac{ca_s^2}{2} + \frac{m}{p}t$. The

⁹A linear damage function implies that every unit of pollution has a similar marginal effect on the environment. As Kolstad (2000, p.187) stresses, CO2 emissions could be assumed to exhibit such constant marginal damage. In line with the literature, we assume that returns to scale in abatement technology are diminishing.

¹⁰*Proof:* The utility level of an outsider when it abates is the following:
 $NB_{n1} = w(pa_s + \sum_{n=p+m+1}^{n=N} a_n) - c_o - \frac{ca_n^2}{2}$. Its utility when it does not abate is:
 $NB_{n2} = w(pa_s + \sum_{n=p+m+2}^{n=N} a_n)$. It is easy to check that $NB_{n1} \geq NB_{n2}$ when $w \geq \sqrt{2c_o c}$.

expression of transfers $(\frac{m}{p}, t)$ comes from the fact that m countries make a transfer payment t and these total payments are equally shared across p recipient countries. At equilibrium, the utility level of a coalition member which does not abate but does make a transfer payment (donor) to other coalition members, is written in the following way: $NB_s = w(pa_s + (N - (m + p))a_n) - t$. We assume that the level of transfers is such that the gains to cooperation are identical across the two groups of signatories.¹¹ This gives us the following level of transfers: $t = \frac{p}{p + m} \left[c_o + \frac{ca_s^2}{2} \right]$. Subsequently, the utility level of each coalition member is the following:

$$NB_s = w(pa_s + (N - (p + m))a_n) - \frac{p}{p + m} \left[c_o + \frac{ca_s^2}{2} \right] \quad (4)$$

As concerns agreement DT, we first determine the abatement levels of the p recipient countries in a coalition of $(p + m)$ members. The regulator of the coalition aims to maximize the joint welfare of all its members:

$$\max_{a_s} \left\{ (p + m) \left[w(pa_s + \sum_{n=p+m+1}^{n=N} a_n) \right] - p \left[c_o + \frac{ca_s^2}{2} \right] \right\} \quad (5)$$

The transfer disappears in the sum. The optimum is achieved for:

$$a_s |_{(p+m)} = \frac{(p + m)w}{c}.$$

The problem is now to determine $(p + m)^*$ the number of countries that sign the IEA when the conditions for the internal stability and the external

¹¹For equity reasons, we assume identical transfers for all members. We could also consider more complex transfer schemes between donor and recipient countries.

stability of the coalition are met.

Definition 2: *An IEA consisting of $(p + m)$ signatories must satisfy the two conditions:*

- *internal stability:* $NB_n(p + m - 1) \leq NB_s(p + m)$
- *external stability:* $NB_n(p + m) \geq NB_s(p + m + 1)$

The internal stability guarantees that a signatory country does not have an incentive to leave the coalition, and the external stability ensures that an outsider country does not have an incentive to join the coalition.

Proposition 7: *If non-signatories abate ($w \geq \sqrt{2c_o c}$):*

1) as concerns agreement U ($m = 0$), the stable number of countries in the coalition is equal to 2 when the total number of countries is 2, and is equal to 3 when the total number of countries is at least 3.

2) as concerns agreement DT ($m \geq 1$), the number of signatories which abate is 2 ($p = 2$) and the number of signatories which make a transfer payment is any positive number m ($m \geq 1$) when the total number of countries is at least 3. The stable number of countries in the coalition is thus at least 3, providing the possibility for a grand coalition.

Proof: see Appendix.

When the fixed cost of abatement is sufficiently low ($w \geq \sqrt{2c_o c}$), our findings as regards Proposition (7.1) replicate the result of Barrett (1994) with linear benefit and quadratic cost functions. In fact, the agreement

that Barrett (1994) takes into account is identical to the U agreement as it is defined in this paper. Here, the existence of a fixed cost in abatement technology is taken into account compared to the model of Barrett (1994). When the fixed cost is sufficiently low, as in this case, the countries even abate in the non-cooperative situation. Therefore, they pay the fixed cost both in cooperation and in non-cooperation. We show, as Barrett (1994), that the U agreement is not able to sustain more than two or three signatory countries. In contrast, under the same condition for the level of fixed cost, agreement DT is able to generate a stable coalition if there is precisely two coalition members who abate and pay the fixed cost, with at least one coalition member which finances their abatement efforts. The higher the number of donor countries in the coalition, the larger the size of the stable coalition. It leads that a grand coalition can emerge in the case of a differentiated agreement.

Given the coalition size for each agreement, we now compare the levels of individual payoff, individual abatement, and global abatement across the U and DT agreements.

Proposition 8: *If non-signatories abate ($w \geq \sqrt{2c_o c}$) and if the total number of countries is at least 3 ($N \geq 3$):*

1) The individual payoff under agreement U (with a coalition size of 3) is higher than that under agreement DT (with a coalition size of $2 + m$, with $m \geq 1$).

2) The individual abatement under agreement DT is higher or equal to that under agreement U.

3) *The global abatement under agreement DT is higher than that under agreement U, if the number of countries which make a transfer payment in agreement DT is at least 5.*

Proof: see Appendix.

When the fixed cost of abatement is sufficiently low ($w \geq \sqrt{2c_o c}$), we know that the number of countries which abate is 3 for agreement U, and it is equal to 2 for agreement DT. As Proposition (8.2) shows: the higher the number of countries which abate, the lower the level of individual abatement. This leads to the reduction of individual abatement costs in agreement U. This reduction in costs is able to offset the lower level of benefits from global abatement, the latter being true for $m \geq 5$ (Proposition (8.3)). Therefore, each signatory country is better off under agreement U than under agreement DT (Proposition (8.1)). The global payoff associated with an agreement is defined as the sum of the payoff of signatories and the payoff of non-signatories. The comparison of the global payoff under agreements U and DT is ambiguous. It depends on the values of the total number of countries, N , and the number of donor countries in agreement DT, m .

Proposition 9: *If non-signatories do not abate ($w < \sqrt{2c_o c}$):*

1) *as concerns agreement U ($m = 0$), when the total number of countries is at least 2, then the stable number of countries in the coalition is equal to 2 if $2 \geq \frac{2c_o c}{w^2}$, zero if not.*

2) *as concerns agreement DT ($m \geq 1$),*

when the total number of countries is 2, then the number of signatories which abate is 1 ($p = 1$) and the number of signatories which make a transfer payment is 1 ($m = 1$) if $3 \leq \frac{2cc_o}{w^2} \leq 4$,

when the total number of countries is at least 3, then the number of signatories which abate is 2 ($p = 2$) and the number of signatories which make a transfer payment must satisfy $m + 2 \geq \frac{2cc_o}{w^2}$.

Proof: see Appendix.

When the fixed cost of abatement is sufficiently high ($w < \sqrt{2c_o c}$), the U agreement is not able to sustain more than two signatory countries, no matter the total number of countries affected by the environmental problem. In contrast, under the same condition for the level of fixed cost, agreement DT is able to generate a stable coalition of at least 3 countries, if the total number of countries is sufficiently high. Again, the higher the number of donor countries in the coalition, the larger the size of the stable coalition. The high level of fixed costs requires, however, more donor countries in this case. If there was no possibility of transfer payments between countries, the stable coalition would contain only two signatories (this falls into the case of agreement U).

Proposition 10: *If non-signatories do not abate ($w < \sqrt{2c_o c}$) and if the total number of countries is 2 ($N = 2$):*

1) *With a coalition size of 2, the individual payoff under agreement U is higher than that under agreement DT.*

2) *The global payoff under agreement U is higher than that under agreement DT if $2 \geq \frac{c_o c}{w^2}$.*

3) *The individual abatement under agreement DT is equal to that under agreement U.*

4) *The global abatement under agreement U is higher than that under agreement DT.*

Proof: see Appendix.

When the fixed cost is sufficiently high ($w < \sqrt{2c_o c}$) and the total number of countries is 2, we know that the number of countries which abate is 2 for agreement U, and it is equal to 1 for agreement DT. Proposition (10.3) shows that, under agreement DT, one of the countries makes the abatement equal to the sum of the abatement levels of countries under agreement U. Since the global abatement is higher for agreement U (Proposition (10.4)), this leads to higher benefits from global abatement under this agreement. Indeed, this increase in benefits under agreement U outperforms the savings in abatement costs allowed by agreement DT. Therefore, individual countries are better off in agreement U (Proposition (10.1)). Remember that the level of fixed cost must be, at the same time, sufficiently low in this case, i.e., $2 \geq \frac{2c_o c}{w^2}$, given by the self-enforcing condition for the U agreement of Proposition 9.1. As concerns global payoff, it is also higher in agreement U if the level of fixed cost is not too high (Proposition (10.2)).

Proposition 11: *If non-signatories do not abate ($w < \sqrt{2c_o c}$)*

and if the total number of countries is at least 3 ($N \geq 3$) :

1) *The individual payoff under agreement DT (with a coalition size of $2 + m$, with $m \geq 1$) is higher than that under agreement U (with a coalition size of 2).*

2) *The global payoff under agreement DT is higher than that under agreement U if $N \geq 4$.*

3) *The individual abatement under agreement DT is higher than that under agreement U.*

4) *The global abatement under agreement DT is higher than that under agreement U.*

Proof: see Appendix.

When the fixed cost is sufficiently high ($w < \sqrt{2c_o c}$) and the total number of countries is at least 3, we know that the number of countries which abate is 2 for agreement U, and it is equal to $2 + m$, with $m \geq 1$, for agreement DT. In this case, each country is better off under agreement DT than under agreement U (Proposition (11.1)). The receipt of transfer payments by the two countries in agreement DT provides each of them with the incentive to abate more (Proposition (11.3)), and, in turn, to abate more at the global level (Proposition (11.4)). This leads to higher benefits from global abatement under agreement DT. Furthermore, the DT agreement is able to divide the abatement costs across the signatory countries by the means of transfer payments. Remember that the level of fixed cost must be, at the same time, sufficiently low in this case, i.e., $m + 2 \geq \frac{2cc_o}{w^2}$, given by the self-enforcing con-

dition for the DT agreement of Proposition 9.2. As concerns global payoff, it is also higher under agreement DT if the number of signatories is at least 4 (Proposition (11.2)).

To sum up, it would appear that with linear benefit and quadratic cost functions, agreement U could sustain at most 3 signatory countries. This confirms Barrett's (1994) result. Hence, agreement U cannot substantially increase overall payoffs when the number of countries affected by the environmental problem is very high. However, agreement DT could lead to a coalition of at least 3 countries. In this case, when the fixed cost is high, the number of donor countries must be sufficiently high. By taking into account fixed costs in abatement technology, it would be possible to alter the pessimistic finding on the small number of signatories of an IEA that is reported in the literature. In this case, the agreement must be designed given the differentiated standards with transfers. That is, some of the countries should abate on behalf of all the other countries and pay the fixed cost of abatement. The remaining coalition partners should then compensate them by transfer payments.

4 Conclusion

We investigate the relative efficiency of an agreement based on a uniform standard without transfers and one based on differentiated standards with transfers when strictly identical countries deal with transboundary pollution.

We especially ask what role fixed cost plays. Two approaches are examined: the Nash bargaining solution, implicating two countries, and the coalition formation framework, implicating numerous countries and emphasizing self-enforcing agreements. In the former, in terms of welfare, strictly identical countries may wish to reduce their emissions in a non-uniform way under the differentiated agreement. For this result to hold, the fixed cost of investment in abatement technology must be sufficiently high. The nature of the threat point of negotiations, however, also plays a crucial role. As concerns global abatement, the two countries abate more under the uniform agreement than under the differentiated one. In terms of coalition formation when numerous countries are involved, a grand coalition could emerge under a differentiated agreement.

Our results highlight the fact that when the level of fixed cost – arising from the installation of an abatement technology – is accounted for, it could be optimal, even for perfectly symmetric countries, to sign an agreement based on differentiated abatement standards. This analysis could apply to similar and future environmental negotiations, in the case where the cost of the initial investment in abatement activities is too expensive for the countries. This cost could be split between the countries concerned by the environmental problem. This type of environmental cooperation is already on the European Union agenda in order to strengthen bilateral environmental projects involving southern and eastern Mediterranean countries within the framework of the European Union Neighborhood Policy.

References

- [1] **d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J.** (1983): *On the stability of collusive price leadership*, Canadian Journal of Economics, vol. 16, 17-25.
- [2] **Barrett, S.** (1994): *Self-enforcing international environmental agreements*, Oxford Economic Papers, vol.46, 878-894.
- [3] **Barrett, S.** (2003): *Environment & Statecraft, The Strategy of Environmental Treaty-Making*, Oxford University Press, New York.
- [4] **Bayramoglu, B., Jacques, J-F.** (2005): *Comparison of negotiated uniform vs. differentiated abatement standards for a transboundary pollution problem*, working paper, Cahiers de recherche EURISCO [2005.15], Université Paris-Dauphine.
- [5] **Boyer, M., Laffont, J-J.** (1999): *Toward a political theory of the emergence of environmental incentive regulation*, Rand Journal of Economics, vol.30, No.1, 137-157.
- [6] **Carraro, C., Siniscalco, D.** (1993): *Strategies for international protection of the environment*, Journal of Public Economics, vol. 52, 309-328.

- [7] **Chander, P., Tulkens, H.** (1995): *A core-theoretic solution for the design of cooperative agreements on transfrontier pollution*, International Tax and Public Finance, vol. 2, 279-294.
- [8] **Copeland, B., Taylor, S.** (2005): *Free trade and global warming: a trade theory view of the Kyoto protocol*, Journal of Environmental Economics and Management, vol. 49, 205-234.
- [9] **de Zeeuw, A.** (2008): *Dynamic effects on the stability of international environmental agreements*, Journal of Environmental Economics and Management, vol. 55, 163-174.
- [10] **Finus, M., Rundshagen, B.** (1998): *Toward a positive theory of coalition formation and endogenous instrumental choice in global pollution control*, Public Choice, vol. 96, 145-186.
- [11] **Finus, M.** (2001): *Game Theory and International Environmental Cooperation*, Edward Elgar, Cheltenham.
- [12] **Finus, M., Rundshagen, B.** (2003): *Endogenous coalition formation in global pollution control: a partition function approach*, in C.Carraro (ed.), *The Endogenous Formation of Economic Coalitions*, Edward Elgar, chapter 6, 199-243.
- [13] **Harstad, B.** (2007): *Harmonization and side payments in political cooperation*, American Economic Review, vol. 97, 871-889.

- [14] **Hoel, M.** (1991): *Global environmental problems: the effects of unilateral actions taken by one country*, Journal of Environmental Economics and Management, vol. 20, 55-70.
- [15] **Hoel, M.** (1992): *International environment conventions: the case of uniform reductions of emissions*, Environmental and Resource Economics, vol. 2, 141-159.
- [16] **Kolstad, C.D.** (1987): *Uniformity versus differentiation in regulating externalities*, Journal of Environmental Economics and Management, vol. 14, 386-399.
- [17] **Kolstad, C.D.** (2000): Environmental Economics, Oxford University Press, New York.
- [18] **Larson, B.A., Tobey, J.A.** (1994): *Uncertain climate change and the international policy response*, Ecological Economics, vol. 11, 77-84.
- [19] **McAusland, C.** (2005): *Harmonizing tailpipe policy in symmetric countries: Improve the environment, improve welfare?*, Journal of Environmental Economics and Management, vol. 50, 229-251.
- [20] **Nash, J.F.** (1950): *The bargaining problem*, Econometrica, vol. 18, 155-162.
- [21] **Rubio, S.J., Casino, B.** (2005): *Self-enforcing international environmental agreements with a stock pollutant*, Spanish Economic Review, vol. 7, 89-109.

- [22] **Schelling, T.C.** (1960): The Strategy of Conflict, Harvard University Press, Cambridge.
- [23] **Welsch, H.** (1992): *Equity and efficiency in international CO₂ agreements*, in E.Hope and S.Strom (eds.), Energy Markets and Environmental Issues, Oslo.

APPENDIX

Proof of Lemma 4

The maximization of the function $N(a_1, a_1, t)$ for agreement U ($a_1 = a_2 = \bar{a}$ and $t = 0$) leads to the following first-order condition: $[2B'(2a) - c'(a)][B(2a) - c_o - c(a) - NB_2^*] = -[2B'(2a) - c'(a)][B(2a) - c_o - c(a) - NB_1^*]$. This condition defines $\bar{a}$. This holds true for the three Nash equilibria which represent the threat points of the negotiation on agreement U.

The maximization of the function $N(a_1, a_1, t)$ for agreement DT ($a_1 \neq 0, a_2 = 0$ and $t \neq 0$) leads to the following first-order conditions: $(B'(a_1) - c'(a_1))[B(a_1) - t - NB_2^*] + B'(a_1)[B(a_1) - c_o - c(a_1) + t - NB_1^*] = 0$ and $[B(a_1) - c_o - c(a_1) + t - NB_1^*] = [B(a_1) - t - NB_2^*]$.

As concerns Type 1 Nash equilibrium, we have $NB_1^* = NB_2^* = B(2\hat{a}) - c_o - c(\hat{a})$. Then, the expression of transfers in agreement DT takes the following form, $t = \frac{(c_o + c(a_1))}{2}$. Consequently, we obtain $(B'(a_1) - c'(a_1))[B(a_1) - \frac{(c_o + c(a_1))}{2} - B(2\hat{a}) + c(\hat{a})] + B'(a_1)[B(a_1) + \frac{c_o}{2} - \frac{c(a_1)}{2} - B(2\hat{a}) + c(\hat{a})] = 0$. This condition leads to a_1 .

As concerns Type 2 Nash equilibrium, we have $NB_1^* = B(\hat{\hat{a}}) - c_o - c(\hat{\hat{a}})$ and $NB_2^* = B(\hat{\hat{a}})$. Then, the expression of transfers in agreement DT takes the following form, $t = \frac{(c(a_1) - c(\hat{\hat{a}}))}{2}$. Consequently, we obtain $(B'(a_1) - c'(a_1))[B(a_1) - B(\hat{\hat{a}}) - \frac{(c(a_1) - c(\hat{\hat{a}}))}{2}] + B'(a_1)[B(a_1) - B(\hat{\hat{a}}) - \frac{[c(a_1) - c(\hat{\hat{a}})]}{2}] = 0$. This condition defines a_1 .

As concerns Type 3 Nash equilibrium, we have $NB_1^* = NB_2^* = 0$. Then, the expression of transfers in agreement DT is the following, $t = \frac{[c_o + c(a_1)]}{2}$.

Consequently, we obtain $(B'(a_1) - c'(a_1))[B(a_1) - \frac{(c_o + c(a_1))}{2}] + B'(a_1)[B(a_1) - \frac{(c_o + c(a_1))}{2}] = 0$. This condition defines a_1 .

As the expressions of transfers in the three cases show, there exists a unique level of transfers such that the two countries are equally well off. This is derived from the definition of the Nash bargaining solution and the assumption that all the countries have the same negotiation powers.

Proof of Proposition 2

$$[B(2\bar{a}) - c_o - c(\bar{a})] < \left[B(a_1) - \frac{[c_o + c(a_1)]}{2} \right] \text{ if and only if } c_o > \bar{c}_o.$$

The (identical) payoff of the countries in agreement U when the threat point is represented by Type 1 Nash equilibrium is the following: $[B(2\bar{a}) - c_o - c(\bar{a})]$, which exceeds their payoff at the threat point $[B(2\hat{a}) - c_o - c(\hat{a})]$ because $\bar{a}$ maximizes the function $B(2x) - c_o - c(x)$.

Proof of Proposition 3

$$[B(2\bar{a}) - c_o - c(\bar{a})] < \left[B(a_1) - \frac{[c(a_1) - c(\hat{a})]}{2} \right] \text{ if and only if } c_o > \bar{\bar{c}}_o.$$

The payoffs of the countries in agreement DT when the threat point is represented by Type 2 Nash equilibrium are the following: $\left[B(a_1) - \frac{[c(a_1) + c(\hat{a})]}{2} - c_o \right]$ for country 1 and $\left[B(a_1) - \frac{[c(a_1) - c(\hat{a})]}{2} \right]$ for country 2. These payoffs exceed those at the threat point ($NB_1^* = B(\hat{a}) - c_o - c(\hat{a})$ and $NB_2^* = B(\hat{a})$), because a_1 maximizes the function $2B(x) - c(x)$. Consequently, we have $B(a_1) - \frac{c(a_1)}{2} > B(\hat{a}) - \frac{c(\hat{a})}{2}$.

Proof of Proposition 4

$$[B(2\bar{a}) - c_o - c(\bar{a})] < \left[B(a_1) - \frac{[c_o + c(a_1)]}{2} \right] \text{ if and only if } c_o > \bar{c}_o.$$

The payoff of the countries in agreement U when the threat point is represented by Type 3 Nash equilibrium is the following: $[B(2\bar{a}) - c_o - c(\bar{a})]$, which exceeds 0 if $c_o < c_{oU}$.

The payoff of the countries in agreement DT when the threat point is represented by Type 3 Nash equilibrium is the following: $\left[B(a_1) - \frac{[c(a_1) + c_o]}{2}\right]$, which exceeds 0 if $c_o < c_{oDT}$.

Proof of Lemma 5

1) $c_{o2} < c_{oU}$ because $B(\hat{\bar{a}}) - c(\hat{\bar{a}}) < B(2\bar{a}) - c(\bar{a})$. This holds true because $B(2x) > B(x)$ for every x , since $B(\cdot)$ is an increasing function. Consequently, the maximum of the function $B(2x) - c(x)$ is higher than that of the function $B(x) - c(x)$.

$c_{oU} < c_{oDT}$ because $B(2\bar{a}) - c(\bar{a}) < 2B(a_1) - c(a_1)$. This holds true because $B(2x) < 2B(x)$ by the concavity of the function $B(\cdot)$ and by the assumption that $B(0) = 0$. Consequently, the maximum of the function $2B(x) - c(x)$ is higher than that of the function $B(2x) - c(x)$.

2) We know that $\bar{c}_o = 2c_{oU} - c_{oDT}$, then $\bar{c}_o < c_{oU}$ because $c_{oU} < c_{oDT}$ (see above). Since the inequality $c_{oDT} > 2c_{oU}$ could exist, $\bar{c}_o$ could be negative.

3) $\bar{\bar{c}}_o < \bar{c}_o$ because $B(2\bar{a}) - c(\bar{a}) - B(a_1) - \frac{[c(\hat{\bar{a}}) - c(a_1)]}{2} < 2(B(2\bar{a}) - c(\bar{a})) - (2B(a_1) - c(a_1)) = 2\bar{\bar{c}}_o + c(\hat{\bar{a}})$.

Proof of Proposition 7

We first study the condition of **internal stability**: $NB_n(p + m - 1) \leq NB_s(p + m)$. These utility levels are defined in the following way:

$$\begin{aligned}
NB_n((p-1)+m) &= w[(p-1)a_s + (N-(m+p-1))a_n] - c_o - \frac{ca_n^2}{2}. \\
NB_s(p+m) &= w[pa_s + (N-(m+p))a_n] - \frac{p}{m+p} \left[c_o + \frac{ca_s^2}{2} \right]. \\
\text{We have } a_s(m+p) &= \frac{(m+p)w}{c} \text{ and } a_n = \frac{w}{c}. \text{ Substituting, we obtain:}
\end{aligned}$$

$$NB_n((p-1)+m) = w \left[(p-1) \left(\frac{(m+p-1)w}{c} \right) + (N-(m+p-1)) \frac{w}{c} \right] - c_o - \frac{c}{2} \left(\frac{w}{c} \right)^2 \quad (6)$$

$$\begin{aligned}
NB_s(p+m) &= w \left[p \left(\frac{(m+p)w}{c} \right) + (N-(m+p)) \frac{w}{c} \right] \\
&\quad - \frac{p}{m+p} \left[c_o + \frac{c}{2} \left(\frac{(m+p)w}{c} \right)^2 \right] \quad (7)
\end{aligned}$$

Then, the condition $NB_n(p+m-1) \leq NB_s(p+m)$ reduces to the following: $\frac{w^2}{2c} [3 + p(p-4) + m(p-2)] \leq \frac{m}{m+p} c_o$.

We now study the condition of **external stability**: $NB_n(p+m) \geq NB_s(p+m+1)$. These utility levels are defined in the following way:

$$\begin{aligned}
NB_n(p+m) &= w[pa_s + (N-(m+p))a_n] - c_o - \frac{ca_n^2}{2}. \\
NB_s(p+m+1) &= w[(p+1)a_s + (N-(m+p+1))a_n] \\
&\quad - \frac{p+1}{m+p+1} \left[c_o + \frac{ca_s^2}{2} \right]. \\
\text{We have } a_s(m+p) &= \frac{(m+p)w}{c} \text{ and } a_n = \frac{w}{c}. \text{ Substituting, we obtain:}
\end{aligned}$$

$$NB_n(p+m) = w \left[p \left(\frac{(m+p)w}{c} \right) + (N - (m+p)) \frac{w}{c} \right] - c_o - \frac{c}{2} \left(\frac{w}{c} \right)^2 \quad (8)$$

$$\begin{aligned} NB_s(p+m+1) = & w \left[(p+1) \left(\frac{(m+p+1)w}{c} \right) + (N - (m+p+1)) \frac{w}{c} \right] \\ & - \frac{p+1}{m+p+1} \left[c_o + \frac{c}{2} \left(\frac{(m+p+1)w}{c} \right)^2 \right] \end{aligned} \quad (9)$$

Then, the condition $NB_n(p+m) \geq NB_s(p+m+1)$ reduces to the following: $0 < \frac{2cc_o}{w^2} \leq \frac{m+p+1}{m}(m(p-1) + p(p-2))$.

Let turn now to the core of the proof of Proposition 7.

1) The condition of **internal stability** is written in the following way in this case (the calculus for agreement U is similar if we put $m = 0$ and $t = 0$ in the calculus above): $NB_s(p) - NB_n(p-1) = \frac{w^2}{c}(-\frac{1}{2}p^2 + 2p - \frac{3}{2}) \geq 0$.

The condition of **external stability** is (the calculus for agreement U is similar if we put $m = 0$ and $t = 0$ in the calculus above): $NB_n(p) - NB_s(p+1) = \frac{w^2}{c}(\frac{1}{2}p^2 - p) \geq 0$.

These conditions are identical to those in Barrett (1994). We will show that only $p = 2$ and $p = 3$ satisfy these two conditions.

$NB_s(p) - NB_n(p-1)$ is a polynomial in p with a maximum attained for $p = 2$. This maximum implies the following value $\frac{w^2}{c}(\frac{1}{2}) > 0$. This polynomial is equal to 0 for $p = 1$ and $p = 3$, it is negative for $p > 3$.

$NB_n(p) - NB_s(p+1)$ is also a polynomial with a minimum attained for $p = 1$. This minimum implies the following value $\frac{w^2}{c}(-\frac{1}{2}) < 0$. It is equal to 0 for $p = 2$ and is strictly positive for $p > 2$.

2) The condition of **internal stability** is written in the following way in this case (see above): $(3 + p(p-4) + m(p-2)) \leq \frac{m}{m+p} \frac{2cc_0}{w^2}$.

The condition of **external stability** is (see above): $(m(p-1) + p(p-2)) \geq \frac{m}{m+p+1} \frac{2cc_0}{w^2}$.

For $p = 1$, we check that the second condition does not hold because $\frac{m}{m+2} \frac{2cc_0}{w^2} > 0$. For $p = 2$, these two conditions are written in the following way: the first condition $-1 \leq \frac{m}{m+2} \frac{2cc_0}{w^2}$ and the second condition $m \geq \frac{m}{m+3} \frac{2cc_0}{w^2}$, or $\frac{w^2}{2cC_0} \geq \frac{1}{m+3}$, which is true for each m strictly positive. We should now prove that a stable coalition could not exist for $p \geq 3$ and $m \geq 1$. Recall the condition of internal stability $(3 + p(p-4) + m(p-2)) \leq \frac{m}{m+p} \frac{2cC_0}{w^2}$. We know that $(3 + p(p-4) + m(p-2)) \frac{m+p}{p}$ exceeds $(3 + p(p-4) + m) \frac{m+p}{p}$ when $p \geq 3$ and $m \geq 1$. This latter expression is strictly superior to 1 for $p \geq 3$ and $m \geq 1$. However, we have $\frac{2cC_0}{w^2} \leq 1$ by assumption. Consequently, the condition of internal stability does not hold for $p \geq 3$ and $m \geq 1$.

Proof of Proposition 8

1) The individual payoff under agreement U (with a coalition size of $p = 3$) is as follows: $NB_s^U(3) = \frac{w^2(2N+3)}{2c} - c_o$.

The individual payoff under agreement DT (with a coalition size of $2 + m$, with $m \geq 1$) is as follows: $NB_s^{DT}(2 + m) = \frac{w^2N}{c} - \frac{2c_o}{m+2}$.

$NB_s^{DT}(2 + m) > NB_s^U(3)$ if $c_o > \frac{3w^2(m+2)}{2cm}$, which is incompatible with

our initial assumption $c_o < \frac{w^2}{2c}$. Therefore, $NB_s^U(3) > NB_s^{DT}(2+m)$.

2) The individual abatement under agreement DT is: $a_S^{DT}(2+m) = \frac{(2+m)w}{c}$.

The individual abatement under agreement U is: $a_S^U(3) = \frac{3w}{c}$. We then have $a_S^U \leq a_S^{DT}$ because $m \geq 1$.

3) The global abatement under agreement U is as follows:

$$A^U = pa_S^U + (N-m-p)a_n = 3\frac{3w}{c} + (N-3)\frac{w}{c} = \frac{w}{c}(6+N).$$

The global abatement under agreement DT is as follows:

$$A^{DT} = pa_S^{DT} + (N-m-p)a_n = 2\frac{(2+m)w}{c} + (N-m-2)\frac{w}{c} = \frac{w}{c}(2+N+m).$$

$A^{DT} > A^U$ if $m \geq 5$.

Proof of Proposition 9

We first study the condition of **internal stability**:

$$NB_n(p+m-1) \leq NB_s(p+m).$$

We have $a_s(m+p) = \frac{(m+p)w}{c}$ and $a_n = 0$. We obtain:

$$NB_n(p+m-1) = w \left[(p-1) \left(\frac{(m+p-1)w}{c} \right) \right] \quad (10)$$

$$NB_s(p+m) = w \left[p \left(\frac{(m+p)w}{c} \right) \right] - \frac{p}{m+p} \left[c_o + \frac{c}{2} \left(\frac{(m+p)w}{c} \right)^2 \right] \quad (11)$$

Then, the condition $NB_n(p+m-1) \leq NB_s(p+m)$ reduces to the following: $\frac{w^2}{2c} [-2 + p(4-p) + m(2-p)] \geq \frac{p}{m+p}c_o$.

We now study the condition of **external stability**: $NB_n(p + m) \geq NB_s(p + m + 1)$. We have $a_s(m + p) = \frac{(m + p)w}{c}$ and $a_n = 0$. We obtain:

$$NB_n(p + m) = w \left[p \left(\frac{(m + p)w}{c} \right) \right] \quad (12)$$

$$NB_s(p + m + 1) = w \left[(p + 1) \left(\frac{(m + p + 1)w}{c} \right) \right] \quad (13)$$

$$- \frac{p + 1}{m + p + 1} \left[c_o + \frac{c}{2} \left(\frac{(m + p + 1)w}{c} \right)^2 \right] \quad (14)$$

Then, the condition $NB_n(p + m) \geq NB_s(p + m + 1)$ reduces to the following: $\frac{w^2}{2cc_o}(m(1 - p) + p(2 - p) + 1) \leq \frac{p+1}{p+m+1}$.

Let turn now to the core of the proof of Proposition 9.

1) The condition of **internal stability** is written in the following way in this case (the calculus for agreement U is similar if we put $m = 0$ and $t = 0$ in the calculus above): $(-p^2 + 4p - 2) \geq \frac{2c_o c}{w^2}$. This inequality does not hold for $p = 1$ because $w^2 < 2c_o c$. It holds for $p = 2$ if $2 \geq \frac{2c_o c}{w^2}$. Neither it can hold for $p = 3$ because $w < \sqrt{2c_o c}$, nor for $p > 3$ because $\frac{2c_o c}{w^2} > 0$.

We should now prove that the condition of **external stability** (the calculus for agreement U is similar if we put $m = 0$ and $t = 0$ in the calculus above) holds for $p = 2$. In the general case, this condition is written in the

following way $(-p^2 + 2p + 1) \leq \frac{2c_o c}{w^2}$, and the case $p = 2$ satisfies this inequality.

2) The condition of **internal stability** is written in the following way (see above): $(m(2 - p) + p(4 - p) - 2) \geq \frac{p}{m+p} \frac{2cc_o}{w^2}$.

The condition of **external stability** is (see above): $(m(1 - p) + p(2 - p) + 1) \leq \frac{p+1}{m+p+1} \frac{2cc_o}{w^2}$.

For $p = 1$, the first condition leads to $(m + 1)^2 \geq \frac{2cc_o}{w^2}$, and the second condition implies $(m + 2) \leq \frac{2cC_0}{w^2}$. Since the total number of countries is 2, the number of signatories which make a transfer is 1 ($m = 1$). Consequently, the two conditions imply the following relationship $3 \leq \frac{2cc_o}{w^2} \leq 4$.

For $p > 2$ and $m \geq 1$, we note that the condition of internal stability $(m(2 - p) + p(4 - p) - 2)$ is lower than $(2 - p) + p(4 - p) - 2 = -p^2 + 3p$. This final expression is a polynomial with an integer maximum attained for $p = 1$ and $p = 2$. These maxima imply the following value (2). This polynomial is inferior or equal to zero for $p \geq 3$. Consequently, the condition of internal stability does not hold for $p \geq 3$.

For $p = 2$, the condition of internal stability is equal to $2 \geq \frac{2}{m+2} \frac{2cc_o}{w^2}$. This holds for a given value of m higher than a threshold level, if the total number of countries is sufficiently high. The condition of external stability is equal to $(1 - m) \leq \frac{3}{m+3} \frac{2cc_o}{w^2}$, which always holds for $m \geq 1$.

Proof of Proposition 10

1) The individual payoff under agreement U (with a coalition size of $p = 2$) is as follows: $NB_s^U(2) = \frac{2w^2}{c} - c_o$.

The individual payoff under agreement DT (with a coalition size of 2) is

as follows: $NB_s^{DT}(2) = \frac{w^2}{c} - \frac{c_o}{2}$.

$NB_s^{DT}(2) > NB_s^U(2)$ if $w^2 < \frac{cc_o}{2}$, which is incompatible with the assumption $c_o \leq \frac{w^2}{c}$ needed to have a coalition size of 2 in agreement U. Therefore, $NB_s^U(2) > NB_s^{DT}(2)$.

2) The global payoff under agreement U (with a coalition size of $p = 2$) is as follows:

$$V^U(2) = (m + p)NB_s^U + (N - m - p)NB_n^U = 2 \left[\frac{2w^2}{c} - c_o \right] = \frac{4w^2}{c} - 2c_o,$$

because there is no non-signatory for $N = 2$.

The global payoff under agreement DT (with a coalition size of $m + p = 2$) is as follows: $V^{DT}(2) = (m + p)NB_s^{DT} + (N - m - p)NB_n^{DT} = 2 \left[\frac{w^2}{c} - \frac{c_o}{2} \right] = \frac{2w^2}{c} - c_o$.

$$V^U(2) > V^{DT}(2) \text{ if } w > \sqrt{\frac{cc_o}{2}}.$$

3) The individual abatements under agreements U and DT are as follows:

$$a_S^U(2) = a_S^{DT}(2) = \frac{2w}{c}.$$

4) The global abatement under agreement U is as follows: $A^U = 2a_S^U(2) = \frac{4w}{c}$.

The global abatement under agreement DT is as follows: $A^{DT} = a_S^{DT}(2) = \frac{2w}{c}$.

We have $A^{DT} < A^U$.

Proof of Proposition 11

1) The individual payoff under agreement U (with a coalition size of $p = 2$) is as follows: $NB_s^U(2) = \frac{2w^2}{c} - c_o$.

The individual payoff under agreement DT (with a coalition size of $2 + m$)

is as follows: $NB_s^{DT}(2+m) = \frac{w^2(m+2)}{c} - \frac{2c_o}{m+2}$.

It is easy to check that $NB_s^{DT}(2+m)$ is always higher than $NB_s^U(2)$.

2) The global payoff under agreement U (with a coalition size of $p = 2$) is as follows: $V^U(2) = 2NB_s^U + (N-2)NB_n^U = 2 \left[\frac{2w^2}{c} - c_o \right] + (N-2) \left[\frac{4w^2}{c} \right] = \frac{4w^2}{c}(N-1) - 2c_o$.

The global payoff under agreement DT (with a coalition size of $2+m$) is as follows: $V^{DT}(2+m) = (2+m)NB_s^{DT} + (N-m-2)NB_n^{DT} = (2+m) \left[\frac{w^2(m+2)}{c} - \frac{2c_o}{m+2} \right] + (N-m-2) \left[\frac{2w^2(m+2)}{c} \right] = \frac{w^2}{c} [2N(m+2) - (m^2 + 4m + 4)] - 2c_o$.

$V^{DT}(2+m) > V^U(2)$ if $N \geq 4$.

3) The individual abatement under agreement U is as follows: $a_S^U(2) = \frac{2w}{c}$.

The individual abatement under agreement DT is as follows: $a_S^{DT}(m+2) = \frac{(m+2)w}{c}$.

We have $a_S^{DT}(m+2) > a_S^U(2)$.

4) The global abatement under agreement U is as follows: $A^U = 2a_S^U(2) = \frac{4w}{c}$.

The global abatement under agreement DT is as follows: $A^{DT} = 2a_S^{DT}(m+2) = \frac{2(m+2)w}{c}$.

We have $A^{DT} > A^U$.