

International environmental agreement : stability, transfer and sequential adhesion

prepared for EAERE, Southampton, June 2001

Preliminary version

Stéphane De Cara*, Pierre-Alain Jayet**

**CARD/FAPRI, Iowa State University*

***UMR en économie publique, INRA Grignon*

January 2001

Abstract

International negotiations on climate change show the importance of reaching agreements which group relatively large numbers of countries as far as global pollution is concerned. This document investigates stability properties of such agreements when the welfare of each of the countries depends on the emissions of all the others. Environmental agreements are modelled along the lines of a cartel of countries deciding to jointly reduce their emissions. A non-cooperative, two-phase game is used. During the first phase each country must decide whether or not to sign the agreement. During the second phase the levels of reduction are decided upon, both by the signatories (jointly) and by the non-signatories (individually). In contrast to Barrett (1994), we focus on the Nash equilibrium in the second phase game. Two polar cases are examined: a good the production of which is responsible for the pollution is distributed on an integrated market (single global price) or on a segmented market (prices determined at the national level). In both cases, the overall equilibrium of the game is calculated. We show that in the symmetrical case, where countries are identical and where there is intense market interdependence (integrated market), no stable agreement can emerge. There can be a stable agreement if the markets are segmented. In this case, an agreement can group a limited number of countries. We show, nevertheless, that if a sufficient number of signing countries can concede transfers to non-signatories, a sequential adherence process becomes a possibility. Extensions are given in the case where the countries entertain differing points of view as to their sensitivity to global pollution.

Introduction

The cycle of international negotiations being run under the auspices of the United Nations Framework Convention on Climate Change (UNFCCC) since the Rio Conference (1992) has resulted in quantifiable goals in the reduction of the gas emissions which are responsible for a modification in the thermal properties of the atmosphere. The main framework governing the adoption of such an agreement is that of State Sovereignty. In fact, each of the parties is free, in principal, to sign or not to sign such an agreement. The Kyoto protocol, signed in 1997, showed a clear separation between, on the one hand, the OCDE and Central and Eastern Europe (referred to as annex 1) which agreed to make an effort in reducing gas emissions and, on the other, countries referred to as annex 2 for which no quantitative targets were set.

Analysis of international accords dealing with the global environment has given birth to extensive literature, most of which draws mainly upon concepts stemming from the games theory. One finds two main approaches to dealing with the question. The first consists in viewing international negotiations on the climate from the angle of sharing the abatement costs amongst the countries sharing the resources. This approach, initiated in the works of Chander et Tulkens (1992), brings to light the possibility of cooperation between the group of States by means of transfers facilitating sequential convergence in the reductions of each country towards a state of the Pareto-Optimal system. The dynamic system of transfers proposed by these authors, based on concepts of cooperation, includes the interesting aspect of leading to a vector of abatement cost sharing which is part of the core of the game. That is to say, it cannot be blocked by any deviant coalition (Chander et Tulkens, 1995; Chander et Tulkens, 1997).

This depiction of international negotiations is based on a "strong" conception of the role played by international authorities. Indeed, as pointed out by Rotillon et Tazdaït (1996), the process supposes the existence of an agency capable of collecting and redistributing the transfers according to the scheme presented. This role, taken to be fully accepted by the set of Parties even before the running of the game, confers upon such an agency a status of supra-national authority which is not, however, authenticated. Stated differently, this hypothesis is based on a criteria of unanimity in the game of adhesion between the States (Carraro et Moriconi, 1998).

The second approach, initiated mainly through the works of Carraro et Siniscalco (1992) and Barrett (1994), highlights the sovereignty of the States when looking at international agreements on the global environment and therefore emphasizes the freedom of each party, in the absence of any supra-national authority, to adhere or not to adhere to an agreement. The privileged agreement concept in these works is that of "*spontaneous*" agreement (*self-enforcing agreement*). The explicit question put forward in contributions to this line of thought is that concerning the possibility of reaching a voluntarily signed agreement between these sovereign countries. In the absence of any institution capable of orientating individual decisions towards a Pareto-improvement, reaching an agreement is viewed as representing the formation of a coalition

between countries which collectively accept to reduce their emissions on the basis of their individual interests.

Formally, for an accord to be *self-enforcing*, it must satisfy three criteria: profitability, internal stability and external stability (Carraro et Siniscalco, 1993). These criteria, stemming from the analysis of cartels in industrial economics (d'Aspremont, Jacquemin, Gabszewicz et Weymark, 1983), correspond to the hypothesis of free adhesion implied by taking the sovereignty of the States into account. It is, thus, no longer a case of using a system of transfers to protect against the possibility of deviation by sub-groups of countries but, rather, one of taking into account the possibility that the countries, on a voluntary basis, will accept to make an effort in reduction so as to take advantage of improved environmental quality (Tulkens, 1997, for a critical comparison of the two approaches).

This reverse logic as compared to the preceding approach leads to different results. The grand coalition cannot be maintained insofar as it will be in the interest of at least one of the countries to deviate unilaterally to take advantage of the efforts made by the signatories to the agreement but without supporting the costs. Stated differently, global optimality generally cannot be reached with a *self-enforcing* agreement (as such optimality in fact requires the participation of all the countries).

Part of the externalities are, however, internalized thanks to the partially stable equilibrium cooperation. The possible gains brought about by this partial cooperation compared to the non-cooperation situation depends mainly on the number of countries accepting the equilibrium agreement. Such an agreement is endogenous to the extent that the conditions of stability described above are required. Work carried out on this aspect has shown that the number of countries taking part in the equilibrium is strongly dependent on the specification of reduction goals the signing countries are ready to assume and, therefore, on the specification of the game between the signatories on the one hand and the non-signatories on the other ¹.

The aim of our contribution is to apply and extend analysis in terms of the formation of environmental agreements which are stable in two directions.

First, in contrast to that used by Barrett (1994), the generic model we use explicitly takes into account the existence of a market on which is exchanged a good for which the production is responsible for the global pollution. This model makes it possible for us to discuss the influence of price fixing on the results of the environmental agreement. We will begin by comparing two polar cases: a global integrated market (section 2) and a market segmented between

¹Using numerical simulations Barrett (1994) brings to light the existence of *self enforcing* agreements which group the set of countries sharing the common resource. Nevertheless, the possibility that such agreements exist implies a Stackelberg game specification between the non-signatories and the signatories which attributes the role of *leaders* to the latter, thus casting the signature of the agreement in a more favorable light. If one abandons this hypothesis and privileges a Nash solution, stable agreements cannot be made between more than two countries. For an interpretation of the Stackelberg game hypothesis, the reader should see Barrett (1997b).

closed economies (section 3). Working within the hypothesis of homogeneous countries, we calculate the results of the game in emissions and examine the conditions of profitability and stability in both cases. We show that in the case of an integrated global market *leakage* prohibits a stable agreement, whereas such an agreement is possible if the economies are closed. Even in the latter case, it remains limited (never larger than 2 in size). In both cases we examine the threshold from which it can be in the interests of a group of agreeing countries to propose mutually advantageous transfer to a non-signatory in exchange for his adhesion. Since the conditions of stability have not been verified, such a transfer requires an engagement on the part of the countries (it is, in fact, question of a sequential engagement examined by Carraro et Siniscalco (1993)). We will show that there is a critical size after which such transfer becomes credible and which could make it possible to prime a "virtuous" engagement dynamic.

The second orientation covered in this work is that of dropping the hypothesis of homogeneous countries. Theoretical analyses on this approach are still quite scarce. Two recent tentatives have been made by Botteon et Carraro (1997) and Barrett (1997a). The former adopt a strategy of fixing the parameters by calibrating them on objective data and limiting the analysis to five regions. They then calculate the possible results according to the different value sharing rules of the agreement. As for Barrett, he privileges a wider-range analysis by foreseeing all types of parameters (but limiting the number of country types to two in the case of *leakage*) and by considering a Stackelberg game between signatories and non-signatories. In both cases, simulations must be used to obtain results in terms of stability.

Within the framework of the generic model we propose, the heterogeneity question (section 4) is approached by distinguishing between two types of country which differ in their sensitivity to damage. Working principally on the basis of simulations, we discuss the results of the game between the States (based, contrary to the Barrett model, on a Nash equilibrium concept).

1 The model

An economy is considered constituted by a set \mathcal{N} (of cardinal N) of countries. Within each country (indicated by $i \in \mathcal{N}$), a representative company produces good y in quantity y_i . The production of this good is the cause of a quantity of emission, referred to as e_i , considered to be a joint production.

1.1 Technology and pollution

For each country i , the production technology has a diminishing returns to scale. Formally, each representative company faces growing and convex cost variables $c_i(y_i)$ of the form:

$$c_i(y_i) = \frac{1}{2c_i}y_i^2 \quad \forall i \in \mathcal{N} \quad (1)$$

Each company is, thus, distinguished by the slope of its marginal costs $\frac{1}{c_i}$. Fixed production costs are taken to be nul. What is more, in the absence of any problems caused by the pollution, competition is taken to be perfect inside each country, making marginal prices and costs equal.

Pollution is taken to vary linearly with production. One therefore has the following relation:

$$e_i(y_i) = \lambda_i y_i \quad \forall i \in \mathcal{N} \quad (2)$$

1.2 Consumer welfare

Each country i is characterized by a national demand function for the good y taken to be linear and equal to $a_i - b_i p_i$ ($\forall i \in \mathcal{N}$) where p_i is the price of good y in country i .

Moreover, consumers sustain damages according to the total amount of pollution (E_N). We consider that the utility function is separable² and that global pollution, for the consumer in country i , results in a growing and convex reduction in his surplus $v_i(\cdot)$ of the form:

$$v_i(E_N) = \frac{1}{2}h_i E_N^2 \quad \forall i \in \mathcal{N} \quad (3)$$

Consumers of each country i are, hence, characterized by parameters a_i and b_i which determine the function of national demand as well as by their sensitivity to environmental damage (h_i).

1.3 Form of markets

Two concurrent hypotheses can be formulated for determining the price of good y . First, if one considers an integrated global market, the equilibrium price of good y will be determined by the confrontation between aggregate global demand and aggregate global supply (in this case, $p_i = p \forall i \in \mathcal{N}$ and p is the price of good y on the global market).

The price of good y can, on the contrary, be determined nationally by the confrontation between national supply and national demand. The latter case corresponds to the hypothesis of closed economies and a segmented market for good y .

²This hypothesis means that both any possible "disgust" effect of the pollution on consumption and any "compensation" are excluded. The first (resp. the second) effect is portrayed through a drop (resp. an increase) in the marginal utility resulting from the consumption of a quantity y_i .

1.4 Goal of each State function

It is considered that each State, when carrying out international negotiations, takes into account the total surplus of national consumers and the profits of the national companies (π_i) diminished by the damages caused by global pollution ($v_i(E_N)$). The goal of each State, referred to as W_i , is therefore given by the following relation:

$$W_i = SC_i - v_i(E_N) + \pi_i \quad (4)$$

$$= \frac{1}{2b_i}(a_i - b_i p_i)^2 - \frac{1}{2}h_i E_N^2 + p_i y_i - \frac{1}{2c_i} y_i^2 \quad (5)$$

Each of the States is considered to be capable of imposing production levels on companies which will correspond to maximization of national welfare, given that the price will be determined by either aggregate global demand or by national demand.

1.5 Specification of the game between States

Having specified the parameters entering into the determination of the goals of each State, in this section we will specify the strategic interactions the States must confront. We define, in particular, the notion of international environmental agreement, the behavior of each of the States both within and outside the agreement and the rules running the game.

In this contribution we will work within the scope of a formalization of international agreements based on the concept of "voluntary" agreement. The initial situation therefore corresponds to that of non-cooperation and States voluntarily decide, in light of the growth of the goal assigned to them, to work together in jointly determining the level of abatement they will accept. This approach is thus clearly different from that proposed by Chander et Tulkens (1997) even if the equilibrium concept used is the same (*Partial Agreement Nash Equilibrium*, cf Tulkens (1997)).

In place of the games proposed by Barrett (1997a), Carraro et Siniscalco (1992) and Botteon et Carraro (1997), we look at the running of negotiations between the States sharing the common resources in the form of a two-phase game. The first corresponds to each country's decision, on the basis of its own interests, to take part in the agreement or not. The second phase corresponds to the emissions game between, on the one hand, the signatories and, on the other, the non-signatories.

As it is understood that the game is run in a context of perfect information, we will examine the perfect sub-game equilibriums obtained by backward recurrence.

1.5.1 Definition of international environmental agreements

The definition 1 clarifies the notion of an international environmental agreement (IEA).

Définition 1 (International Environmental Agreement) *An IEA is a set, denoted \mathcal{S} ($\mathcal{S} \subseteq \mathcal{N}$) and of cardinal s , of countries which jointly decide upon the level of their emissions so as to maximize their total gains ($W_{\mathcal{S}} = \sum_{i \in \mathcal{S}} W_i$).*

Cooperation is, thus, here understood as a set of coordinated actions within a sub-group of countries. An IEA is therefore both the overall data for the set of participating countries (signataries) and an emission vector associated with each of them individually.

Définition 2 (Agreement structure) *An agreement structure is a partitioning of the set of countries $\mathcal{P}(\mathcal{N}) = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$.*

1.5.2 Second phase of emissions game

In order to define the second phase game, it is appropriate to specify the nature of the behavior of the signataries vis-à-vis the other States.

Définition 3 *Take an agreement structure $\mathcal{P}(\mathcal{N})$. Let it be understood that the value of the agreement associated with $\mathcal{P}(\mathcal{N})$ is the total maximum national gains ($V(\mathcal{S}, \mathcal{P}(\mathcal{N}))$) of the member States to an agreement $\mathcal{S} \in \mathcal{P}(\mathcal{N})$ obtained whenever the other agreements of the partition function in their best interests.*

The equilibrium concept we have retained and which underlies the definition of value for the agreement thus corresponds to a Nash equilibrium. Each player (here the set of countries participating in a given agreement) makes the best possible reactions with reference to the other players.

In particular agreement structures which consist of only one agreement \mathcal{S} and singletons, one finds the equilibrium concept used by Tulkens (1997) of *Partial Agreement Nash Equilibrium with respect to \mathcal{S}* corresponding to the solution of the following system:

$$P_{\mathcal{S}} \begin{cases} \max_{(e_i)_{i \in \mathcal{S}}} \sum_{i \in \mathcal{S}} W_i \\ \max_{e_i} W_i \quad \forall i \notin \mathcal{S} \end{cases}$$

At this stage of the game, prices p_i of good y are taken to be in the domain of common knowledge and fixed by demand on each of the national markets or on the global, depending on the hypothesis retained. Solving the emissions game is possible in a relatively general case (see annex 1).

1.5.3 Adhesion game of the first phase

In this phase each State decides individually as to its adherence or non-adherence to an agreement. It is understood that at this level each player can perfectly foresee the results of the second phase of the game and takes only the effect of his own deviation into account.

The solution of the game relies on using the following N internal and external stability conditions in the case of an agreement structure composed of a sole agreement \mathcal{S} of size $S \geq 2$ and $N - S$ singletons. The individual value function after solving the emissions game is denoted V_i .

$$V_i(\mathcal{S}, \{j\}_{j \notin \mathcal{S}}) \geq V_i(\mathcal{S} \setminus \{i\}, \{j\}_{j \notin \mathcal{S} \setminus \{i\}}) \quad \forall i \in \mathcal{S} \quad (6)$$

$$V_k(\mathcal{S}, \{j\}_{j \notin \mathcal{S}}) \geq V_k(\mathcal{S} \cup \{k\}, \{j\}_{j \notin \mathcal{S} \cup \{k\}}) \quad \forall k \in \mathcal{N} \setminus \mathcal{S} \quad (7)$$

The following profitability conditions S should be added to this condition:

$$V_i(\mathcal{S}, \{j\}_{j \notin \mathcal{S}}) \geq V_i(\{j\}_{j \in \mathcal{N}}) \quad \forall i \in \mathcal{S} \quad (8)$$

At this level it must be taken that each country knows the rule of value sharing amongst the members of an agreement³.

2 Homogenous countries and an integrated global market

We are placing ourselves within the framework of an integrated market for good y where $p_i = p \quad \forall i \in \mathcal{N}$. We are also limiting ourselves to the analysis of agreement structures composed of a sole agreement between singletons ($\mathcal{P}(\mathcal{N}) = \{\mathcal{S}, \{j\}_{j \notin \mathcal{S}}\}$).

The price of good y is taken to be determined by demand on the global market for the level of overall production. It is therefore the same for each of the companies.

Finally, in this section we retain a homogeneity hypothesis for the countries with reference to parameters $a_i, b_i, c_i, h_i, \lambda_i$ (as these parameters are taken to be homogenous between the countries, the indices will not be given in this section). Moreover, λ is normalized to 1. This final hypothesis makes it possible for us to link the overall emissions level directly to the overall production level.

³In the case of homogenous countries, agreement value sharing amongst the members poses no problem insofar as equal sharing is the obliged solution to the negotiations. The problem only exists when the characteristics of the countries are different. In section 4 - where the consequences of dropping the homogeneity hypothesis are examined - no account is taken of possible lateral transfers between the signatories to an agreement. Individual gains are determined entirely by the goal attributed to them and which maximizes the welfare of the joint members of the agreement. Everything happens as if the signing countries totally delegated the choice of emission levels to the set of members of the agreement.

2.1 Non-cooperative situation

The initial non-cooperative situation corresponds to the partitioning of \mathcal{N} into N singletons. The results of the game between the States is therefore given by a Nash equilibrium between the N players

Calculating the intersection of the best answer functions of the N countries gives the following overall results:

$$Y_N^{NC} = E_N^{NC} = \frac{Nac}{b + c + Nbch} \quad (9)$$

Since the companies are taken to be identical and operating within the framework of an integrated global market, each of the companies produces (and emits) an identical quantity equal to $\frac{ac}{b+c+Nbch}$. That is to say, the presence of the overall externality ($h > 0$) implies that national productions are such that $y_i \leq c_i p$.

In the non-cooperative situation only the external effects caused by national pollution on national welfare are taken into account.

2.2 Total cooperation

The cooperative situation corresponds to the global optimum insofar as the set of crossed externalities is internalized and the sum of national gains is maximized. Under the hypotheses governing this section, one reaches the following results:

$$Y_N^{FC} = E_N^{FC} = \frac{Nac}{b + c + \tau} \quad (10)$$

with:

$$\tau = N^2bch \quad (11)$$

$$(12)$$

It is obvious that this cooperation makes it possible to reduce the level of emissions more than in the non-cooperative situation, and all the more so depending on how high N is raised.

2.3 Partial cooperation

Here we analyze the intermediary situations corresponding to partitions of \mathcal{N} in the form $\mathcal{P}(\mathcal{N}) = \{\mathcal{S}, \{j\}_{j \notin \mathcal{S}}\}$.

2.3.1 Study of the emissions game

Here we place ourselves after the first phase of the game. The agreement structure has been fixed by the results to the adherence game. The results of the emissions game is given by solving the program (P_S).

After a few algebraic manipulations, the global and individual emissions can be expressed by the equilibrium:

$$Y_N = E_N = Nac \frac{X_S}{(b+c)X_S + \tau Z_S} \quad (13)$$

$$\forall i \in S : y_i = \frac{Y_N}{N} - \frac{acQ_S}{(b+c)X_S + \tau Z_S} \quad (14)$$

$$\forall i \notin S : y_i = \frac{Y_N}{N} + \frac{acU_S}{(b+c)X_S + \tau Z_S} \quad (15)$$

with the notations :

- $X_S = N(Nb + c) + (N - S)(S - 1)c$
- $Z_S = Nb + c + (S - 1)(Sb + c)$
- $\tau = N^2bch$
- $Q_S = (N - S)(S - 1)\tau$
- $U_S = S(S - 1)\tau$

Noting that $X_N = Z_N$, one immediately finds the cooperative solution obtained by making $S = N$ in the equations (13) to (15) once again. Moreover, the non-cooperative solution presents itself as a particular case of the final formula by making⁴ $S = 1$ ($X_1 = NZ_1$).

In the limit case $h = 0$, *i.e.* the case where pollution has no effect on consumers, the equilibrium quantity $Y_N = \frac{Nac}{b+c}$ corresponding to the $\frac{a}{b+c}$ equilibrium price is found once again.

One can also see that, as expected, the countries within the agreement emit less than the average of the set of countries whereas the non-signing countries emit more. Finally, total emissions decreases as the size of the agreement increases and, therefore, a larger part of the externalities are internalized. These properties are illustrated⁵ in figure 1

2.3.2 Study of the adherence game

Examining the results of the adherence game necessitates preliminary analysis of the form of the individual gains of each country according to their decision to

⁴With the Nash equilibrium used in this paper, it is obvious that the non-cooperative solution can be obtained indifferently by making $S = 1$ or $S = 0$, since the two situations are strictly equivalent. It would be different if we had, like Barrett (1994), hypothesized a Stackleberg game for the second phase.

⁵In the simulations presented, the following values are retained for the parameters of the model: $N = 180$, et, $\forall i \in \mathcal{N}$, $a_i = 1$, $b_i = 0.01$, $c_i = 0.05$, $h_i = 0.004$

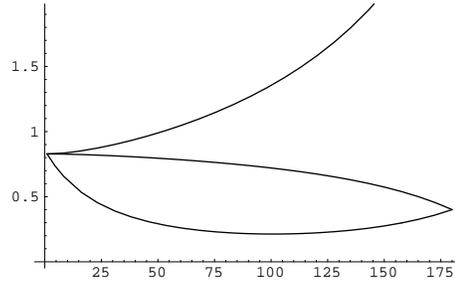


Figure 1: Comparison of emissions of non-signing countries, emissions of signing countries and average emission.

sign or not sign the agreement. This analysis is summed up by the proposition 1 and illustrated in Figure 2.

Proposition 1 *The minimum individual gain for a country in the agreement is found for an S size between 1 and N . The gain of a non-signing country increases strictly within the $[1, N]$ interval. Average gain for the set of countries increases strictly according to the number of signatories.*

Proof: This proposition results from the study of gain functions given in annex 2. ■

As we have just seen, once sensitivity to damage (h) has become strictly positive, the gain of a signing country drops whenever the S size of the agreement increases, for initial S values, until it reaches a minimum for a size strictly positioned between 1 and N . The reduction in a signing player's gain for the first S values corresponds to the presence of a *leakage*. The latter is due to the possibility of the non-signatories to revise their production choices (and therefore of emissions) depending on the decisions of the signatories. If such is the case, the overall reduction in pollution is no longer sufficient to offset the costs born by the signatories (Botteon et Carraro, 1997; Barrett, 1997b).

We also mentioned the fact that the total gain, and therefore the individual average as well, increases strictly with the number of signatories. Hence it is obvious that the individual gain of the grand coalition is strictly superior to the gain a country would obtain in a singletons game. There is therefore an S size for the agreement, less than N , beyond which a country will obtain a higher gain than it would obtain in a singletons game. We sum up this result in the following proposition.

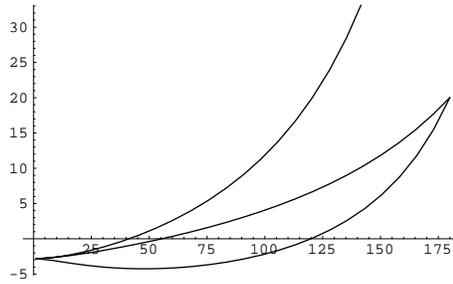


Figure 2: Gains of signing countries, non-signing countries and average gain.

Proposition 2 *There is an agreement size, less than N , after which the agreement is profitable.*

Proof: This proposition results from the examination of the profitability condition (inequality 8) due to the study of the $\Gamma(S)$ function:

$$\Gamma(S) = P(S) - Q(0)$$

or⁶ $P(S) = V_{i \in \mathcal{S}}(\mathcal{S}, \{j\}_{j \notin \mathcal{S}})$ and $Q(S) = V_{i \notin \mathcal{S}}(\mathcal{S}, \{j\}_{j \notin \mathcal{S}})$. ■

An agreement is stable whenever the internal (inequality 6) and external (inequality 7) stability conditions are met. That is to say, as Carraro et Siniscalco (1992) points out, an agreement is stable as long as the matrix of gains in the partial adherence game between the two potential signatories is of the same type as in a *chicken game*.

We sum up these conditions in the study of the sign of the $\Delta(\cdot)$ function:

$$\Delta(S) = P(S) - Q(S - 1) \tag{16}$$

Whenever $\Delta(S) < 0$, it is in the interest of a signing State to leave the agreement. Whenever $\Delta(S + 1) > 0$, it is in the interest of a non-signing State to join the agreement. We can therefore give the following definition:

Définition 4 *An agreement of size S^* is stable if and only if $S^* > 1$ verifying $\Delta(S^*) > 0$ and $\Delta(S^* + 1) < 0$.*

⁶To simplify, here we express the functions as $P(\cdot)$ and $Q(\cdot)$ depending on S , not \mathcal{S} . As the countries are taken to be identical, an agreement can be assimilated to its size with no problem.

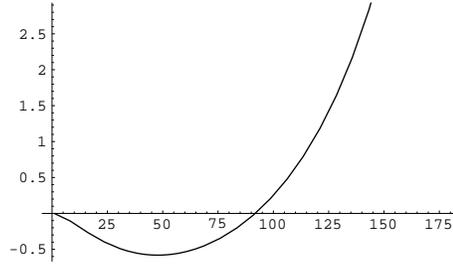


Figure 3: Profitability function $\Gamma(S)$.

The following proposition indicates that within the framework of hypotheses retained in this section no stable agreement obtained on a voluntary basis can appear. The instability of the potential agreements is illustrated in figure 4.

Proposition 3 *Within the framework of an integrated global market and homogenous countries, there are no stable agreements.*

Proof: Using the gains functions given in annex 2, we show that the stability function is always strictly negative for $S > 1$. ■

In the simulation presented, it is obvious that the incitation to deviate increases with size of the agreement.

2.4 Transfers and sequential engagement

We have shown the possible existence of a profitable agreement. Stated differently, beyond a certain threshold defined by the parameters of the model it becomes possible for each of the members of the agreement to improve his welfare as compared with the non-cooperative solution. Unfortunately, such an agreement, once made, would not be stable (proposition 3) as long as the individual deviation would be profitable to a signing State.

The situation can also be viewed in a somewhat different light. Let us imagine that an agreement \mathcal{S} has already been reached. It can be in the interest of the members of the agreement to propose a transfer to a non-signatory in exchange for his adherence. This transfer must necessarily offset the losses incurred by the new entree ($P(S+1) - Q(S)$) which is negative due to the instability

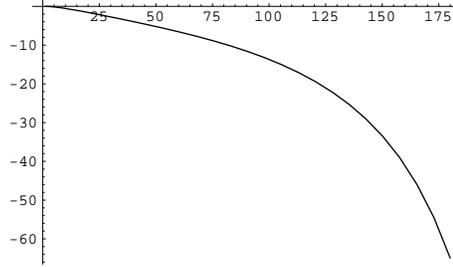


Figure 4: Stability function $\Delta(S)$.

explained above). It must as well be self-financed by the members of the existing agreement. That is to say, it must not exceed the supplementary gain obtained by the S signatories following the entrance of a new State ($S(P(S + 1) - P(S))$). We will therefore study the sign of the $\Theta(\cdot)$ function:

$$\Theta(S) = P(S + 1) - P(S) - \frac{Q(S) - P(S + 1)}{S} \quad (17)$$

As soon as the agreement size reaches a state where $\Theta(S)$ is positive, the members of the agreement have the possibility of offering a transfer to a non-signatory which will make mutual improvement of individual gains possible. Once the new entree has joined the agreement, the process can be repeated towards another non-signatory.

We are thus using the sequential concept of engagement proposed by Carraro et Siniscalco (1993) here. As these authors point out, the enlargement of an agreement by means of such a transfer does not make it possible to make up for the instability problem explained above. The problem of the "lone rider" remains. For cooperation to appear and be maintainable, it must be understood that the countries taking part in this process accept a certain form of engagement.

Nevertheless, to the extent that such a system of transfers is credible, it can make it possible to enlarge the agreement to the set of countries through a succession of mutually profitable improvements. In particular, for the enlargement to be credible, the size of the initial agreement must be such that it makes transfers of the ($\Theta(S) > 0$) type possible. Proposition 4 indicates that there is such a threshold.

Proposition 4 *Generally there is an agreement threshold size \bar{S} beyond which the member States can propose to an exterior State to join the group with any of the States sustaining any loss.*

Proof: We show that the transfer function (obtained using the expressions of the gains functions given in annex 2) is negative for $S = 1$ and positive for $S = N$. ■

Figure 5 illustrates this proposition. It in fact gives particular importance to the existence of ratification thresholds in international environmental agreements⁷. If the agreement is matched with a minimal ratification condition which is equal to or less than \bar{S} , the system of transfer can allow for a "virtuous" adherence dynamic.

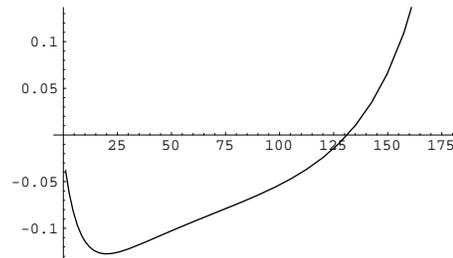


Figure 5: Transfers from signatories to non-signatories ($\Theta(S)$).

3 Homogenous countries, segmented markets

In this section we explore a different hypothesis as far as determining the price of good y is concerned. We now leave the framework of an integrated global market to examine what influence a hypothesis of segmented markets has where the price is determined mainly by national demand.

We will follow the same procedure as in the preceding section to illustrate how these results contrast with those of the hypothesis of an integrated global market. We will, in particular, retain the hypothesis of homogenous countries.

⁷This condition belongs to the Kyoto agreement. A justification is differently given by Black, Levi et De Meza (1993).

3.1 Partial cooperation

We take a direct look at the results of the emissions game for a partition of \mathcal{N} with an agreement of size S compared to $N - S$ singletons. Indeed, the countries are taken to be identical, meaning that the non-cooperative and the cooperative situations lead to the same results as in the case of an integrated global market (see respectively equations (9) and (10)) since these two extreme cases are completely symmetrical.

It is therefore only in the situation where the behavior of the countries leads them to split into two groups (signatories and non-signatories) that the hypothesis of a different price on national markets plays an important role.

3.1.1 Study of the emissions game

In the case of segmented markets, Nash equilibrium between the S signatories and the $N - S$ non-signatories results in the following overall results:

$$Y_N = E_N = \frac{Nac}{b + c + (N + S(S - 1))bch} \quad (18)$$

One can once again express the individual emission levels in relation to the average emissions (the signatories are indexed by i and the non-signatories by k):

$$y_i = \frac{Y_N}{N} \left(1 - \frac{(N - S)(S - 1)bch}{b + c} \right) \quad (19)$$

$$y_k = \frac{Y_N}{N} \left(1 + \frac{S(S - 1)bch}{b + c} \right) \quad (20)$$

As before, the signing countries emit less than the average while the non-signing countries emit more. Figure 6 clearly suggests much less substantial *leakage* in the case of segmented markets. For the same parameter values, the emissions of the non-signatories react much less to the size of the agreement than in the preceding case.

3.2 Study of the adherence game

Individual gains of each State in the situation where the agreement S is reached are given by the following equations:

$$P(S) = \frac{a^2c}{2b(b + c)} \left(1 - \frac{N^2bch(b + c + S^2bch)}{(b + c + bch(N + S(S - 1)))^2} \right) \quad (21)$$

$$Q(S) = \frac{a^2c}{2b(b + c)} \left(1 - \frac{N^2bch(b + c + bch)}{(b + c + bch(N + S(S - 1)))^2} \right) \quad (22)$$

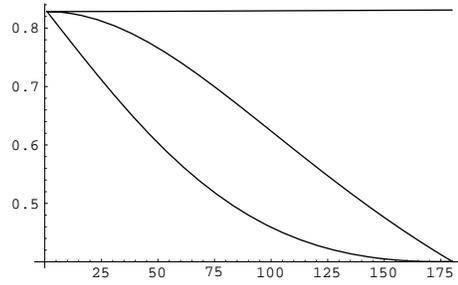


Figure 6: Emissions of non-signing countries, signing countries and average emissions (in the case of segmented markets).

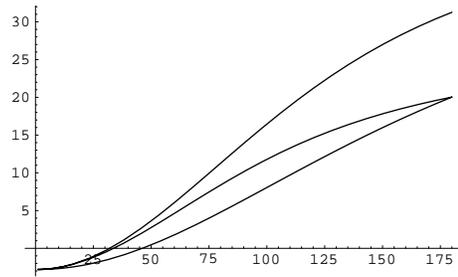


Figure 7: Gains of countries outside the agreement or within the agreement and average gain (in the case of segmented markets).

The simulations presented in Figure 7 clearly indicate that the profitability threshold is reduced. This is a direct consequence of the major reduction in *leakage*. The reduction in the agreement members' gains is thus lower and makes it possible for a stable agreement to emerge.

Proposition 5 *In the case of segmented markets, there can be a stable agreement. the latter cannot include more than two countries.*

Proof: Using equations (21) and (22) we show that $\Delta(S) < 0$ for $S \geq 3$. ■

As an illustration, the $\Delta(S)$ stability function is shown in graph 8. It is annulled for a value of $2 < S < 3$, indicating a stable agreement of size 2 for the parameters retained.

The same general result can be reached using the model developed by Barrett (1994) and privileging the hypothesis of a Cournot-Nash game (rather than a Stackelberg game) between the signatories and non-signatories. The interest here lies in the fact that we show the result is linked to the extent of the *leakage* caused by the form of the markets.

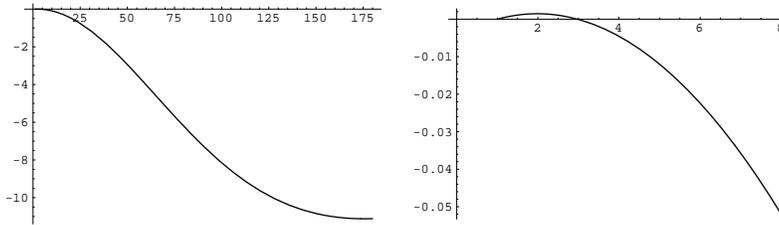


Figure 8: Stability function (in the case of segmented markets).

3.3 Transfers and sequential engagement

The *leakage* reduction shown in the preceding section leads to a substantial modification in the transfers which can be offered to non-signatories by the members of an agreement. It is expressed notably through a major reduction in the threshold beyond which such transfer makes mutual improvement of individual gains possible. This point is illustrated by Figure 9.

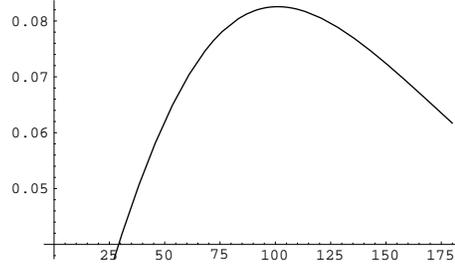


Figure 9: Transfers from the signatories to non-signatories (in the case of segmented markets).

4 Heterogeneity of sensitivity to damage

Henceforth we consider that the countries differ in the sensitivity to damage. N_1 (respectively N_2) countries are characterized by a parameter h_1 (respectively h_2) of sensitivity to emissions such that $N = N_1 + N_2$. Without leaving the general scope, we understand $h_1 > h_2$. The market is taken to be integrated so that there is a sole price for good y .

An agreement of size S will be composed of S_1 type 1 countries and S_2 type 2 countries, so that $S = S_1 + S_2$. We indicate:

$$\begin{aligned}
 H &= N_1 h_1 + N_2 h_2 \\
 H_S &= S_1 h_1 + S_2 h_2 \\
 G_S &= (b + c)X_S + Nbc((Nb + c)H + N(S - 1)bH_S) \\
 E_{S,i} &= \frac{(Nb + c)H + N(S - 1)bH_S - X_S h_i}{Nb + c}
 \end{aligned}$$

In this section we must accept that the complexity of the calculations leaves us guessing rather than with clearly established propositions⁸.

⁸Simulations are realised with splitting the global set of countries as ($N_1 = 90, N_2 = 90$) respectively for the two country types, and the valuations of externality are $h_1 = 0.006$ et $h_2 = 0.002$. Curves in figures strongly depends of these values. But no simulation was found against the conjectures announced in this section.

4.1 Solving the emissions game

The level of emissions is expressed in the following manner:

$$E_N = Y_N = Nac \frac{X_S}{G_S} \quad (23)$$

$$\forall j \in S : e_j = y_j = \frac{Y_N}{N} + \frac{acN^2bc(H - (N - S + 1)H_S)}{G_S} \quad (24)$$

$$\forall k \notin S : e_k = y_k = \frac{Y_N}{N} + \frac{acN^2bcE_{S,i}}{G_S} \quad i = 1, 2 \quad (25)$$

The gains expressions are given in annex 3. From the preceding expressions one can easily deduce the following proposition.

Proposition 6 *Considering the set of countries, average emission drops and average gain rises whenever the number of signatories increases, not matter what the composition of the agreement. Moreover, the emission of a signing country, while depending on the composition of the agreement, does not depend on country type.*

Through simulation one can generally verify that the emission of a non-signing country increases whenever the size of the agreement increases. On the other hand, and contrary to the result obtained within the homogenous framework, it is no longer certain that the emission of a signing country will drop whenever the size of the agreement increases (see Figure 10).

This result is intuitive. In fact, the higher the number of countries sensitive to environmental damage within the agreement, the larger the effort of abatement taken on by the signatories. If, on the contrary, the proportion of countries less sensitive to damage increases, the level of emission of a signatory has a tendency to increase and the agreement proves less demanding in terms of emissions reduction.

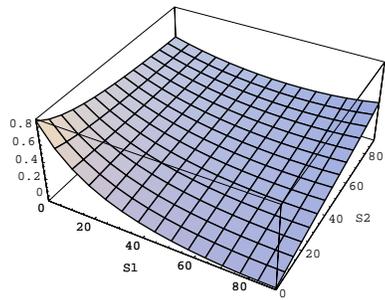
4.2 Adherence game

The gain of a type i signatory to an agreement (S_1, S_2) is denoted $P_i(S_1, S_2)$ and the gain of a type k ($k = 1, 2$) non-signatory is denoted $Q_k(S_1, S_2)$.

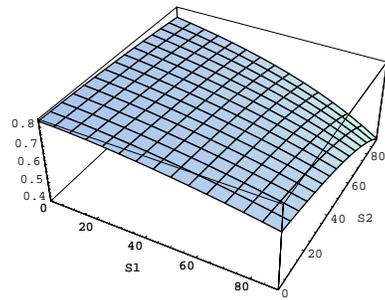
4.2.1 Profitability

From the point of view of a type i country, profitability of adherence is analyzed as the difference between gains obtained by adherence and by the situation of non-cooperation, respectively:

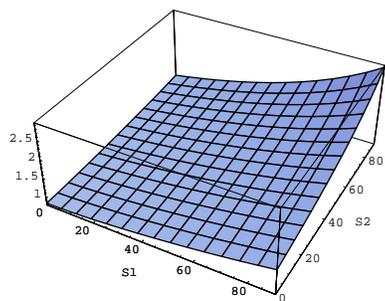
$$\Gamma_i(S_1, S_2) = P_i(S_1, S_2) - Q_i(0, 0) \quad (26)$$



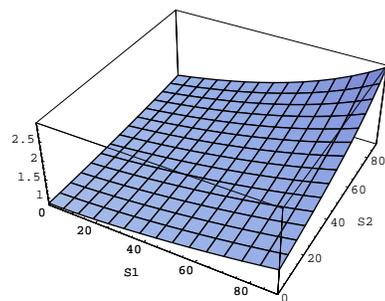
Signatory country



Average emission

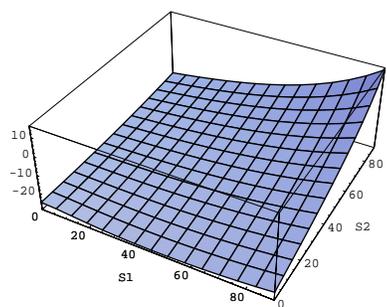


Type 1 non-signatory country

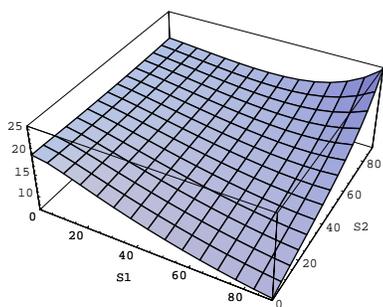


Type 2 non-signatory country

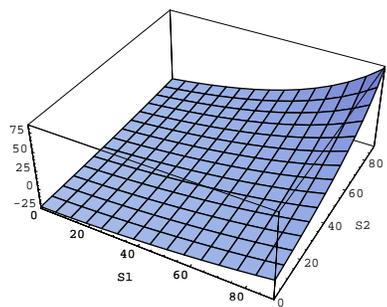
Figure 10: Levels of emissions in relation to the composition of the agreement (S_1 et S_2).



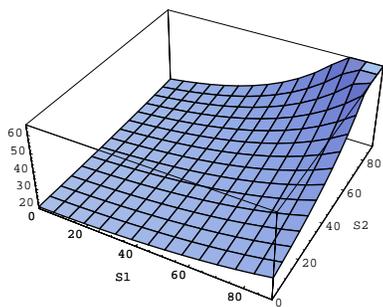
Type 1 signatory country



Type 2 signatory country

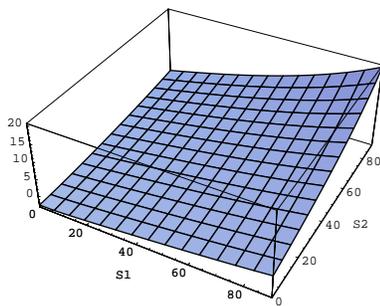


Type 1 non-signatory country



Type 2 non-signatory country

Figure 11: Individual gains in relation to the composition of the agreement (S_1 et S_2).



includegraphics[width = 5cm]/home/public2/combette/these/idee/strasbourg99/gainWT.eps

Figure 12: Average gain in relation to the composition of the agreement (S_1 et S_2).

It seems that countries which are less sensitive to environmental damage (type 2) have less incentive to participate in the agreement. A type 2 country will be penalized by its adherence if the agreement is composed of a large number of type 1 countries and a small number of type 2 countries. The effort it must make is therefore too high compared to the gain it can obtain through the improvement of environmental quality (see Figures 11 and 12).

Figure 13 illustrates this point. Agreements which are profitable to none of the signing countries figure in zone 1. Agreements in zone 2 are profitable from the point of view of a type 1 country but not from that of a type 2 country. In fact, the only agreements which appear to be profitable from the points of view of both types of country figure in zone 3. In this final zone are found the agreements which have a large enough number of signatories and for which the composition is not sufficiently equilibrated to penalize type 2 countries.

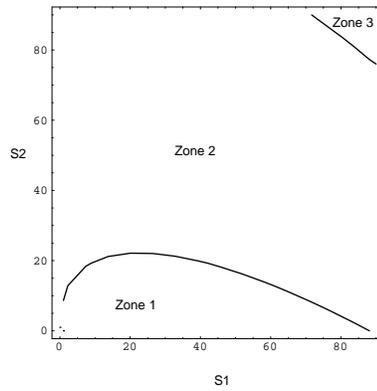


Figure 13: Site of profitable (S_1 and S_2) agreements.

4.2.2 Stability

The simulations developed for the very divers values of the N_1 and N_2 effectives and the h_1 and h_2 externality parameters make it possible to suggest that one could extrapolate the result obtained within a homogenous framework. This result, (see Figures 14) is founded on the four conditions generalizing the stability definition within a homogenous framework based on the following stability functions:

$$\text{for } i = 1 \text{ or } 2 : \Delta_i(S_i, S_{-i}) = V_{i \in S}(S_i, S_{-i}) - V_{i \notin S}(S_i - 1, S_{-i})$$

Définition 5 *An agreement with an (S_1^*, S_2^*) composition is stable if:*

$$\Delta_1(S_1^*, S_2^*) > 0 \tag{27}$$

$$\Delta_2(S_1^*, S_2^*) > 0 \tag{28}$$

$$\Delta_1(S_1^* + 1, S_2^*) < 0 \tag{29}$$

$$\Delta_2(S_1^*, S_2^* + 1) < 0 \tag{30}$$

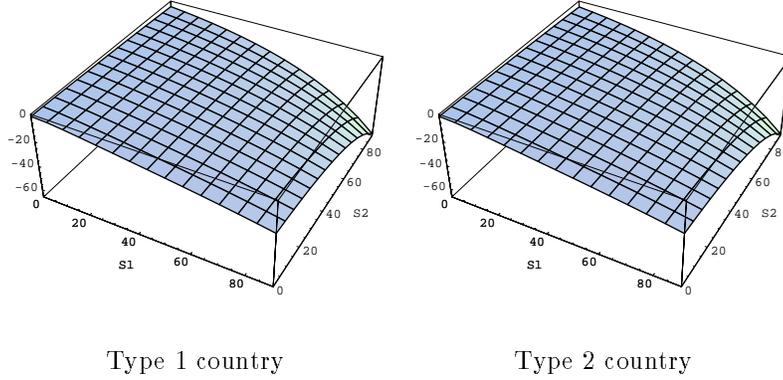


Figure 14: Site of possible transfers to new entrees for an $(S_1$ and $S_2)$ agreement.

4.3 Transfer and sequential engagement

Sequential engagement in the agreement presents substantial differences depending on whether one looks at it from the point of view of a type 1 or type 2 country and whether the potential entree is a type 1 or type 2 country. Here, too, one finds the same type results as those found in section 4.2.1.

The characteristic functions associated with this engagement problem are also based on the mode of transfer sharing. Here we will consider only an egalitarian mode, meaning that all members of the agreement, no matter what their type, contribute in the same manner to compensating the losses sustained by the entree. Let us define functions Θ_{ij} as the gains, nets of transfer, of type i

countries belonging to the agreement whenever the entree is a type j country:

$$\Theta_{11}(S_1, S_2) = P_1(S_1 + 1, S_2) - P_1(S_1, S_2) - \frac{Q_1(S_1, S_2) - P_1(S_1 + 1, S_2)}{S_1 + S_2} \quad (31)$$

$$\Theta_{12}(S_1, S_2) = P_1(S_1, S_2 + 1) - P_1(S_1, S_2) - \frac{Q_2(S_1, S_2) - P_2(S_1, S_2 + 1)}{S_1 + S_2} \quad (32)$$

$$\Theta_{21}(S_1, S_2) = P_2(S_1 + 1, S_2) - P_2(S_1, S_2) - \frac{Q_1(S_1, S_2) - P_1(S_1 + 1, S_2)}{S_1 + S_2} \quad (33)$$

$$\Theta_{22}(S_1, S_2) = P_2(S_1, S_2 + 1) - P_2(S_1, S_2) - \frac{Q_2(S_1, S_2) - P_2(S_1, S_2 + 1)}{S_1 + S_2} \quad (34)$$

Let us consider an agreement (S_1, S_2) which is profitable for each of the members (i.e. $\Gamma_i(S_1, S_2) > 0$, $i = 1, 2$). The engagement of the members of this agreement vis-à-vis a type j entree will not be effective unless the following four conditions are all met:

$$\Theta_{ij}(S_1, S_2) > 0 \quad i, j = 1, 2 \quad (35)$$

$$(36)$$

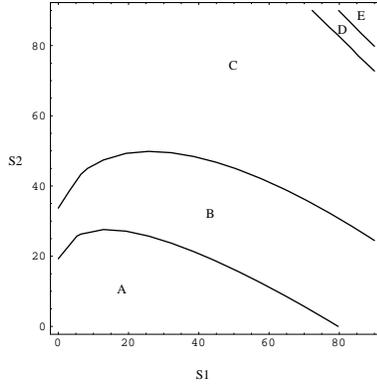


Figure 15: Site of $(S_1$ and $S_2)$ agreements for which the transfers are positive.

The simulations show that the engagement is only possible on the basis of a composite agreement generally composed of a large number of type 1 and type 2 countries (Figure 15). In zone A, none of the signatories are ready to

finance the adherence of a non-signatory. In zone B, type 1 signatories (and only them) are ready to propose a positive transfer to non-signatories of the same type (but only to them). It is only in zone D that type 2 signatories are ready to finance the adherence of type 2 countries. Finally, in zone E they will consent to positive transfers to type 1 countries.

Conclusion

The different cases examined in this contribution show the difficulty in reaching an environmental agreement on the sole basis of voluntary signing of an agreement by countries sharing a common environmental resource. In the case of a segmented market a partial agreement can be envisaged. However, in the case where there is a globally-fixed price for the good the production of which is responsible for the pollution, the incentive for a signatory to deviate is such that no stable agreement is likely to emerge.

International cooperation thus takes on a new significance and must be understood as the expression of an irrevocable engagement on the part of certain countries. Under this condition it is possible to envisage a transfer system which will make it possible to enlarge the agreement from a threshold size.

In the case of heterogeneity, the preceding conclusions are not modified. It remains difficult to reach a stable agreement. Still, this difficulty is compounded by the antagonistic interests of the different types of countries making up the agreement. The results of the simulations presented show that the possibility that the agreement could be preferable to the situation of non-cooperation for *each* of the signatories is quite small.

This final conclusion poses the problem of agreement value sharing amongst the members. In this contribution, we have taken the hypothesis that a signatory's gain results solely from the reduction goal which has been assigned to it. The introduction of lateral transfers susceptible of reducing the antagonisms between the members of the agreement should be the subject of further research.

Annexes

1 Emissions game within an integrated homothetic framework

The general case corresponds to the case where supply, demand and externality parameters differ from one country to another and where agreements grouping countries in some sort of partition can exist. We will see that $\forall \mathcal{S} \subseteq \mathcal{N}$:

- $A_{\mathcal{S}} = \sum_{i \in \mathcal{S}} a_i$, $B_{\mathcal{S}} = \sum_{i \in \mathcal{S}} b_i$, $C_{\mathcal{S}} = \sum_{i \in \mathcal{S}} c_i$, $H_{\mathcal{S}} = \sum_{i \in \mathcal{S}} h_i$, $Y_{\mathcal{S}} = \sum_{i \in \mathcal{S}} y_i$

- $\varepsilon_S = \frac{C_S}{B+C_S}$, $T_S = (B+C) \sum \varepsilon_S + BC \sum H_S \varepsilon_S$

In the particular case where $\mathcal{S} = \mathcal{N}$, we will successively accept: $A = A_{\mathcal{N}}, B = B_{\mathcal{N}}, C = C_{\mathcal{N}}, H = H_{\mathcal{N}}$.

The following hypotheses are formulated:

- The global economy is integrated on the market of the goods in question. Therefore: $A - Bp = Y_{\mathcal{N}}$.
- The national economies are homothetic or, more precisely: $i \neq j \Rightarrow \frac{a_i}{a_j} = \frac{b_i}{b_j} = \frac{c_i}{c_j} = \frac{h_i}{h_j}$. One also sees that by making the parameters a, b, c, h appear: $= \frac{a_i}{a} = \frac{b_i}{b} = \frac{c_i}{c} = \frac{h_i}{h} = k_i$.
- Parameter values will be taken to be such that the levels of production (and emissions) will remain strictly positive within each of the economies, no matter what the coalitions formed.

In the specific case of homogenous countries, it is sufficient to consider that: $k_i = 1$.

Taking account of the hypothesis of positive productions (i.e. emissions), the sole Nash equilibrium between the coalitions is characterized by the price and emission levels hereunder:

$$p = A \frac{\sum \varepsilon_S + C \sum H_S \varepsilon_S}{T_S} \quad (37)$$

$$Y_{\mathcal{N}} = AC \frac{\sum \varepsilon_S}{T_S} \quad (38)$$

$$\forall S : Y_S = A \varepsilon_S \frac{BC(\sum_{S'} H_{S'} \varepsilon_{S'} - H_S \sum \varepsilon_{S'}) + (B + C_S) \sum \varepsilon_{S'}}{T_S} \quad (39)$$

$$i \in S : y_i = \frac{c_i}{C_S} Y_S \quad (40)$$

By construction, the highest sum of national gains is obtained by the grand coalition. The (\mathcal{N}, W) is essential.

Proposition 7 *Within a general framework of homothetic economies, the level of production is lower than the level of equilibrium in the absence of any externality: $y_i \leq c_i p$.*

In the absence of any externality, one sees that the productive sector proper to each economy produces in such a way that marginal cost and price are equal ($y_i = c_i p$).

2 Gains upon adherence to the agreement

One must remember the following notations:

- $X_S = N(Nb + c) + (N - S)(S - 1)c$
- $Z_S = Nb + c + (S - 1)(Sb + c)$
- $\tau = N^2bch$
- $R_S = (b + c)X_S + \tau Z_S$
- $Q_S = (N - S)(S - 1)\tau$
- $U_S = S(S - 1)\tau$

After a few manipulations, the gains can be expressed in the following manner:

$$\forall i \in S : V_{i \in S} = \frac{a^2c(b + c - \tau)X_S^2 + 2\tau Z_S(X_S - Q_S) - bQ_S^2}{2b R_S^2} \quad (41)$$

$$\forall i \notin S : V_{i \notin S} = \frac{a^2c(b + c - \tau)X_S^2 + 2\tau Z_S(X_S + U_S) - bU_S^2}{2b R_S^2} \quad (42)$$

$$\frac{W_N}{N} = \frac{a^2c(b + c - \tau)X_S^2 + 2\tau Z_S X_S - bQ_S U_S}{2b R_S^2} \quad (43)$$

By-products of individual gains makes it possible to analytically study their evolution in relation to the size of the agreement.

$$\frac{\partial V_{i \in S}}{\partial S} = \frac{a^2c}{b} \frac{\tau^2}{R_S^3} (Nb + c)^2 [N^2(S - 1)b - (S^2 - 3NS + N^2 + N)Sc - S(N - S^2)\tau] \quad (44)$$

$$\frac{\partial V_{i \notin S}}{\partial S} = \frac{a^2c}{b} \frac{\tau^2}{R_S^3} (Nb + Sc)(S^2c + 2NSb - Nb) [N^2b + (2N - 1)c + \tau] \quad (45)$$

3 Gains in the emissions game in the heterogeneous case

Solving the first phase in the adherence game leads to simplified expressions using certain of the previously defined expressions, as well as the following intermediary expressions:

- $D_S = H - (N - S + 1)H_S$
- $\sigma = Nbc$
- $F_S = (Nb + c)H + N(S - 1)bH_S$

- $G_S = (b + c)X_S + \sigma F_S$
- $E_{S,i} = \frac{F_S - X_S h_i}{Nb + c}$

State gains are thus expressed in the following manner:

$$\forall i \in S : V_{i \in S} = \frac{a^2 c (b + c - N \sigma h_i) X_S^2 + 2 \sigma F_S (X_S + N \sigma D_S) - N^2 \sigma^2 b D_S^2}{2b G_S^2} \quad (46)$$

$$\forall i \notin S : V_{i \notin S} = \frac{a^2 c (b + c - N \sigma h_i) X_S^2 + 2 \sigma F_S (X_S + N \sigma E_{S,i}) - N^2 \sigma^2 b E_{S,i}^2}{2b G_S^2} \quad (47)$$

References

- Barrett, S. (1994). Self-Enforcing International Environmental Agreements, *Oxford Economic Papers* **46**: 878–894.
- Barrett, S. (1997a). Heterogeneous International Environmental Agreements, in C. Carraro (éd.), *International Environmental Negotiations. Strategic Policy Issues*, Edward Elgar, Cheltenham, UK, pp. 9–25.
- Barrett, S. (1997b). Towards a Theory of Environmental Cooperation, in C. Carraro et D. Siniscalco (éds), *New Directions in the Economic Theory of the Environment*, Cambridge University Press, Cambridge, pp. 239–280.
- Black, J., Levi, M. et De Meza, D. (1993). Creating a Good Atmosphere : Minimum Participation for Tackling the Greenhouse Effect, *Economica* **60**: 281–293.
- Botteon, M. et Carraro, C. (1997). Burden-Sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries, in C. Carraro (éd.), *International Environmental Negotiations. Strategic Policy Issues*, Edward Elgar, Cheltenham, pp. 26–55.
- Carraro, C. et Moriconi, F. (1998). Endogenous Formation of Environmental Coalitions, *mimeo*, Fondazione Eni Enrico Mattei (FEEM).
- Carraro, C. et Siniscalco, D. (1992). The International Dimension of Environmental Policy, *European Economic Review* **36**(2/3): 379–387.
- Carraro, C. et Siniscalco, D. (1993). Strategies for the International Protection of Environment, *Journal of Public Economics* **52**(3): 309–328.

- Chander, P. et Tulkens, H. (1992). Theoretical Foundations of Cost Sharing in Transfrontier Pollution Problems, *European Economic Review* **36**(2/3): 388–399.
- Chander, P. et Tulkens, H. (1995). A Core-theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution, *International Tax and Public Finance* **2**: 273–294.
- Chander, P. et Tulkens, H. (1997). The Core of an Economy with Multilateral Environmental Externalities, *International Journal of Game Theory* **26**: 379–401.
- d’Aspremont, C. A., Jacquemin, A., Gabszewicz, J. J. et Weymark, J. (1983). On the Stability of Collusive Price Leadership, *Canadian Journal of Economics* **16**: 17–25.
- Rotillon, G. et Tazdaït, T. (1996). Jeux, Coopération et Problèmes Environnementaux Globaux, *Cahiers d’Economie et de Sociologie Rurales* **39-40**: 252–268.
- Tulkens, H. (1997). Cooperation versus Free Riding in International Environmental Affairs : Two Approaches, *Working Paper 9752*, Center for Operational Research and Econometrics (CORE), Université Catholique de Louvain.