Financial transfers to ensure cooperative international optimality in stock pollutant abatement

by M. Germain[1], Ph.L. Toint[2] and H. Tulkens[3]

Rapport 97/4 March 3, 1997

[1] CORE, Université Catholique de Louvain.
Email : germain@core.ucl.ac.be

[2] Department of Mathematics, Facultés Universitaires ND de la Paix,
61, rue de Bruxelles, B-5000 Namur, Belgium, EU
Email : pht@math.fundp.ac.be
Current reports available by anonymous ftp from the directory
"pub/reports" on thales.math.fundp.ac.be

[3] CORE and IRES (Institut de Recherches Economiques et Sociales),
Université Catholique de Louvain,
and Facultés Universitaires Saint-Louis, Bruxelles.

Financial transfers to ensure cooperative international optimality in stock pollutant abatement

Marc Germain* Philippe Toint† Henry Tulkens‡

January 1997

Abstract

It is well known that the transnational character of many environmental problems requires cooperation amongst the countries involved, if a social optimum is at all to be achieved. Most of the numerous contributions dealing with the problems raised by the cooperation issue bear only on pollutants that do not accumulate: they are thus only relevant in a static, or a-temporal, context. On the other hand, many contributions which deal with the dynamic dimension of the problem when the pollutant accumulates leave aside the issue of the voluntary implementation of the international optimum. The aim of the present contribution is to overtake the two above limitations. Using both cooperative and differential game theories, we design a scheme of financial transfers between countries such that aggregate abatement costs are covered in a way that makes cooperation both individually rational and strategically stable.

1 Introduction

It is well known that the transnational character of many environmental problems (e.g. greenhouse gas emissions, acid rains, pollution of international waters) requires cooperation amongst the countries involved, if a social optimum is at all to be achieved. The issues raised thereby have often been addressed, in the economic literature, in terms of concepts borrowed from cooperative game theory (see e.g. Mäler, 1989).

However, most of these contributions have only dealt with static (one shot) games, which are only appropriate for flow pollution models. When the environmental damage arises from the presence of stock pollutants that accumulate (and possibly decay), the problem acquires a

*Center for Operational Research and Econometrics (CORE), Université Catholique de Louvain. Email: germain@core.ucl.ac.be . Germain thanks the Fonds de Développement Scientifique of the Université Catholique de Louvain (projet n° 7-29039).

†Département de Mathématique, Facultés Universitaires Notre-Dame de la Paix, Namur. Email: pht@math.fundp.ac.be .

‡Center for Operational Research and Econometrics (CORE) and IRES (Institut de Recherches Economiques et Sociales), Université Catholique de Louvain, and Facultés Universitaires Saint-Louis, Bruxelles. Email : tulkens@core.ucl.ac.be .

dynamic and intertemporal dimension. In this case, differential game theory is a more appropriate tool for the analysis of cooperation, as is done in, e.g., van der Ploeg and de Zeeuw (1992), Kaitala, Pohjola et Tahvonen (1992), Hoel (1992) and Tahvonen (1993). Such analysis proceeds by evaluating the gain to be obtained from cooperation, at an international optimum, in comparison to the non cooperative state of affairs modeled as a Nash non cooperative equilibrium (open loop or closed loop).

Most often, these contributions leave aside the issue of the voluntary implementation of the international optimum. This is an important drawback because in this context no supranational authority can be called upon to impose the optimum. In view of ensuring such implementation, it has often been suggested that financial transfers between the countries involved might reinforce incentives towards cooperation. This property, understood in the sense of the theory of the core of a cooperative game, has in effect been demonstrated by Chander and Tulkens (1995, 1996), who propose a particular transfer scheme based on parameters reflecting the relative intensities of the countries' environmental preferences. The result is obtained, however, for flow pollutants only, that is, in a static game model.

The present contribution establishes that the same property can be obtained for stock pollutants, within the wider context of differential games. It also shows that the transfer scheme implies a sharing rule between the countries of the aggregate emission abatement cost. Moreover, with this cost sharing interpretation, the transfers appear to be a strategically stable form of joint implementation of the international optimum.

The paper is structured as follows. Section 1 presents the stock pollutant model and characterizes emissions that correspond to an international optimum without transfers. In Section 2, emissions corresponding to a non cooperative (open loop) equilibrium are similarly characterized and contrasted with the optimal ones. In Section 3, financial transfers are formulated that are shown to induce voluntary cooperation of each country individually, whereas in Section 4 these transfers are further specified as to achieve cooperation in the core theoretic sense, i.e. for all subsets of countries. Finally, Section 5 considers some issues that arise when the time profile of the transfers is examined, and not only their aggregate present value.

2 The model

The model is formulated in discrete time. There are n regions or countries, indexed $i \in \mathcal{N} = \{1, \ldots, n\}$. Economic activity in each country entails pollutant emissions: let $\mathbf{E}_t = (E_{1t}, \ldots, E_{nt})'$ be the vector of these emissions at date t. They spread across the various countries and contribute to the formation of a pollutant stock S (≥ 0) which is determined by the equation

$$S_t \;=\; [1-\delta]S_{t-1} + E_t \tag{1}$$

where by definition $E_t = \sum_{i=1}^{n} E_{it}$ is the sum of the emissions and δ is the rate of natural decay of the stock $(0 < \delta < 1)$ [1].

[1]Such as e.g. atmospheric CO_2. Pollution is here assumed to be global, in the sense that it spreads uniformly across countries. Some pollutions, such as acid rains or river water pollution, do not obey that equation: they are

The pollutant stock implies damages for each country i: at period t, and in monetary terms, these amount to $D_i(S_t)$, where D_i is supposed to be positive, increasing and convex ($D_i > 0$ if $S > 0, D_i \geq 0$ if $S = 0, D'_i > 0, D''_i \geq 0$). The only means the countries have to keep the stock of pollutant under control is at its source, that is, by reducing their emissions. More precisely, associate to country i an abatement function $C_i(E_i)$ positive, decreasing and strictly convex ($C_i \geq 0, C'_i < 0, C''_i > 0$), that expresses the total costs incurred by the industries of country i due to the fact that total emissions in that country are restricted to E_i. The decreasing character of the function reflects the obvious phenomenon of the increase in aggregate costs when emissions are abated.

The international optimum obtained through cooperation corresponds to the minimisation of the sum of the discounted total costs of all countries involved; in other terms, to the solution of the problem

$$\min_{\{\mathbf{E}_t\}_{t\in\mathcal{T}}} \sum_{t=1}^{T}\sum_{i=1}^{n} \beta^t [C_i(E_{it}) + D_i(S_t)] \tag{2}$$

$$\text{such that} \begin{cases} S_t = [1-\delta]S_{t-1} + E_t;\ S_0 \text{ given} \\ E_{it} \geq 0,\ \forall t \in \mathcal{T},\ \forall i \in \mathcal{N} \end{cases}$$

where β is the discount factor ($0 < \beta \leq 1$), $\mathcal{T} = \{1, \ldots, T\}$, T being the time horizon of the problem (T positive, integer and possibly infinite) [2].

The necessary conditions for an interior minimum of problem (2) yield optimal trajectories for the emissions $\{\mathbf{E}^*_t\}_{t\in\mathcal{T}}$ and for the pollutant stock $\{S^*_t\}_{t\in\mathcal{T}}$ which satisfy the following equations: (the details of the computations are provided in Appendix 1):

$$C'_i(E^*_{it}) + \sum_{\tau=t}^{T} \beta^{\tau-t}[1-\delta]^{\tau-t} \sum_{j=1}^{n} D'_j(S^*_\tau) \ = \ 0,\ \forall i \in \mathcal{N},\ \forall t \in \mathcal{T}. \tag{3}$$

This expression means that at every moment of time, the marginal abatement cost of country i must be equal to the sum over all countries of their respective marginal damage costs over the entire horizon remaining. In this setting, the higher the damages (either intrinsically *via* the D'_i and the stock, or *via* the weight attributed to these in the future, that is, *via* β), the more important it is to abate and the more restricted are the optimal emissions; similarly, the slower the pollutant stock depreciates (the smaller δ is).

In the particular case where the damage functions are linear, that is, when

$$D_i(S_t) = \pi_i S_t, \quad i \in \mathcal{N}, \tag{4}$$

called "directional". However, the results presented here could be extended to directional pollutions, by introducing other forms of transfer functions between emissions and depositions of pollutant (as e.g. in ¡Germain, Toint et Tulkens [1996], where we use such a function in the framework of a discrete model without pollutant accumulation).

[2] In the case where T is infinite, stronger conditions must be satisfied by the parameters and by the functions C_i et D_i to guarantee convergence of the objective and of the optimality conditions. In particular, as far as (2) is concerned, it is sufficient that $\beta < 1$ and that the sets $\{E_i : C_i(E_i) = k\}$ and $\{S : D_i(S) = k\}$ ($\forall i \in \mathcal{N}$) be bounded for any constant k.

optimal emissions obey the following equations:

$$C_i'(E_{it}^*) + \pi_{\mathcal{N}} \frac{1-\beta^{T+1-t}[1-\delta]^{T+1-t}}{1-\beta[1-\delta]} \;=\; 0,\; i \in \mathcal{N} \tag{5}$$

where by definition $\pi_{\mathcal{N}} = \sum_{i=1}^{n} \pi_i$. With an objective function such as (2), we get the usual result that with linear damages, optimal emissions do not depend upon the initial stock (Tahvonen [1993]).

Under the same linearity hypothesis, we observe that because the functions $C_i(E_i)$ are decreasing and convex over their domain of definition, optimal emissions are increasing with time if T is finite. Indeed, when T is finite, (3) suggests that the incentive to abate (as measured by the discounted sum of future damages) decreases as one gets closer to the horizon.

This does not apply when T is infinite: this explains that optimal emissions are constant over time in this case.

3 The open-loop non cooperative Nash equilibrium

In an open-loop non cooperative Nash equilibrium, each country chooses once and for all, at time 0, its emissions for all future periods, given the existing pollutant stock at the starting point S_0 and the emissions of the other countries. The open-loop characteristic is that, once its strategy is chosen, a country no longer changes it, and thus does not re-adjusts it according to the evolution of the stock over time.

Formally, each country i ($i \in \{1, \ldots, n\}$) solves the following problem:

$$\min_{\{E_{it}\}_{t\in\mathcal{T}}} \sum_{t=1}^{T} \beta^t [C_i(E_{it}) + D_i(S_t)] \tag{6}$$

$$\text{such that} \left\{ \begin{array}{l} S_t = [1-\delta]S_{t-1} + E_t;\; S_0 \text{ given} \\ E_{it} \geq 0,\; \forall t \in \mathcal{T};\; E_{jt}\; (j \neq i) \text{ given.} \end{array} \right.$$

The open-loop non cooperative Nash equilibrium is obtained by solving simultaneously the n problems above. The emissions and stock trajectories $\{\mathbf{E}_t^N\}_{t\in\mathcal{T}}$ and $\{S_t^N\}_{t\in\mathcal{T}}$, respectively, relative to that equilibrium are the solution of (see Appendix 2):

$$C_i'(E_{it}^N) + \sum_{\tau=t}^{T} \beta^{\tau-t}[1-\delta]^{\tau-t} D_i'(S_\tau^N) = 0,\; \forall i \in \mathcal{N},\; \forall t \in \mathcal{T}. \tag{7}$$

Contrary to what is the case at an optimum, country i takes account of the impact of its decisions only on its own environment (hence the presence of $D_i'(S_t^N)$ in lieu of $\sum_{j=1}^{n} D_j'(S_t^*)$ in (7)).

In the particular case of linear damage costs, the countries' emissions obey the following equations:

$$C_i'(E_{it}^N) + \pi_i \frac{1-\beta^{T+1-t}[1-\delta]^{T+1-t}}{1-\beta[1-\delta]} \;=\; 0,\; i \in \mathcal{N}. \tag{8}$$

As it was the case at the international optimum, and for reasons that are quite similar (see previous section), emissions are increasing if T is finite and constant if T is infinite.

In view of the fact that $\forall i,\ \pi_i < \pi_{\mathcal{N}}$, we notice , by comparing (5) to (8), that if the D_i's are linear, the Nash equilibrium emissions of country i are always higher that the internationally optimal ones. This is only natural since at a Nash equilibrium, the abatement decisions of each country are based on their impact on their own environment only.

4 Transfers inducing an individually rational international optimum

Let us recall that $\mathbf{E}_1^*, \ldots, \mathbf{E}_T^*$ and $S_0, S_1^*, \ldots, S_T^*$ are internationally optimal emissions and stock trajectories. Then,

$$W(S_0) \ = \ \sum_{t=1}^{T} \beta^t \sum_{i=1}^{n} [C_i(E_{it}^*) + D_i(S_t^*)] \tag{9}$$

is the optimal aggregate total cost of all countries, and

$$W_i(S_0) \ = \ \sum_{t=1}^{T} \beta^t [C_i(E_{it}^*) + D_i(S_t^*)] \tag{10}$$

is the share of $W(S_0)$ borne by country i ($i \in \mathcal{N}$). Of course $W(S_0) = \sum_{i=1}^{n} W_i(S_0)$.

Let us remind also that $\mathbf{E}_1^N, \ldots, \mathbf{E}_T^N$ and $S_0, S_1^N, \ldots, S_T^N$ are the emissions and stock trajectories at the Nash non cooperative equilibrium. Then:

$$V_i(S_0) \ = \ \sum_{t=1}^{T} \beta^t [C_i(E_{it}^N) + D_i(S_t^N)] \tag{11}$$

is the discounted total cost of country i at the Nash equilibrium and

$$V(S_0) \ = \ \sum_{i=1}^{n} V_i(S_0) \tag{12}$$

s the sum of these total costs over the set of all countries. Clearly, by definition of the optimum, $V(S_0) \geq W(S_0)$.

If $\forall i \in \{1, \ldots, T\} : W_i(S_0) \leq V_i(S_0)$, then international cooperation through the optimal trajectories is individually rational, since all countries have an interest in cooperating. By contrast, a country i for which $W_i(S_0) > V_i(S_0)$, has no such interest in cooperating, even if cooperation is globally favourable for all countries taken as a whole. In order to gain the cooperation of such a country, some form of compensation must be designed, e.g. in the form of financial transfers.

Borrowing from Chander and Tulkens [1995, 1996] who dealt with such kind of issue in a static context (as opposed to the intertemporal framework considered here), we propose financial transfers between the countries that are of the following form:

$$\tau_i(S_0) \ = \ -[W_i(S_0) - V_i(S_0)] + \mu_i[W(S_0) - V(S_0)], \ i \in \mathcal{N} \tag{13}$$

where the parameters μ_i are arbitrary fixed values chosen between 0 and 1 and satisfy

$$\sum_{i=1}^{n} \mu_i \ = \ 1. \tag{14}$$

This last condition guarantees that the financial transfers defined by (13) be balanced, in the sense that

$$\sum_{i=1}^{n} \tau_i(S_0) \;=\; 0. \tag{15}$$

If with international cooperation, country i receives a financial transfer equal to $\tau_i(S_0)$, then its discounted total cost along the optimal trajectory becomes

$$\tilde{W}_i(S_0) = W_i(S_0) + \tau_i(S_0). \tag{16}$$

It is easy to verify that this optimal cost *with transfers* borne by i is smaller than or equal to the cost it would bear at the Nash equilibrium. Indeed, it follows from (13) and (16) that

$$\tilde{W}_i(S_0) \;=\; V_i(S_0) + \mu_i[W(S_0) - V(S_0)] \leq V_i(S_0), \;\forall i \in \{1, \ldots, n\} \tag{17}$$

because the μ_i are positive and $W(S_0) - V(S_0) \leq 0$ by definition of an optimum. Thus with the transfers defined by (13), international cooperation becomes individually rational.

5 Transfers inducing rationality in the sense of coalitions

The fact that the parameters μ_i must lie between 1 and 0 is not enough to make them unique. The aim of this section is to use the degrees of freedom left so as to obtain certain properties of "rationality in the sense of coalitions" as suggested by the "core" concept in cooperative game theory. To do so we will adapt to our present dynamic framework the approach in terms of "global" games proposed by Chander and Tulkens (1995, 1996) in a static framework.

A cooperative game in characteristic function form (with transferable utility)is defined by the pair $[\mathcal{N}, w(\cdot; S_0)]$, where $\mathcal{N} = \{1, \ldots, n\}$ is the set of players (i.e. the n countries) and w is the characteristic function. The space of the players' strategies on which this function is defined is specified as follow: for each country i, this space is the interval of all possible emission levels, i.e. $[0, \infty[$. For any coalition $U \subseteq \mathcal{N}$, it is the set product of these intervals over the members of U.

The characteristic function of the game may then be defined by using the concept of *partial Nash equilibrium with respect to a coalition* proposed by Chander and Tulkens (1995, 1996), subject to some adaptation to the present framework. Specifically, the function specifies as follow the vectors $\mathbf{E}_t$ ($t \in \mathcal{T}$) of the strategies adopted by all players when a coalition $U \subseteq \mathcal{N}$ forms:

(i) For the members of U, the trajectories are described by $\{E_{it}^U : i \in U; t \in \mathcal{T}\}$ and are solution of

$$\min_{\{E_{it}\}_{\substack{i \in U \\ t \in \mathcal{T}}}} \sum_{t=1}^{T} \sum_{i \in U} \beta^t [C_i(E_{it}) + D_i(S_t)] \text{ s.t. } (1) \tag{18}$$

where $\forall j \in \mathcal{N} \setminus U, \forall t \in \mathcal{T}, E_{jt} = E_{jt}^U$ as defined by *(ii)*;

(ii) For the players out of the coalition U, the trajectories $\{E_{jt}^U : t \in \mathcal{T}\}$, $j \in \mathcal{N} \setminus U$, are solution of the simultaneous resolution of the problems

$$\min_{\{E_{jt}\}_{t \in \mathcal{T}}} \sum_{t=1}^{T} \beta^t [C_j(E_{jt}) + D_j(S_t)] \text{ s.t. } (1), \;\; j \in \mathcal{N} \setminus U \tag{19}$$

where $\forall i \in U, \forall t \in \mathcal{T}, E_{it} = E_{it}^U$ as defined by *(i)*.

Thus one supposes that if a coalition forms, its members minimize *together* the sum of their total discounted costs, while each country outside of it reacts by minimizing its own total discounted cost. The hypothesis about this last behaviour justifies the expression of "partial Nash equilibrium" that has just been used.

On this basis, the characteristic function[3] may be written

$$w(U; S_0) \;=\; \sum_{t=1}^{T} \sum_{i \in U} \beta^t [C_i(E_{it}^U) + D_i(S_t^U)], \;\; U \subseteq \mathcal{N} \tag{20}$$

where $S_t^U = [1-\delta] S_{t-1}^U + E_t^U, \forall t \in \mathcal{T}$ (with $S_0^U \equiv S_0$ given). Because of (18) and (19), one notes that $w(\mathcal{N}; S_0)$ is equal to $W(S_0)$ defined by (9), in other words to the optimal total cost of all countries.

For the game $[\mathcal{N}, w(\cdot; S_0)]$, any n-dimensional and non negative vector whose components sum up to $w(\mathcal{N}; S_0)$ is called an *imputation* of the game. Thus an imputation may be seen as a share of the optimal total cost between the different players.

The vector $(W_1, (S_0), \ldots, W_n(S_0))$, where $W_i(S_0)$ is defined by (10) constitutes such an imputation, where each country bears itself the abatement and damage costs induced by the optimal strategy $\{\mathbf{E}_t^* : t = 1, \ldots, T\}$. But the possibility of financial transfers between countries implies that there exists (an infinity of) other imputations associated with the same strategy. Indeed, all vectors $(\tilde{W}_1(S_0), \ldots, \tilde{W}_n(S_0))$ defined by (16) such that (15) is verified is an imputation.

A *solution* of the game is an imputation that verifies certain properties. Among the imputations we have just defined thanks to the transfers $\tau_i(S_0)$, those who verify the condition

$$\sum_{i \in U} \tilde{W}_i(S_0) \;\leq\; w(U; S_0), \forall U \subseteq \mathcal{N}, \forall S_0 > 0 \tag{21}$$

are said *to belong to the core* of the game. The core is thus the set of imputations that attribute to any coalition a share of the aggregat cost $W(S_0)$ less than or equal to the mimimal cost $w(U; S_0)$ this coalition could achieve by itself.

We will call "rational in the sense of coalitions" any imputation which belongs to the core of the game described above: with such a share indeed, no coalition has an interest to form because its members, taken as a whole, would bear a total cost greater than what is proposed to them.

We now show that when *the damage functions are linear* (cfr. (4)), such an imputation exists and can be exhibited in terms of specific values of the parameters μ_i appearing in (13).

Theorem: Let $\{\mathbf{E}_t^* : t = 1, \ldots, T\}$ be the trajectory of the optimal emissions solution of problem (2), and $\{\mathbf{E}_t^N : t = 1, \ldots, T\}$ be the trajectory of emissions at the Nash equilibrium solution of problems (6). Let $(W_1, (S_0), \ldots, W_n(S_0))$ and $(V_1(S_0), \ldots, V_n(S_0))$ be the vectors of total discounted costs of each country along these two trajectories (as defined in (10) and (11)). Then,

[3] Labelled in this case by Chander and Tulkens (1995, 1996) as a "γ-characteristic function". Other specifications are possible, as discussed by these authors.

under the assumption of convexity of functions C_i and linearity of functions D_i $(i \in \mathcal{N})$, the imputation $(\tilde{W}_1(S_0), \ldots, \tilde{W}_n(S_0))$ defined by

$$\tilde{W}_i(S_0) \;=\; W_i(S_0) + \tilde{\tau}_i(S_0), \imath = 1, \ldots, n \tag{22}$$

where

$$\tilde{\tau}_i(S_0) \;=\; -[W_i(S_0) - V_i(S_0)] + \frac{\pi_i}{\pi_{\mathcal{N}}}[W(S_0) - V(S_0)] \tag{23}$$

and $W(S_0)$ and $V(S_0)$ are respectively defined by (9) and (12), belongs to the core of the cooperative game $[\mathcal{N}, w(\cdot; S_0)]$.

Proof: cfr. appendix 3.

It is easy to see that the coefficients defined by $\delta_i = \pi_i/\pi_{\mathcal{N}}$ (where the π_i are the coefficients of the damage functions (4)) are contained between 0 and 1 and verify (14), and that the financial transfers $\tau_i(S_0)$ $(i \in \mathcal{N})$ are balanced in the sense of (15). By analogy with individual rationality, we call the optimal trajectory using transfers (23) "strategically stable" or "rational in the sense of coalitions".

One may note also that this property is satisfied whatever are the initial stock S_0 and the temporal horizon T (which may be infinite).

6 The time dimension of the financial transfers

The transfers introduced in Sections 3 and 4 have been defined on a global basis: they are given as total sums related to the complete period $\mathcal{T} = \{1, \ldots, T\}$ where cooperation takes place. As a consequence, the question of their precise allocation in time during this period remains open, in particular when the horizon is infinite[4]. Even in the case where it is finite, the countries may benefit from parts of the cooperation dividends before the horizon is attained. The purpose of the present section is to briefly analyze some simple situations that may occur.

The first observation is that, if the international optimum globally dominates the Nash equilibrium (in other words $W(S_0) < V(S_0)$, where W and V are defined by (9) and (12)), this domination is not *monotonic*, in the sense that it does not imply that, for each elementary period (a year, for instance) $t \in \mathcal{T}$, the global cost for this period is inferior at the international optimum to that at the Nash equilibrium. Let us define

$$J_t^* = \sum_{i=1}^{n} J_{it}^* = \sum_{i=1}^{n}[C_i(E_{it}^*) + D_i(S_t^*)]\beta^t, \;\forall t \in \mathcal{T}$$

$$J_t^N = \sum_{i=1}^{n} J_{it}^N = \sum_{i=1}^{n}[C_i(E_{it}^N) + D_i(S_t^N)]\beta^t, \;\forall t \in \mathcal{T}.$$

In order to focus the exposition, we will only consider here the three following assumptions.

[4]This question has also been considered by Zaccour(1994). This author distinguishes three possible decomposition modes of the global transfer, of which one has the interesting feature that it ensures a gain for each party at each period, provided the parties are not myopic.

Assumption 1: $J_t^* \ \leq \ J_t^N, \ \forall t \in \mathcal{T}$.

The international optimum monotonically dominates the Nash equilibrium. In this case, the global surplus generated by cooperation increases every year and the global transfers defined by (13) may be decomposed in annual transfers of the form

$$\Theta_{it} = -[J_{it}^* - J_{it}^N] + \mu_i[J_t^* - J_t^N], i \in \mathcal{N}, t \in \mathcal{T}. \tag{24}$$

One easily verifies from (9)-(13) and Assumption 1 that $\tau_i(S_0) = \sum_{t=1}^T \Theta_{it}$.

Assumption 2: $\exists\ M : 1 \leq M < T : \begin{cases} J_t^* < J_t^N \text{ if } 1 \leq t \leq M, \\ J_t^* > J_t^N \text{ if } M < t \leq T. \end{cases}$

The international optimum dominates the Nash equilibrium until a certain year M, starting from which the domination order is reversed. In this case, one verifies that

$$W(S_0) - V(S_0) \ = \ \sum_{t=1}^{M}[J_t^* - J_t^N] + \sum_{t=M+1}^{T}[J_t^* - J_t^N] > \sum_{t=1}^{M}[J_t^* - J_t^N]. \tag{25}$$

The global surplus is inferior *in absolute value* to the surplus obtained during the sub-period $\{1, \ldots, M\}$ (both sides of (25) indeed represent negative cost differences). If one applies the annual transfer formula (24), country $i (i \in \mathcal{N})$ will save $\mu_i \sum_{t=1}^M [J_t^* - J_t^N]$ at the optimum compared to the Nash equilibrium during the first M years, and will loose $\mu_i \sum_{t=M+1}^T [J_t^* - J_t^N]$ during the each of the years $\{M+1, \ldots, T\}$. Since the whole procedure is individually rational, the gain in the first period is necessarily superior (in absolute value) to the loss in the second. In this context, if this country wishes to restore, as far as it is concerned, the domination of the international optimum over the Nash equilibrium, it is enough that it levies the loss of the second period from its gains at the first, and then transfer the resulting sum to the second period. Compared to the Nash equilibrium, it will then benefit, at the international optimum with transfers, of a gain of $\mu_i[W(S_0) - V(S_0)]$ during the period $\{1, \ldots, M\}$ and of a zero gain during period $\{M+1, \ldots, T\}$.

Assumption 3: $\exists\ M : 1 \leq M < T : \begin{cases} J_t^* > J_t^N \text{ si } 1 \leq t \leq M, \\ J_t^* < J_t^N \text{ si } M < t \leq T. \end{cases}$

The international optimum is dominated by the Nash equilibrium until year M, after which the order of domination is reversed. This is the most plausible situation, in that the optimum implies abatement costs higher than those of the Nash equilibrium, costs which must financed from the date where they are decided while the gains in damages that they induce are obtained only later[5]. In this context, the countries obviously cannot transfer future gains to cover present losses, in order to ensure the domination of the international optimum over the Nash equilibrium during the early sub-period $\{1, \ldots, M\}$. However, if the countries globaly loose until $t = M$, the impact

[5]As we have noticed in Section 2, the depollution costs will certainly be higher in the case where the damage functions are linear. But, this linearity condition does not imply that Assumption 3 is automatically satisfied, even it is the most reasonable, as is shown by the following example.: with quadratic depollution costs $((C_i(E_i) = \frac{\gamma_i}{2}[\overline{E}_i - E_i]^2, i \in \mathcal{N}$, where γ_i and E_i are positive parameters), linear damages $((D_i(S) = \pi_i S, i \in \mathcal{N}))$ and an

of this loss may be very different from country to country. The most affected countries could then benefit from financial transfers from the less affected ones.

7 Conclusion

Using the theory of non-cooperative games in the framework of a model of intertemporal transboundary pollution, we have constructed a sharing scheme for the abatement costs between the countries involved which makes their cooperation coalitionally rational. This sharing scheme is "strategically stable" in the sense that no coalition of countries can guarantee to its members a total cost lower than that obtained at the international optimum with financial transfers.

The interest of the result is however limited by two restrictive assumptions: the first is the linearity of the damage function, and the second is that the Nash equilibrium considered as reference is one that remains fixed over time. Removing either of these assumptions provides interesting research projects for the future.

8 Acknowledgements

The authors wish to thank C. d'Aspremont, J.-P. van Ypersele and Y. Nesterov for their helpful comments.

Bibliography

Chander P. and H. Tulkens [1992], "Aspects stratégiques des négociations internationales sur les pollutions transfrontières et du partage des coûts de l'épuration", *Revue Economique*, 43, 755-768.

Chander P. and H. Tulkens [1995], "A core-theoretic solution for the design of cooperative agreements on transfrontier pollution", *International Tax and Public Finance*, 2, 279-293.

Chander P. and H. Tulkens [1996], "The core of an economy with multilateral environmental externalities", CORE discussion paper n° 9550 (revised: March 1996), to appear in *International Journal of Game Theory.*

Germain, M., P. Toint and H. Tulkens [1996], "Calcul économique itératif pour les négociations internationales sur les pluies acides entre la Finlande, la Russie et l'Estonie", *Annales d'Economie et de Statistique*, 43, 101-127.

infinite horizon, one shows that, for the first year,

$$J_1^N - J_1^* = \frac{1}{2}\frac{1}{1-\beta[1-\delta]}\sum_{i=1}^{n}\frac{\pi_i^2-\pi_{\mathcal{N}}^2}{\gamma_i} + \sum_{i=1}^{n}\frac{\pi_{\mathcal{N}}[\pi_{\mathcal{N}}-\pi_i]}{\gamma_i}.$$

The first term of the right-hand side is negative and the second positive (because $\pi_i < \pi_{\mathcal{N}}, \forall i$). If $\beta[1-\delta]$ is close enough to 1 (which is reasonable in the case of atmospheric CO_2), then the optimum is dominated by the Nash equilibrium for $t = 1$. On the other hand, if $\delta = 1$ (the case where the pollutant doesn't accumulate), then $J_1^N > J_1^*$.

Hoel, M. [1992], "Emission taxes in a dynamic international game of CO_2 emissions", in R. Pethig (ed.): *Conflicts and cooperation in managing environmental resources*, Microeconomic Studies, Springer Verlag, Berlin.

Jepma, C. (ed.) [1995], *The feasibility of joint implementation*, Environment and Policy series, Kluwer Academic Publishers, Boston.

Kaitala, V., M. Pohjola and D. Tahvonen [1992], "Transboundary air pollution and soil acidification: a dynamic analysis of an acid rain game between Finland and the USSR", *Environmental and Resource Economics*, 2, 161-181.

Mäler K. [1989], "The acid rain game", in H. Folmer and E. van Ierland (eds.), *Valuation Methods and Policy Making in Environmental Economics*, Amsterdam, Elsevier.

Sage, A. and C. White [1977], *Optimal systems control*, Prentice-Hall, New-York, 2e édition.

Tahvonen, O. [1993], "Carbon dioxyde abatement as a differential game", discussion paper in economics n° 4, University of Oulu.

van der Ploeg F. and de Zeeuw A. [1992], "International aspects of pollution control", *Environmental and Resource Economics*, 2, 117-139.

Zaccour, G. [1994], "Side payments in a dynamic game of environmental policy coordination", in M. Breton et G. Zaccour (eds.), preprint volume of the 6th international symposium on dynamic games and applications, St Jovite, Québec, Canada - Juillet 1994.

Annex 1: Computation of the optimal trajectory

Rewrite the problem (2) in standard form (Sage and White, 1977, ch. 6), defining $X_{t+1} = S_t$. This yields:

$$\min_{\{bfE_t\}_{t\in\mathcal{T}}} \sum_{t=1}^{T} \beta^t \sum_{i=1}^{n} [C_i(E_{it}) + D_i([1-\delta]X_t + E_t)] \qquad (A.1)$$

$$\text{s.t.} \left\{ \; X_{t+1} = [1-\delta]X_t + E_t, \;\; X_1 \text{ given}(\text{A.2}) \;\; E_{it} \geq 0, \forall t \in \mathcal{T}, \forall i \in \mathcal{N}. \right.$$

The Hamiltonian of this problem is then given by

$$H_t \;=\; \beta^t [\sum_{i=1}^{n} C_i(E_{it}) + \sum_{i=1}^{n} D_i([1-\delta]X_t + E_t)] + \lambda_{t+1}[[1-\delta]X_t + E_t] \qquad (A.3)$$

where λ_t ($t \in \mathcal{T}$) are the Lagrange multipliers associated with the constraints (A.2). If one restricts ones attention to an interior optimum, the necessary conditions for a minimum may be written as

$$\lambda_t = \frac{\partial H_t}{\partial X_t} = \beta^t [1-\delta] \sum_{i=1}^{n} D_i'([1-\delta]X_t + E_t) + \lambda_{t+1}[1-\delta], \;\; \lambda_{T+1} = 0 \qquad (A.4)$$

$$0 = \frac{\partial H_t}{\partial E_{it}} = \beta^t [C_i'(E_{it}) + \sum_{j=1}^{n} D_j'([1-\delta]X_t + E_t)] + \lambda_{t+1}, \; i \in \mathcal{N} \tag{A.5}$$

together with the constraints (A.2).

Combining (A.4) and (A.5) and observing that $S_t = [1-\delta]X_t + E_t$, we obtain that, for all $i \in \mathcal{T}$,

$$C_i'(E_{it-1}) + \sum_{i=1}^{n} D_i'(S_{t-1}) \;=\; \beta[1-\delta] C_i'(E_{it}), \quad t \in \mathcal{T}. \tag{A.6}$$

The relations (A.4) and (A.5) then imply that, for $t = T$, $C_i'(E_{iT}) = \sum_{i=1}^{n} D_i'(S_T)$. This final condition and (A.6) then give that

$$C_i'(E_{it}) + \sum_{\tau=t}^{T} \beta^{\tau-t} [1-\delta]^{\tau-t} \sum_{i=1}^{n} D_i'(S_\tau) = 0, \quad t \in \mathcal{T} \tag{A.7}$$

which is (3) above.

In the special case where the damage functions are *linear*, that is when

$$D_i(S_t) \;=\; \pi_i S_t, \; i \in \mathcal{N},$$

(A.7) may be rewritten as

$$C_i'(E_{it}) + \pi_{\mathcal{N}} \frac{1 - \beta^{T+1-t}[1-\delta]^{T+1-t}}{1 - \beta[1-\delta]} \;=\; 0, \quad i \in \mathcal{N}, \tag{A.8}$$

where $\pi_{\mathcal{N}} = \sum_{i=1}^{n} \pi_i$ by definition. If T tends to infinity, this last relation yields

$$C_i'(E_{it}) + \frac{\pi_{\mathcal{N}}}{1 - \beta[1-\delta]} \;=\; 0, \quad i \in \mathcal{N} \tag{A.9}$$

since $0 < \beta[1-\delta] < 1$.

Annex 2: Computation of the Nash equilibium.

Applying the same reasoning as in Annex 1, we see that the non-cooperative equilibium is characterized by (A.2) together with the conditions

$$\lambda_{it} \;=\; \beta^t [1-\delta] D_i'([1-\delta]X_t + E_t) + \lambda_{it+1}[1-\delta], \; \lambda_{iT+1} = 0 \tag{A.10}$$

$$0 \;=\; \beta^t [C_i'(E_{it}) + D_i'([1-\delta]X_t + E_t)] + \lambda_{it+1} \tag{A.11}$$

$\forall i \in \mathcal{N}$, $\forall t \in \mathcal{T}$, where we have supposed that the optimum for each country is interior. The λ_{it} are the Lagrange multipliers associated with the problem for country i. Combining the above relations, we obtain equation (7) of the text.

Annex 3: proof of Theorem 3.

The proof's structure is similar to that of Chander and Tulkens (1995), and results from the two following lemmas.

Lemma 1: Let $\{\mathbf{E}_t{}^* : t = 1, \ldots, T\}$ be the trajectory of the internationally optimal solution, $\{\mathbf{E}_t{}^N : t = 1, \ldots, T\}$ of the solutions corresponding to the Nash equilibrium, and $\{\mathbf{E}_t{}^U : t = 1, \ldots, T\}$ the trajectory of the solutions corresponding to a partial Nash equilibrium with respect to the coalition $U \subseteq N$ (solution of problems (18) and (19)). If we assume, as in (4), that the damage functions are linear, then

$$E_{it}^N \;\leq\; E_{it}^U \;\leq\; E_{it}^*,\ \forall i \in \mathcal{N},\ \forall t \in \mathcal{T}. \qquad (A.13)$$

Proof: Applying the same reasoning as in Annex 1, we obtain that the Nash equilibrium with respect to the coalition U is characterized by the conditions

$$C_i'(E_{it}^U) + \pi_U \frac{1 - \beta^{T+1-t}[1-\delta]^{T+1-t}}{1 - \beta[1-\delta]} \;=\; 0,\ i \in U \qquad (A.14)$$

$$C_j'(E_{jt}^U) + \pi_j \frac{1 - \beta^{T+1-t}[1-\delta]^{T+1-t}}{1 - \beta[1-\delta]} \;=\; 0,\ j \in \mathcal{N} \setminus U \qquad (A.15)$$

where, by definition, $\pi_U = \sum_{i \in U} \pi_i$ and where we have assumed that all optima are interior. The conditions generalize (5) and (8) in the cases where $U = N$ (cooperative optimum) and $U = \{i\}$ (Nash equilibrium). As $\forall i \in U,\ \pi_i \leq \pi_U \leq \pi_{\mathcal{N}}$, (A.13) immediately results from the convexity of $C_i,\ (i \in \mathcal{N})$, (A.14) and (A.15). □

Proving Theorem 3 is equivalent to showing that the imputation defined by (22) and (23) satisfies condition (21). We proceed by contradiction and suppose that this latter condition is violated. We may then prove the following lemma.

Lemma 2: Assume that condition (21) fails for coalition U, and consider the imputation $(\hat{W}_1(S_0), \ldots, \hat{W}_n(S_0))$ defined by

$$\hat{W}_i(S_0) \;=\; W_i(S_0) + \hat{\tau}_i(S_0) \qquad (A.16)$$

such that

$$\hat{\tau}_i(S_0) \;=\; -[W_i(S_0) - w_i(U; S_0)] + \frac{\pi_i}{\pi_{\mathcal{N}}}[W(S_0) - \sum_{j=1}^{n} w_j(U; S_0)] \qquad (A.17)$$

where, by definition,

$$w_i(U; S_0) \;=\; \sum_{t=1}^{T} \beta^t [C_i(E_{it}^U) + \pi_i = S_t^U],\ i \in \mathcal{N} \qquad (A.18)$$

is the discounted cost for country i induced by the strategy $\{\mathbf{E}_t{}^U : t = 1, \ldots, T\}$. This imputation then dominates $(\tilde{W}_1, (S_0), \ldots, \tilde{W}_n(S_0))$ (defined by (22) and (23) in the sense that

$$(i) \qquad \sum_{i\in U} \hat{W}_i(S_0) \; < \; \sum_{i\in U} \tilde{W}_i(S_0) \tag{A.19}$$

$$(ii) \qquad \hat{W}_i(S_0) \; \leq \; \tilde{W}_i(S_0), \; \forall i \in \mathcal{N} \setminus U. \tag{A.20}$$

Proof: We first prove *(i)* and observe that (A.16) and (A.17) imply that

$$\hat{W}_i(S_0) \;=\; w_i(U;S_0) + \frac{\pi_i}{\pi_{\mathcal{N}}}[W(S_0) - \sum_{j=1}^{n} w_j(U;S_0)] \leq \; w_i(U;S_0) \tag{A.21}$$

since $W(S_0)$ characterizes the cooperative optimum. We then obtain (A.19) from (A.21) and the assumption that (21) is violated, because

$$\sum_{i\in U} \hat{W}_i(S_0) \leq w(U;S_0) = \sum_{i\in U} w_i(U;S_0) < \sum_{i\in U} \tilde{W}_i(S_0) \qquad (A.22).$$

We next prove *(ii)*, which is equivalent to showing that $\hat{\tau}_i(S_0) \leq \tilde{\tau}_i(S_0), \forall i \in \mathcal{N} \setminus U$. Combining (4), (9), (12) and (23), we deduce that

$$\tilde{\tau}_i(S_0) \;=\; -\sum_t \beta^t[C_i(E^*_{it}) - C_i(E^N_{it})] + \frac{\pi_i}{\pi_{\mathcal{N}}} \sum_{j=1}^{n} \sum_t \beta^t[C_j(E^*_{jt}) - C_j(E^N_{jt})]. \tag{A.23}$$

Proceeding similarly with (4), (9), (10), (A.17) and (A.18), we obtain that

$$\hat{\tau}_i(S_0) \;=\; -\sum_t \beta^t[C_i(E^*_{it}) - C_i(E^U_{it} =)] + \frac{\pi_i}{\pi_{\mathcal{N}}} \sum_{j=1}^{n} \sum_t \beta^t[C_j(E^*_{jt}) - C_j(E^U_{jt})]. \tag{A.24}$$

Showing that $\hat{\tau}_i(S_0) \leq \tau_i(S_0)$ is therefore equivalent to showing that

$$-\sum_t \beta^t[C_i(E^N_{it}) - C_i(E^U_{it})] + \frac{\pi_i}{\pi_{\mathcal{N}}} \sum_{j=1}^{n} \sum_t \beta^t[C_j(E^N_{jt}) - C_j(E^U_{jt})] \leq 0. \tag{A.25}$$

Now (8) and (A.15) imply that $E^N_{it} = E^U_{it}$ if $i \in \mathcal{N} \setminus U$. As a consequence, (A.25) reduces to

$$\frac{\pi_i}{\pi_{\mathcal{N}}} \sum_{j\in U} \sum_t \beta^t[C_j(E^N_{jt}) - C_j(E^U_{jt})] \; \leq \; 0. \tag{A.26}$$

But this inequality clearly holds because of (A.13) and the non-increasing nature of C_j, $(j \in \mathcal{N})$. □

The proof of Theorem 3 now immediately follows from Lemma 2. Indeed, we obtain from (22), (A.16), (A.19) and (A.20) that

$$\sum_{i=1}^{n} \hat{\tau}_i(S_0) \; < \; \sum_{i=1}^{n} \tilde{\tau}_i(S_0). \tag{A.27}$$

But this inequality is impossible since (23) and (A.17) imply that the transfers specified by the vectors $(\tilde{\tau}_1(S_0), \ldots, \tilde{\tau}_n(S_0))$ and $(\hat{\tau}_1(S_0), \ldots, \hat{\tau}_n(S_0))$ are balanced, i.e. they satisfy (15). Hence condition (21) must hold and the theorem is proved.