CORE DISCUSSION PAPER
9832

TRANSFERS TO SUSTAIN CORE-THEORETIC COOPERATION IN INTERNATIONAL STOCK POLLUTANT CONTROL

Marc Germain[1] Philippe Toint[2] Henry Tulkens[3]
and Aart de Zeeuw[4]

May 1998

[1]CORE, Université catholique de Louvain. E-mail: germain@core.ucl.ac.be . Germain thanks the Fonds de Développement Scientifique of the UCL for its support.

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

[3]CORE, Université Catholique de Louvain, and Facultés Universitaires Saint-Louis, Bruxelles. Email: tulkens@core.ucl.ac.be

[4]Département of Economics and CentER, Tilburg University. Email: A.J.deZeeuw@kub.nl

This paper (CORE Discussion Paper no9832) was initiated as part of the research project "Caractère adéquat des engagements et applications conjointes: deux modalités-clef de la mise en oeuvre de la Convention de Rio sur les changements climatiques" (project 7-29039) financed by the Fonds de Développement Scientifique of the UCL. It was completed as part of the subsequent research project "Changements climatiques, Négociations internationales et Stratégies de la Belgique" (CLIMNEG), financed at UCL by the Belgian Government (Services du Premier Ministre, Services fédéraux des Affaires scientifiques, techniques et culturelles (SSTC). The fourth author's contribution was made possible thanks to the Human Capital Mobility programme "Designing Economic Policy for the Management of Natural Resources and the Environment" (CHRX CT93) of the European Union. The authors gratefully acknowledge from comments from Claude d'Aspremont, Jean-Pascal van Ypersele, Pierre-André Jouvet, Philippe Michel and Yurii Nesterov.

Abstract

International environmental agreements aiming at correcting negative externalities generated by transboundary pollution are difficult to achie- ve for many reasons. Important obstacles arise from asymmetry in costs and benefits, and instability may occur due to the fact that coalitions of countries may attempt to do better for themselves outside of any proposed agreement. In a static context it has already been shown that it is possible to achieve stability in the sense of the core of a cooperative game, by means of appropriately defined transfers between the countries involved. However, the transboundary pollution problems that are most important are caused by accumulated pollutants so that a dynamic analysis is required. This paper provides a transfer scheme that yields a core property in a dynamic context. The possibility of computing such transfers numerically is discussed.

Keywords: transfrontier pollution; stock pollutant; dynamic cooperative games; coalitions; core solution.

JEL Classification: C73, D62, F42, H4, Q3

Contents

1 Introduction

This paper presents a cooperative game theoretic analysis of the economics of international agreements on transnational pollution control, when the environmental damage arises from stock pollutants that accumulate (and possibly decay). The issues raised by the necessity of cooperation amongst the countries involved, if a social optimum is to be achieved, have already been addressed in the literature in terms of game theory concepts; 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. With stock pollutants the problem acquires an intertemporal and therefore dynamic dimension. In this case, differential game theory is a more appropriate tool for the analysis, as is done in e.g. van der Ploeg and de Zeeuw (1992), Kaitala , Pohjola and Tahvonen (1992), Hoel (1992), Tahvonen (1993), Petrosjan and Zaccour (1995).

With exception of the last paper, these contributions leave aside the issue of the *voluntary* implementation of the international optimum. This is an important drawback since no supranational authority can be called upon to impose the optimum in a context where countries' interest in cooperation diverges strongly between one another, and especially if some countries loose when the social optimum is implemented. In view of ensuring such implementation, it has often been suggested that financial transfers between the countries involved would provide incentives towards cooperation[1]. This property, understood in the sense of the core of a cooperative game, has in effect been demonstrated by Chander and Tulkens (1995, 1997), who propose a particular transfer scheme based on parameters reflecting the relative intensities of the countries' environmental preferences.

This result, established for flow pollutants only (that is, in a static game model), has been extended by Germain, Toint and Tulkens (1998) to the larger context of open-loop differential games. Despite its interest, this extension suffers from the fact that it rests on the restrictive assumption that negociations take place once and for all. Agreements are binding until the end of the planning horizon. The aim of the present paper is precisely to relax this

[1] In a 2-countries framework, Petrosjan and Zaccour (1995) use financial transfers to obtain a cooperative solution as a Nash equilibrium. However, being limited to 2 countries, they cannot deal with the issue of coalitions, which is one of our aims here. Kverndokk (1994) deals both with coalitions and financial transfers for the greenhouse problem, but he does neither resort to differential games, nor to cooperative ones.

assumption : *cooperation is renegotiated at each period*, i.e. players reevaluate at each time the interest of cooperation taking account of the current stock of pollutant.

In this context, financial transfers that induce cooperation in the core-theoretic sense are considered again. All countries can sign an international agreement designed on that basis because it would be stable in the sense that incentives for the formation of objecting coalitions are eliminated.

Other approaches concentrate on the stable number of signatories by considering the trade-off for an individual country between joining some coalition and staying out of that coalition (see Carraro and Siniscalco (1993) and Barrett, (1994)). A critical comparison of the two approaches is given in Tulkens (1995).

The structure of the paper is as follows. Section 2 presents the stock pollutant model and then characterizes the international optimum as well as the feedback Nash non-cooperative equilibrium. In section 3, financial transfers are formulated so that each country is not worse off when it participates, in a finite horizon framework; this result is also extended to the infinite horizon case. In section 4, transfers are further specified so as to achieve stability in the core theoretic sense. Section 5 discusses the possibility of computing numerical solutions and section 6 is a conclusion.

2 International optimality and non-cooperative equilibrium

Our economic model is written in discrete time. Consider n countries indexed by $i \in \mathcal{N} = \{1, 2, ..., n\}$ and some planning period $\mathcal{T} = \{1, 2, ..., T\}$ (T a positive integer, possibly infinite). In each country, pollution is entailed by economic activity : let $\mathbf{E_t} = (E_{1t}, ..., E_{nt})'$ denote the vector of the different countries' emissions of a certain pollutant at time t. These emissions spread uniformly in the atmosphere and contribute to a stock of pollutant S according to the equation

$$S_t = [1 - \delta]S_{t-1} + \sum_{i=1}^{n} E_{it} \tag{1.1}$$

where the initial stock of pollutant S_0 is given and where δ is the pollutant's natural rate of degradation ($0 < \delta < 1$). This model describes, for example,

the basics of the climate change problem where all emissions of greenhouse gases add to the stock of greenhouse gases, which only gradually assimilates and which is the cause of the climate change.

This stock of pollutant causes damages to each country's environment. For country i ($i \in \mathcal{N}$), these damages during period t are given in monetary terms by $D_i(S_t)$, where D_i is supposed to be a differentiable, increasing and convex function of the current stock S_t ($D_i' > 0, D_i'' \geq 0$). ,As described in (1.1), this current stock is a function of the inherited stock S_{t-1} and of the current emissions $\mathbf{E}_t$. The only way to control the stock of pollutant is through the control of emissions[2]. More precisely, each country i can reduce its own emissions, and the cost of doing this is described by a differentiable, decreasing and stricly convex function $C_i(E_i)$ ($C_i' < 0, C_i'' > 0$). $C_i(E_i)$ measures the total costs incurred by country i from limiting its emissions to E_i. The convexity of the function reflects the intuitive idea that the marginal cost of reducing emissions is higher for lower levels of emissions.

From equation (1.1), it is clear that the damages to the environment of country i will depend on the emissions of all countries. In what follows, we will consider two different modes of behaviour of the countries. A first one assumes that they behave in a cooperative way, i.e. that each of them takes account of the impact of its pollution on itself as well as on all other countries. In this case, countries *jointly* choose at each period their emission levels in order to minimize total discounted costs, i.e. $\forall t \in \mathcal{T}$, they solve the following problem :

$$\min_{\{\mathbf{E}_s\}_{s\in\{t,\dots,T\}}} \left\{ \sum_{s=t}^{T} \sum_{i=1}^{n} \beta^s [C_i(E_{is}) + D_i(S_s)] \right\}$$

subject to the constraints (1.1) and $E_{it} \geq 0$ ($\forall t \in \mathcal{T}$, $\forall i \in \mathcal{N}$). β is the discount factor ($0 < \beta \leq 1$). According to Bellman's principle of optimality, the solution can be found by solving the dynamic programming equations

$$W(T, S_{T-1}) = \min_{\mathbf{E}_T} \left\{ \sum_{i=1}^{n} [C_i(E_{iT}) + D_i(S_T)] \right\} \tag{1.2}$$

[2]I.e. reducing pollution is done through the reduction of emissions, and not through the cleaning of the environment.

$$W(t, S_{t-1}) = \min_{\mathbf{E}_t} \left\{ \sum_{i=1}^{n} [C_i(E_{it}) + D_i(S_t)] + \beta W(t+1, S_t) \right\}, \; t = 1, 2, ..., T-1 \tag{1.3}$$

subject to the constraints (1.1) and $E_{it} \geq 0$ ($\forall t \in \mathcal{T}, \; \forall i \in \mathcal{N}$). In (1.2) and (1.3), W is called the *value function*. As the total costs of all countries are minimized, the resulting trajectories of emissions and stock constitute the *international optimum*. The convexity of the functions C_i and D_i ($\forall i \in \mathcal{N}$) suffices to guarantee that the minimum exists and is unique.

In an alternative mode of behaviour, one may assume that countries behave non-cooperatively in the sense of a Nash equilibrium, where each of them minimizes at each period only its own discounted costs given the emissions of the other countries. I.e. $\forall t \in \mathcal{T}$, country i ($i \in \mathcal{N}$) solves the following problem :

$$\min_{\{E_{is}\}_{s \in \{t,...,T\}}} \left\{ \sum_{s=t}^{T} \beta^s [C_i(E_{is}) + D_i(S_s)] \right\}, \; t \in \mathcal{T}$$

subject to the constraints (1.1), $E_{it} \geq 0$ ($\forall t \in \mathcal{T}, \; \forall i \in \mathcal{N}$), and E_{jt} $j \neq i$ given. Within the framework of dynamic programming this leads to the value functions

$$N_i(T, S_{T-1}) = \min_{E_{iT}} \; \{[C_i(E_{iT}) + D_i(S_T)])\}, \; i \in \mathcal{N} \tag{1.4}$$

$$N_i(t, S_{t-1}) = \min_{E_{it}} \; \{[C_i(E_{it}) + D_i(S_t)] + \beta N_i(t+1, S_t)\}, \; i \in \mathcal{N}, \; t = 1, 2, ..., T-1 \tag{1.5}$$

under the constraints (1.1) and $E_{it} \geq 0$ ($\forall t \in \mathcal{T}, \; \forall i \in \mathcal{N}$). The convexity of the functions C_i and D_i suffices to guarantee that the Nash equilibrium exists and is unique. A trajectory so derived is called a *non-cooperative feedback Nash equilibrium* (Başar and Olsder, 1998)[3]. The value functions N_i characterize the costs-to-go at this equilibrium for each country i. The emissions are functions of the current stock of pollutant and the equilibrium has the property of *Markov perfectness* (Maskin and Tirole, 1988), in the sense that it remains an equilibrium if the game is restarted at any intermediate year t

[3]See Mäler and de Zeeuw (1998) and van der Ploeg and de Zeeuw (1992) for applications of such an equilibrium to acid rains and climate change respectively.

from any value of the stock of pollutant.

At the optimum, and contrary to what happens at the Nash equilibrium, each country takes account of the impact of its pollution on the environment of all other countries. Therefore, from a collective point of view, the international optimum is better than the feedback Nash equilibrium. Nothing ensures, however, that this is also true at the individual level. Indeed, countries being different, it is possible that some country at some time t is better off at the non-cooperative equilibrium than at the optimum, so that cooperation is not profitable for this country (at least at time t). The same can occur for subsets of countries - i.e. coalitions - in the sense that, by limiting cooperation to such coalitions, the members of the latter could be better off than at the international optimum. The aim of this paper is to show that financial transfers between countries can make each one of them interested in cooperating at all periods t (individual rationality), and that in addition no sub-group of countries has never an incentive to form a coalition (coalition rationality).

To make the above argument precise, it is important to state explicitly what the fall-back position is for each country when no transfers occurs. At each time t one could take the non-cooperative feedback Nash equilibrium from t onwards as such point of reference, and determine the transfers accordingly. However, one should not neglect the fact that countries know that later on, thanks to the cooperative transfers to which they will have access, they can be better off than at the non-cooperative Nash equilibrium. Hence, a more rationally expected point of reference at time t is: *non-cooperation at time t, followed by cooperation afterwards.*This idea was already put forward in Houba and de Zeeuw (1995), and we shall make use of it in the framework of dynamic programming. In this spirit, we deal first with individual rationality and then with coalitional rationality.

3 Transfers to sustain individual rationality at the international optimum

3.1 The transfers at the final time T

Let us start by determining which transfers yield gains for all countries when they cooperate in the last period, for any level of the stock of pollutant S inherited from the past. In the non-cooperative equilibrium the countries are supposed to solve problem (1.4). Country i's total cost is then

$$N_i(T,S) = C_i(E_{iT}^N) + D_i(S_T^N),\ i \in \mathcal{N} \tag{2.1}$$

where E_{iT}^N denotes the emission equilibrium level and S_T^N denotes the resulting stock of pollutant given by

$$S_T^N = [1-\delta]S + \sum_{i=1}^{n} E_{iT}^N. \tag{2.2}$$

If countries cooperate, they jointly solve problem (1.2). Country i's total cost is

$$W_i(T,S) = C_i(E_{iT}^*) + D_i(S_T^*) \tag{2.3}$$

where E_{iT}^* is the optimal emission level and S_T^* is the optimal stock of pollutant, given by

$$S_T^* = [1-\delta]S + \sum_{i=1}^{n} E_{it}^*. \tag{2.4}$$

By definition of the optimum, one verifies that

$$W(T,S) \stackrel{\triangle}{=} \sum_{i=1}^{n} W_i(T,S) \leq \sum_{i=1}^{n} N_i(T,S) \stackrel{\triangle}{=} N(T,S). \tag{2.5}$$

The difference between the two sides of this inequality measures the *ecological surplus* resulting from international cooperation.

If $\forall i \in \mathcal{N}$ one has that $W_i(T,S) \leq N_i(T,S)$ then international cooperation is *individually rational*, in the sense that each country has an interest to participate. But if $\exists i \in \mathcal{N}$ such that $W_i(T,S) > N_i(T,S)$, then country i will not cooperate without financial compensation. Since dynamic programming reduces the choice of emissions to one period at the time, one can use the transfers formula proposed by Chander and Tulkens (1997) in a static framework. Let

$$\theta_i(T,S) = -[W_i(T,S) - N_i(T,S)] + \mu_{i,T}[W(T,S) - N(T,S)] \tag{2.6}$$

be the transfer (< 0 if received, > 0 if paid) to country i at time T, where $\mu_{i,T} \in]0,1[,\ \forall i \in \mathcal{N}$ and $\sum_{i=1}^{n} \mu_{i,T} = 1$. Then country i's total cost *including* transfers becomes

$$\tilde{W}_i(T,S) = W_i(T,S) + \theta_i(T,S). \tag{2.7}$$

By construction, the budget of the transfers defined by (2.6) is balanced (i.e. $\sum_{i=1}^{n} \theta_i(T,S) = 0$). Since

$$\tilde{W}_i(T,S) - N_i(T,S) = \mu_{i,T}[W(T,S) - N(T,S)] \leq 0, \ \forall i \in \mathcal{N} \tag{2.8}$$

cooperation with transfers is individually rational at time T, whatever the inherited stock of pollutant S [4].

3.2 The transfers at time $T-1$

Countries know that, whatever they do at $T-1$, financial transfers exist (defined by (2.6)) that make the international optimum at T preferable for each of them with respect to the non-cooperative equilibrium. Let us assume that these transfers induce cooperation[5], and that countries therefore expect, at $T-1$, that they will cooperate in period T. The problem to be considered next is the one of constructing transfers that make them interested to cooperate at $T-1$ as well.

In the absence of cooperation at $T-1$, each country i minimizes its own discounted total costs over the two periods $T-1$ and T, expecting cooperation and transfers in T. Thus, given the emissions of the other countries, country i solves problem (1.5) for $t = T-1$ with N_i under the *min* operator replaced by $\tilde{W}_i$. This leads to

$$V_i(T-1,S) = \min_{E_{i,T-1}} \left\{ [C_i(E_{i,T-1}) + D_i(S_{T-1})] + \beta \tilde{W}_i(T, S_{T-1}) \right\}, \ i \in \mathcal{N} \tag{2.9}$$

under the constraints (1.1) and $E_{i,T-1} \geq 0, \ \forall i \in \mathcal{N}$. This yields an equilibrium characterized at time $T-1$ by emission levels $\mathbf{E}^V_{T-1}$ as functions of the initial stock S at time $T-1$. The value function

$$V_i(T-1,S) = C_i(E^V_{i,T-1}) + D_i(S^V_{T-1}) + \beta \tilde{W}_i(T, S^V_{T-1}), \ i \in \mathcal{N} \tag{2.10}$$

[4] The fact that $\mu_{i,T}$ cannot be equal to 0 ensures that country i will benefit from cooperation if $W(T,S) < N(T,S)$. The fact that $\mu_{i,T}$ cannot be equal to 1 excludes that country i monopolizes all the gains of cooperation.

[5] Note that, following Chander and Tulkens (1997, section 5), one could indeed obtain the cooperative optimum with transfers as an equilibrium, called *ratio-equilibrium*.

where

$$S^V_{T-1} = [1-\delta]S + \sum_{i=1}^{n} E^V_{i,T-1} \tag{2.11}$$

denotes country i's discounted equilibrium costs. We will call this equilibrium the *fallback non-cooperative equilibrium* at time $T-1$.

In the case where all countries cooperate, they solve problem (1.3) for $t = T-1$. Optimal levels of emissions and of the resulting stock of pollutant are denoted by $\mathbf{E}^*_{T-1}$ and S^*_{T-1} respectively. Both are functions of S. This yields

$$W_i(T-1,S) = C_i(E^*_{i,T-1}) + D_i(S^*_{T-1}) + \beta \tilde{W}_i(T, S^*_{T-1}) \tag{2.12}$$

which is country i's part in the optimal total discounted costs, *taking into account the transfers expected in* T.

As in period T (see (2.5)), one verifies that

$$W(T-1,S) \stackrel{\triangle}{=} \sum_{i=1}^{n} W_i(T-1,S) \leq \sum_{i=1}^{n} V_i(T-1,S) \stackrel{\triangle}{=} V(T-1,S). \tag{2.13}$$

$V(T-1,S) - W(T-1,S)$ measures the ecological surplus induced by extending cooperation to period $T-1$, with respect to the alternative scenario where cooperation is limited to T. If $\forall i \in \mathcal{N}$ one observes that $W_i(T-1,S) \leq V_i(T-1,S)$, then international cooperation is individually rational, in the sense that each country has an interest to participate. But if $\exists i \in \mathcal{N}$ such that $W_i(T-1,S) > V_i(T-1,S)$, then country i will not want to extend cooperation to period $T-1$ without financial compensation.

To induce country i to participate in $T-1$, we proceed as in period T. Let

$$\theta_i(T-1,S) = -[W_i(T-1,S) - V_i(T-1,S)] + \mu_{i,T-1}[W(T-1,S) - V(T-1,S)] \tag{2.14}$$

be the transfer paid or received by country i at time $T-1$, where $\mu_{i,T-1} \in]0,1[$, $\forall i \in \mathcal{N}$ and $\sum_{i=1}^{n} \mu_{i,T-1} = 1$. Then country i's total cost *including* transfers becomes

$$\tilde{W}_i(T-1,S) = W_i(T-1,S) + \theta_i(T-1,S). \tag{2.15}$$

By construction, the budget of the transfers defined by (2.14) is balanced (i.e. $\sum_{i=1}^{n} \theta_i(T-1,S) = 0$). Furthermore, since

$$\tilde{W}_i(T-1,S) - V_i(T-1,S) = \mu_{i,T-1}[W(T-1,S) - V(T-1,S)] \leq 0, \ \forall i \in \mathcal{N} \tag{2.16}$$

cooperation with transfers is individually rational at time $T-1$, whatever the inherited stock of pollutant S.

3.3 The transfers at earlier periods

The analysis of subsection 2.2 can be repeated for all earlier periods. Assume that the fallback non-cooperative equilibrium exists and is unique at $t+1$, $t+2$, ..., T [6]. At period t, countries face the alternative of whether or not to cooperate, knowing that there exists tranfers that induce cooperation from time $t+1$ onwards. The final result will be that the countries cooperate in each period. This determines the emission levels in each period and also the trajectory of the stock of pollutant, given its initial value S_0. In turn this trajectory determines the values of the functions V_i, W_i and $\tilde{W}_i$, and therefore also the values of the transfers θ_i. In section 3 it will be shown that, under certain assumptions on the functions C_i and D_i, and with specific values of μ_i, these transfers yield an interesting property from the perspective of cooperative game theory.

3.4 The infinite horizon case

In the infinite horizon case, the backward reasoning considered in the preceding subsections applies no more. However, we can consider the stationary solution by taking advantage of the fact that the cost functions C_i and D_i as well as the sharing parameters μ_i do not depend directly on time. The functional forms of the solutions thus only vary in time through the varying stock of pollutant S. The structure of the problem is therefore very similar to the finite horizon case.

[6] Unlike what happens for the non-cooperative feedback Nash equilibrium, convexity of the functions C_i and D_i does not guarantee that the fallback non-cooperative equilibrium exists and is unique. Indeed, it is not sure that the objective of problem (2.9) is convex, because of the presence of $\tilde{W}_i$, which contains the financial transfers. However, one can verify that convexity prevails when the damage functions D_i are linear (see Germain, Tulkens and de Zeeuw, 1998), or when the functions C_i and D_i are quadratic with some restrictions on the parameters.

When the countries cooperate, one has to solve the following stationary dynamic programming equations

$$W(S) = \min_{\mathbf{E}} \{\sum_{i=1}^{n} [C_i(E_i) + D_i(\bar{S})] + \beta W(\bar{S})\} \tag{2.17}$$

under the constraint

$$\bar{S} = [1-\delta]S + \sum_{i=1}^{n} E_i, \tag{2.18}$$

with $E_i \geq 0,\ \forall i \in \mathcal{N}$, where S is the stock inherited from the past. Then

$$W(S) = \sum_{i=1}^{n} [C_i(E_i^*) + D_i(S^*)] + \beta W(S^*) \tag{2.19}$$

denotes the total cost of all countries, where $\mathbf{E}^*$ and S^* are optimal emission and stock levels respectively, with

$$S^* = [1-\delta]S + \sum_{i=1}^{n} E_i^*. \tag{2.20}$$

The same reasoning can be applied for the countries' individual functions V_i and $\tilde{W}_i$. In an infinite horizon framework, by similarity with (2.6), (2.7) and (2.9), country i's total cost characterizing the fallback non-cooperative equilibrium (non-cooperation today, cooperation in the future) will be

$$V_i(S) = \min_{E_i} \{C_i(E_i) + D_i(\bar{S}) + \beta[V_i(\bar{S}) + \mu_i[W(\bar{S}) - V(\bar{S})]]\} \tag{2.21}$$

under the constraints (2.18), $E_i \geq 0,\ S$ and $E_j\ (\forall j \neq i)$ given. At the fallback non-cooperative equilibrium, total costs of country i will be

$$V_i(S) = C_i(E_i^V) + D_i(S^V) + \beta[V_i(S^V) + \mu_i[W(S^V) - V(S^V)]],\ i \in \mathcal{N} \tag{2.22}$$

where $\mathbf{E}^V$ and S^V are the emissions and stock equilibrium levels with

$$S^V = [1-\delta]S + \sum_{i=1}^{n} E_i^V. \tag{2.23}$$

If $\forall i \in \mathcal{N}$ one has that

$$W_i(S) = C_i(E_i^*) + D_i(S^*) + \beta \tilde{W}_i(S^*) \leq V_i(S) \tag{2.24}$$

i.e. if the cost induced to country i by the optimal strategy is less than the one it would incur in absence of cooperation, then this optimal strategy is individually rational in the sense that each country has an interest to participate. If the contrary happens, one may apply a reasoning similar to the one of the preceding subsection and propose the following structure of transfers :

$$\theta_i(S) = -[W_i(S) - V_i(S)] + \mu_i[W(S) - V(S)],\ i \in \mathcal{N} \tag{2.25}$$

where by definition, $V(S) = \sum_i V_i(S)$ and where $\mu_i \in]0,1[$, $\forall\ i \in \mathcal{N}$, and $\sum_i \mu_i = 1$. By construction, these transfers are balanced ($\sum_i \theta_i(S) = 0$). Furthermore, if country i receives $\theta_i(S)$ in case of cooperation it follows that

$$\tilde{W}_i(S) = W_i(S) + \theta_i(S) = V_i(S) + \mu_i[W(S) - V(S)] \leq V_i(S),\ \forall i \in \mathcal{N} \tag{2.26}$$

so that cooperation is individually rational whatever the inherited stock of pollutant S.

4 Transfers to sustain coalitional rationality at the international optimum

The only condition so far on the μ_i's appearing in (2.25) is to take values between 0 and 1 and to sum up to 1 [7]. The aim of this section is to utilize the degrees of freedom left to obtain the property of "coalitional rationality" suggested by the core concept in cooperative game theory. To do so, we adapt to the present intertemporal context the approach proposed by Chander and Tulkens (1995,1997) in a static framework.

4.1 The cooperative game associated with the economic model

Usually, a cooperative game (with transferable utility) is defined by the pair $[\mathcal{N}, w]$, where $\mathcal{N} = \{1, \cdots, n\}$ is the set of players (i.e. the n countries), and w is the characteristic function[8]. Such a game is associated with our economic

[7] This is of course also true for $\mu_{i,T}$ and $\mu_{i,T-1}$ appearing in (2.6) and (2.14) respectively. In the following, we will limit the discussion of coalitional rationality to the infinite horizon case. The discussion for the finite horizon case would be very similar.

[8] See e.g. Osborne and Rubinstein (1994), p.357. These authors use the more recent - and actually more appropriate - alternative terminology of "coalitional game".

model at each time t by specifying the characteristic function as $w(., S)$, where S is the inherited stock of pollutant at that time. This function is defined on the space of strategies of the players, which for each country i is the interval of possible emission levels E_i at time t, i.e. $[0, \infty[$, and for each coalition $U \subseteq \mathcal{N}$ the product of $|U|$ of these intervals, where $|U|$ denotes the cardinality of U.

In order to take into account the externality generated, through S, on all players by each individual emission level E_i, the characteristic function needs to be specified in a particular way, that we formulate as follows using a concept proposed by Chander and Tulkens (1995,1997) for static games. When some coalition $U \subseteq \mathcal{N}$ forms, let the vector $\mathbf{E}^U$ of the strategies adopted by all players be such that :

(i) the emissions of the coalition members be described by the vector $\{E_i^U : i \in U\}$ which is the solution of the optimization problem

$$\min_{\{E_i\}_{i \in U}} \sum_{i \in U} [C_i(E_i) + D_i(\bar{S}) + \beta \tilde{W}_i(\bar{S})] \text{ s.t. } (2.18) \tag{3.1}$$

where, $\forall\ j \in \mathcal{N} \setminus U$, $E_j = E_j^U$ as defined by (ii) below[9];

(ii) the emissions $E_j^U, j \in \mathcal{N} \setminus U$, of the countries out of the coalition, be the solutions that simultaneously solve the optimization problems

$$\min_{E_j}\ C_j(E_j) + D_j(\bar{S}) + \beta \tilde{W}_j(\bar{S}), j \in \mathcal{N} \setminus U, \text{ s.t. } (2.18) \tag{3.2}$$

where, $\forall\ i \in U$, $E_i = E_i^U$ as defined by (i) above[10].

Thus the idea is that, if a coalition forms, its members together minimize the sum of their discounted total costs and each country outside of it reacts by minimizing its own individual discounted total cost. In fact, the vector $\mathbf{E}^U$ is of the nature of a Nash equibrium of a non-cooperative game where the players are the subset U and the elements of its complement $\mathcal{N} \setminus U$. The equilibrium so defined is called by Chander and Tulkens a *partial agreement Nash equilibrium with respect to the coalition* U. The authors show that such an equilibrium exists and is unique under assumptions similar to those of our economic model.

[9] Indeed, the variables $E_j,\ j \in \mathcal{N} \setminus U$, do enter (3.1) through the stock variable $\bar{S}$ (recall (2.18)).

[10] Again, the E_i's, $i \in U$ enter (3.2) through $\bar{S}$ and (2.18).

On this basis, we write the characteristic function[11] as

$$w^{\gamma}(U;S) = \sum_{i \in U} [C_i(E_i^U) + D_i(S^U) + \beta \tilde{W}_i(S^U)] \tag{3.3}$$

where $S^U = [1-\delta]S + \sum_{i=1}^{n} E_i^U$. Note that $w^{\gamma}(\mathcal{N};S)$ is equal to $W(S)$ defined by (2.19), i.e. to the optimal total cost for all countries together.

For the game $[\mathcal{N}, w^{\gamma}(\cdot;S)]$ so defined, any n-dimensional vector whose components sum up to $w^{\gamma}(\mathcal{N};S)$ is called an *imputation* of the game. An imputation can thus be seen as a way of sharing of the optimal total cost between the players.

The vector $(W_1(S), \cdots, W_n(S))$, where $W_i(S)$ is defined as in the left hand side of the inequality in (2.24), is an example of such an imputation, where each country bears its own abatement and damage costs as induced by the optimal strategy $\{\mathbf{E}^*\}$. However, the possibility of financial transfers between the countries implies that (an infinite number of) other imputations exist, associated with the same strategy. Indeed, all vectors $(\tilde{W}_1(S), \cdots, \tilde{W}_n(S))$ defined as in the left hand side of the inequality in (2.26) and such that $\sum_i \theta_i(S) = 0$ are imputations also.

A *solution* of the game is an imputation that satisfies certain properties; and among the imputations defined by (2.26), the ones that satisfy

$$\sum_{i \in U} \tilde{W}_i(S) \leq w^{\gamma}(U;S), \quad \forall\, U \subseteq \mathcal{N} \tag{3.4}$$

are said *to belong to the core* of the game. In words, the core is the set of imputations with the property that each possible coalition bears a fraction of the total costs $W(S)$ less than or equal to $w^{\gamma}(U;S)$, i.e. the least cost that this coalition can guarantee for itself. Imputations belonging to the core of the game defined above are therefore called "rational in the sense of coalitions" since the members of any coalition would suffer a total cost greater than the one at the optimum with transfers.

[11] Called "γ-characteristic function" by Chander and Tulkens (1995, 1997), because of item (ii) in the definition. Other specifications (namely α- and β-characteristic functions) are offered in the literature; but they are shown by these authors to be inadequate for environmental externalities.

4.2 Transfers inducing a γ-core imputation

To summarize, recall from section 2.4 that $\mathbf{E}^*$ was defined as the vector of optimal emission levels and S^* as the optimal stock level; that $\mathbf{E}^V$ was the vector of emission levels and S^V the stock characterizing the fallback non-cooperative equilibrium of section (2.4); and that $(W_1(S), \cdots, W_n(S))$ and $(V_1(S), \cdots, V_n(S))$ are, respectively, the vectors of discounted total costs at the optimum and at the fallback non-cooperative equilibrium (cfr. (2.24) and (2.22)), with $W(S)$ and $V(S)$ the sum of these vectors' components.

For the game $[\mathcal{N}, w^\gamma(\cdot; S)]$, consider now the imputation $(\tilde{W}_1(S), \cdots, \tilde{W}_n(S))$ defined by

$$\tilde{W}_i(S) = W_i(S) + \tilde{\theta}_i(S), \ \forall i \in \mathcal{N} \tag{3.5}$$

where $\tilde{\theta}_i(S)$ is a transfer of the form

$$\tilde{\theta}_i(S) = -[C_i(E_i^*) - C_i(E_i^V)] + \tilde{\mu}_i(S^*) \sum_{j=1}^{n} [C_j(E_j^*) - C_j(E_j^V)], \tag{3.6}$$

with S^* depending on S through (2.20) and where

$$\tilde{\mu}_i(S) = \frac{F_i'(S)}{\sum_{j=1}^n F_j'(S)} \tag{3.7}$$

with

$$F_i(S) = D_i(S) + \beta \tilde{W}_i(S). \tag{3.8}$$

We can now state the following

THEOREM: The imputation (3.5) belongs to the core of the game $[\mathcal{N}, w^\gamma(\cdot; S)]$, if one of the following two conditions is satisfied :
(i) $\forall i \in \mathcal{N}, \ D_i$ is linear or
(ii) $\forall i \in \mathcal{N}, \ \tilde{W}_i$ is monotonically increasing and convex and, $\forall U \subseteq \mathcal{N}$,

$$\sum_{i \in U} [C_i(E_i^U) - C_i(E_i^V)] \geq 0 \tag{3.9}$$

where $\mathbf{E}^U$ is the vector of emission levels at the partial agreement Nash equilibrium with respect to coalition U.

Proof: Case (i) - linear damage costs - is proven in Germain, Tulkens and de Zeeuw (1998). The proof of case (ii) is given in the appendix.

4.3 Interpretation

The transfers (3.6) used in the theorem are the sum of two components. The first term in the right-hand side of (3.6) is either (if > 0) an amount received by country i equal to the increase of its abatement costs due to its cooperating behaviour during this period, or (if < 0) a payment made by country i equal to its savings in abatement costs, in the case where cooperation allows for an increase in this country's emissions. The second term (always < 0) is country i's contribution to cover the total cost of the aggregate abatement effort of the cooperating countries.

Notice that the transfers (3.6) are formulated differently than in (2.25). The first part of the proof in the appendix shows that the former with (3.7) and (3.8) are a special case of the latter.

Further remarks concern condition (ii). Notice that it is a sufficient, not a necessary one. It requires, with (3.9), that if a coalition forms, the total of the abatement costs of its members be larger than what this total would be at the fallback non-cooperative equilibrium. Thus, coalitions for which (3.9) does not hold might credibly pretend to do better for their members than what the imputation (3.5) yields them. How likely are such coalitions?

In order for a coalition U to form - irrespective to its power to oppose (3.5) - it should at least satisfy

$$\sum_{i\in U}[C_i(E_i^U) - C_i(E_i^V)] + \sum_{i\in U}[F_i(S^U) - F_i(S^V)] \leq 0 \tag{3.10}$$

since otherwise its members are all better off at the fallback non-cooperative equilibrium. In (3.10) the magnitudes in the bracketed differences under the second summation sign are always negative, because S^U is smaller than S^V. Thus, a coalition U that forms can only oppose (3.5) if in addition the first sum in (3.10) is negative, i.e. if the aggregate abatement costs of its members is

lower than what it is at the fallback non-cooperative equilibrium[12]. However, one may think that the existence of such coalitions is unlikely because, as is shown in the appendix (see the corollary after Lemma 1), the sum of the emissions of the members of any coalition is always lower than what is at the fallback non-cooperative equilibrium.

5 Computability issues

5.1 Computability with constant sharing parameters and quadratic cost functions

If the cost functions C_i and D_i are quadratic in E_i and S, respectively, and if the sharing parameters μ_i are constant in S, the analysis of section 2 can be further elaborated, leading to simple algorithms to calculate the value functions.

First consider the finite horizon case. At the final time T both optimal and equilibrium emission levels are linear functions of the inherited stock of pollutant S, so that the functions N_i and W and the value functions $\tilde{W}_i$ are quadratic functions of S at T. This implies that the optimal and equilibrium emission levels at time $T-1$ are also linear in S, so that again the functions V_i and W_i and the value functions $\tilde{W}_i$ at $T-1$ are quadratic functions of S. Backwards induction implies that this hold for all periods.

Formally, suppose that for each $i \in \mathcal{N}$ the cost functions are given by

$$C_i(E_i) = \frac{\gamma_i}{2}[E_i - \bar{E}_i]^2,\ 0 \leq E_i \leq \bar{E}_i \tag{4.1}$$

and

$$D_i(S) = \frac{\pi_i}{2}S^2,\ 0 \leq S \tag{4.2}$$

where γ_i, π_i and $\bar{E}_i$ are strictly positive parameters. At the final time T, given the inherited stock of pollutant S and assuming that the constraints on the E_i's and on S are not binding, the Nash equilibrium has to satisfy the conditions

[12] In a static framework, Chander and Tulkens (1997) propose a condition on the marginal damages that ensures that all members of a coalition decrease their emissions w.r.t. the non-cooperative equilibrium. Adapted to our framework, this condition is that $\forall U \subset \mathcal{N},\ U \neq \mathcal{N},\ |U| \geq 2,\ \sum_{i \in U} F_i'(S^*) \geq F_j'(S^V),\ j \in U$. This condition guarantees (3.9).

$$\gamma_i[E_i^N - \bar{E}_i] + \pi_i \left[[1-\delta]S + \sum_{j=1}^{n} E_j^N \right] = 0, i \in \mathcal{N} \tag{4.3}$$

which yields the Nash equilibrium emission levels

$$E_{iT}^N = \bar{E}_i - \frac{\pi_i[[1-\delta]S + \sum_{j=1}^n \bar{E}_j]}{\gamma_i[1 + \sum_{j=1}^n \gamma_j^{-1}\pi_j]},\ i \in \mathcal{N} \tag{4.4}$$

with total costs

$$N_i(T,S) = \frac{k_{iT}^N}{2}S^2 + g_{iT}^N S + h_{iT}^N,\ i \in \mathcal{N} \tag{4.5}$$

where the parameters of these quadratic forms depend on the parameters of the problem. (4.3) and (4.4) assume that contraints are not binding. If some constraint is binding, the present framework is no more quadratic and one has to resort to the more complex numerical methods discussed in subsection 4.2 [13].

If the countries cooperate, the first-order conditions at the final time T are

$$\gamma_i[E_i^* - \bar{E}_i] + \left[\sum_{j=1}^{n} \pi_j \right] \left[[1-\delta]S + \sum_{j=1}^{n} E_j^* \right] = 0, i \in \mathcal{N} \tag{4.6}$$

which yield the optimal emission levels

$$E_{iT}^* = \bar{E}_i - \left[\sum_{j=1}^{n} \pi_j \right] \frac{[1-\delta]S + \sum_{j=1}^n \bar{E}_j}{\gamma_i[1 + [\sum_{j=1}^n \gamma_j^{-1}][\sum_{k=1}^n \pi_k]]},\ i \in \mathcal{N} \tag{4.7}$$

with total costs

$$W(T,S) = \frac{k_T^W}{2}S^2 + g_T^W S + h_T^W \tag{4.8}$$

where the parameters of these quadratic forms can again be expressed in terms of the parameters of the problem. It follows that the value functions $\tilde{W}_i$, including the transfers, are also quadratic functions of S and are given by

[13] As the π_i are positive, (4.4) shows that emissions will always be below their maximum value $\bar{E}_i$ if $S \geq 0$. Thus a contraint may only be binding from below. This is the less plausible to happen the smaller are the π_i's. Indeed, the smaller are the π_i's, the less is a country incited to depollute and the closer E_i^N will be to $\bar{E}_i$.

$$\tilde{W}_i(T,S) = \frac{k_{iT}^W}{2}S^2 + g_{iT}^W S + h_{iT}^W,\ i \in \mathcal{N} \tag{4.9}$$

with

$$k_{iT}^W = k_{iT}^N + \mu_i \left[k_T^W - \sum_{j=1}^{n} k_{jT}^N \right],\ i \in \mathcal{N} \tag{4.10}$$

and similarly for g_{iT}^W and h_{iT}^W.

At time $T-1$ the problem is quadratic again and the analysis at this period is the same as at T, except that the parameters of the quadratic forms do not only depend on the parameters of the problem but also on the parameters of the value functions at time T. This induction step can be repeated for all earlier periods and the whole solution unravels by backwards induction.

The derivation of the stationary solution in the infinite horizon case starts from quadratic forms in S for the functions V_i and W. An analysis similar to the induction step above for the finite horizon case yields quadratic functions in S which have to be the same as these quadratic forms because of the stationarity. They are identical for all S if the parameters are the same which yields a system of three algebraic equations that can easily be solved. The optimal stationary emission levels and stock of pollutant follow immediately from the values of these parameters.

5.2 Computability with non-constant sharing parameters

The problem is, as was shown in section 3, that in order to get coalitional rationality the sharing parameters μ_i cannot be constant in S but become rational functions of S (cfr. (3.7)). Then, even with quadratic cost functions, the quadratic structure breaks down and one has to resort to more complex numerical methods.

There is unfortunately no guarantee that the general problem can be numerically solved. The main reason is that the general dependence of the sharing parameters on S may cause the minimization problem at one or more periods to have multiple solutions. As a result, the imputations $\tilde{W}_i(S)$ may, in the worst case, no longer be well-defined functions of S, but be point-to-set mappings instead[14]. One could possibly think of circumventing this unique-

[14]Of course, what is then meant by these imputations is unclear in this case.

ness problem by choosing the global minimizer at each period, but this is in general computationally very difficult, unless the cost and damage functions are very simple. In effect, the difficulties mentioned here are common not only to our framework, but to a large class of bi- or multi-level optimization problems. Whether they make practical computations impossible is altogether a different question. Indeed, a large class of nonconvex optimization problems do have unique solutions, and one may hope that the special structure of the problem at hand can make numerical solutions both practical and reliable.

If one assumes that the theoretical problems associated with multiple solutions do not occur, one is nevertheless faced with a complex optimization calculation, mostly because the objective function for the emissions is recursively defined and cannot be expressed in closed form. Two different technical approaches may then be developed.

The first one is to consider the determination of the emission at the first period as an optimization problem whose objective function depends on the optimal values of the emissions for the second period. These optimal values depend themselves on the stocks at the end of the first period, which can be determined once the values of the emissions at the first period are fixed, and on the optimal values of the emissions at the third period. One can then compute these optimal values recursively, because the forward dependence on optimal emissions stops at the last period.

As an illustration of this procedure, consider the problem of computing the optimal values $\mathbf{E}^*$ defined by (1.2) and (1.3), and assume that there are only two periods ($T = 2$). Start by considering the problem of solving (1.3) for $E^*_{i,1}$ (that is minimizing $W(1, S_0)$, given that the stock at the beginning of the first period, S_0, is known). In order to solve this problem, one may apply an optimization algorithm which will attempt to evaluate its objective function for given values of $E_{i,1}$. But this evaluation requires the knowledge of $W(2, S_1)$. Observe now that the value of S_1 only depends on S_0 and the $E_{i,1}$, which are all known. Thus S_1 can be evaluated. It therefore remains to evaluate $W(2, S_1)$ itself. But this is a well-defined optimization problem to which one may apply the same optimization algorithm that is used for period 1. This induces a "recursive" use of this algorithm, since it calls itself to evaluate its objective function. The situation is conceptually identical for the computation of the non-cooperative feedback Nash equilibrium, except that one must apply an equilibrium finding algorithm instead of a minimizing one.

Similarly, the values of the fallback non-cooperative equilibrium solutions can be determined by applying an equilibrium finding algorithm at period 1, which calls upon a minimizing algorithm for period 2 (see (2.9)). Furthermore, the transfers at the second period can be computed from (2.6) and the transfers at period 1 from (2.14). If more than two periods are considered, the calculation becomes more involved since one has to consider the fallback non-cooperative equilibria for each period but the last in order to determine the transfers. Of course, this procedure is typically computationally very intensive, especially if the number of periods is large.

A second approach is to avoid the use of recursive algorithms by constructing explicit approximations for the surfaces $W(t, S_{t-1})$ and $V_i(t, S_{t-1})$ as functions of S_{t-1}. This can be done by using classical interpolation schemes and the accuracy of these approximations can be refined in the neighbourhood of the values corresponding to the optimal sequence of stocks.

Which of these two procedures or which combination of them performs numerically best remains to be seen. Research is currently being carried out to assess the reliability and real difficulty of these approaches in the context of a realistic climate modelling exercise.

6 Conclusion

This paper develops transfer schemes that yield both individual and coalitional rationality for the design of international agreements that seek to achieve world optimality in stock pollutant control problems. It also shows that optimal emission levels, stocks of pollutant and transfers can easily be calculated for quadratic cost functions when only individual rationality is required, whereas the calculations appear to be much more complex if one also wants coalitional rationality. Ideas for numerical methods whereby this complexity could be mastered are put forward but a detailed elaboration is left for further research.

7 Bibliography

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

Başar T. and G.-J. Olsder (1995), *Dynamic non-cooperative game theory*. Academic Press, London, 2nd edition.

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

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 (1997), "The core of an economy with multilateral environmental externalities", *International Journal of Game Theory*, 26, 379-401.

Germain M., Ph. Toint and H. Tulkens (1998), "Financial transfers to ensure cooperative international optimality in stock pollutant abatement", CORE discussion paper n° 9701. To appear in Duchin F., Faucheux S., Gowdy J. and Nicolaï I. (eds), *Sustainability and firms: technological change and the changing regulatory environment*, Edward Elgar, London.

Germain M., H. Tulkens and A. de Zeeuw (1998), "Transferts financiers dans le cadre d'un jeu dynamique de pollution transnationale avec effets de stock", IRES discussion paper n° 9701, Université Catholique de Louvain, to appear in *Revue Economique*(Paris).

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.

Houba H. and A. de Zeeuw (1995). "Strategic bargaining for the control of a dynamic system in state-space form", *Group Decision and Negotiation*, 4, 71-97.

Kaitala V., Pohjola M. and O. 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.

Kverndokk S. (1994), "Coalitions and side payments in international CO_2 treaties". In Van Ierland (ed.) : *International environmental economics, theories, models and application to climate change, international trade and acidification*, Developments in Environmental Economics, vol. 4, Elsevier, Amsterdam.

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

Mäler, K.-G. and A. de Zeeuw (1998), "The Acid Rain Differential Game", nota di lavoro 7.95, Fondazione Eni Enrico Mattei, Milan. To appear in *Environmental and Resource Economics*.

Maskin E. and J. Tirole (1988) : "A theory of dynamic oligopoly, I : Overview and quantity competition with fixed wage costs", *Econometrica*, 56(3), 549-569.

Osborne M. and A. Rubinstein (1994) : *A course in game theory*, Cambridge MA, MIT Press.

Petrosjan L. and G. Zaccour (1995). "A multistage supergame of downstream pollution", G-95-14, GERAD, Ecole des Hautes Etudes Commerciales, Université de Montréal.

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

Tahvonen, O. (1993). "Carbon dioxyde abatement as a differential game". Discussion Paper in Economics n^o 4, University of Oulu (Finland).

Tulkens, H. (1995), "Cooperation *vs.* free riding in international environmental affairs: two approaches", Invited keynote speech at the Sixth Meeting of the European Association of Environmental and Resource Economists, Umea, Sweden, June 22, 1995; revised version (July 1997): *FEEM Nota di Lavoro* 47.97, *Fondazione Eni Enrico Mattei*, Milan and *CORE Discussion Paper* n 9752, Université Catholique de Louvain, Louvain-la-Neuve. To appear in N. HANLEY and H. FOLMER (eds), *Game Theory and the Environment*, Elgar, London.

8 Appendix : proof of the theorem

PART 1 First the transfers (3.6) are rewritten in a form like (2.25). The total decrease of costs generated by the extension of the cooperation to the current period (the negative of the ecological surplus) is equal to

$$W(S) - V(S) = \sum_{i=1}^{n} \{[C_i(E_i^*) - C_i(E_i^V)] + [D_i(S^*) - D_i(S^V)] + \beta[\tilde{W}_i(S^*) - \tilde{W}_i(S^V)]\} \quad (A.1)$$

D_i and $\tilde{W}_i$ are supposed to be monotonically increasing, so that F_i defined by (3.8) is also monotonically increasing. It follows that the $\tilde{\mu}_i$ defined by (3.7) are positive. They also sum up to one. Because F_i are supposed to be convex, it follows that

$$F_i(S^*) - F_i(S^V) \leq F_i'(S^*)[S^* - S^V] = F_i'(S^*) \sum_{j=1}^{n} [E_j^* - E_j^V] \quad (A.2)$$

Let

$$G_i(E_i, S) \stackrel{\triangle}{=} C_i(E_i) + E_i \sum_{j=1}^{n} F_j'(S) \quad (A.3)$$

$$R_i(\mathbf{E}, S) \stackrel{\triangle}{=} F_i(S) - F_i'(S^*) \sum_{j=1}^{n} E_j \quad (A.4)$$

(A.1) can then be rewritten as

$$W(S) - V(S) = \sum_{i=1}^{n} [G_i(E_i^*, S^*) - G_i(E_i^V, S^*)] + \sum_{i=1}^{n} [R_i(\mathbf{E}^*, S^*) - R_i(\mathbf{E}^V, S^V)] \quad (A.5)$$

G_i is convex in E_i since C_i is, so that $G_i(E_i^*, S^*) - G_i(E_i^V, S^*) \leq \frac{\partial G_i}{\partial E_i}(E_i^*, S^*)[E_i^* - E_i^V]$, which is equal to 0 because

$$C_i'(E_i^*) + \sum_{j=1}^{n} [D_j'(S^*) + \beta \tilde{W}_j'(S^*)] = C_i'(E_i^*) + \sum_{j=1}^{n} F_j'(S^*) = 0, \ i \in \mathcal{N} \quad (A.6)$$

characterizes the optimum. Furthermore, $R_i(\mathbf{E}^*, S^*) - R_i(\mathbf{E}^V, S^V) \leq 0$ because of (A.2). It follows that (A.5) is the sum of two non-positive terms.

Define transfers in a form like (2.25) as

$$\tilde{\theta}_i(S) = -[W_i(S) - V_i(S)] + \tilde{\mu}_i(S^*) \sum_{j=1}^{n} [G_j(E_j^*, S^*) - G_j(E_j^V, S^*)]$$

$$+[R_i(\mathbf{E}^*, S^*) - R_i(\mathbf{E}^V, S^V)] \qquad (A.7)$$

which are balanced (given (A.5)) and make cooperation individually rational. It is easy to verify, using the definitions of W_i, V_i, F_i, G_i, R_i and $\tilde{\mu}_i$ (cfr. (2.24), (2.22), (2.26), (3.8), (A.3), (A.4) and (3.7)), that the transfers given by (A.7) can be rewritten as (3.6), so that indeed (3.6) has a form like (2.25).

PART 2 The second part of the proof of the theorem is inspired by Chander and Tulkens (1997), and is based on the following two lemmas.

Lemma 1 : Let $\mathbf{E}^U$ and S^U be the vector of emission levels and the stock characterizing the partial agreement Nash equilibrium w.r.t. coalition $U \subseteq \mathcal{N}$ (cfr. (3.1-2)). Let $\mathbf{E}^V$ and S^V be the vector of emission levels and the stock characterizing the fallback non-cooperative equilibrium (solution of (2.21)). If $\forall i \in \mathcal{N},\ \forall S \geq 0,\ \tilde{W}_i'(S) \geq 0$ then
(i) $S^U \leq S^V$ and
(ii) $\forall i \in \mathcal{N} \backslash U,\ E_i^U \geq E_i^V$.

Proof : (i) First-order conditions characterizing the partial agreement Nash equilibrium w.r.t. coalition U are (cfr.(3.1-2) and (3.8)) :

$$C_i'(E_i^U) + \sum_{k \in U} F_k'(S^U) = 0,\ i \in U \qquad (A.8)$$

$$C_j'(E_j^U) + F_j'(S^U) = 0,\ j \in \mathcal{N} \backslash U \qquad (A.9)$$

Because D_i and $\tilde{W}_i$ are convex and increasing functions of S, F_i is also a convex and increasing function of S.

Suppose that the lemma is false. Then $S^U > S^V$ implies that
$\forall i \in \mathcal{N},\ F_i'(S^U) \geq F_i'(S^V)$, which implies that

$$\begin{cases} -C_i'(E_i^U) = \sum_{k \in U} F_k'(S^U) \geq F_i'(S^U) \geq F_i'(S^V) = -C_i'(E_i^V),\ \forall i \in U \\ -C_j'(E_j^U) = F_j'(S^U) \geq F_j'(S^V) = -C_j'(E_j^V),\ \forall j \in \mathcal{N} \backslash U \end{cases}$$

so that $C_i'(E_i^U) \leq C_i'(E_i^V),\ \forall i \in \mathcal{N}$ implies that $E_i^U \leq E_i^V,\ \forall i \in \mathcal{N}$ which yields a contradiction with $S^U = (1-\delta)S + \sum_i E_i^U$, $S^V = (1-\delta)S + \sum_i E_i^V$ and $S^U > S^V$.

(ii) $S^V \geq S^U$ implies that $F_i'(S^V) \geq F_i'(S^U),\ \forall i \in \mathcal{N}$ which implies that $-C_i'(E_i^V) = F_i'(S^V) \geq F_i'(S^U) = -C_i'(E_i^U),\ \forall i \in \mathcal{N}\backslash U$ so that $E_i^V \leq E_i^U,\ \forall i \in \mathcal{N}\backslash U$. QED

As a corollary of Lemma 1 it follows that, unless everything stays the same, the coalition as a whole emits less than in the fall-back non-cooperative equilibrium, because the stock decreases according to (i) and the countries outside the coalition emit more according to (ii).

Lemma 2 : Consider the imputation $(W_1^U(S), \cdots, W_n^U(S))$ defined by

$$W_i^U(S) = W_i(S) + \theta_i^U(S), \tag{A.10}$$

with

$$\theta_i^U(S) = -[W_i(S) - w_i(U;S)] + \tilde{\mu}_i(S^*)\sum_{j=1}^{n}[G_j(E_j^*, S^*) - G_j(E_j^U, S^*)]$$

$$+[R_i(\mathbf{E}^*, S^*) - R_i(\mathbf{E}^U, S^U)], \tag{A.11}$$

where

$$w_i(U;S) = C_i(E_i^U) + D_i(S^U) + \beta\tilde{W}_i(S^U) = C_i(E_i^U) + F_i(S^U) \tag{A.12}$$

is the total discounted cost of country i induced by strategy $\mathbf{E}^U$.

Suppose that (3.4) does not hold for some coalition U while condition (ii) of the theorem holds. Then, the imputation (A.10-11) dominates the imputation $(\tilde{W}_1(S), \cdots, \tilde{W}_n(S))$ (defined by (3.5-6)) in the sense that

$$\text{(i)} \sum_{i\in U} W_i^U(S) < \sum_{i\in U} \tilde{W}_i(S) \tag{A.13}$$

$$\text{(ii)} \sum_{i\in \mathcal{N}\backslash U} W_i^U(S) \leq \sum_{i\in \mathcal{N}\backslash U} \tilde{W}_i(S). \tag{A.14}$$

Proof : (i) It follows from (A.10) and (A.11) that

$$W_i^U(S) = w_i(U;S) + \tilde{\mu}_i(S^*)\sum_{j=1}^{n}[G_j(E_j^*, S^*) - G_j(E_j^U, S^*)]$$
$$+[R_i(\mathbf{E}^*, S^*) - R_i(\mathbf{E}^U, S^U)] \leq w_i(U;S) \qquad (A.15)$$

(cfr. Part 1 of the appendix). Because it was assumed that (3.4) does not hold, it follows that

$$\sum_{i\in U} W_i^U(S) \leq \sum_{i\in U} w_i(U;S) = w(U;S) < \sum_{i\in U} \tilde{W}_i(S). \qquad (A.16)$$

(ii) Comparing (3.5) and (A.10),(A.14) is equivalent to

$$\sum_{i\in\mathcal{N}\setminus U} \theta_i^U(S) \leq \sum_{i\in\mathcal{N}\setminus U} \tilde{\theta}_i(S). \qquad (A.17)$$

In Part 1 of the appendix, it was shown that $\tilde{\theta}_i(S)$ can be written as (3.6) or as (A.7). Analogously, one can show that (A.11) can be rewritten as

$$\theta_i^U(S) = -[C_i(E_i^*) - C_i(E_i^U)] + \tilde{\mu}_i(S^*)\sum_{j=1}^{n}[C_j(E_j^*) - C_j(E_j^U)]. \qquad (A.18)$$

It follows that (A.17) is equivalent to

$$\sum_{i\in\mathcal{N}\setminus U}[C_i(E_i^U) - C_i(E_i^V)] + \left[\sum_{i\in\mathcal{N}\setminus U}\tilde{\mu}_i(S^*)\right]\sum_{j=1}^{n}[C_j(E_j^V) - C_j(E_j^U)] \leq 0 \quad (A.19)$$

or

$$\sum_{i\in\mathcal{N}\setminus U}[C_i(E_i^U) - C_i(E_i^V)] + \left[\sum_{i\in\mathcal{N}\setminus U}\tilde{\mu}_i(S^*)\right]\sum_{j\in\mathcal{N}\setminus U}[C_j(E_j^V) - C_j(E_j^U)]$$
$$+\left[\sum_{i\in\mathcal{N}\setminus U}\tilde{\mu}_i(S^*)\right]\sum_{j\in U}^{n}[C_j(E_j^V) - C_j(E_j^U)] \leq 0 \qquad (A.20)$$

or

$$\left[\sum_{i\in\mathcal{N}\setminus U}[C_i(E_i^U) - C_i(E_i^V)]\right]\left[1 - \sum_{i\in\mathcal{N}\setminus U}\tilde{\mu}_i(S^*)\right]$$

$$+\left[\sum_{i\in\mathcal{N}\setminus U}\tilde{\mu}_i(S^*)\right]\sum_{j\in U}[C_j(E_j^V)-C_j(E_j^U)]\leq 0. \qquad (A.21)$$

The first factor of the first term is a sum of non-positive terms because of Lemma 1 (ii). The second factor is non-negative since $0\leq\tilde{\mu}_i(S^*)\leq 1\,\forall i\in\mathcal{N}$, and $\sum_{i=1}^n\tilde{\mu}_i(S^*)=1$. The same holds for the first factor of the second term. The second factor of the second term is non-positive because of condition (ii) of the theorem. QED

Proof of the theorem : It has to be shown that the imputation defined by (3.5-6) belongs to the core of the game or that (3.4) holds. Suppose that (3.4) does not hold for some coalition U, then Lemma 2 defines an imputation, given by (A.10-11) satisfying (A.13) and (A.14), so that

$$\sum_{i=1}^n\theta_i^U(S)<\sum_{i=1}^n\tilde{\theta}_i(S) \qquad (A.22)$$

This yields a contradiction, because according to (3.6) and (A.18) both sides of (A.22) are equal to 0. QED