CESifo *Working Paper Series*

THE KYOTO PROTOCOL: AN ECONOMIC AND GAME THEORETIC INTERPRETATION

Parkash Chander
Henry Tulkens
Jean-Pascal van Ypersele
Stéphane Willems*

Working Paper No. 229

December 1999

CESifo
Poschingerstr. 5
81679 Munich
Germany
Phone: +49 (89) 9224-1410/1425
Fax: +49 (89) 9224-1409
http://www.CESifo.de

---

* The authors are especially indebted to their MIT colleague Denny Ellerman who kindly informed them of the state of his work and provided key numerical information. They also thank their colleague Jean Gabszewicz for a careful reading of an earlier version. The research on which it is reported here is part of the program "Changements climatiques, Négociations internationales et Stratégies de la Belgique" (CLIMNEG), supported by the Belgian State's Services du Premier Ministre, Services fédéraux des Affaires scientifiques, techniques et culturelles (SSTC), Brussels. Chander's work in Belgium was financed by the CLIMNEG program.

CESifo *Working Paper No. 229*
December 1999

THE KYOTO PROTOCOL: AN ECONOMIC AND GAME THEORETIC INTERPRETATION

Abstract

Calling upon both positive and normative economics, we attempt to characterize the issues at stake in the current international negotiations on climatic change. We begin (section 2) by reviewing the main features of the Protocol. Then (Section 3), we identify by means of an elementary economic model the main concepts involved: optimality, non cooperation, coalitional stability. We observe (Section 4) that "business-as-usual", "no regrets" and other domestic policies are alternative ways to conceive of the non cooperative equilibrium prevailing before the negotiations. Which one should be retained? Data suggest that the prevailing situation is a mixed one, exhibiting characteristics of several of these policies. We then turn (Section 5) to interpreting the Protocol. While there is no firm basis to assert that the emission quotas chosen at Kyoto correspond to optimal emissions (although they are a step in the right direction), economic and game theoretical arguments are put forward to support the view that for achieving these emission quotas, trading ensures efficiency, as well as coalitional stability for the agreement provided it is adopted at the largest scale *i.e.* worldwide. Finally, it is argued in Section 6 that beyond the Kyoto Protocol, the achievement of coalitionally stable optimality at the world level is a real possibility with trading, provided agreement can be reached in the future as to appropriate reference emission levels, in particular as far as developing countries are concerned.

JEL Classification: C7, D74, F2, H4

Parkash Chander
Indian Statistical Institute (ISI), New Delhi and
Université Catholique de Louvain
CORE

Henry Tulkens
Université Catholique de Louvain, CORE
Voie du Roman Pays 34
1348 Louvain-la-Neuve
Belgium

Jean-Pascal van Ypersele
Université Catholique de Louvain
Institut d'Astronomie et de Géophysique
Georges Lemaître (ASTR)

Stéphane Willems
Task Force Développement Durable (TFDD)
Bureau fédéral du Plan
Brussels

Outline

1. Introduction: Cooperation at the world level, from Rio to Kyoto

Our central theme in this paper is the one of cooperation at the world level on the issue of climatic change. We start from the facts and then try to enlighten them by means of ideas provided by economics and game theory.

The negotiations on climate change that have been taking place since the late 1980's within the United Nations institutions[6] are obviously a quasi-worldwide process, judging by the length of the list of countries[7] participating. But these negotiations, prior to the Kyoto meeting, had led only to a "framework convention", signed in 1992 in Rio de Janeiro, that was little more than a declaration of intentions. The real issue was then: are the continuing negotiations eventually going to lead to a sustainable agreement bearing on effective actions that is also worldwide? Or will they lead to a breaking up of the countries into separate blocks, each acting to the best of its own interests?

The Kyoto Protocol, signed in December 1997, is the major development in the post-Rio evolution of these negotiations. Its importance lies mainly in the fact that it bears on effective actions to be taken by countries, actions that are recognized as binding commitments by them.

However, according to the Protocol, not all countries have to take specific actions. As our summary presentation will report more in detail below, commitments to quantified emissions reduction or limitation are mentioned only for the so-called "Annex 1" parties[8]. The rôle of the other countries in the agreement, while not ignored, is much less precisely specified.

The natural question that then arises is whether the Kyoto Protocol is to be considered as just an "Annex I" Protocol; or is it to be seen, after further thought and beyond the appearances, as a worldwide Protocol? Below, we defend the second thesis, first on the basis of our own conviction, but also because we think we can support it by means of well established conceptual tools of economics and game theory.

---

[6] For a thorough account of the scientific evidence on the state of the problem, the reader is referred to the work of the Intergovernmental Panel on Climate Change (IPCC), and in particular to the contribution of its Working Group III (see under IPCC 1995 in our references below).

[7] 178 in Rio, 159 in Kyoto and 161 in Buenos Aires.

[8] "Annex 1" (to the Rio Convention text) countries are the OECD countries, those of the former Soviet Union and those of the eastern European economies in transition.

For the applied economist, the Kyoto Protocol offers an exceptionally rich combination of opportunities to put theory into use, both in a positive and a normative way — that is, for explaining events as well as for advising on decision making. Indeed, several strands of theory are involved as we shall show: externalities of the Samuelsonian public good type, Nash non cooperative equilibria, worldwide Pareto optimality, cooperative solution concepts, and finally Walrasian market equilibria with their Edgeworthian coalitional stability... . All these are involved!

2. Main features of the Protocol

We briefly state here what the main features of the Protocol are, from the point of view of our arguments to follow:

(i) Dated *emission quotas*, expressed in percentages of 1990 emissions, are established for Annex I countries, to be met around 2010.

(ii) The principles of (a) *emissions trading* by countries (or by their nationals) and of (b) *joint implementation* are established for Annex I countries.

(iii) A *clean development mechanism* (CDM) is established as a way to involve non Annex-1 countries (especially developing ones) in some particular form of joint implementation and emissions trading.

No explicit provision in the Protocol mentions the introduction of targets for non Annex-1 countries. But it is expected that this will take place in the future through the general review clauses of the Protocol and of the Convention.

Trade in emissions will be allowed only among countries who do ratify. It is also expected that it will not be allowed with the countries that would not fulfill their obligations under the Protocol.

Finally, the Protocol comes automatically into force[9] only if *(i)* at least 55 Parties to the Convention have ratified it, and *(ii)* these 55 Parties include a number of Annex 1 Parties accounting for at least 55% of the base year CO_2 emissions of all Annex I Parties to the Convention.

---

[9] In Kyoto, the text of the protocol was adopted unanimously by the delegates of the 159 countries that participated in the negotiations. Signature of the text by governments and ratification by parliaments are the following stages of the process.

While parties are committing themselves to proceed to enforcement within their country, no sanctions are specified if a ratifying country does not fulfill its obligations under the Protocol, except for the above provision on being excluded from emissions trading. A compliance regime, including possible sanctions for non-compliance, is yet to be specified in the process of future negotiations.

3. Economics of the issues at stake

Consider the n countries of the world (indexed below i = 1, ...,n), who each enjoy an aggregate consumption level x_i , equal to the aggregate value of their production activities y_i , minus the damage D_i consisting in lost production that results from global pollution. Each country i's productive activity entails indeed some amount of polluting emissions e_i, that are related with production according to the increasing and strictly concave production function[10] $y_i = g_i(e_i)$. Damages in each country are generated by the total of such emissions, $\sum_{i=1}^{n} e_i$; they affect production possibilities in each country[11] in a way that is usually represented by an increasing damage cost function $D_i = d_i(\sum_{j=1}^{n} e_j)$ that for simplicity we assume to be linear. Each country's consumption possibilities are thus given by the expression

$$x_i = g_i(e_i) - d_i \sum_{j=1}^{n} e_j, \tag{1}$$

where $d_i > 0$ is thus the damage cost per unit of emission or, equivalently, the benefit per unit of abatement (for decreasing $S\, e_j$).

3.1 World optimality

Ignoring distributional issues, world consumption optimality can then be represented by the consumption levels that maximize $\sum_{i=1}^{n} x_i$ with respect to the n variables $e_1, \dots ,e_n$. Let $(e_1^*, \dots ,e_n^*)$ be the vector of emission levels in the n countries that achieves such a world optimum. First order conditions for a maximum are given by the following system of equations expressing equality

---

[10] We may think of e_i either as the energy input in the production, or as the pollution emission, assuming that a unit of energy generates a unit of pollutant as a byproduct. Accordingly, $g'_i(e_i)\ (= dg_i(e_i)/de_i)$ may be interpreted either as the marginal product of energy or (for decreasing e_i) the marginal cost of abatement, depending upon the context.

[11] Numerical estimates of damages in some regions of the world are given in Table 1 below.

between the marginal cost of global damages and the marginal abatement cost of each party i :

$$g_i'(e_i^*) - \sum_{j=1}^{n} d_j = 0 , \qquad i=1,\dots,n. \qquad (2)$$

We shall develop our arguments below under the assumption that climate change negotiations are aiming at achieving such a world optimum. Attaining it requires coordination among the countries, so as to ensure that each one of them does take into account the effect of its emissions on the other countries as reflected by their damage cost functions.

3.2 Non cooperative equilibrium

It is indeed often argued[12] that in the absence of coordination countries choose emission policies that best suit their own interest, taking as given what the other countries do. This leads to consider that a non cooperative equilibrium of some sort prevails between countries if no negotiations take place.

How would a country determine its best emissions levels? The answer is not immediate since imposing to itself low emissions implies low net production according to the function g_i, whereas allowing for high emissions entails high damage costs according to the function D_i . Classical economic reasoning suggests that a rational domestic optimum for each country would be one that best balances these two aspects; it is achieved by maximizing its own consumption level x_i with respect to e_i as defined in (1), taking as given all variables e_j with $j \neq i$. If all countries adopt such a behavior, a Nash-type equilibrium between countries prevails, that we represent by the vector of emissions[13] $(\bar{e}_1, \dots, \bar{e}_n)$. For each country i, the first order condition of its maximizing behavior is given by the equation

$$g_i'(\bar{e}_i) - d_i = 0 \qquad (3)$$

while its achievable consumption level is

$$\bar{x}_i = g_i(\bar{e}_i) - d_i \sum_{j=1}^{n} \bar{e}_j .$$

---

[12] see e.g. CHANDER and TULKENS 1992

[13] Uniqueness of this vector is ensured under our assumptions of concavity of the functions g_i and of linearity of the functions D_i .

Two characteristics of the non cooperative equilibrium so defined are essential for our purposes: *(i)* the equilibrium emissions $(\bar{e}_1, \ldots, \bar{e}_n)$ are clearly not a world optimum as can be seen from comparing (2) and (3): that is why negotiations are necessary; and *(ii)* $\bar{e}_i > e_i^*$ for each i since g_i is concave and $\sum_{j=1}^{n} d_j > d_i$ for all i's ; thus, world optimal emissions are lower than those prevailing at the non cooperative equilibrium.

3.3 Coalitional stability for the treaty

The basic reason for the non optimality of the Nash Equilibrium is a well known externality argument. Each country decides its emission level $\bar{e}_i$ without concern for the effects on other countries: it thus equates its marginal abatement cost, $g'_i(\bar{e}_i)$, to its own marginal damage cost, d_i, whereas a world optimum requires each country to equate its marginal abatement cost to the aggregate world marginal damage cost, $\sum_{i=1}^{n} d_i$.

Furthermore, a world optimum may require from the various countries different levels of abatement $\bar{e}_i - e_i^*$, entailing costs and benefits that are *a priori* by no means identical across them: some may have high abatement costs while having only small damage costs to avoid, whereas other countries may have low abatement costs while facing high damage costs. To have the world optimum voluntarily agreed upon by all countries requires in addition that *for each country* and *for each group of countries* the benefits exceed the costs of abatement. Because of the asymmetries just mentioned, this can be achieved only by means of appropriately designed resource transfers from the net gainers to the net losers.

To that effect CHANDER and TULKENS 1997 have proposed that the optimal emission levels $(e_1^*, \ldots, e_n^*)$ specified in the treaty be accompanied by a scheme of transfers $(T_1, \ldots, T_n)$ which are of the form:

$$T_i = \left\{g_i(\bar{e}_i) - g_i(e_i^*)\right\} - \frac{d_i}{\sum_{i-1}^{n} d_j}\left\{\sum_{j=1}^{n} g_j(\bar{e}_j) - \sum_{j=1}^{n} g_j(e_j^*)\right\}, \quad i=1,\ldots,n, \tag{4}$$

where $T_i > 0$ if the transfer is received by country *i*, while $T_i < 0$ if the transfer is paid[14]. The first expression within braces is equal to the abatement cost borne

---

[14] The transfers are expressed here in units of physical goods. The issue whether it is preferable that such transfers be financial rather than in real terms is an important one, that we cannot deal with here.

by country i from moving from its Nash equilibrium level of emissions $\bar{e}_i$ to the level e_i^* prescribed by world optimality. As this amount is positive in (4), the rôle of this part of the transfer appears to be to cover that cost increase for i. The second expression within braces is the *world total* over all countries of their emissions abatement cost from the Nash equilibrium levels to the world optimal ones— also a positive magnitude. With the ratio $d_i / \sum_{j=1}^{n} d_j$ and taking account of the negative sign, the second term in (4) thus determines a contribution of country i, which is specified as a fraction of the aggregate abatement cost.

Clearly $\sum_{i=1}^{n} T_i = 0$, so that these transfers would ensure a balanced budget if an international agency were established for implementing them. Notice also the role played by the reference emission levels $\bar{e}_i$ in the design of the transfers — a feature whose importance will be highlighted below.

The "coalitional stability" property claimed above for the Chander-Tulkens proposal of optimality with transfers is that, in addition to making each country *individually* better off compared to the Nash equilibrium, it also makes every *group* of countries better off, compared to what they could get by adopting any alternative arrangement among themselves, be it in terms of emissions, or transfers, or both. For further reference in our arguments below, let us be more precise on this property. Let $W = \{i = 1, \dots, n\}$ denote the set of all countries of the world and $S \subset W$ be any subset, or "coalition" of countries. Then the best outcome that the members of S could obtain by making arrangements among themselves only — to be called a "partial agreement Nash equilibrium with respect to S" (PANE w.r.t. S) — is the one resulting from the emissions policy $(\tilde{e}_1, \dots, \tilde{e}_n)$ defined by, for the members of S,

$$(\tilde{e}_i)_{i \in S} = arg\,max \left(\sum_{i \in S} g_i(e_i) - (\sum_{i \in S} d_i)(\sum_{i \in S} e_i + \sum_{j \in W \setminus S} \tilde{e}_j) \right),$$

and for the countries not in S:

$$\tilde{e}_j = arg\,max \left(g_j(e_j) - d_j(e_j + \sum_{i \neq j} \tilde{e}_i) \right), \quad j \in W \setminus S.$$

A PANE w.r.t. S is thus a Nash equilibrium between the countries in S acting jointly and the remaining countries acting individually. It can be characterized by the first-order conditions:

$$\left.\begin{aligned} g_i'(\tilde{e}_i) &= \sum_{j \in S} d_j, \quad i \in S \\ \text{and} \qquad & \\ g_i'(\tilde{e}_i) &= d_i, \quad i \in W \setminus S \end{aligned}\right\}$$

A comparison of these conditions with (2) implies $\sum_{i \in S} \tilde{e}_i \leq \sum_{i \in S} \bar{e}_i$ and $\tilde{e}_j = \bar{e}_j, j \in W \setminus S$. Since in a PANE w.r.t. S the countries within the coalition coordinate their emissions so as to take into account of their effect on each other, their emissions are lower compared to the Nash equilibrium. The emissions of the countries outside of the coalition are however not lower. In fact, they might be higher if the damage function is convex but not linear[15]. Moreover, since total world emissions are lower in a PANE w.r.t. S, the countries outside the coalition are better off, although that is not the intention of the coalition.

3. 4 Statics *vs.* dynamics

Thus far, and for most of the sequel, the above quantities x_i and e_i are considered to be flows per unit of time. The damages from climate change are however induced less by the flow of greenhouse gas emissions than by the increase[16] ΔS in their accumulated stock S in the atmosphere. At each time t, ΔS_t is thus determined by a relation of the form

$$\Delta S_t = -d\, S_{t-1} + k \sum_{i=1}^{n} e_{it}\,, \qquad t=1,2,\ldots$$

where according to climatic science common wisdom $d \cong 0.01$ and k is of the order of .5 (and slowly increasing over time).

The issues at stake have thereby an inherently dynamic component that is by no means ignored in the economics literature on climate change; see *e.g.* NORDHAUS and YANG 1996, or, for our part, GERMAIN, TOINT, TULKENS and de ZEEUW 1998. One might therefore consider that world optimality is not to be defined in terms of just one period emission, production and consumption levels as we have done but, instead, of multiperiod emission trajectories $\left\{(e_{1t}, \ldots e_{nt})_{t=1,2,\ldots}\right\}$ and similarly for production and consumption.

---

[15] This might also happen when the countries outside the coalition are not acting rationally but following the business-as-usual policy, since the abatement by the coalition S might result into lower energy prices in the rest of the world. Ellerman and Decaux 1998 observe this phenomenon in their computable general equilibrium model, and call it "leakages".

[16] usually taken with respect to pre-industrial times.

While this more elaborate modeling has its merits, it turns out to be unnecessary for our purposes. Indeed, one may have noted that the specific object of the Kyoto Protocol is *not* trajectories of emissions: it is emissions *levels* at some point in time (around[17] 2010). We therefore feel justified in working, in the present paper, with the usual "static" or one period model[18].

4. "No regrets", "business-as-usual" and other possible domestic non cooperative policies at the pre-agreement stage

The non cooperative behavior described in section 3.2 is not the only one conceivable of this kind. Indeed, the fulfillment of conditions (3) that characterize it requires domestic policies to be designed and implemented, involving an energy tax or appropriately priced pollution permits, so that the energy price including the tax or the permits unit price be equal to the domestic marginal damage cost d_i . These belong to the class of what is often called "no regret policies". However, not all countries can be said to have adopted such a nationally rational course of action.

For instance, industrial firms in some countries may have strong lobbying power and use it so as to obtain low energy prices. While still choosing, as profit maximizers, energy use and emissions so that g_i' be equal to the price of energy (denoted henceforth as p_i), this results into emissions $\bar{\bar{e}}_i$ higher than $\bar{e}_i$ and such that $g'_i(\bar{\bar{e}}_i) < d_i$, thus successfully preventing the nationally rational policy to be adopted. If this behavior is assumed to prevail in all countries, a different equilibrium — equally non cooperative — results, called by NORDHAUS and YANG 1996 the "market solution" ("business-as-usual", according to others).

Alternatively, energy importing countries facing balance of payments problems may have introduced high taxes and domestic prices of energy: their emission levels $\bar{\bar{e}}_i$ are then likely to be such that $g'_i(\bar{\bar{e}}_i) > d_i$.

Finally, another reason why a nationally rational policy may not come about is that firms in a country may simply not be profit maximizers, as it is particularly the case with large public sector enterprises of non market economies. In such cases, the domestic equilibria are neither of the "market" nor of the "nationally rational" type, and energy prices do not induce any well

---

[17] Actually an average over the years 2008-2012.

[18] Nash equilibrium and optimal trajectories are determined and discussed in GERMAIN, TOINT, TULKENS and de ZEEUW 1998.

defined emissions policy — except for the fact of a generally low concern for economical use of energy.

Our point in this section is that in the situation prevailing at the pre-negotiation stage, all three types of country behavior are likely to be present, and we wish to illustrate this empirically with the data in Table 1 below.

We first note a similarity between the structure, across some major countries, of the average energy prices for three kinds of fossil fuels (first three columns) on the one hand, and of the marginal abatement costs (fourth column) on the other hand. In particular, it is seen that the energy prices in the US are systematically lower and so is the marginal cost of abatement[19]. Moreover, for the three developed regions US, EU and Japan which are also market economies, the higher the energy prices, the higher the marginal abatement costs[20]. For the other countries we cannot say much, not only because of lack of data but also because they are either non market or less developed, or both.

Second, we have an opportunity to characterize some domestic policies by using equation (3) — according to which in countries that choose their emission levels rationally, *i.e.* in the "no regrets sense", the marginal damage cost from emissions must be equal to the marginal abatement cost and also to the average energy prices. Indeed, the data in the table reveal that marginal abatement costs are lower in the US compared to the EU and Japan (they are even lower than those of a developing country like India). Now, it can hardly be the case that the marginal damage cost for the US, the largest economy, be lower than that of the EU or of Japan. Therefore, we have an indication that in the US, decisions regarding emissions are determined by the "business-as-usual" policy rather than optimized at the national level.

We indicate in the last column the type of pre-negotiation domestic equilibrium we conjecture from the data to prevail in each region .

What is the relevance of these observations for our purposes in this paper? While the optimum emissions $(e_1^*, \dots, e_n^*)$ are, as seen from (2), independent of those at the pre-negotiation stage, the transfers T_i defined in (4) may have to be modified with the $\bar{e}_i$'s substituted by the actual emission levels of each country i as they are described here. Does such a substitution affect the

[19] In case of Japan, the marginal cost of abatement may look exceptionally high, but this is because of its large dependence on nuclear energy and natural gas.

[20] Coal in Japan is a noticeable exception; but its use there is considerably lower.

coalitional stability property of the transfers? The answer is no[21], as long as one can assume that coutries do adopt the same behavior at the pre-agreement stage and at a PANE when not in the colaition. For the sake of simplicity however, we will continue to consider the $\bar{e}_i$'s as Nash equilibrium emission levels.

Table 1 — **Retail prices (in US$ per unit) of industrial fossil fuels, marginal abatement cost and damage cost in selected countries or regions**

	Heavy fuel oil for industry* (per ton)	Steam coal for industry* (per ton)	Natural gas for industry* (per 10kcalGCV)	Marginal abatement cost/ton of carbon for first 100Mton reduction **	Annual damage cost as % of GDP***	Type of domestic equilibrium
US	138.00	35.27	136.62	$ 12	1.3	$g_i'(\bar{e}_i) = p_i < d_i$
EU	187.4	76.0	182.0	$ 40	1.4	$g_i'(\bar{e}_i) = p_i \geq d_i$
Japan	172.86	49.90	423.12	$ 350	1.4	$g_i'(\bar{e}_i) = p_i \geq d_i$
India	191.15	19.36	na	$ 22	na	?
FSU	na	na	na	$ 22	0.7	?
China	na	na	na	$ 3.5	4.7	?

*Source: *Energy Prices and Taxes* 1996
**Source: ELLERMAN and DECAUX 1998
***Source: FANKHAUSER 1995

[21] In technical terms, because levels of the $\bar{e}_i$'s higher than those of a Nash equilibrium induce a larger core for the game whereby Chander and Tulkens 1997 establish the coalitional stability property of the transfers (4).

5. Kyoto quotas, worldwide trading and coalitional stability

5.1 The Kyoto quotas: not optimality, but a step in the right direction

While it is straightforward to define and characterize a world optimum in theory, as we have done in section 3, implementing it is undoubtedly difficult in practice, for several reasons among which we identify four. First, determining optimal emissions at the world level requires knowledge of, and agreement on, what the aggregate marginal damage costs $\sum_{i=1}^{n} d_i$ are, as well as the countries' marginal abatement costs $g_i'(e_i)$. While "objective" technical studies can provide some of that information, one can expect that, due to the huge interests at stake in many segments of all concerned economies, pressures are exerted for either concealing or simply not collecting the statistical material required.

Secondly, because the achievement of a stable world optimum may require, as noted earlier, resource transfers between the countries to compensate those for which net benefits,*i.e.* benefits minus costs, are low or negative, institutions or mechanisms that hardly exist today are needed to implement such transfers.

Thirdly, the reference emission levels $\bar{e}_i$ — that play a rôle in the design of the transfers (4) ensuring coalitional stability — may themselves be considered unfair, typically by those countries that are in the early stages of their economic development: they currently have comparatively low emission levels, while developed countries have high ones. In the future, when they will be developed, currently developing countries will have higher emissions and they might argue that these should be used as reference levels instead of those of today.

Finally, if reductions in emissions $\bar{e}_i - e_i^*$, are very large (as proposed by some countries), they are simply not politically feasible, at least in the short run. In fact, the Kyoto Protocol only requires relatively small reductions for the immediate future (the next fifteen years), leaving further reductions for later periods.

For all these reasons it is difficult to assess whether the emissions reductions chosen by the Kyoto signatories correspond to world optimal emissions.

Yet, countries in Kyoto have agreed upon *some* scheme of quotas on their emissions. Denote this scheme by the vector $(\hat{e}_1,\dots,\hat{e}_n)$ where $\hat{e}_i$ is the quota on emissions of country i and write $\hat{e} = \sum_{i=1}^{n} \hat{e}_i$ for the so induced aggregate reduced emissions[22]. Because $\hat{e}$ is lower than $\bar{e} = \sum_{i=1}^{n} \bar{e}_i$, that is, the total sum of emissions in 1990, these aggregate reduced emissions are for sure a step in the right direction since irrespective of whether 1990 emissions are business-as-usual or no regrets policies, both do imply too large emissions with respect to the world optimum.

5.2 Efficiency of emissions trading

If the Kyoto aggregate emissions reduction to $\hat{e}$ is achieved by letting each country abide to its emissions quota and simply emit up to $e_i = \hat{e}_i$, the ensuing aggregate gross world production, $\hat{y} \underset{def}{=} \sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} g_i(\hat{e}_i)$, may not be the highest achievable level. If so, the national policies $e_i = \hat{e}_i$ for each i would be inefficient. Alternative specifications of the countries' emissions e_i, all achieving $\hat{e}$, are conceivable. In fact, recalling (1), the highest possible world consumption levels compatible with $\hat{e}$ would be those given by the vector $\hat{e}^* = (\hat{e}_1^*,\dots,\hat{e}_n^*)$ that solves the problem

$$Max \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} [g_i(e_i) - d_i \sum_{j=1}^{n} \hat{e}_j] \tag{5}$$

$$\text{subject to } \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \hat{e}_i . \tag{6}$$

How are these efficient emission levels to be determined? With appropriate information on the production (or abatement cost) functions g_i, this could be done by computation. However, having argued above that such information is hard to come by, it is likely that strong opposition would arise against the computed emission levels, and in particular against those that would be larger than $\hat{e}_i$, which is indeed a possibility!

We want to show presently that the desired efficient emission levels are precisely those that a competitive market equilibrium in tradable emission quotas would determine; in other words, that tradability of quotas automatically solves the problem (5)-(6).

---

22 Notice that for all non Annex 1 countries, we have $\hat{e}_i = \bar{e}_i$.

To this effect, notice first that the vector $\hat{e}^* = (\hat{e}_1^*, \ldots, \hat{e}_n^*)$ we are interested in identically solves the problem of maximizing aggregate gross production:

$$Max \sum_{i=1}^{n} [g_i(e_i)] \quad \text{subject to} \quad \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \hat{e}_i \ , \tag{7}$$

because the dropped terms $d_i \sum_{j=1}^{n} \hat{e}_j$ are constants in (5).

Next, define a *competitive emissions trading equilibrium with respect to* $(\hat{e}_1, \ldots, \hat{e}_n)$ as a vector of national emissions $(\hat{e}'_1, \ldots, \hat{e}'_n)$ and a price $\hat{g} > 0$ for CO_2 (expressed in units of consumption goods per unit of CO_2 emission) such that for each country $i = 1, \ldots, n$,

$$\hat{e}'_i = arg\ max\ [g_i(e_i) + \hat{g}(\hat{e}_i - e_i)] \ , \tag{8}$$

and

$$\sum_{i=1}^{n} \hat{e}'_i = \sum_{i=1}^{n} \hat{e}_i . \tag{9}$$

In such a competitive emissions trading equilibrium, the countries (typically their firms, but conceivably also other economic agents) freely trade in their pollution rights, equal to their emissions quotas $(\hat{e}_1, \ldots, \hat{e}_n)$, at the given price $\hat{g}$, and at that price, demand and supply of pollution rights are equal[23]. The magnitudes $\hat{g}(\hat{e}_i - \hat{e}'_i)$ represent the value, in private goods, of payments for the purchase, at world price $\hat{g}$, of quotas if $(\hat{e}_i - \hat{e}'_i)$, the amount purchased, is negative, or of receipts from the sale of quotas if $(\hat{e}_i - \hat{e}'_i)$, the amount sold, is positive.

Clearly the vector $(\hat{e}'_1, \ldots \hat{e}'_n)$ defined by (8)-(9) is also the one that solves (7), hence (5)-(6) and thereby maximizes world consumption, since

$$\sum_{i=1}^{n} g_i(\hat{e}'_i) + \hat{g} \sum_{i=1}^{n} (\hat{e}_i - \hat{e}'_i) = max \sum_{i=1}^{n} g_i(e_i) \quad \text{subject to} \quad \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} \hat{e}_i .$$

As a confirmation, it can be seen from the first order condition for (8) that at the price $\hat{g}$ the equality $g_i(\hat{e}_i') = \hat{g}$ is satisfied for all i's, implying that $g_i(\hat{e}_i') = g_j(\hat{e}_j')$ for all $i, j = 1, \ldots, n$, a necessary condition for a solution of (5)-(6).

---

[23] Existence and uniqueness of a competitive emissions trading equilibrium follow from concavity of the functions g_i and continuity arguments.

Trading thus allows countries to achieve the aggregate emissions reduction $\hat{e}$ with the highest level of world consumption compatible with this reduction or, in other words, at the lowest opportunity cost for the world. This holds even if for some countries $(\hat{e}_i - \hat{e}_i')$ is negative: the point is indeed that world consumption be maximized and not that all countries necessarily emit $\hat{e}_i'$ lower than $\hat{e}_i$.

5.3 Coalitional stability of competitive trading

If trade in emissions is allowed another question arises: shall there be blocks of countries forming in emissions trading? We answer the question in this section by means of a simple argument based on the theory of market games.

Let $S \subset W$ be a block of countries whose members would decide, given the vector $(\hat{e}_i)_{i \in S}$ of their individual Kyoto quotas, to adopt some joint policy of their own for meeting their aggregate quota, $\sum_{i \in S} \hat{e}_i$, such as *e.g.* trading only among themselves, or engaging in other bilateral or multilateral agreements that fulfill the same condition. To characterize the economic effect of the formation of such a block , define

$$v(S) \;=\; \text{Max}\; \Sigma_{i \in S} g_i(e_i) \quad \text{subject to} \quad \Sigma_{i \in S} e_i = \Sigma_{i \in S} \hat{e}_i, \tag{10}$$

that is, is the maximum total gross[24] output that the countries in the block S can hope to jointly achieve given their aggregate emissions constraint.

Consider now again $(\hat{e}'_1, \ldots \hat{e}'_n)$, the world competitive emissions trading equilibrium with respect to $(\hat{e}_1, \ldots, \hat{e}_n)$. If we can show that the members of S are better off at that worldwide competitive equilibrium than at their best actions as a separate block (as identified in (10)), we shall have established that block S has no interest to form, thus answering in the negative the question raised in this section.

This is in fact straightforward. Indeed, with our notation it amounts to show that

$$\Sigma_{i \in S} g_i(\hat{e}_i') \geq v(S), \tag{11}$$

that is, using (8), that

---

[24] We need not subtract damages as they are already fixed in the aggregate by the aggregate emissions constraint.

$$\sum_{i\in S}\left[g_i(\hat{e}_i')+\hat{g}\,(\hat{e}_i-\hat{e}_i')\right]\geq \sum_{i\in S} g_i(\tilde{e}_i),$$

where $(\tilde{e}_i)_{i\in S}$ is the solution to (10). Clearly, we have $\Sigma_{i\in S}\hat{e}_i = \Sigma_{i\in S}\tilde{e}_i$. Hence we must show that

$$\sum_{i\in S} g_i(\hat{e}_i')+\hat{g}\,(\sum_{i\in S}\tilde{e}_i-\sum_{i\in S}\hat{e}_i')\geq \sum_{i\in S} g_i(\tilde{e}_i).$$

But this inequality is true since from concavity of g_i we have for each *i* in *S*

$$g_i(\hat{e}'_i)+\hat{g}\,(\tilde{e}_i-\hat{e}'_i)\geq g_i(\tilde{e}_i)\ , \tag{12}$$

using the fact that $\hat{g} = g_i'(\hat{e}_i')$ at the world competitive emissions trading equilibrium[25].

Repeating this argument for any conceivable block of countries leads to the conclusion that no block has an interest to form[26], once a competitive emissions trading equilibrium prevails at the world level.

We have thus shown that the outcome of competitive trade in emissions among the countries cannot be improved upon by the formation of coalitions of countries, such as *e.g.* trading blocks. We are thereby rediscovering — in fact, just applying — a general property of market equilibria known as their "core" property, which says that such equilibria belong to the core of some appropriately formulated cooperative game[27].

---

[25] and irrespective of whether $(\tilde{e}_i - \hat{e}'_i)$ is positive or negative.

[26] Not only no block *S* taken *in the aggregate*, as formulated in (11), but also *each member* of the block, as (12) shows

[27] In technical game theoretical terms, the expression *v(S)* defined above is the payoff that *S* can achieve for its members and any vector $(e_i)_{i\in S}$ that meets the condition $\sum_{i\in S} e_i = \sum_{i\in S}\hat{e}_i$ is an emissions strategy for *S*. Then the pair $[W, v]$ satisfies the definition of a *n*-person game in characteristic function form where v is the function $v : 2^W \varnothing \leftarrow$ defined by (10). A strategy for the grand coalition W, $(e_i)_{i\in W}$, is said to be in the core of this game if for each $S\subset W$, $\Sigma_{i\in S} g_i(e_i)\geq v(S)$. That there exists such a strategy, that is, that the core of our game is non empty can be asserted in general terms by showing that the game is balanced (in the sense of SHAPLEY 1967). But we provide above the same positive answer in an economically more interesting way by exhibiting an actual strategy for W — namely the equilibrium outcome of worldwide competitive emissions trading — that we show to belong to the core. Notice that the cooperative game defined here (and hence its core) is *not* the same as the one proposed in CHANDER and TULKENS 1995,1997. The present game is a pure market game where externalities play no rôle since, once the quotas are fixed, the public good aspect of the problem disappears. One is left with only the private goods-type problem of allocating the emissions between the countries. It is worth pointing out, finally, that the game is one for an economy with production, and not of the usual pure exchange type.

5.4 Desirability of worldwide and competitive trading

While the Kyoto Protocol can be seen as allowing for trading in emissions among the Annex 1 or more parties, it leaves open the questions of the extent and nature of such trading[28]. Economic and game theoretic considerations can be further called upon to resolve these questions.

As to the extent of trading, that is, the number of participants in the trade, market equilibrium theory makes a case in favor of emissions trading *with the largest number of traders* possible. Thus, *worldwide* emissions trading is desirable. This is implied by the previous argument on subgroups, be they trading blocks or any other form of "coalitions". Indeed, if it is not to the benefit of any such subgroup of countries to form and act independently of the other countries, the outcome is also not more beneficial *for these other countries*, if a subgroup were to form This is because their only best actions would be to act also as a subgroup, and for this subgroup the inequality (11) also applies.

We claim on that basis that it is in the world's overall economic interest that non Annex 1 countries, whose emissions are not subject to quotas, be nevertheless allowed to participate in the trading process. The clean development mechanism (CDM) contains provisions to that effect. A policy implication of our claim is that this mechanism be designed so as to make it as open as possible to the largest number of countries. The fact that no quotas was assigned to many countries is irrelevant to the beneficial property, both for the world in general and for those countries in particular, of a worldwide emissions trading equilibrium.

As to the nature of trading, the same body of theory advocates that the institutions governing the trades be designed so as to ensure that they be as *competitive* as possible — competitiveness meaning here that all participants behave as price takers. It is indeed only for markets with that property that efficiency, coalitional stability and worldwide maximal benefits are established.

Regulatory provisions that would result in restricting competitiveness in the emissions trading process are thus to be avoided, just as well as the absence of regulations designed to prevent restrictions to competition. Such are, for instance, provisions allowing for market power to be exerted by some traders so as to influence price formation to their advantage, as well as regulatory

[28] To be addressed at the Conference of Parties in Buenos Aires in November 1998.

controls that would impede sufficient price flexibility; or still, as proposed by some, the capping of the quantities that participants are allowed to trade on.

As is well known, the larger the number of participants, the more competitive the market is likely to be: our argument favoring a large extent of the market is thus also one that favors competition[29]. Large numbers are admittedly neither the only way nor a sufficient condition to ensure the competitive character of a market, but they are a powerful factor.

5.5 A numerical illustration

Using the carbon emissions reduction commitments made by the Annex 1 parties to the Kyoto Protocol, as well as the marginal cost abatement curves generated by MIT's EPPA model (which is a multi-regional, multi-sectoral computable General Equilibrium model of economic activity, energy use and carbon emissions), ELLERMAN and DECAUX 1998 develop a method for estimating quantitatively the outcome in 2010 of various trading regimes, including the world competitive emissions trading equilibrium. They highlight the substantial differences in the outcome of the various trading regimes — confirming our theoretical claim of maximal efficiency of worldwide trading, but they leave open the question of which one might be agreed upon by the parties to the Protocol.

Our analysis above brings an answer to this question, again in favor of world competitive emissions trading, based on showing that strategic behaviour of coalitions of countries cannot be more beneficial to them than worldwide emissions trading. For illustrative purposes we reproduce here (see Table 2) Ellerman and Decaux's estimate of the world competitive emissions trading equilibrium in 2010 and of its price $\hat{g}$ at that time which is US$/ton 24.75.

6. Beyond the Kyoto quotas: towards a world coalitionally stable optimum

The outcome of the competitive emissions trading equilibrium with respect to the Kyoto quotas $(\hat{e}_1, \ldots, \hat{e}_n)$ is described in the last row of Table 2. It is seen that it results into monetary transfers among the countries and equalizes their marginal costs of abatement. This equilibrium thus very much looks like

[29] With large numbers, our previous argument on the rôle of markets to achieve coalitional stability is also reinforced by a central result in economic theory (due to DEBREU and SCARF 1963, elaborating on EDGEWORTH 1881) according to which the *only* coalitionally stable outcome (in our case, the only emissions allocation with that property) is the competitive one.

the worldwide treaty described above (Chander and Tulkens 1997) which also requires transfers among the countries and equalizes their marginal costs of abatement (see (2) and (4)), except for the fact that that treaty leads to the worldwide optimal emissions $(e_1^*, \dots, e_n^*)$ while the Kyoto quotas do not.

This prompts our final question: could an appropriate emission quotas and trading scheme of the Kyoto type nevertheless be used to reach a world optimum with the same coalitional stability property as ensured by the Chander-Tulkens transfers?. The answer is yes, because that optimum can be shown to be equivalent to an emission quotas and trading scheme.

To that effect define quotas $(\hat{e}_1^*, \dots, \hat{e}_n^*)$ from the optimal emissions $(e_1^*, \dots, e_n^*)$ and the reference emissions $(\bar{e}_1, \dots, \bar{e}_n)$ such that for each country i,

$$(\hat{e}_i^* - e_i^*) \sum_{j \in W} d_j = g_i(\bar{e}_i) - g_i(e_i^*) - \frac{d_i}{\Sigma_{j \in W} d_j} \left(\Sigma_{j \in W} g_j(\bar{e}_j) - \Sigma_{j \in W} g_j(e_j^*) \right). \qquad (13)$$

The left hand side of this expression is what country i pays (or receives) if it buys (sells) emission rights in amount $(\hat{e}_i^* - e_i^*)$ at a price $\hat{g} = \sum_{j \in W} d_j$. This suggests that $(e_1^*, \dots, e_n^*)$ and $\hat{g} = \sum_{j \in W} d_j$ are nothing else than a competitive emissions trading equilibrium with respect to the quotas $(\hat{e}_1^*, \dots, \hat{e}_n^*)$, in the sense of (8)-(9). And the right hand side is precisely the Chander and Tulkens transfer T_i advocated in Section 3 (see (4)) above to achieve optimality in a coalitionally stable way.

Notice that while the world optimum emissions $(e_1^*, \dots, e_n^*)$ as defined in (2) are independent of the reference emission levels $(\bar{e}_1, \dots, \bar{e}_n)$ as defined in (3), the emission quotas $(\hat{e}_1^*, \dots, \hat{e}_n^*)$ as defined in (13) are not. In fact, since the optimal emissions are independent of reference emissions, there is a one-to-one correspondence between $(\hat{e}_1^*, \dots, \hat{e}_n^*)$ and $(\bar{e}_1, \dots, \bar{e}_n)$. This means that if the reference emission levels $(\bar{e}_1, \dots, \bar{e}_n)$ are not in dispute, then the emission quotas $(\hat{e}_1^*, \dots, \hat{e}_n^*)$ along with competitive emissions trading would also be acceptable to all countries since by definition these would not only lead to the optimum emissions $(e_1^*, \dots, e_n^*)$ but also to transfers that would make each country or group of countries better-off compared to $(\bar{e}_1, \dots, \bar{e}_n)$.

The significance of this shift in perspective lies in the fact that, as noted earlier, the currently considered reference emission levels $(\bar{e}_1, \dots, \bar{e}_n)$ are felt to be unfair, typically by the countries that are in the early steps of economic development with comparatively low emission levels. Therefore, the emissions of such countries may not be subjected to quotas, as agreed upon at Kyoto, at least until the time when their emission levels become comparable to those of Annex 1 countries. With time their emissions will rise as a result of economic development and those of the Annex 1 countries will fall as a result of abatement. While the Kyoto Protocol is a step in the right direction in terms of the actual emissions, we are suggesting here that the effective ultimate aim should in fact be to reach an agreement on appropriate reference emission levels (or pollution rights) $(\bar{e}_1,\ldots,\bar{e}_n)$ at some future round of negotiations.

The discussion in the preceding paragraph clarifies that once an agreement is reached regarding reference emissions $(\bar{e}_1,\ldots,\bar{e}_n)$ then an agreement would also be reached regarding the target emission quotas $(\hat{e}_1^*,\ldots,\hat{e}_n^*)$ and competitive emissions trading which by definition lead to optimum emissions $(e_1^*,\ldots,e_n^*)$ and transfers that ensure coalitional stability.

Such an agreement requires the countries first to agree on equity principles to be adopted, as for instance per capita or per unit of GDP emissions. The currently considered baselines of business-as-usual Nash equilibrium or historically grandfathered emissions are known to be problematic. Something else seems to be required, making explicit room, for instance, to principles like the one of "common but differentiated responsibilities". If such new reference levels can be agreed upon, our analysis suggests that a quotas and trading scheme of the kind pioneered in Kyoto is an appropriate tool to reach stable world optimality in the future.

References

CHANDER, P. and TULKENS, H. 1992, "Theoretical foundations of negotiations and cost sharing in transfrontier pollution problems", *European Economic Review*, 36 (2/3), 288-299 (April).

CHANDER, P. and TULKENS, H. 1995, "A core-theoretic solution for the design of cooperative agreements on transfrontier pollution", *International Tax and Public Finance* 2 (2), 279-294.

CHANDER, P. and TULKENS, H. 1997 "The core of an economy with multilateral environmental externalities", *International Journal of Game Theory* 26, 379-401

DEBREU, G. and SCARF, H. 1963, "A limit theorem on the core of an economy", *International Economic Review* 4, 235-246.

EDGEWORTH, F.Y 1881, *Mathematical Psychics*, Paul Kegan, London.

ELLERMAN, A. D. and DECAUX, A., 1998, "Analysis of Post-Kyoto CO_2 Emissions Trading Using Marginal Abatement Curves", Joint Program on the Science and Policy of Global Change, mimeo, Massachusetts Institute of Technology (August).

Energy Prices and Taxes , 1998.

FANKHAUSER, S. 1995, *Valuing Climate change: The Economics of the Greenhouse*, Earthscan Publications, London

GERMAIN, M., TOINT, Ph., TULKENS, H. and de ZEEUW, A. 1998, "Transfers to sustain cooperation in international stock pollutant control", *CLIMNEG Working Paper* n°6, Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve (also available as CORE *Discussion Paper* n°9832, and *Report* 98/08, Publications du Département de Mathématiques, Facultés universitaires de Namur).

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC) 1995, *Climate Change 1995: The Economic and Social Dimensions of Climate Change,* Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press.

NORDHAUS, W.D. and YANG, Z. 1996, "A Regional Dynamic General Equilibrium Model of Alternative Climate -Change Strategies", *American Economic Review* 86(4), 741-765.

SHAPLEY, L. 1967, On balanced sets and cores", *Naval Research Logistics Quarterly* 14, 453-460.

Table 2 — Ellerman and Decaux characterization of the world competitive emissions trading equilibrium with respect to the Kyoto quotas

	USA	JPN	EEC	OOE	EET	FSU	EEX	CHN	IND	DAE	BRA	ROW	World
Reference noncooperative emissions in 2010 (Mton) $\bar{e}_i$	1838.25	424.24	1063.72	472.04	394.76	873.32*	927.39	1791.96	485.76	308.32	97.27	531.61	9208.63
Kyoto quotas of permitted emissions (Mton) $\hat{e}_i$	1266.67	280.05	756.51	300.66	247.45	873.32	927.39	1791.96	485.76	308.32	97.27	531.61	7866.95
Post-trading emissions reductions (Mton) $\bar{e}_i - \hat{e}'_i$	186.22	12.33	74.96	60.07	52.98	213.36	52.54	447.93	104.87	42.78	2.50	91.07	1341.61
Emission permits (Mton) imported (+)/exported (-) $\hat{e}'_i - \hat{e}_i$	385.36	131.86	232.25	111.31	94.33	-213.36	-52.54	-447.93	-104.87	-42.78	-2.50	-91.07	0.07
Marginal cost of abatement ($/ton) $\hat{\gamma} = g'_i(\hat{e}'_i)$	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75	$ 24.75
Total cost of own abatement ($ billion) $g_i(\bar{e}_i) - g_i(\hat{e}'_i)$	1.77	0.15	0.76	0.44	0.46	0.86	0.57	4.49	1.01	0.47	0.03	0.86	11.86
Cost (+)/receipt (-) of emission permits exports/imports($ billion) $\hat{\gamma}(\hat{e}'_i - \hat{e}_i)$	9.54	3.26	5.75	2.75	2.33	-5.28	-1.30	-11.09	-2.60	-1.06	-0.06	-2.25	0.00

Source: ELLERMAN and DECAUX 1988, Table G.

Annex 1 countries: USA, Japan (JPN), European Union, 12 countries (EEC); other OECD countries (OOE), Eastern Economies in Transition (EET), Former Soviet Union (FSU).

Non Annex 1 countries: Energy Exporting Countries (EEX), China (CHN), India (IND), Dynamic Asian Economies (DAE), Brazil (BRA), Rest of the World (ROW).

For non Annex 1 countries, Kyoto quotas of permitted emissions $\hat{e}_i$ have been taken to be equal to their estimated non cooperative emissions in 2010, *i.e.* $\bar{e}_i$, since it was agreed that their emissions need not be caped in this round of negotiations.

* For FSU, we have taken the reference emissions, $\bar{e}_i$, to be equal to the Kyoto commitment (873.32), although the actual emissions have been estimated to be only (762.79). This is equivalent to giving credit for emission reductions that would happen in any case.