ISSN 0265-8003

STRATEGIES FOR THE INTERNATIONAL PROTECTION OF THE ENVIRONMENT

Carlo Carraro and Domenico Siniscalco

Discussion Paper No. 568
August 1991

Centre for Economic Policy Research
6 Duke of York Street
London SW1Y 6LA
Tel: (44 71) 930 2963

This Discussion Paper is issued under the auspices of the Centre's research programme in **Applied Microeconomics**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Leverhulme Trust, the Esmée Fairbairn Trust, the Baring Foundation, the Bank of England and Citibank; these organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

CEPR Discussion Paper No. 568

August 1991

ABSTRACT

Strategies for the International Protection of the Environment*

This paper provides a general framework for studying the profitability and stability of international agreements to protect the environment in the presence of trans-frontier or global pollution. *N* countries are assumed to bargain on emission control. Each country decides whether or not to coordinate its strategy with other countries. A coalition is formed when both profitability and stability (no free riding) conditions are satisfied. The analysis shows that such coalitions exist but that only a small number of countries decide to cooperate. The paper thus explores the possibility of expanding such coalitions through transfers that induce other countries to cooperate. It is shown that large stable coalitions exist when low environmental interdependence exists and/or when the environmental damage functions are near-separable with respect to domestic and imported emissions. It is also shown that there are cases in which environmental negotiations can achieve substantial emission control even if countries behave non-cooperatively.

JEL classification: 024, 026, 720
Keywords: coalitions, environmental agreements, global commons

Carlo Carraro
Dipartimento di Scienze Economiche
Ca' Foscari
Università degli Studi di Venezia
30123 Venezia
ITALY
Tel: (39 4) 1529 8363

Domenico Siniscalco
Director
Fondazione Eni Enrico Mattei
Via S. Sofia 27
20122 Milano
ITALY
Tel: (39 2) 5203 6950

*We are grateful to K Arrow, K Maler and participants at the IEA conference on 'Managing the International Commons', Siena, 25–27 April 1991, for helpful comments and suggestions on a previous version of this paper. Ken Arrow's remarks also helped to improve the demonstration of Proposition 3.

Submitted 19 June 1991

NON-TECHNICAL SUMMARY

Every year large amounts of pollutants are discharged into the environment as a result of human activity. Some emissions are transported in the atmosphere and in water affecting other countries as well as the global environment. There are many examples of such trans-frontier pollution. The current policy debate centres on atmospheric emissions and the associated damages: carbon dioxide and other greenhouse gases in relation to global warming; CFC in connection with the ozone layer; and sulphur dioxide and nitrous oxide, which are mainly responsible for acid depositions.

The transportation of pollutants in the atmosphere and in water is a source of substantial interdependence among countries: each country receives benefits from using the environment as a receptacle for emissions and is damaged by environmental abuse. While the benefits are related to domestic emissions only, in the presence of trans-frontier and global pollution, the damage is related to both domestic and foreign emissions and to the foreign emissions which reach the country from other countries. The problem is therefore one of international externalities, which in the absence of supra-national institutions can only be solved by international agreements between sovereign countries.

Several difficulties arise, however, because of asymmetries between the countries concerned: attitudes towards the environment differ between countries according to their preferences, their levels of development and their environmental endowments. An additional difficulty is the incentive to free-ride: each country would like to enjoy a cleaner global environment without paying for it. There is therefore an intrinsic instability in any agreement.

This paper presents a game-theoretical framework for analysing the profitability and stability of international agreements to protect the environment in the presence of trans-frontier or global pollution. The players are sovereign countries which interact in a common environment which fulfils many competing functions. The interaction among countries depends on a very small set of crucial parameters related to preferences, to technologies and to the transportation of emissions in the environment. In turn, these parameters are closely related to the specific pollutant involved. This shows that one should not analyse environmental negotiations or policies 'as such', but should place the analysis into a context: the analysis also provides a kind of taxonomy of the different cases, which also helps to explain some conflicting results in the recent literature.

Starting from this framework the main results of our analysis can be summarized as follows:

1) For some pollutants, when the interdependence of emissions is very high, effective environmental protection can be achieved not only by cooperative agreements, but also by non-cooperative emission control. As interdependence decreases, environmental protection can only be achieved by cooperative agreements.

2) In the presence of trans-national transportation of pollutants, cooperation among all countries in environmental policy is profitable *vis-à-vis* non-cooperative behaviour. In the absence of binding agreements, however, the free-rider problem makes full cooperative outcomes unstable.

3) In many cases there exist small coalitions among sub-groups of countries, which are profitable and stable even if they are second best in terms of aggregate emissions. In such cases the interaction among players does not correspond to a Prisoner's Dilemma: other models (chicken or coordination games) can be more appropriate.

4) The gains from the partial cooperation in 3) can be used to expand a coalition by self-financed utility transfers, if environmental policy is backed by other policy instruments and if a minimum degree of commitment is introduced into the game.

To sum up, partial cooperation in environmental policy can be profitable and sometimes stable. The resulting coalitions tend to involve a small number of countries. Wider profitable agreements on pollution control cannot be reached by environmental negotiations alone (if emissions are the only strategic variable). By changing the strategy (including an extra policy instrument such as trade, development or debt policy) and by allowing some change in the rules of the game (for example, allowing a minimum degree of commitment among a small sub-group of players in the original sub-coalition), partial coalitions can be expanded. The paper provides examples of how far a small coalition can be expanded and under what conditions a partial coalition can 'buy' full cooperation.

Our results have some direct policy implications. The common method for reaching agreement on the protection of the global or international environment is comprehensive negotiation. Aside from the Montreal protocol on CFCs, the most important example of this is the current negotiation on climate change, called for by the UN and involving 137 countries and the EC. These negotiations which seek full cooperation are usually very difficult and despite the difficulties, the recent trend has been to make negotiations even more comprehensive.

Our results provide an alternative blueprint for such negotiations: countries (or groups of countries) with stronger environmental preferences should begin by forming small coalitions and then try to 'buy' other countries instead of aiming for a comprehensive agreement at the outset. This blueprint may be relevant if one considers that the key players in environmental negotiations are actually few (say the EC, the US, USSR, China and Brazil) and that some of them appear to be already committed.

In addition, the analysis shows that in some cases non cooperative emission control can be the only effective way and explains, accordingly, why there are many 'agreements' which basically formalize emission reductions which would have been attained without them.

STRATEGIES FOR THE INTERNATIONAL PROTECTION OF THE ENVIRONMENT

Carlo Carraro and Domenico Siniscalco

I.Introduction

Every year a large amount of pollutants is discharged in the environment, as a result of human activity in each country. Some emissions are transported in the atmosphere and in the water and affect other countries as well as global environment. The trans-frontier diffusion of pollutants creates a problem of "international externalities", with negative consequences on welfare and optimality. A further problem confronting policy-makers is that, in the present institutional setting, there is no such an institution to impose super-national regulations to protect the international environment.

This paper explores the possibilities to protect trans-national commons, such as the atmosphere or the oceans, by means of agreements among sovereign countries. These agreements can be cooperative or can simply reflect non-cooperative behaviour. To analyse this issue we characterize international negotiations as bargaining games in which sovereign countries bargain to set emission limits. We show how the alternative outcomes crucially depend on a small set of fundamentals, such as preferences and technology, and on the damage related to the transportation of the different pollutants in the environmental media. The analysis makes it possible to re-interpret and systematize some recent and often controversial results set forth in the literature with reference to specific pollutants, such as carbon dioxide, sulphur dioxide, and CFC (Cf. Barrett, 1989,1990; Bohm, 1990; Hoel, 1990, 1991; Kaitala-Pohjola-Tahvonen, 1990; Maler,1989, 1990;

Newbery, 1990; Van Der Ploeg–De Zeeuw, 1990). In addition it explores new strategies to protect the international or the global environment.

The main conclusions can be summarized as follows:

i) the strategic interaction among countries in a common environment does not lead necessarily to the "tragedy of commons", but there is a full range of voluntary agreements to control emissions, even if cooperation among all countries is typically unstable;

ii) beyond non–cooperative emission control, in particular, there exist partial cooperative agreements among sub–group of countries which are not undermined by free riding;

iii) the partial coalitions mentioned above tend to be small; in such cases, however, gains from partial cooperation can be used to sustain much broader coalitions by inducing other countries to cooperate through self–financed utility transfers. To sustain broader coalitions, however, two conditions must be met: environmental policy must be backed by other policy instruments, and a minimum degree of commitment must often be introduced into the game.

The main policy implications of the analysis are straightforward: partial cooperation in environmental policy can be profitable and, sometimes, stable. The resulting coalitions tend to involve a small number of countries. These coalitions however can be expanded through appropriate changes in the strategies to sustain environmental cooperation. This route to environmental protection seems to be much simpler to achieve and maintain than a full cooperative agreement among all countries. It could therefore provide an alternative and more pragmatic blueprint for environmental negotiations, which usually seek full cooperation.

The paper is divided into four sections. Section two introduces a general framework and defines the main kind of agreements which can lead to pollution control; it also provide three main results on stable coalitions. Section three specialises the results to a class of benefit and damage functions, showing that the various kind of agreements proposed in section two may correspond to possible outcomes in plausible situations; section four, finally, discusses the main results and policy implications of the proposed analysis, toghether with some extensions and scope for furher work.

II. The Analytical Framework

II.1 Players, Payoffs, Strategies

Consider n countries ($n \geq 2$) that interact in a common environment, and bargain over emissions of specific pollutants. Each country i benefits from using the environment as a factor of production and as a receptacle for emissions. Its welfare, however, is negatively affected both by its own emissions x_i and by a given share α_i of other countries' emissions $X-x_i$, where $X=x_1+...+x_i+...+x_n$. Parameter α_i is computed using a pollutant-specific transportation model; α_i can be interpreted as an aggregate parameter derived from a general transportation model (e.g. the RAINS model) that identifies the amount of pollutant a_{ij} transported from country j to country i. We focus on a single, aggregate parameter for simplicity's sake.

Country i's benefit and damage enter a welfare function $P_i(x) = B_i(x_i) - D_i(x_i,\alpha_i(X-x_i))$, where $B_i(x_i)$ denotes benefits arising from the use of environment for production and consumption activities, $D_i(x_i,\alpha_i(X-x_i))$ denotes damages (utility losses) determined by pollution emissions, and $x = (x_1...x_n)$.

Consider the benefit function $B_i(x_i)$: a reduction in pollution, which can be achieved through domestic environmental policies, is costly and reduces benefits. The benefit function, which depends on abatement cost, is country–specific and is related to technology, economic structure, the level of development and environmental endowments. By technology we mean more than the mechanical process of turning inputs into outputs; we mean useful knowledge and experience, institutions and organizational form, norms and values that impinge upon and govern the processes of production and exchange.

The damage function $D_i(x_i,\alpha_i(X-x_i))$ depends on country's perceived effects of emissions of a given pollutant, as well as on the evaluation of such effects. It is thus mainly based on a subjective, country–specific, evaluation of environmental goods. The parameter α_i, $0\leq\alpha_i\leq1$, reflects pollution externalities; it is equal to zero if the pollutant has only local effects, i.e. when foreign emissions do not reach country i; it can be equal to one if domestic and imported emissions have equal weight for country i, as in the case of global pollutants such as greenhouse gases. The specific functional form of $D_i(.)$ can be determined using appropriate models of Environmental Impact Evaluation (EIE). Such models usually contain both the measurement of the relevant physical damages and their evaluation, the two being summarized in an index.

Let δ_i be the maximum level of pollution emission for country i; it is computed by maximising environmental benefits $B_i(x_i)$ without taking into account the associated costs evaluated through the function $D_i(.)$; δ_i is also a measure of country dimension and development.

The "emission game" between the n countries is thus defined by a triple (N,S,P), and by appropriate rules; as usual, $N=\{1...n\}$ is the set of players,

$S=S_1x...xS_n$, where $S_i=[0,\delta_i]$, is the strategy space, $P=(P_1(x)...P_n(x))$ is the payoff vector. Complete information is assumed. Problems arising in the presence of asymmetric information will briefly be mentioned in section IV.

In this context, a country may decide whether or not to cooperate with other countries in order to reduce total emissions (by taking into account reciprocal externalities). Cooperative agreements are assumed not to be binding. As in the actual practice, countries consider one pollutant at a time. The decision whether or not to cooperate is the outcome of a "metagame" in which each country anticipates the choice (cooperative or non–cooperative) of the other countries, and the relative outcomes in terms of emission levels.

We restrict the analysis to one–shot games. Analysing repeated games would be relevant (in terms of additional equilibrium outcomes) only if appropriate trigger or stick/carrot strategies could sustain cooperation as an equilibrium outcome. Emissions, however, can hardly be conceived as a trigger variable which is expanded in response to other countries' defection: emission reduction, in most cases such as CO2 or CFC, involves substantial irreversible investments; expanding emissions as a retaliation, moreover, would generate an environmental damage primarily to the triggering country; finally, emission expansions can hardly be used as a selective punishment. Other effective punishments (e.g. trade protectionism) could be even more costly for the triggering country and therefore not credible. For these reasons we believe that trigger or stick/carrot strategies are of little help in sustaining cooperation, and that the equilibria of the repeated game would coincide with the equilibria of the one–shot game. We thus concentrate our attention on the latter.

Let us solve the game by analysing its outcomes under alternative strategic combinations.

First we assume that countries play simultaneously, and non-cooperatively. In this case, country i's optimal level of emissions is determined by equating marginal benefits and marginal costs, given the emission levels set by the other countries. The solution of the system of first-order conditions determines the Nash equilibria of the game. For simplicity's sake, we assume the equilibrium to be unique.

The Nash equilibrium of the non-cooperative game can also be determined by computing the fixed point of countries best-reply functions. Let $R_i(x)$, $x=(x_1..x_n)$, be country i's best reply function, where $R_i(x) = \{x_i : P_i(x_i, X-x_i) \geq P_i(s_i, X_s-s_i)$, $X_s=x_1+...+x_{i-1}+s_i+x_{i+1}+...+x_n$, for all $s_i \in S_i\}$. The non-cooperative equilibrium x^o is defined by $x^o = R(x^o)$, where $R(x) = (R_1(x)...R_n(x))$.

Alternatively, countries can decide to set emissions cooperatively. In this case, we assume that a bargaining process takes place in order to achieve a Pareto optimal outcome. The bargaining process may lead to the formation of a coalition among j countries, where j goes from 2 (the smaller feasible coalition) to n (when all countries set emissions by taking into account reciprocal externalities). We name full cooperation a coalition formed by n countries.

In this work we determine the cooperative outcome of the game by using the Nash bargaining solution. This is meant to capture a dynamic bargaining process in which countries alternate offers until an agreement is reached, and in which the time-interval between successive offers is arbitrarily short. As argued by Binmore-Rubinstein-Wolinsky(1986), the outcome of such a dynamic bargaining process coincides with the Nash bargaining solution of the one-shot game.

Moreover, we use the non-cooperative equilibrium $x^o = (x^o_1...x^o_n)$ as the threat point of the bargaining process. This means to interpret the alternating offers model as a model in which players face a risk that, if the agreement is delayed, then the opportunity they hope to exploit it jointly may be lost.

More formally, using the Nash bargaining solution corresponds to say that, when j countries reach an agreement, they set emissions in order to maximise the joint product of the difference between $P_i(x)$ and P^o_i, the non-cooperative welfare.

Before setting emission levels each country must therefore decide whether to act cooperatively or not. We model this decision problem by defining a "metagame" in which countries choose between the cooperative and the non cooperative strategy anticipating the outcomes of the related emission game. Most environmental studies model this "metagame" as a one-shot "Prisoner's Dilemma" in which non-cooperation is the dominant strategy. As we show in the next section, there exist many environmental problems that do not correspond to a "Prisoner's Dilemma".

II.2 Profitability and Stability

Let $P_i(j)$ be the welfare obtained by country i when it decides to cooperate, and $Q_i(j)$ be its welfare when it does not join the coalition formed by j countries. Moreover let J be the set of cooperating countries, whereas J^o denotes the set of countries that play non-cooperatively.

Let us suppose, for simplicity, that all countries are symmetric, that is the welfare function $P_i(x)$ is not country specific. We do not therefore index the welfare functions P and Q and their parameters.

The minimum requirement to be imposed for an environmental coalition to be formed is that the welfare of each country signing the cooperative agreement be larger than its welfare under no cooperation. In other words, country i gains from joining the coalition, with respect to the non-cooperative welfare, if $P(j) > P^o$. This leads to:

Definition 1: *A coalition formed by j players is* <u>profitable</u> *if $P(j) > P^o$ for all countries belonging to J.*

This is of course a minimum requirement that may not suffice to induce countries to sign the agreement. As stated in the introduction, the main problem preventing the formation of any coalition is the possibility of free-riding by some countries. The usual explanation is the following: as one country can profit from the reduction of emissions by cooperating countries, it has an incentive to let other countries to sign the cooperative agreement. If all countries are symmetric, no cooperation takes place. In other words, the "metagame" in which countries choose between cooperation and non-cooperation is represented as a Prisoner's Dilemma. In such game, cooperation is profitable, but each country has an incentive to defect once the other countries cooperate. This leads all countries not to cooperate. As we will see, however, this representation of countries' strategic choice may not be appropriate if countries' best-reply function are near-orthogonal.

Let us define the problem formally. For each country, the crucial comparison is between $P(j)$, the payoff it gets if it join the coalition, and $Q(j-1)$, the country's payoff when it chooses not to sign the cooperative agreement. Let us define by $L(j)$ the function denoting country's incentive to defect from a coalition formed by j players, i.e. $L(j) = Q(j-1) - P(j)$. Moreover, let $P(j+1) - Q(j) = -L(j+1)$, be the incentive for a non-cooperating country to join a j coalition (that therefore becomes a j+1 coalition).

Definition 2: *A coalition formed by j players is* <u>stable</u> *if there is no incentive to defect, i.e.* $Q(j-1) - P(j) < 0$, *for all countries belonging to J, and there is no incentive to broaden the coalition, i.e.* $P(j+1) - Q(j) < 0$, *for all countries belonging to* J^o.

This leads directly to:

Proposition 1: *A stable coalition is defined by the largest integer j lower than or equal to* $j^* = \arg \min_j [\Sigma_j L(j)]$

Proof: This proposition can easily be proved by noting that a coalition is stable for all j^* such that $L(j^*) < 0$, and $L(j^*+1) > 0$.

We will later show that for given, sufficiently general, welfare functions, stable coalitions exist. This is not, however, a satisfactory response to the problem of protecting international commons because, as shown in section III, stable coalitions are generally formed by $j^* \leq n$ players, where j^* is a small number, whatever n. We are therefore concerned with the following question: can the j^* players who cooperate expand the coalition through self-financed utility transfers to the remaining players that, by definition, have no incentive to join it?

In order to add one player to a j-coalition, the gain that the j players obtain from moving to a j+1-coalition must be larger than the loss in which incurs the j+1 player by entering it, i.e.

$$(1) \quad \sum_{i \in J} (P_i(j+1) - P_i(j)) > Q_{j+1}(j) - P_{j+1}(j+1)$$

This condition makes it possible to self–finance an enlarged coalition. Is the broadened coalition stable? The j+1 player does not defect if the transfer is larger than $Q_{j+1}(j)-P_{j+1}(j+1)$; however, as by definition of stable coalition, $P_i(j+1) < Q_i(j+1)$, $i\in J$, the j players of the original coalition have an incentive to defect; this incentive is larger because of the transfer to the j+1 player. Hence the j+1 coalition is unstable. This leads to the following conclusion: *utility transfers from countries belonging to a stable coalition to non–cooperating countries cannot be used to expand the initial coalition (because the larger coalition would be unstable), unless the rules of the game are changed.*

Suppose therefore that some countries, but not all, are *committed* to carry on the environmental policy. This obviously changes the rules of the game. The questions to be answered are the following: if all players of the initial, stable, coalition are committed to cooperation, how many other countries can be induced to join the coalition through appropriate monetary transfers from the initial coalition to the newly entered countries? Which is the minimum number of countries that must be committed to cooperation if full cooperation (a situation in which all countries sign the cooperative agreement) is to be achieved? The answer is provided by the following two propositions:

Proposition 2: *If j countries are committed to carry on the cooperative agreement whatever the number of countries in the coalition, and if* $P(j+s)>P(j)$, $Q(j+s)>Q(j)$ *for all positive s,* $s\leq n-j-1$, *then at most r countries can be induced to join the initial coalition, where r is the largest integer satisfying:*

(2) $r < j[P(j+r)-P(j)]/[Q(j+r-1)-P(j+r)] = j/(\Phi_1 - 1)$

where $\Phi_1 = [Q(j+r-1)-P(j)]/[P(j+r)-P(j)]$

Proof: The initial j countries can use their gain from broadening the coalition to finance other countries'cooperation. This gain is $j[P(j+r)-P(j)] > 0$ if $P(j+r)>P(j)$. For transfers to be self–financed, it must be larger then the incentive to defect for the r countries that have to enter the coalition. This is $r[Q(j+r-1)-P(j+r)]$. Hence:

(2') $j[P(j+r)-P(j)] > r[Q(j+r-1)-P(j+r)]$

Moreover, the maximum transfer $j[P(j+r)-P(j)]$ must be larger than the loss that the total entering countries suffer; this is $r[Q(j)-P(j+r)]$, i.e.

(3) $j[P(j+r)-P(j)] > r[(Q(j)-Q(j+r-1))+(Q(j+r-1)-P(j+r))]$

Notice that $(Q(j)-Q(j+r-1))<0$ and $(Q(j+r-1)-P(j+r))>0$. Hence, (2') implies (3). As (2) implies (2'), the proposition is proved. The newly entered countries has no incentive to defect, and gain from joining the coalition; the initial cooperating countries gain from expanding the coalition, and are committed to cooperation. The new equilibrium constitutes a Pareto improvement.

Proposition 3: *If $Q(n-1) > Q(j)$ for all positive $j<n-1$, the* minimum fraction *of countries that must be committed to a cooperative strategy for all n countries to*

cooperate is defined by the lowest ratio j/n such that:

$$(4) \qquad j/n > [Q(n-1)-P(n)]/[Q(n-1)-P(j)] = 1/(1+\Phi_2)$$

where $\Phi_2 = [P(n)-P(j)]/[Q(n-1)-P(n)]$

Proof: Assume j countries are committed to carry on the cooperative strategy whatever the coalition that is formed. Given the assumed bargaining process, their emissions are still a function of the fraction of players in the coalition. Suppose the j players accept to transfer part or all of the gain from moving to a n–coalition to the n–j players that do not cooperate. This transfer should compensate the n–j players for the loss from joining the coalition, and should also offset their incentive to defect from the n–coalition. In order to compensate the n–j players for the loss from joining the n–coalition we must have:

$$(5) \quad j(P(n)-P(j)) > (n-j)(Q(j)-P(n))$$

This condition ensures that the enlarged coalition is self–financed. It can be re–written as:

$$(5') \quad j[Q(j)-P(j)] > n[Q(j)-P(n)]$$

In order to offset the incentive to deviate from the n–coalition the gain P(n) plus the transfer $j(P(n)-P(j))/(n-j)$ must be larger than the defector's welfare Q(n–1), i.e. re–arranging the equation:

(6) $j(Q(n-1)-P(j)) > n(Q(n-1)-P(n))$

which is equivalent to (4). Notice that both sides of the equation are positive. Eq. (6) can be re-written as:

(6') $j[(Q(n-1)-Q(j))+(Q(j)-P(j))] > n[(Q(n-1)-Q(j))+(Q(j)-P(n))]$

Let us show that (6') implies (5'). Assume that (6') holds as an equality and solve it with respect to $Q(j)-P(j)$. Then replace this expression into eq. (5'). We get $(n-j)[Q(n-1)-Q(j)] > 0$ which is satisfied for all positive $j < n-1$. Hence, condition (4) garantees that both the financing condition (5), and the no-defection condition (6) are satisfied. As a consequence, the n-j players joining the initial coalition have no incentive to defect. The initial j players are instead committed to cooperate. Notice that the fraction of players that must be committed to cooperation decreases as the gain from full cooperation increases, and the incentive to deviate from full cooperation decreases. Finally, the move to the n-coalition is a Pareto improvement. This proves the proposition.

There exist another possibility to expand an environmental coalition. As non-cooperating countries gain when the cooperative agreement is broadened (because they receive less emissions), they could use their own additional gains to induce some countries to enter the coalition. Suppose therefore that *non-cooperating countries agree to finance environmental cooperation* (emission reduction in other countries). Which is the largest number of countries that can be induced to join a stable j coalition?

Proposition 4: *A stable* <u>outside supported</u> *coalition formed by j+r players exists if P(j+s)>P(j) and Q(j+s)>Q(j) for all positive s≤n−j−1, and:*

(7) $(j+r)/n < 1/(1+\theta)$ *where* $\theta = [Q(j+r-1)-P(j+r)]/[Q(j+r)-Q(j)]$

Proof: Assume a stable j–coalition. Countries who do not join the coalition gain from financing, through appropriate transfers, a larger coalition, if Q(j+r) –their payoff when the j+r–coalition is formed – less (Q(j+r−1)−P(j+r))(j+r)/(n−j−r) – the transfer to the j+r cooperating countries – is larger than Q(j) – their payoff before broadening the coalition. This is true if (7) holds. Moreover, Q(j+r)−(Q(j+r−1)−P(j+r))(j+r)/(n−j−r) must be larger than P(j+r+1), i.e. no more countries want to join the coalition. This is true if:

(8) $(n-j-r)[Q(j+r)-P(j+r+1)] > (j+r)[Q(j+r-1)-P(j+r)]$

which can be written as:

(8') $n[Q(j+r)-P(j+r+1)] > (j+r)[Q(j+r)-P(j+r)+Q(j+r-1)-P(j+r+1)]$

Comparing (7) and (8'), it is easy to see that (7) implies (8') (and therefore (8)).

We are left with the proof that the players in the j+r coalition have no incentive to defect. This is true if P(j+r) – the welfare when the j+r coalition is formed – plus Q(j+r−1)−P(j+r) – the transfer each cooperating country receives – is not lower than $Q(j+r-1)$ – the welfare that each cooperating country would receive by defecting from the coalition. This implies $Q(j+r-1) \geq Q(j+r-1)$, which obviously holds (we assume that when a country is indifferent between cooperation and defection, it cooperates). Moreover, $P(j+r)+[Q(j+r-1)-P(j+r)]$ is larger than $P(j)$, the welfare that countries in the stable coalition received before its expansion, because $P(j+r)>P(j)$ by assumption and $Q(j+r-1)-P(j+r)>0$ by the instability condition; it is also larger than $Q(j)$, the welfare that countries entering the coalition received before, because $Q(j+r-1)>Q(j)$ by assumption.

As a consequence all players in the j+r coalition do not defect, all players outside the coalition do not want to join it, and the move to a j+r coalition constitutes a Pareto improvement. This completes the proof.

We have thus proposed two ways of expanding a stable coalition. In the first one, we determine the minimum number of countries that must commit themselves to a cooperative behaviour in order to achieve an equilibrium in which more countries cooperate. We call this <u>minimum commitment</u> to cooperation. In the next section, we show that the commitment of few countries may lead to a stable coalition formed by all countries. In the second one, we consider the incentive that non cooperating countries may have to finance other countries' cooperative behaviour (emission reductions). We call this <u>outside support</u> to cooperation. In the next section, we show that with this type of monetary transfers the number of players belonging to a stable coalition may be doubled or even tripled.

Two important remarks must be made in order to clarify some implications of previous results:

Remark 1: When the coalition is sustained by a minimum degree of commitment, and when it is sustained by transfers from non-cooperating to cooperating countries, an additional policy instrument has to be introduced. Utility transfers are indeed impossible and/or inefficient if exclusively based on emission contractions. *Large profitable and stable coalitions can therefore be obtained only if countries bargain over different policy instruments*: some examples could be the coordination of environmental and trade policies, or environmental and debt policies when LDC countries are concerned.

Remark 2: If a stable coalition, however sustained, exists, the metagame in which countries decide whether or not to cooperate is not a Prisoner's Dilemma. Assume countries are symmetric. Assume that a stable coalition is formed: no incentive to defect exists; all countries, however, have an additional incentive not to cooperate: as non-cooperating countries gain from the others' cooperative behaviour, each country has an incentive to let other countries form the coalition. This is not a Prisoner's Dilemma because the situation in which a group of countries cooperate and the others do not cooperate is an equilibrium of the "metagame", as shown by the following 2x2 representation:

Table 1

		Country h	
		C	N
Country i	C	3,3	2,4
	N	4,2	1,1

In this table, C and N denotes the cooperative and non–cooperative strategy respectively, and the payoff pair (y_1,y_2) indicates countries' welfare (the ranking is ordinal). The table represents a situation in which j–1 countries cooperate. A stable coalition is formed by j countries. Countries i and h do not belong to the set of j–1 cooperating players. Both have an incentive to join the coalition (by definition of stability). However, country i's most preferred outcome is the one in which it does not cooperate, but the other country does cooperate, thus forming a stable j coalition. However, if country h does not act cooperatively, country i chooses to cooperate in order to belong to the stable coalition (by definition of stability). Hence non–cooperation is not the dominant strategy.

This game is known as a *chicken game*[1]. There are two equilibria (N,C) and (C,N), but all players, who prefer not to cooperate, have an incentive to convince the others to cooperate. The game has no dominant strategy; each country may commit himself to non–cooperation, thus achieving the worst outcome (N,N). The cooperative outcome (C,C) is not Pareto optimal; it cannot thus be sustained by the usual trigger mechanism in repeated games.

The empasse is solved by the introduction of asymmetries into the game. If countries have different preferences, technology or environmental endowment, it is possible to figure out which countries are likely to form a coalition. For example, in the case of outside supported stable coalitions, countries with higher abatement costs are likely to finance emission reductions in countries with lower abatement

---

[1] In the political science literature, it is currently debated whether the game described above is a chicken game rather than a coordination game. As we want to emphasize the potential instability, at least in the symmetric case, arising from the incentive to commit not to cooperate, we prefer to identify the game as a chicken game.

costs who therefore form the coalition. In the case of stable coalitions with minimum commitment, countries in which environmental policy is part of a package of coordinated policies, or large countries that heavily affect the global environment, are more likely to commit themselves to cooperation, thus attracting other cooperators. If the game were repeated, countries with higher discount rate would be more likely to form a coalition.

III. Reaction Functions and Stable Coalitions

III.1 The Damage Function

In this section we show that there exist *stable coalitions*, *outside supported coalitions*, and *coalitions with minimum commitment* that are formed by more than two players. Moreover, we show that the existence of such coalitions crucially depends on the elasticity of countries' best-reply functions.

For analytical reasons, we restrict ourselves to linear reaction functions. As we will see, however, linearity is not crucial for the results.

Let the benefit function be represented by a concave function which exhibits decreasing returns of environment exploitation. For example:

$$(9)\quad B_i(x_i) = k_i[\delta_i x_i - x^2_i/2] \qquad i=1,\ldots,n$$

When emissions have no cost, the optimal emission level is δ_i, which therefore denotes the maximum level of country i's emissions. This parameter depends on country technology, economic structure, development and environmental endowment; k_i parametrizes total benefit from emissions (the larger k_i, the larger the benefit): it can be seen as a technological parameter.

The damage function is more difficult to specify, being strictly related to the specific pollutant, to adaptation costs, and to country's preferences. We choose the following functional form: $D_i(x_i,\alpha_i(X-x_i)) = ½m_i(x_i + 2\alpha_i(X-x_i))(x_i)^{\phi}$. This function accounts for the previously described features: m_i parametrizes the level of perceived damage from pollution, α_i is the share of imported emissions ($0 \le \alpha_i \le ½$), and ϕ is the degree of separability of the damage function. When $\phi = 0$, the function is separable; as ϕ grows, the negative effects of local emissions are progressively amplified by imported emissions. Notice that, when $\phi = 0$, countries' best reply functions are orthogonal. As ϕ grows, they become increasingly sloped. Parameter ϕ is pollutant specific; it thus assumes the same value for all countries.

For analytical reasons, we consider two cases: in the first one, the damage function is separable (ϕ=0); in the second one, local and imported emissions have negative, equally weighted, multiplicative effects on national welfare (ϕ=1):

Case 1 (ϕ=0): $D_i(x_i,\alpha_i(X-x_i)) = \frac{1}{2}m_i(x_i + 2\alpha_i(X-x_i))$

Case 2 (ϕ=1): $D_i(x_i,\alpha_i(X-x_i)) = \frac{1}{2}m_ix_i(x_i + 2\alpha_i(X-x_i))$

These two cases have different implications. As previously noticed, in Case 1, countries' best-reply functions are orthogonal, whereas in Case 2, countries' best-reply functions are negatively sloped[2]. This implies a different impact of free-riding behaviour on countries' welfare, and, therefore, different conclusions about stability of cooperative environmental agreements.

III.2 Non-cooperative equilibrium

Let us consider first the non-cooperative equilibrium. As a monotone transformation of payoffs does not affect the equilibrium outcomes of the game, we divide country i's payoff function by k_i. In Case 1, country i's best-reply function is:

---

[2] In Case 2, we could have chosen a fully quadratic damage function $2m_i(x_i+\alpha_i(X-x_i))^2$; This would not change our conclusions, because the reaction function would still be linear and negatively sloped.

(10') $x_i = \delta_i - m_i/2k_i \equiv \tau_i \qquad i=1,\dots,n$

whereas, in Case 2,

(10") $x_i = \delta_i/h_i - [\beta_i(n-1)/h_i]x^a_i \qquad i=1,\dots,n$

where $h_i = 1 + m_i/k_i$, $\beta_i = \alpha_i m_i/k_i$, and $x^a_i = (X-x_i)/(n-1)$ is the average emissions of the other countries (which are taken as given when computing the best-reply function).

Eq. (10') denotes a set of n orthogonal best-reply functions, whereas (10") represents a set of negatively sloped linear functions. The slope decreases as α_i and m_i/k_i become small, i.e. countries are less interdependent when imported emissions are low, and when the perceived damage is small with respect to abatement costs. This is a sufficient, non-necessary condition: as Case 1 ($\phi=0$) shows, the best-reply function can be orthogonal even when α_i and m_i/k_i are very large. Figures 1a and 1b show the reaction functions in Case 1 and 2.

More generally, consider a separable damage function $D_i(x) = d_i(x_i) + C_i(\alpha_i(X-x_i))$ and a general benefit function $B_i(x_i)$. If we linearise country i's best-reply function around the equilibrium point, we get

(11) $x_i = c_1/[\frac{1}{2}(B"_i - d"_i)] \qquad i=1,\dots,n$

where the constant c_1 and the second order derivatives $B"_i$, $d"_i$ are computed at the non-cooperative equilibrium x^o. Notice that the linearised best-reply function are orthogonal, whatever the functions $d_i(x_i)$ and $C_i(\alpha_i(X-x_i))$, and whatever the transportation parameter α_i. Case 1 therefore represents a linear approximation of a more general case with separable damage functions.

Consider now a non-separable damage function $D_i(x_i,\alpha_i(X-x_i))$ and a general benefit function $B_i(x_i)$. Linearizing country i's best reply function around the equilibrium vector x^o, we get:

$$(12) \qquad x_i = \frac{c_2}{B''_i - [\delta D'_i/\delta x_i]} + \frac{\alpha_i(n-1)[\delta D'_i/\delta x^a_i]}{B''_i - [\delta D'_i/\delta x_i]}\, x^a_i$$

where $D'_i = \delta D(x_i,\alpha_i(X-x_i))/\delta x_i$, and the constant c_2 and the second derivatives $[\delta D'_i/\delta x_i]$, $[\delta D'_i/\delta x^a_i]$ are computed at the non-cooperative equilibrium. If $[\delta D'_i/\delta x^a_i]$ is negative and $B''_i - [\delta D'_i/\delta x_i]$ is positive, we obtain a set of negatively-sloped linerarised best-reply functions, of which eqs (8") are just a particular case. Case 2 is therefore a linear approximation of a genaral case with non-separable damage function where $(m_i/k_i)/(1+m_i/k_i)$ is equal to $\{[\delta D'_i/\delta x^a_i]/[B''_i - \delta D'_i/\delta x_i]\}$. Notice that, in the general non-separable case, the slope of the reaction function increases as foreign emissions increasingly affect the home country (α_i grows), as the slope of the marginal benefit function decreases, and as the perceived marginal damage from local and foreign emissions increases.

Having clarified the relationship between the general case and the two examples that we consider for analytical purposes, let us compute the non-cooperative equilibrium of the game. Assume, as a benchmark, that all countries are symmetric, that is the benefit and damage functions are not country-specific. We have:

Case 1.

(13') $x^o_i = \tau \quad P^o_i = \frac{1}{2}\tau^2 - \beta(n-1)\tau \qquad i=1...n$

Case 2.

(13") $x^o_i = \delta/[h+\beta(n-1)] \quad P^o_i = \delta^2 h/[2(h+\beta(n-1))^2] \quad i=1...n$

where P^o_i denotes the equilibrium welfare for each country; the index i is no longer used to identify the parameters of the model. It is easy to check that $\delta > x^o_i > 0$ in both Cases. This implies that emission reductions can be achieved even non-cooperatively, namely a non-cooperative environmental policy can reduce emissions by $(\delta - x^o_i) > 0$. This difference is large whenever the ratio m/k (the relative perceived damage) is large. This result seems to be in line with some recent experience (e.g. the Montreal protocol on CFC) where a non-cooperative reduction of emissions occurred in the presence of low abatement costs, high perceived damage, and high transportation. As we will see, these episodes are better explained as non-cooperative emission reductions rather than as the outcome of cooperative agreements.

III.3 Cooperative outcomes

If a country decides to join a coalition, it sets emissions by maximising the joint product of the difference between $P_j(x)$ and P^o_j, $j \in J$. As previously explained, this could be interpreted as the outcome of a bargaining process in which the time-interval between subsequent offers is arbitrarily small. Country i's (Nash) cooperative emissions, and the relative welfare, are:

Case 1.

(14') $x_i(j) = \tau - \beta(j-1) \quad P_i(j) = P^o_i + \frac{1}{2}\beta^2(j-1)^2 \quad i \in J$

Case 2.

(14") $x_i(j) = \delta(h-\beta)/\theta > 0 \qquad i \in J$

$$P_i(j) = \delta^2(h-\beta)[h^2+\beta h(2j-3)-2\beta^2(j-1)]/2\theta^2 \qquad i \in J$$

where $\theta = h^2 + \beta h(n+j-3)-\beta^2[2(n-1)-j(n-j)]$ can be shown to be positive for all $n,j \geq 2$ and $h > \beta$ (we recall that $h-\beta=1+(1-\alpha)m/k>0$ for all $\alpha \in [0,1/2]$). It is easy to check the following proposition:

Proposition 5: *Countries' maximum welfare is achieved when all players cooperate ($j=n$). In this case the cooperative welfare $P_i(n)$ is larger than the non-cooperative welfare P^o_i, $i=1...n$, for all values of the transportation parameter $\alpha \in [0,1/2]$ and of the cost/benefit parameter $m/k>0$.*

Proof: The first part of the proposition is proved by showing that, in both cases, $P_i(j)$ is an increasing function of j when $j>j^*$. Hence, $P_i(j)$ is maximum for $j=n$ if

$P(n)>P(2)$. In Case 1, $\delta P_i(j)/\delta j = \beta^2(j-1)$, which is positive for $j>j^*\equiv 1$. In Case 2,

$$\delta P_i(j)/\delta j = \delta^2\beta^2[h(j-1)+\beta(3j^2-j(n+4)+2)](h-\beta)^2/\theta^3$$

which is positive for $j > j^* \equiv [\beta(n+4)-h+(h^2-2\beta h(n-2)+\beta^2(n^2+8n-8))^{1/2}]/6\beta$, where $1 < j^* < n$, for all $n\geq 2$. Moreover, it can be easily checked that $P(n)>P(2)$ if $h>\beta$.

The second part of the proposition is shown by looking at:

Case 1.

$$P_i(n)-P^o_i = \tfrac{1}{2}\beta^2(n-1)^2 \qquad i=1...n$$

Case 2.

$$P_i(n)-P^o_i = \delta^2\beta^2(n-1)^2/2[h+2\beta(n-1)][h+\beta(n-1)]^2 \quad i=1...n$$

which are positive for all $\alpha\in[0,1/2]$ and $m/k>0$.

It is important to assess profitability and stability of coalitions formed by j players. Let us denote by $y_i(j)$, $i\in J^o$, the emissions of countries that do not join the coalition. $Q_i(j)$, $i\in J^o$, is their welfare. We have:

Case 1.

(15') $y_i(j) = \tau \quad Q_i(j) = P^o_i + \beta^2 j(j-1) \quad i \in J^o$

Case 2.

(15") $y_i(j) = \delta[h+\beta(j-2)]/\theta > 0$ for all $j \geq 1 \quad i \in J^o$

$$Q_i(j) = \delta^2 h[h+\beta(j-2)]^2/2\theta^2 \qquad i \in J^o$$

Notice that, when j=0 (no player cooperates), eqs. (15) reproduce the non-cooperative equilibrium.

We can also clarify the difference between the two benchmark cases (ϕ=0 and ϕ=1) by comparing the behaviour of players who do not cooperate. This is done by the next Proposition:

Proposition 6: *For all $j \geq 2$ and all n, countries' emissions can be ranked in the following way:*

Case 1.

$$\delta > x^o_i = y_i(j) > x_i(j) > x_i(n) > 0$$

Case 2.

$$\delta > y_i(j) > x^o_i > x_i(j)$$

$x_i(j) > x_i(n)$ for all $2 \leq j \leq n$, iff $h > \beta(n-1)$

Proof. Case 1 is trivial. The ranking for Case 2 can be shown by looking at the following equations:

$$y_i(j)-x^o_i = \delta\beta^2 j(j-1)/\theta(h+\beta(n-1)) > \text{ for } j\geq 2$$

$$x^0_i-x_i(j) = \delta\beta(j-1)[h\beta(n-j-1)]/\theta(h+\beta(n-1)) > 0 \text{ for } j\geq 2$$

$$x_i(j)-x_i(n) = \delta\beta(n-j)(h-\beta j)/\theta(h+\beta(n-1)) \geq 0 \text{ for } h>\beta j$$

Remark 3: This proposition shows that countries that do not join the coalition increase their emissions, with respect to the non-cooperative case, only when the damage function is non separable. This is relevant because in the separable case, cooperating countries do not suffer the damage originating from increased emissions by non-cooperating countries. This makes cooperation more profitable (and stable, as shown later) in the separable case. The problem is further clarified by Figures 2a and 2b that show the "best-reply functions" of an average cooperating country vs. an average non-cooperating countries, in the two Cases. In Case 1, the reaction functions are orthogonal, whereas in Case 2 they are negatively sloped. The four equations are:

Case 1.

$x_i(j) = \tau - \beta(j-1) \qquad y_i(j) = \tau$

Case 2.

$$x_i(j) = [\delta - \beta(n-j)y_i(j)]/[h+2\beta(j-1)]$$

$$y_i(j) = [\delta - \beta j x_i(j)]/[h+2\beta(j-1)]$$

Notice that, in Case 2, the slope decreases as j, the coalition dimension, increases (in Fig. 2, full lines denote reaction curves when j is small, whereas broken lines denote reaction curves when j is close to n). Therefore, the welfare gain that a free-rider can achieve is larger the larger the number of cooperating countries (in other words, the difference between $y_i(j)-x_i(j)$ increases as j increases). Moreover, the two reaction curves tend to be orthogonal as α and m/k become small.

III.4 Coalitions

We now have all information to verify the existence of profitable and stable coalitions in the two benchmark cases we are analysing. The profitability of a coalition can be assessed by comparing $P_i(j)$ and P^o_i, whereas stability can be checked by showing that the sign of $L_i(j) = Q_i(j-1)-P_i(j)$ changes at some $j^* \leq n$.

We show that:

Proposition 7: *In case 1, $P_i(j) > P^o_i$ for all $j > 1$, whereas, in Case 2, $P_i(j) > P^o_i$ if j/n, the share of countries that cooperate, is larger than $s^*(j)$, where*

$$(16)\qquad s^*(j) = \frac{\beta j[h(j+1)-2\beta]}{j(h-\beta)[h(h+2\beta(j-1))]^{\frac{1}{2}}-(h^2-\beta h(j(j-1)+3)+2\beta^2)} > 0$$

Moreover, $\delta s^*(j)/\delta j < 0$, and $\delta s^*(j)/\delta g < 0$, where $g = h/\beta$, $\delta g/\delta\alpha < 0$, $\delta g/\delta m < 0$.

Proof: In Case 1, $P_i(j)-P^o_i = \frac{1}{2}\beta^2 j(j-1) > 0$ for $j>1$, which proves the first part of the Proposition. In Case 2, the numerator of the difference $P_i(j)-P^o_i$ is:

$$(17)\quad n^2\beta^2[2\beta-h(j+1)]-2n\beta[2\beta^2-\beta h(j^2-j+3)+h^2]+[2\beta^3-\beta^2 h(j^2(j+1)-3j+5)+$$
$$+2\beta h^2(j^2-2j+2)+h^3(j-1)]$$

whereas the denominator is positive. Solving eq. (17), we get condition (16). Moreover, $s^*(j)$ is positive because the numerator of (16) is positive for all $j \geq 1$, wherease the denominator is positive because it is positive for $j=2$ and it is a growing function of j for all h and β. Finally, tedious algebra show that $s^*(j)$ decreases as j increases, and as α and m decreases.

Remark 4: In Case 2, Proposition 7 proposes a condition $j/n>s^*(j)$, in which both sides depend on j, the number of cooperating players. This may raise some doubts on the existence of a critical level j^*, $2 \leq j^* < n$, that satisfies $j^*/n = s^*(j^*)$. However, j/n is a growing function of j, whereas $s^*(j)$ is a decreasing function of j. Moreover, 2/n is lower than $s^*(2)$, whereas $n-1/n$ is larger than $s^*(n-1)$. Hence, there exists a unique j^* such that $j^*/n = s^*(j^*)$ (because $s^*(j)$ is monotone); moreover, j^* is smaller as α and m/k become smaller, i.e. as the reaction curves become orthogonal.

Proposition 8: *In both Case 1 and Case 2, there exists a stable coalition. In Case 1, the stable coalition is formed by 3 players, whereas in Case 2 there exists a stable coalition if* $n \leq n^*$, *where*

$$(18) \quad n^* = [5g-2-g^2+2(g-1)(g(g+2))^{1/2}]/(3g-2) \qquad g=h/\beta \geq 2$$

In Case 2, the number of players forming the coalition is 2 for all finite g.

Proof: In Case 1, $Q_i(j-1) - P_i(j) = \frac{1}{2}\beta^2(j-3)(j-1)$ which is equal to zero when $j=3$ (the smallest coalition is formed by two countries. i.e. $j \geq 2$). Hence, there is a unique stable coalition formed by three players. In Case 2, we show the existence of a stable coalition by using Weierstrass theorem. Assume for a moment that j is a real number; $L_i(2) < 0$ if $n<[5g-2-g^2+2(g-1)(g(g+2))^{1/2}]/(3g-2)$, whereas $L_i(n)>0$ for $g\geq 2$. Moreover, the function $L_i(j)$ is continuous. Hence, by Weierstrass theorem, there exists a real number $j^\#$ such that $L_i(j^\#)=0$. The stable coalition is formed by the largest integer j smaller than or equal to $j^\#$. However, $L_i(3) > 0$ for $n > [3g+4-g^2+(g-1)(g(g+4))^{1/2}]/2g$ which is implied by $n\geq 3$, whatever the value of $g<\infty$. Hence, for all $n^*\geq n\geq 3$ there exists a unique stable coalition formed by two players only.

Remark 5: This proposition shows that in both Cases there exists a stable coalition. In Case 2, the existence condition (16) depends on g: in particular, $\delta n^*/\delta g > 0$. As g is a decreasing function of α, this means that a stable coalition is more likely to exists when α is small, i.e. when the reaction functions tend to be orthogonal. The condition is not binding when $\alpha=0$ ($g=\infty$), that is when orthogonality holds. When g is large (above 100), condition (18) can be approximated by the following inequality:

$$(19) \qquad n \leq g/3$$

which clarifies that a stable coalition cannot exists if there are "too many" players in the game, and/or the slope of countries' best-reply functions is "too" negative (the slope absolute value increases as g decreases).

Remark 6: The results of Proposition 7 confirms that the "metagame" in which countries decide whether to cooperate or not is not a Prisoner's Dilemma. However, *the number of cooperating players is very small* (2 or 3); our results cannot therefore be considered a satisfactory solution to the problem of reducing global emissions. We must explore other ways of achieving stable (and large) coalitions by introducing the possibility of utility transfers (through an additional policy instrument) and partial commitment.

III.5 Expanding coalitions.

The first possibility to expand a coalition proposed in the previous section is based on the commitment of some players belonging to the original coalition. What we want to determine is: (i) the largest coalition that can be attained if the j members of the original stable coalition are committed to cooperation; (ii) what is the minimum commitment which is necessary to achieve full cooperation.

This analysis could be relevant when some countries are committed to carry out a cooperative environmental policy (whatever the behaviour of other countries) because environmental cooperation is part of a wider program of cooperative policies (for example, EEC countries).

Consider first the case in which j countries belonging to a stable coalition are committed to cooperation. Using Proposition 2, we compute the following tables, that show the largest attainable coalition in Case 1 and Case 2, as a function of the number of players of the game:

Table 1: Case 1.

Countries	n=10	n=100	n=1000
Coalition	$j+r\leq 7$	$j+r\leq 7$	$j+r\leq 7$

In Case 2, the coalition dimension is also a function of reaction functions' slope, as summarised by the parameter $g \geq 2$:

Table 2: Case 2.

Elasticity	g=2	g=10	g=100
n=10	j+r≤2	j+r≤2	j+r≤4
n=100	j+r≤2	j+r≤2	j+r≤2

Notice that, in Case 1, the largest coalition is independent of n, the number of players of the game; by contrast, in Case 2, as n grows, the incentive to defect becomes larger: this makes it irrelevant the original j countries' commitment and their monetary transfers to other countries. The largest coalition is still formed by two countries only.

Secondly, we consider the minimum degree of commitment that has to be introduced to achieve full cooperation. Using Proposition 3, we get:

Table 3: Case 1.

Countries	n=10	n=100	n=1000
Coalition	j≥5	j≥60	j≥618

In Case 2, the number of committed countries also depends on reaction functions' slope:

Table 4: Case 2.

Elasticity	g=2	g=10	g=100
n=10	j≥9	j≥7	j≥6
n=100	j≥98	j≥91	j≥68

Notice that, unless the reaction curves are orthogonal (Case 1) or near-orthogonal (Case 2, g=100), almost all countries ought to commit themselves to cooperation in order to achieve an equilibrium in which all countries cooperate.

Remark 7: Previous tables show that, in the two benchmark cases, the commitment of j^* countries is sufficient to lead to full cooperation, if the j^* countries are ready to transfer gains from cooperation to the other countries. It is however necessary that: (i) there exists an additional policy instrument with which utility transfers can be carried out; hence, environmental policy should better be studied in connection with other policies; (ii) for the countries committed to cooperation, there exist gains coming from other types of policy coordination that compensate for the loss originating from the monetary transfers and the commitment to cooperation. Again, therefore, environmental policy ought to be analysed considering its relationship with other policy decisions.

The second possibility proposed in the previous section is based on the non-cooperating countries' incentive to finance cooperative emission reductions. This is the case, for example, of countries with large abatement costs that may finance emission reductions in countries with low abatement costs. The possibility exists even when countries are symmetric. Using Proposition 4, we obtain the following tables, that show the largest stable coalition attainable through monetary transfers from non–cooperating countries:

Table 5: Case 1.

Countries	n=10	n=100	n=1000
Coalition	$j+r\leq 7$	$j+r\leq 67$	$j+r\leq 666$

Table 6: Case 2.

Elasticity	g=2	g=10	g=100
n=10	$j+r\leq 2$	$j+r\leq 6$	$j+r\leq 7$
n=100	$j+r\leq 38$	$j+r\leq 45$	$j+r\leq 60$

These results show that outside support to environmental cooperation (monetary transfers from non–cooperating to cooperating countries) can be a powerful instrument to achieve large stable coalitions. Even in the non–separable case (Case 2), large stable coalitions can be attained.

Remark 8: The possibility that some countries finance other countries' cooperative emission reductions can have relevant effects on the global environment. However, even in this case, an additional policy instrument, through which utility transfers are carried out, has to be designed; moreover, feedback effects on environmental coordination ought to be explored, unless this policy instrument is independent of emissions. Finally, the non-cooperating countries (at the environmental level) should coordinate their decisions on the second policy instruments, in order to rule out free-riding on utility transfers.

IV Main Results and Policy Implications

IV.1 Non-cooperative vs. Cooperative Emission Control

The analysis presented so far has been carried out into two steps. The first step introduces the main definitions and concepts by means of a general model, whithout assuming any particular form for the benefit and damage functions. The second step shows that the various kinds of outcomes examined in general terms may correspond to plausible situations.

Consider first non-cooperative emission control. Countries which interact in a common environment with mutual externalities set their emissions by equating their own marginal benefit to marginal damage, given emissions set by the other countries. In this context, country i's actual emissions are lower than emissions δ_i which maximise the benefit; moreover, non-cooperative emissions are increasingly reduced if the transportation of the pollutant is high, the damage is high, the abatement costs are low. The non-cooperative reduction of emissions, in other words, is greater the higher is interdependence, i.e. whenever the best-reply functions are mutually elastic and negatively sloped.

There are reasons to believe that some international protocols to protect the commons are to be obtained (or have been obtained) as non-cooperative outcomes, and that the corresponding negotiations can be seen as pre-play communications in order to reach a Nash equilibrium.

Consider now cooperation. Countries, in this case, bargain over emission levels in order to achieve an optimal aggregate outcome; they set emissions by taking into account reciprocal externalities. As showed in section III, cooperation among all countries is profitable and optimal, but it is intrinsically undermined by free-riding behaviour. If the best-reply functions are orthogonal or near orthogonal, however, there is scope for cooperation, and partial cooperative agreements among small groups of countries can be profitable and stable.

The above considerations suggest that, in environmental negotiations, there is a sort of trade-off. When the best-reply functions are negatively sloped there is a high degree of interdependence and non-cooperative emission control can lead to substantial results. However, if one or more countries unilaterally or cooperatively reduce emissions below the non-cooperative level, this contraction is offset by an expansion by non-cooperating countries. This kind of interaction undermines cooperation, because the free-riding behaviour implies a substantial loss for countries who wish to cooperate.

With orthogonal or near orthogonal best-reply functions the situation is somehow opposite. Non-cooperative emission control leads to small emission reductions, but the scope for cooperation is now greater: if a number of countries cooperatively reduce their emissions, this reduction is not offset by free-riders, who simply enjoy a better environment without paying for it. In this case, gains from partial cooperation can be used to finance various cooperative agreements. We explored two possibilities: first, a small coalition can use the gains from partial cooperation to induce other countries to enter the coalition, by compensating gains from free-riding. In this case, a minimum degree of commitment to cooperation is required. Secondly, a group of non-cooperating countries can subsidize a cooperative reduction of emissions in other countries.

The expanded coalitions, which can be self-financed up to a certain number of players, require that environmental policy be backed by other instruments (e.g. trade or financial policy) to transfer welfare; this requires a change of strategy vis à vis the typical environmental negotiation, where countries bargain only on emissions. In addition to this, expanded coalition may also require a commitment by a certain number of countries; this commitment, which is much less demanding than the commitment by all countries, may even lead to full cooperation.

The choice between cooperation and non-cooperation, as well as the level of emissions in each case, are the key decisions of the game. As we just recalled, they crucially depend on the slope of countries' best-reply functions.

IV.2 The Determinants of Interdependence

As the relative slope of the best-reply functions is so important, let us reconsider, from the economic point of view, the main conditions which lie behind the different cases.

Let us first consider a non-separable damage function, and the related payoff and best-reply functions as in Case 2 of section III. In this case, the reaction functions tend to become mutually elastic if α is large, and/or if $g = h/\beta$ is small (i.e. m/k is large), and/or ϕ is large. The first condition is quite easy to understand: with a high transportation parameter α, a great proportion of foreign emissions hit country i, which therefore changes its action in response to greater foreign emissions.

The second condition (m/k large) needs some discussion. For any given damage, if the abatement cost is very high, country i will respond to a contraction of foreign emissions by expanding its own. By the same token, the higher is the damage the greater will be the contraction in response to greater foreign emissions (and viceversa): hence the high interdependence.

The above examples, of course, are only sufficient conditions, because orthogonal best-reply functions can be obtained even under the third condition (ϕ small): in this case, if ϕ is sufficiently small, the best-reply functions are orthogonal even when α and m/k are very large. This is the case, for example, of a damage function separable in domestic and foreign emissions, as in Case 1 of section III. This specification, which is rather common in the literature (Cf. Maler,1990; Hoel,1990), implies that, at the margin, country i will not respond to other countries' emission variations (each country has a dominant strategy in terms of emissions).

Over and above the examples, it is important to understand that the pattern of environmental interaction among countries depends on the mutual elasticity of the best-reply functions, which determines the interdependence of decisions.

How often can environmental interdependence be captured by orthogonal, rather than negatively sloped, reaction functions? Given the relevance of their slope, we believe that the crucial parameters and functions in a model of environmental policy coordination can only come from serious applied work.

Estimates of the international transportation of pollutants can be obtained with specific models, such as the RAINS model of SO2 and NOX developed by IIASA; they can also be based on judgemental evaluations: in the case of pollutants with only global effects, such as CO2 or CFC, for example, parameter α implies full transnational effect.

Estimates of the abatement costs of various emissions (i.e. the estimate of the parameters entering the benefit function) do not involve particular difficulties, and can be based on single-country or multi-country econometric models, which consider a full range of macro-economic feedbacks. A interesting OECD paper (Holler-Dean-Nicolaisen, 1990) surveys a dozen of recent models which assess the cost of abating greenhouse gases emissions. Beyond the data, the paper is relevant because it contains a very detailed discussion of the problems which can be met in such an assessment. To our purpose, it is sufficient to recall that the benefits from pollution and the associated abatement costs are related to the level of development, the economic structure and, above all, to the specific pollutant involved (suffice it to compare the abatement costs of CFC and CO2: reducing CFC is a matter of changing technologies in sprays and refrigerating devices; reducing CO2 affects energy consumption, which is closely related to development and to life-style).

The most serious difficulties arise in estimating a damage function. On the empirical ground, a damage function can be obtained econometrically, by estimating the first order conditions, under several, rather heroic, assumptions. Alternatively, it can be calculated by means of appropriate environmental evaluation models, which summarize in an index the physical damage of a given pollutant and its evaluation. Taking into consideration several damages, adaptation costs and the like, the indexes produced by ecological models embody a sort of utility function, which simply reflects the model builder's (or the user's) preferences, but are far from being "objective". In this field, therefore, much work still has to be done to merge the economic and the ecological approach.

Despite these difficulties, we believe that a meaningful analysis of international interactions in environmental policy cannot be carried out without empirical data. Only empirical work can justify alternative specifications of countries' interdependence. Only empirical work, moreover, can support the intuition we submit: while in the traditional case of common-property goods (fisheries, pastures, forests, etc.) the best-reply functions are non-orthogonal, in the case of some global pollutants, e.g. CO2, the best-reply functions are probably orthogonal (or near-orthogonal), because the damage function is plausibly nonlinear but separable. In plain words, the damage resulting from various emissions is associated to the total amount of pollutant in country i, as a sum of domestically produced and imported emissions.[3]

This intuition of course needs further investigation on the empirical ground. But if it proves correct, the analysis of cooperative agreements carried out in section III may become relevant for some current policy discussions. The fruitfulness of the analysis is enhanced if one takes a few large groups of countries – such as the EEC, US, USSR, Eastern Europe, China the NICS and the LDCs– as the relevant players. In this case a stable coalition of three committed players could even lead to full cooperation (see section III above).

To elaborate this point, assume that a small group of big players – say the seven big groups mentioned above – negotiate over the reduction of a specific pollutant, say CO2, in the presence of near-orthogonal best-reply functions. Non

---

[3] Non separable effects probably occur if we consider that various pollutants produce combined effects (for example the damage of CO and SO2 is combined). This sort of considerations is usually ruled out in environmental negotiations which consider one pollutant only.

cooperative control, in this case, implies a limited reduction of emissions. There is some scope, however, for partial cooperation. Let three countries form a stable and profitable coalition. If these countries commit themselves to cooperation they can "buy" all the other players and reach full cooperation by self financed utility transfers.

Another possibility is to create coalitions that are sustained by transfers from non–cooperating to cooperating countries. In this case too, utility transfers are viable and efficient only if environmental policy is backed by other policy instruments. This case provides a rationale to the recent European proposal of subsidizing a pollution reduction in some areas, e.g. Eastern Europe or the LDCs, especially if they have relatively lower abatement costs. This proposal has been the object of much analytical work (e.g. Newbery, 1989; Maler, 1990).

It may be interesting to note that the possibility of partial agreements, which is usually discarded in theoretical models, is often advocated in applied policy prooposals (Nitze, 1990). This case gives a rationale to the advocates of a progressive extension of environmental cooperation from Europe, to OECD, to China, Russia and the NICS and the LDCS. To extend cooperation, however, environmental policy ought to be linked to other policies, such as technology transfers, debt policy, development aid, trade policy, etc.

IV.3 Conclusions and scope for further work

In the next few years, the international protection of the environment will increasingly rely on international agreements, which however seem to involve substantial difficulties as they seek to reach a cooperative agreement among a large number of countries. To what extent can the proposed analysis be useful, and how can it be extended?

We believe that our results show that "tragedy" is not the inevitable outcome for the trans-national commons in the absence of a supernational institution; a Leviathan does not seem to be the only way to protect the international environment whenever property rights are not applicable, as it was claimed in the early literature on the commons (Hardin, 1968; Hardin-Baden, 1977; for a recent discussion Cf. McCay-Acheson, 1987; Ostrom, 1990). On the contrary, our analysis shows that there is a range of possible voluntary agreements among sovereign countries to protect the atmosphere and the oceans.[4]

Our model makes it possible to systematize different "models of agreement" which are often discussed in the applied policy literature (e.g. Bohm, 1989; Grubb, 1989; Nitze, 1990). It shows that non-cooperative emission control can be appropriate in some cases of high interdependence. It shows that in other cases they are almost ineffective, but there is room for small coalitions. In the latter case, it explores mechanisms which can sustain and expand cooperative agreements. The message of this paper is that this sort of agreements are much easier to achieve than full cooperation, which is difficult to reach and typically unstable.

If we want to use our framework for policy analysis, however, other problems must be preliminarily solved. First of all, asymmetric players ought to be introduced. The empirical analysis, incidentally, would become much easier if we

---

[4] This does not mean, of course, that a super-national environmental agency is not desirable. On the contrary, cooperative agreements can be seen as ways of setting such an institution.

allowed for countries' asymmetries in preferences, damages and abatement costs. These asymmetries can help explaining who starts a coalition, and who has an incentive to expand it. Secondly, it would be essential to provide a sort of taxonomy relating the various pollutants to appropriate damage functions. Only in this case, will it be possible to contextualize policy analysis, and to obtain meaningful results for each case. Thirdly, it would be necessary to re-appraise the instruments to implement cooperation. Emissions, in many cases, are very difficult to monitor. The various economic instruments to implement an agreement, therefore, must be designed in order to prevent cheating. So far, the literature compared the various agreements in terms of efficiency, i.e. maximum profitability. Our analysis proposes another criterion: an instrument must be efficient, but it must also be effective in preventing or discouraging free-riding. In other words, it must be designed also to promote the stability of the agreements.

Finally, we should work on two extensions. Firstly, asymmetric information. As we mentioned already, preferences cannot be observed. If we remove the assumption of complete information, each country that is induced to enter a coalition would be tempted to overstate the cost and claim for greater incentives. The solution to this problem is to embody an appropriate information or self-selection premium in the incentive to each country that enters the coalition. Secondly, the benefit function must account for the effects of transfers. This prevents the analysis of environmental policy as such, and requires to integrate it in a wider analysis which considers also trade, development and other economic variables in the payoff of each country.

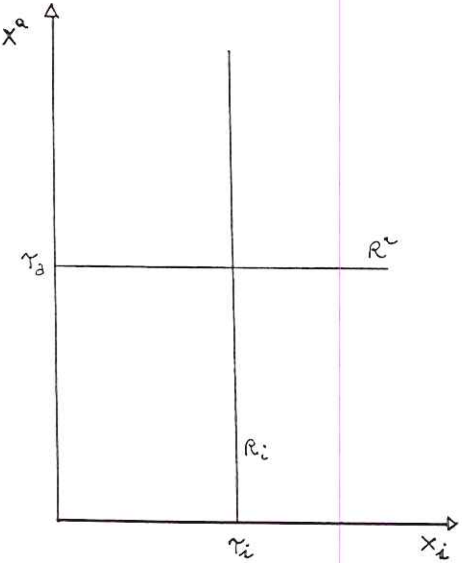

FIG. 1a

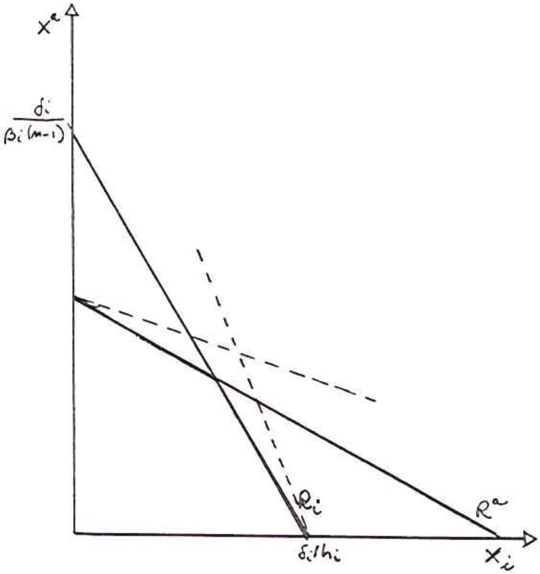

FIG. 1b

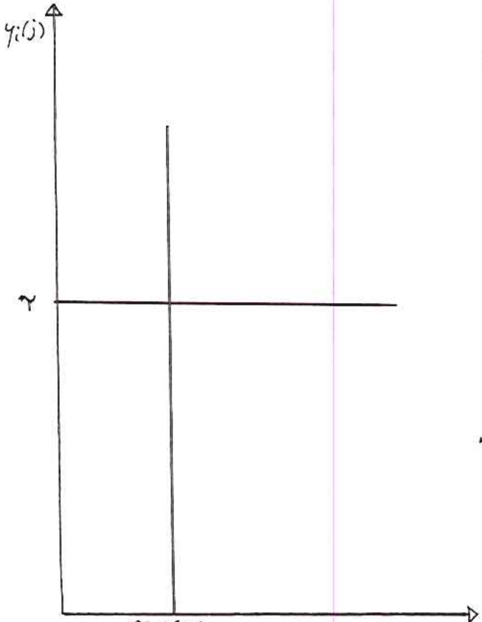

FIG. 2a

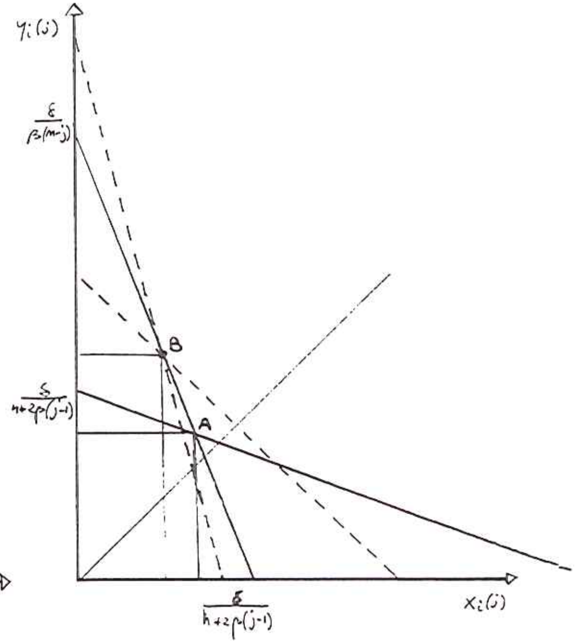

FIG. 2b

REFERENCES

Barrett, S. (1989), "On the Nature and Significance of International Environmental Agreements", University College London, May, mimeo.

Barrett, S. (1990), "International Environmental Agreements as Games", paper presented at the Symposium on Conflicts and Cooperation in Managing Environmental Resources, Siegen, Nov. 1990, mimeo

Binmore, K., Rubinstein, A and A. Wolinski (1986), "The Nash Bargaining Solution in Economic Modelling", Rand Journal of Economics, 17, 176-188.

Bohm, P. (1990), "Efficiency Aspects of Imperfect Treatises on Global Public Bads: Lessons From the Montreal Protocol", mimeo.

Grubb, M. (1989), The Greenhouse Effect: Negotiaing Targets, The Royal Institute of International Affairs", London, pp.56

Hardin, G. (1968) "The Tragedy of Commons", Science, n.162, pp. 1243-1248.

Hardin , G and Baden, J. (1977), Managing the Commons, New York, Freeman & Co, pp. 294.

Hoel, M (1990) "Emission Taxes in a Dynamic Game of CO2 Emissions", Memorandum n 7, University of Oslo, Dept. of Economics, March 90.

Hoel (1991) "Global Environmental Problems: the Effects of Unilateral Actions Taken by One Country", Journal of Environmrental Economics and Management, vol 20, n.1 , pp. 55-70

Holler, P, Dean, A. and J. Nicolaides (1990), "A Survey of Studies of the Costs of Reducing Greenhouse Gas Emissions", OECD Economics and Statistics Department, Working paper n.89.

Kaitala, V. , M. Pojola and O. Tahvonen (1990) "Trans-boundary Air Pollution and Soil Acidification: a Dynamic Analysis of Acid Rain Game Between Finland and the USSR", Helsinki, ETLA Discussion Paper n. 334. Forthcoming in Environmental Resource Economics.

Maler K. G. (1989) "The Acid Rain Game", in H Folmer and E Ierland eds., Valuation Methods and Policy Making in Environmental Economics, New York, Elsevier, 1989.

Maler K. G. (1990) "Environmental Issues in the New Europe" paper presented at the IEA Conference "Economics for the New Europe", Venice, November 1990. Forthcoming in the Conference volume, London, MacMillan, 1991.

Newbery, D. M. (1990) "Acid Rains", Economic Policy, n.1, 1991.

Nitze, W. A. (1990) The Greenhouse Effect: Formulating a Convention, The Royal Institute of International Affairs" London, 66.

Ostrom, E. (1990), Governing the Commons, Cambridge, CUP.

Van Der Ploeg, F. and A. De Zeeuw (1990) "International Aspects of Pollution Control" paper presented at the 1st EAERE Conference, Venice, April, mimeo.