

COOPERATION, STABILITY AND SELF-ENFORCEMENT IN INTERNATIONAL ENVIRONMENTAL AGREEMENTS: A CONCEPTUAL DISCUSSION

by

Parkash CHANDER

Henry TULKENS

National University of Singapore,
Singapore

CORE, Université Catholique de Louvain
Louvain-la-Neuve, Belgium

July 2005

This version: August 2007¹

CORE DISCUSSION PAPER N° 2006/03

ABSTRACT

In essence, any international environmental agreement (IEA) implies cooperation of a form or another. The paper seeks for logical foundations of this. It first deals with how the need for cooperation derives from the public good aspect of the externalities involved, as well as with where the source of cooperation lies in cooperative game theory. In either case, the quest for efficiency is claimed to be at the root of cooperation.

Next, cooperation is considered from the point of view of stability. After recalling the two competing concepts of stability in use in the IEA literature, new insights on the nature of the gamma core in general are given as well as of the Chander-Tulkens solution within the gamma core. Free riding is also evaluated in relation with the alternative forms of stability under scrutiny.

Finally, it is asked whether with the often mentioned virtue of “self enforcement” any conceptual gain is achieved, different from what is meant by efficiency and stability. A skeptical answer is offered, as a reply to Barrett’s (2003) attempt at giving the notion a specific content.

¹ Paper prepared for the David Bradford Memorial Conference “The Design of Climate Policy” held at the Venice Summer Institute organized in San Servolo island by CESifo, Munich, July 22-23, 2005. It will be published as chapter 8 in Roger Guesnerie and Henry Tulkens, eds., *The Design of Climate Policy*, CESifo Seminar Series, The MIT Press, Cambridge MA. The present version was completed during a visit of the second author at the National University of Singapore Department of Economics in March-April 2006. The support of SCAPE at NUS is gratefully acknowledged.

Outline

1. Introduction	p.3
2. Cooperation	p.4
2.1 <i>The economic rationale for cooperation</i>	p.5
2.2 <i>Cooperative games</i>	p.6
2.3 <i>Games with externalities and the core</i>	p.7
2.4 <i>Coalition formation</i>	p.8
2.5 <i>An axiomatic approach</i>	p.11
3. Stability	p.13
3.1 <i>Preliminaries</i>	p.13
3.2 <i>The alternative stability concepts</i>	p.13
3.3 <i>On the nature of γ-core solutions for the IEA game</i>	p.15
3.4 <i>The rationale for a game with a particular coalition structure</i>	p.14
3.5 <i>Free riding and stability</i>	p.18
3.5.1 <i>Two forms of free riding</i>	p.18
3.5.2 <i>NP free riding and I-E free riding vs. γ-core stability</i>	p.20
3.5.3 <i>PR free riding and the particular CT core solution</i>	p.20
4. Self-Enforcement	p.21
5. Conclusion	p.23
References	p.25
Figures 1-4	p.26-29

1. Introduction

This paper is not addressed to game theorists -- unless they are interested to learn something about how their products are being used. The paper is addressed, instead, to those economists who make use of game theory notions in the process of analyzing and advising on climate change negotiations.

In 1995 one of us presented a paper² entitled "Cooperation vs. Free Riding in International Affairs: Two Approaches" where the main question was: can a grand — worldwide — coalition prevail in climate decisions, or is the problem of such a logical structure that treaties involving only small groups of countries will ever be signed? The answer was in the form of an advocacy of the former thesis.

After 10 years, that debate is, to say the least, not closed. Is the present exercise then just a remake? Or has progress been made? While receiving and selecting papers for this conference, we felt that yes, there is progress, but further clarifications are still called for, let alone for ourselves. This motivates our present contribution, whose structure is clear enough from its title.

Let us introduce some notation for easier reference below. With N the set of all countries of the world, indexed $i = 1, \dots, n$, let p_i denote the amount (flow³) of pollutant emissions in country i , let the value of the increasing function (with an upper bound) $g_i(p_i)$, denote the level of country i 's GDP, and let the function $\pi_i(\cdot)$ measure the total cost of damages caused in country i by the aggregate emissions Σp_i .

2. Published thereafter as Tulkens 1998.

3. The specific problems raised by stock externalities will not be considered in this discussion, although such are indeed the externalities generated by greenhouse gas emissions. Our immediate excuse is that they are not dealt with either in the literature we consider. More fundamentally, we think the issues at stake need to be clarified first

In this setting, we shall call a “treaty” a joint choice by several countries of an abatement policy, that is, a level of p_i for each of them, as well as of possible transfers of resources among them. It is also, in general equilibrium terms, a state — or an “allocation” — of the simple international economy specified above.

In the absence of treaty, we assume that each country chooses the policy that suits it best, given the policies of the other countries, this resulting in a state of the international economy which is also a Nash equilibrium of the noncooperative game that can be associated with the above elements.

Efficiency for a group of countries, be it N or any subset S of it, is a joint policy of the members of the group that maximizes the group’s aggregate welfare W . In the case of N , this objective reads

$$W_N = \sum_N [g_i(p_i) - \pi_i(\sum_N p_j)]$$

where all summation signs refer to indexes running from 1 to n . If the group is a (proper) subset S of N however, the maximand is denoted W_S with the first summation in the above expression including only the members of S , whereas the second one still bears on all countries in N . This difference characteristically makes the IEA problem one of externalities.

2. Cooperation

On the theme of cooperation concerning IEAs, we may distinguish two views. One is economic theoretic, the other is game theoretic.

within flow (static) models before being tackled in the dynamic context required by stock externalities.

2.1 The economic rationale for cooperation

The economic view finds its justification in the public good or diffuse characteristic of the externality generated by the emissions that cause climate change. Because the public good is global, that is, world wide, elementary public goods theory (Samuelson 1954) teaches us that efficiency (in the Pareto sense) can be reached only if all concerned parties are involved in the process of resource allocation required to master the externality in question. Thus, getting all parties involved — be it by sharing cost, or by revealing preferences, or both, or still by other means — is an essential requirement for efficiency. Economically, the social objective of efficiency entails the necessity of cooperation. Samuelson saw only the State as an appropriate actor for this purpose.

Independently of the public good characteristic just highlighted for the diffuse externalities here under discussion, another economic argument for cooperation in the sense of engaging in bargaining in the presence of externalities is provided by the Coase theorem. IEAs may be viewed as outcomes of voluntary negotiations between generators and recipients of externalities, as described by Coase 1960. His view concludes at an efficient outcome irrespective of how the rights to exert the externality or not are assigned between the parties. In environmental international affairs, pollutant emitting countries arrogate themselves these rights first, but then spontaneous negotiations strive at an efficient (or at least Pareto superior) outcome . Whether the Coasian argument applies in this context has been recently questioned, however, in a series of papers by Ray and Vohra⁴ who conclude at the possible existence of “robust inefficient outcomes”. The issue is discussed further below.

⁴ Namely Ray and Vohra 1997, 1999 and 2001.

2.2 *Cooperative games*

The game theoretic perspective is the one offered by the theory of cooperative games. This theory flourished in the 60s and 70s prominently within the Jerusalem school of game theory and produced a wealth of “solution concepts” meant to describe the outcome of games (social interactions, in a more recent and better adapted vocabulary) when coalitions of players are the object of analysis. These developments occurred quite independently of public goods and externality theory.

It happens to be difficult, though, to find in this literature arguments explaining and justifying the phenomenon of cooperation. Section 8.1 of Myerson’s 1991 book is entitled “Noncooperative foundations of cooperative game theory” should provide an answer to this query, but the author cautions the reader on how “subtle” the concept is. The attractive idea of giving a noncooperative foundation to cooperative game theory (the “Nash program”) hits at a basic difficulty: the multiplicity of equilibria of the noncooperative games that might support cooperative solutions concepts. Criteria are discussed at length for explaining how selection from among these equilibria might logically occur (focal arbitration of Schelling, institutions, contracts). Somehow, these criteria are one way or another inspired by the notion of efficiency: cooperation finds its *raison d’être* in the efficiency it allows to achieve. It can be given its root in the outcome of some process of bargaining among the cooperating players⁵.

These arguments hardly explain, however, *how* groups are formed, as admitted by the author. At any rate, all game theory textbooks, when they come to their cooperative games chapters (if any), consider that the

⁵. Pushing this view one step farther, some authors consider cooperative games as normative social science, as opposed to noncooperative games being positive science. This is an oversimplification.

theory bears on formed groups taken as given, without enquiring on how they got formed, and why the joint objective of striving for efficiency within the group is attributed to them.

2.3 Games with externalities and the core

Turning to cooperative games for analyzing IEAs is nevertheless quite justified. The theory has indeed provided very compelling arguments in support of competitive market exchanges, arguments pointing to the so-called strategic stability of market equilibria because they belong to the core of cooperative games associated with market exchanges.

It is thus quite natural to ask whether the core concept, if applied to international economies with externalities, can offer similar properties for Coasean agreements between generators and recipients. This question was raised in the early 70s but it was not clearly dealt with all along the 70s and 80s, probably due to imprecise, unrealistic, or *ad hoc* representations of the externality phenomenon itself⁶. Typically, the core theory in these applications was oscillating between results of non existence and problems of non convexities; moreover, the cooperative games under consideration were in fact not really bearing on the multilateral and diffuse form of externalities that is commonly used nowadays and recalled in the above introduction.

It may be argued that this last formulation, called “environmental externalities”, became a standard one in the early 90s because of its appropriateness for dealing with international environmental agreements, in particular due to its close connection with the public good concept. IEAs appeared to be an ideal field of application and their analysis started to develop very quickly at that time, after some early contributions such

⁶ See for instance how widely different are the formulations of externalities by Shapley and Shubik 1969 in their “lake game” and their “garbage game”, or still by Scarf 1972.

as those of Tulkens 1979 and Mäler 1989. This formulation also allowed for game theoretic concepts to yield results in this field. In addition to the Nash equilibrium, used in the two papers just mentioned, the core of a cooperative game was adapted to environmental externalities under the name of “ γ -core “ by Chander and Tulkens 1995 and 1997.

Thus, at least one major concept of cooperative game theory was imported in the IEA literature. One may wonder why, and regret that, other such concepts from the Jerusalem school alluded to above — the bargaining set, the kernel, the nucleolus, the Shapley value or the von Neumann Morgenstern stable sets — have not been similarly more explored in the externalities context⁸.

2.4 Coalition formation

At the beginning of the 90s however, there appeared in the IEA literature (Carraro and Siniscalco 1993, 1995 and Barrett 1994) another category of arguments, bearing on the *formation* of coalitions of countries and inspired from earlier cartel formation models, that some authors later on called a noncooperative approach to IEAs.

This theory is built around the idea that a group (coalition) S forms or does not form depending upon whether or not the payoffs of all players are such that they pass the following two-sided test, called “internal and external stability” :

7. For easier reference in the developments to follow, we briefly remind the reader that the γ -core of a cooperative game with externalities is defined as the core in the usual sense for the γ -characteristic function, which is such that the worth of each coalition S is determined by both the joint payoff maximization of the members of S and the payoff maximizing strategies of the other players assumed to act individually, that is, as singletons. As for each S these simultaneous maximizations induce what the authors call a “partial agreement Nash equilibrium with respect to coalition S ”, the γ -characteristic function is defined over the set of all such partial agreements with respect to coalitions.

(a) $\forall i \in S, W_S^i > W_{S \setminus \{i\}}^i$ (internal stability of S)

and

(b) $\forall i \notin S, W_S^i > W_{S \cup \{i\}}^i$ (external stability of S),

where for any $i \in N$ and any subset $S \subseteq N$, W_S^i denotes the payoff that i obtains when S forms and i is a member of S as in condition (a) or and i is not a member of S as in condition (b)⁹.

The 1995 paper mentioned at the outset criticized this concept because the definition, as just stated, does not make precise how the payoff of player i is determined when i is not in S , namely $W_{S \setminus \{i\}}^i$ in (a) and W_S^i in (b). This has been clarified later on by specifying that players not in a coalition S are assumed to maximize their individual payoffs $g_i(p_i) - \pi_i(\sum_N p_j)$ (that is, to act as singletons), just as the members of S are assumed to maximize their joint payoffs $\sum_{i \in S} [g_i(p_i) - \pi_i(\sum_N p_j)]$. But this is nothing else than what defines a “partial agreement equilibrium with respect to a coalition S ” (PANE wrt S) introduced by Chander and Tulkens 1995-1997 and recalled in footnote 6. Thus, internal-external stability of a coalition S appears to be a property of the PANE with respect to that S .

Further progress has occurred with the introduction in the IEA literature¹⁰ of the notion of *games in partition function form*. With this tool,

8. A notable recent exception is to be found in the work of Van Steenberghe 2004 who deals with the nucleolus and the Shapley value of our externality game, using the g -characteristic function that allowed to define the core.

9. This is reminiscent of von Neumann and Morgenstern “stable sets”, as expounded by Osborne and Rubinstein 1994, p. 279, but is not identical however.

10. See Finus 2001, chapter 15.

the all players set N is split into a family of non overlapping and collectively exhaustive subsets, which defines what is called a *coalition structure*: each partition is such a structure. A coalition structure is an *equilibrium* coalition structure if it is shown to be a Nash equilibrium between the elements of the partition¹¹. Other expressions are “multi-coalitional equilibrium”, or still a “fragmented equilibrium”. If in addition the above internal-external (I-E) stability test is passed by each coalition of the structure, the equilibrium structure is called *I-E stable*. The motivation here is to assert that only coalitions that belong to a stable coalition structure are likely to form.

No analytic conditions¹² ensuring the existence of I-E stable equilibrium coalition structures for the standard IEA model have been provided yet, to the best of our knowledge. However, Eyckmans and Finus 2006a and 2006b have explored the issue by means of numerical simulations with the specific CWS integrated assessment model¹³. They take all conceivable partitions of the set of six regions of the world that the model treats as “countries”, they compute a multi-coalitional equilibrium for each structure and they check for which ones of these structures all coalitions pass the I-E stability test. Similarly, Buchner and Carraro 2005 examine with simulations on the FEEM-RICE model¹⁴ the I-E stability of some conceivable coalition structures.

Having thus taken stock of the state of the art in coalition formation theory, the following question arises. In a multi coalitional equilibrium,

11. Thus, a PANE wrt any S is an equilibrium coalition structure.

12. In games in partition function form, the partition function plays a role similar to the characteristic function in standard cooperative games. Conditions for results should thus hinge on properties of that function. Notice that the γ -characteristic function of Chander and Tulkens is a special case of a partition function, for which the property of balancedness as established by Helm 2001 confirms non emptiness of the core.

13. The CWS model was introduced by Eyckmans and Tulkens 2003.

14. The FEEM-RICE model was introduced in Buonanno, Carraro and Galeotti, 2003.

each coalition S is assumed to achieve efficiency within itself among its members. However, in the standard IEA model here under scrutiny, it is well known that efficiency at the world level can only be achieved by the grand coalition of all countries. While each coalition thus strives for internal efficiency, one could ask why is this quest limited to the members of S ? Why do coalitions not strive for external efficiency, that is, contact other coalitions and adopt mutually beneficial and still more efficient strategies?

If it can be shown that the resulting merged coalition is not I-E stable, then there is an argument for asserting that the merge will not take place. But if it happened to be I-E stable, then why should the merged coalition not form? We would have multiple equilibria, with the merged equilibrium coalition Pareto dominating the equilibrium coalitions it is made of. Giving precedence to the equilibrium with the merge over the one without it would be based on efficiency domination of the former. So, we are driven back to a reasoning on coalition formation essentially led by efficiency considerations¹⁵.

Finally, it should be clear that, no more than what cooperative game theory has to offer, I-E stability criteria do not teach much on the process of *how* stable coalitions are formed.

2.5 An axiomatic approach

On the theme of coalition formation in games with externalities, Maskin 2003 has brought about a contribution based on other arguments. He (courageously) tackled the sequential process of discussions between players on whether or not they will act jointly. His analysis is grounded in an explicit axiomatics that bears (in part) on communication between the

players. One of these axioms specifies that at some point any player is allowed to break communication lines between himself and (some of) the other players. On that basis, the conclusion is derived that the grand coalition will not form.

But doesn't that axiom contain the conclusion? Leaving aside this objection, it is to be praised that Maskin introduces the important factor of communication between the players as a determinant of cooperation. He acknowledges¹⁶, however, that without this axiom, the grand coalition would form in the public good game of his paper (which is very close to the environmental model we deal with in IEA literature), and thus efficiency would prevail.

* * *

To summarize on the complementary themes of cooperation and coalition formation in games with externalities, we have the following: on the one hand, there is a γ -core theory which derives cooperation from the undomination property (in efficiency terms) of the core solution applied to the grand coalition N . On the other hand, (1) the Nash program, being incomplete, cannot explain cooperation; (2) the I-E stability theory only explains the non formation of some coalitions (among which N) and thus only supports partial cooperation; and (3) alternatively, a axiomatic communication breakdown argument attempts to explain the non formation of the grand coalition.

3. Stability

¹⁵. Repeating this reasoning on further mergers might well end up with N as the only coalition!

3.1 Preliminaries.

A first point to be clarified at the outset is the following: stability of what? Of coalitions or of allocations? In many of its formulations in the IEA literature the focus has been more on coalitions than on allocations. This is due, we surmise, to the systematic use of the symmetric players assumption¹⁷ by the authors, which leads them to state their results in terms of a single number, namely the number of signatories, with no mention whatsoever of the ensuing state of the economy or of the environment. The oversimplification of the economic model has made one loose the object of interest. When it comes to derive policy (*i.e.* normative) statements it is not sure that such a slender basis provides a strong enough justification.

Another point to be made is that in the I-E version, the term stability is used in a conceptually very different sense than the one it was given for at least two decades in cooperative game theory with the expressions of “strategic” and “coalitional” stability attached to solution concepts such as the core or the bargaining set. There are thus now two different concepts of stability that we wish to confront here somewhat systematically.

3.2 *The alternative stability concepts*

Let us recall that for a game in general (and an IEA game in particular), the core property of a strategy for all players (respectively, of a proposed

16. Private communication, after the Coalition Theory Network meeting in Paris, January 2005, where the paper was presented and discussed.

17. As well as the rudimentary description of environmental phenomena; but this is acceptable because no model will ever describe reality entirely.

treaty¹⁸ for the coalition of all countries) is that (i) it be Pareto efficient (in terms of countries' emissions) and (ii) that if any individual or group of parties consider deviating from it, the best they can do is less attractive for them than what they get in the said strategy (resp. in the proposed treaty). In the IEA game, if the first condition is met but the second is not, transfers among countries can be devised¹⁹ to ensure that it be fulfilled²⁰. Stability (called strategic or coalitional in this case) is thus a property of robustness of a strategy for all players against the alternatives that any coalition, smaller than N , might look for.

By contrast, I-E stability criteria apply to coalitions of any size, they do not require that the allocation(s) to which they are applied be Pareto efficient and they bear only on individual deviations from any possible coalition. Stability is in this case a property of lesser scope.

18. Incidentally, there is no a priori reason to believe that there is only one allocation (or treaty) that could belong to the core of the IEA game. In other words, the core is not a unique point solution concept, neither in general no in the particular case of IEAs.

19. In the notation of this paper, the Chander-Tulkens 1997 formula for these transfers T_i (>0 if received, <0 if paid) reads:

$$T_i = -(g_i(p_i^*) - g_i(p_i^-)) + \frac{\pi_i'^*}{\sum_{j \in N} \pi_j'^*} \left(\sum_{j \in N} g_j(p_j^*) - \sum_{j \in N} g_j(p_j^-) \right)$$

where p_i^* and p_i^- are, respectively, the world efficient and the Nash equilibrium emission levels of country i and $\pi_i'^*$ is the derivative of the damage cost function $\pi_i(\cdot)$

at the Pareto efficient point $\sum_{j \in N} p_j^*$.

20. It is important always to recall (from Chander, Tulkens, Van Ypersele and Willems 2003, section 5) that the same allocation can be achieved with the transfers being substituted by initial allowances of tradable emission permits, provided that the amounts of these allowances be such that the resulting competitive equilibrium on the permits market induces the g-core allocation just defined. This point is of major importance when discussing the connection between the theories presently examined and actual treaties, such as the Kyoto Protocol for instance, where there are no explicit transfers specified. But the treaty's allowances play the role of the transfers. For a further and thorough exploration of this substitution, see Van Steenberghe 2004.

A further basic difference is in the treatment of deviations: when an individual or a group of parties considers deviating, γ -core theory assumes that the other parties abandon any form of cooperation and act to the best of their interest as singletons — whereas I-E stability theory assumes that the non deviating players keep cooperating among themselves. The rationale for these alternative assumptions will be discussed shortly.

Before doing that, we present with the help of diagrams some apparently unnoticed properties of the γ -core solution, which are independent of the assumptions just mentioned.

3.3 On the nature of γ -core solutions for the IEA game

In the standard multilateral externality model used for dealing with IEAs, each player (country) i is at the same time a polluter and a pollutee²¹. In an effort to disentangle which roles each one of these two functions plays in the determination of the solution, let us consider successively the following elementary, actually unilateral, forms of the model, successively with two, and then three parties.

In the first instance, we have just one polluting country — the polluter, indexed r , which is not polluted, and one polluted country — the pollutee, indexed e , which is not a polluter. Think of a simple upstream-downstream river pollution situation. In the Edgeworth box - type of diagram appearing in Figure 8.1, that one of us introduced in 1974²², the core of this two agents economy consists of all points on the segment A-B if the polluter has the right to pollute (this proviso implying

21. This is why it is called multilateral.

22. See Tulkens and Schoumaker 1975, pp. 247 ff., probably independently redrawn in Varian 1990, pp. 539 and 542. The diagram can be deduced from the simplified version of the IEA model sketched out below the figure. Full details are given in the paper cited.

that point M is the Nash equilibrium in the absence of negotiation); it is also the locus of all allocations that may be reached by Coasean bargaining under the rights allocation just mentioned. This segment A-B is reminiscent of the gain from trade in exchange interpretations of the Edgeworth box. The figure illustrates vividly how the externality is somehow an object of exchange in this setting, yielding what may be called an “ecological surplus”.

Among the core points, the Chander-Tulkens (CT) solution is point A. It is seen to be a Pareto optimum, individually rational with respect to M, and implies a transfer KL such that the polluter is compensated by the pollutee for his abatement cost — but nothing more.

Let us now enlarge this economy with one more pollutee, with the two pollutees being indexed e_1 and e_2 respectively. Figure 8.2 reproduces Figure 8.1 except that the second pollutee’s indifference curve has been added horizontally to the right of the first pollutee’s curve, so that at each point along the resulting curve MN the slope of the tangent measures the *sum* of the marginal rates of substitution between environmental pollution and the numeraire y .

Here, p_r^* is some Pareto efficient level of emission for which the line DE is the core relative to M of the economy and the core point D is the CT solution. At that allocation, the polluter is compensated just for the cost of his abatement, and no more. The bargaining gain (DE) is entirely appropriated by the pollutees as if, in the process, they had acted as a single party.²³

What this illustration makes clear is threefold:

²³. Unfortunately, the picture does not lend itself to show easily how, at the CT solution, the coverage of the polluter’s abatement cost KL is shared between the two pollutees e_1 and e_2 , and thus how the Coasean gain is shared amongst them.

- (i) it shows what the bargaining gain is made of with several pollutees, that is, with several recipients of the externality;
- (ii) it identifies this gain with the core of the game;
- (iii) it shows the particular nature of the CT solution within the core: the polluter is deprived of any pure bargaining gain which goes entirely to the pollutees; however all of his abatement cost is covered.

Note that other core points are conceivable and reachable, all more beneficial to the polluter. If they were reached, it would be out of pure bargaining power between r and the set of e 's; free riding on the part of r would play no role in this respect, for the simple reason that for r , there is nothing to free-ride about!

The relative positions of the players *qua* polluters *vs.* *qua* pollutees in the IEA game, in the core and at the CT solution are further highlighted with diagrams such as those appearing in Figures 8.3 and 8.4. They show how large the γ -core can be as well as the strongly pollutees-favoring character of the CT solution. All of the ecological surplus goes to them. Yet, this is specific to the CT solution: other solutions in the core of the game may benefit the polluters, as suggested by point R on Figure 8.4 where the two polluters succeed to reap part of the bargaining gain (they would reap it all if R was located on the c-d line).

3.4 *The rationale for a game with a particular coalition structure*

As already mentioned in Section 2.4, the γ -characteristic function of an IEA game is a function defined on particular partitions (or coalition structures) of the set N . The question may be raised of why limiting oneself to just that kind of structure and not considering all conceivable partitions? This would transform our IEA game, which is thus far treated

in fact as one in characteristic function form, into a game in partition function form.

We have two reasons for not exploring this extension. One is, as mentioned above, the paucity of results on outcomes, and even of treatment, of such games in the literature²⁴, that we could transpose to our IEA model. The other reason is one of substance: Not all coalition structures can be considered as rational ones, or equally likely to emerge.

Indeed, an argument developed in Chander 2003 establishes that when a coalition forms against a proposed γ -core strategy, it is rational in the sense of an equilibrium strategy for the other players to break up in singletons and thereby induce the defectors to accept the proposed γ -core strategy. Thus, instead of considering mechanically all conceivable structures, taking into account the rationality of the collective behavior of the non-members of S leads one to rather select a well-justified structure.

The said equilibrium strategy is one of a repeated game of coalition formation. Thus, coalition formation theory comes here as a support of the γ -core and the formation of the grand coalition.

3.5 Free riding and stability

3.5.1 Two forms of free riding

Originally, the expression of free riding was used by Samuelson 1954 to describe the behavior of economic agents who conceal their preferences with respect to a public good²⁵ vis-à-vis a single producer — this producer

24. Thrall and Lucas 1963 is an early source, limited to $n \leq 3$.

25. Correct revelation is necessary to be able to check whether efficiency is obtained; but if that information is used to determine the individuals' contributions to the financing of the public good, they will be tempted to understate their preferences and production will be suboptimal, whereas if no connection is made between what they reveal and what

being necessarily the State because of the impossibility of selling the good. On the public good production side, there was no question of leaving or joining coalitions, neither in that paper, nor in the following public goods literature — until the international environmental problems were taken up in the late sixties and early seventies.

Here, the necessarily voluntary character of the provision of the public good, that is, abatement of the environmental externality, together with the fact that the externality is multilateral, shifted the attention from the issue of individual consumers revealing preferences to a planning authority²⁶ to the problem of having several States participate or not in international voluntary agreements on a global externality. The expression of free riding reappeared here not as a preference revelation problem but instead as a way to behave in the face of such agreements²⁷.

There are thus two forms of free riding, that we propose to call “preference revelation (PR) free riding” and “non participatory (NP) free riding”, respectively. Notice that the two forms are not mutually exclusive, but we are not aware of any work that treats them simultaneously. We shall consider here essentially the latter, with occasional allusions to the former.

they have to pay, they will overstate their preferences and production will be larger than optimal.

²⁶. While Samuelson himself in 1954 wrote that only some smart game theorist could master the preference revelation problem raised by the free riding behavior he had identified, the challenge was successfully taken up by game theoretically minded economists fifteen years later in a series of papers written in the context of decentralized planning procedures, starting with Drèze and de la Vallée Poussin 1971, pursued by Roberts 1979, Henry 1979, Groves and Ledyard 1973, and culminating with Champsaur and Laroque 1981. This literature may be seen as one of the main sources of the mechanism design stream of thought that developed subsequently.

²⁷. A third notion of free riding has been put forward by FINUS 2001, namely the behavior that consists in signing an agreement and then not complying with it. We are not sure this wording is appropriate. Non compliance is breaching an agreement that was adhered to. In the two senses described above, free riding is either not signing the agreement, or being part of it under favorable conditions because of an information bias.

3.5.2 NP free riding and I-E vs. γ -core stability

NP free riding is a special form of instability of a group. Depending upon the stability concept one uses, what free riding designates will vary. Thus, when a γ -core allocation is declared not to be internally stable, implying that some i prefers to leave the grand coalition, the non stability statement rests on the assumption that if i leaves N , $N \setminus \{i\}$ remains a coalition (possibly re-optimizing its strategy), and *tolerates* i 's free riding²⁸, that is, it tolerates the global inefficiency induced by i 's defection.

By contrast, the assumption behind the γ -core stability is that $N \setminus \{i\}$ *counters* free riding by reacting, not in an extreme punishing way as it would be the case with the α -characteristic function²⁹ but rather in a way just sufficient for making the free rider realize that he might be put in a situation in which he would prefer what he gets with the grand coalition, as argued in Chander 2003.

The strength of the γ -core concept in dealing with (actually, solving) the NP free rider problem thus lies in the farsighted rationality of the threat it assumes. The weakness of I-E stability is, instead, a myopia that eventually legitimates NP free riding.

3.5.3 PR free riding and the particular CT core solution.

While the CT solution has all the core stability properties just outlined, it allows one in addition to see the effect of a player i joining but incorrectly

28. Eyckmans and Finus 2004 do even reward it, calling it "ideal". Note that to offer that compensation, it is needed to know the preferences of the free rider. Is there any reason to believe that he/she will reveal truthfully while bargaining on a possible defection?

29. The α -characteristic function, which dates back to von Neumann and Morgenstern is, in contrast with the γ -characteristic function, such that for each coalition S the players not in S choose the joint strategy which is the worst for S .

revealing preferences, through the $\frac{\pi_i^{1*}}{\sum_{j \in N} \pi_j^{1*}}$ coefficients in the transfers formula. Understating π_i^{1*} implies a lesser contribution of i to the coverage of the aggregate abatement cost. But that lower value of π_i^{1*} also induces a less than optimal level of aggregate abatement since the optimality criterion is based on the sum of the π_i 's. Thus, the CT solution to the IEA game is vulnerable to PR free riding, at least away from the optimum³⁰.

To conclude, we are back again to the motivation behind seeking stability: from a normative point of view, the reason for avoiding free riding is essentially that it prevents the achievement of efficiency.

4. Self-Enforcement

“Self-enforcement” is an intuitively quite attractive expression, when dealing with international agreements. It evokes the absence of an external authority, which is at the root of the problems raised by this type of agreements. It also contains an implicit reference to incentives. After its introduction by Barrett 1994, the appearance of a book (Barrett 2003) entirely devoted to that idea has positioned the author as its most articulate advocate.

For cooperative game-minded theorists like us, there is a bit of mystery with it: it is difficult to find in the standard literature a commonly received definition of self-enforcement. It does usually not appear in the index of game theory textbooks, and when it does (*e.g.* in Myerson 1991),

30. When the optimum is reached, there is an argument due to Drèze and de la Vallée Poussin 1971 (section 3) establishing that it is a Nash equilibrium of a preference revelation game that all parties do reveal correctly their preferences. Away from the optimum, this is not the case anymore, but the bias in misrepresentation can be identified (see Roberts 1979).

it is only to refer to a property of occasional interest. More importantly, in what sense is self-enforcement more than efficiency, or more than core or I-E stability? Is it an additional concept that we should add to our tool box for IEA analysis?

We feel that while the answer to the last question is definitely yes, the answers to the previous question are difficult to make precise.

Self-enforcement is a property of a treaty that “must satisfy three conditions: individual rationality, collective rationality and fairness” (Barrett 2003, pp. xiii-xiv). Apart from the first one, which is used in its standard sense, the other expressions are given a special meaning. On the one hand, collective rationality is redefined successively in chapters 7 and 11 as a property of a treaty implying not only efficiency for the group under consideration, but in addition free riding deterrence (p. 213)³¹, which is given on p.294 two possible forms (strong and weak collective rationality, respectively). A formal definition is offered in section 11.4, unfortunately with a model of identical players which is hardly convincing. On the other hand, fairness is not formally dealt with, but presented as a requirement that the treaty “be perceived by the parties as being legitimate” (p.xiv).

While potential readers, fond of precise definitions and rigorous developments of sufficiently rich and realistic models, are likely to be sometimes disappointed, the book offers nevertheless a remarkable intellectual challenge to theorists dealing with IEAs.

The one we like to highlight here is the theme of chapter 11, which describes a possible trade-off between the breadth of international cooperation (in terms of the number participants in a treaty) and its depth

³¹. We have responded above to Barrett’s criticism of the γ -core whereby he introduces his collective rationality concept: we claim that the threat he considers as non credible is in fact a farsighted rational one, as proved in Chander 2003.

(in terms of the size of the actions agreed upon by the parties): is a “broad but shallow” treaty better than a “narrow but deep” one?

A shallow treaty would be one that does not achieve full efficiency among the participating countries, *e.g.* by abating less than optimal; this would be the price, so to speak, for having it signed by many countries. The outcome is called by Barrett a “consensus treaty”, asserted elsewhere to be self-enforcing. That this is better than the opposite (deep and narrow) is claimed to be established (p.302) by means of an ingenious symmetric countries model. But we have already voiced the opinion that such a basis is itself quite shallow for transforming into scientific truth this conjecture.

Yet, the trade-off brought to light remains an important intellectual challenge: while it surely deserves scrutiny by means of better adapted, and therefore more elaborate game theoretic tools, it illustrates once more that before proving an idea to be true, it must be generated. This is a major merit of many ideas in Scott Barrett’s book.

Let us close this point with a perhaps timely question: would the David Bradford scheme presented at this conference be self-enforcing?

5. Conclusion

Neither stability nor cooperation are desirable *per se*. Both are there to achieve efficiency, because the welfare of people derives primarily from allocations, not from their stability or from cooperation. The virtues of Barrett’s self-enforcement notion eventually point in the same direction, admitting that otherwise, no treaty would be signed at all.

At a less general level, the analysis revealed that there is much to gain in understanding if one distinguishes more explicitly between the involvement of countries as polluters from their involvement as pollutees.

In fact, this is already done, to some extent, within the Kyoto Protocol: the motivations behind the *aggregate* quotas that have been negotiated are essentially those of the pollutees: they result from their preferences; by contrast, the working of flexible mechanisms is of concern essentially for the polluters. What is less clear, though, is how the bargaining gain turns out to be shared among these two categories of parties.

References

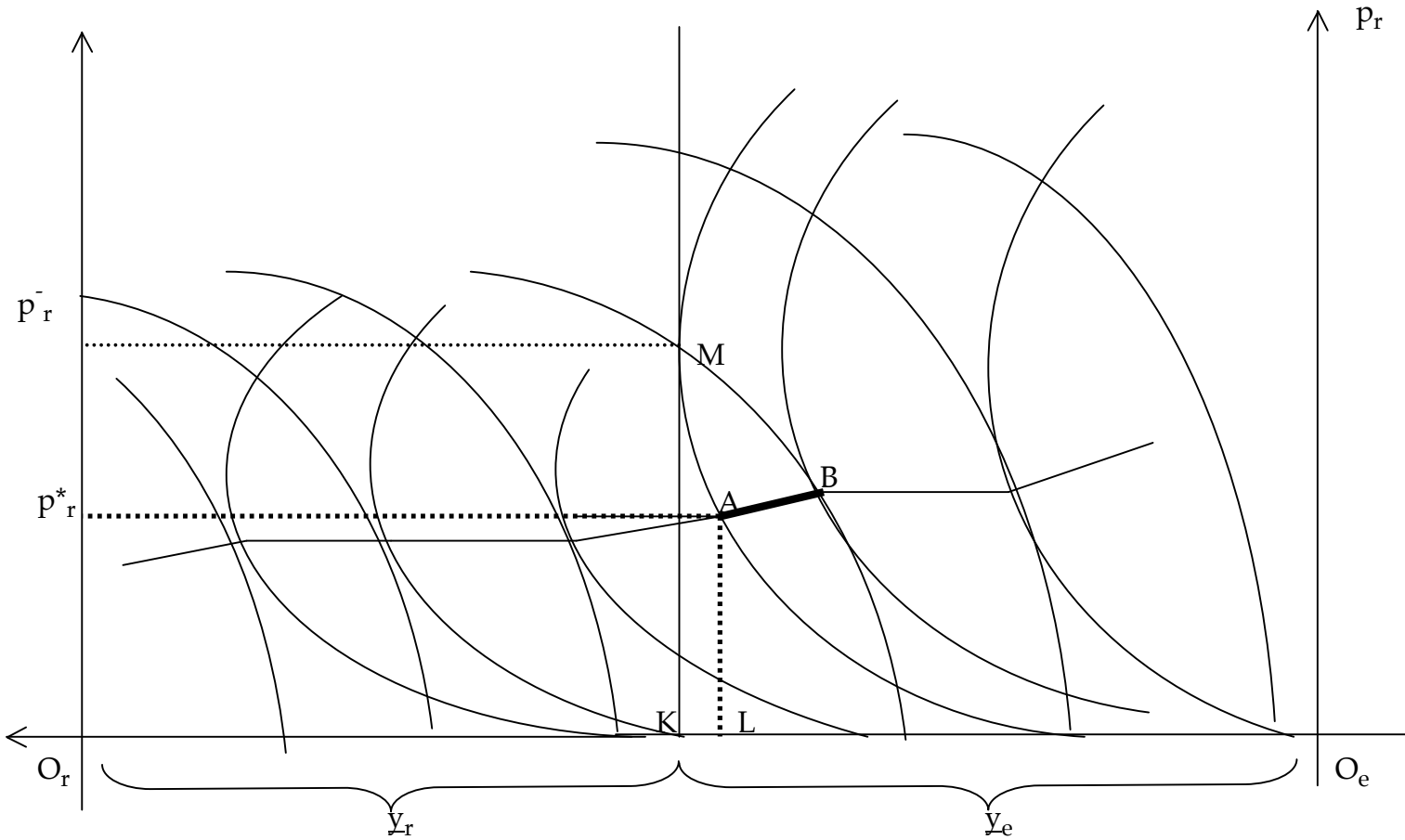
- Barrett, S. 1994, "Self enforcing international environmental agreements", *Oxford Economic Papers* 46, 878-894.
- Barrett, S. 2003, *Environment and Statecraft: The strategy of Environmental Treaty-Making*, Oxford University Press, Oxford.
- Buchner, B. and Carraro, C. 2005, "Climate Blocs: Incentives to Cooperation in International Climate Negotiations", this volume.
- Buonanno, P., Carraro, C., Galeotti, M. 2003, "Endogenous induced technical change and the costs of Kyoto", *Resource and Energy Economics* 25 (2003) 11-34.
- Carraro, C. and Siniscalco, D. 1993, "Strategies for the international protection of the environment", *Journal of Public Economics* 52, 309-328.
- Carraro, C. and Siniscalco, D. 1995, "International coordination of environmental policies and stability of global environmental agreements", chapter 13 in Bovenberg, L. and Cnossen, S. (eds), *Public economics and the environment in an imperfect world*, Kluwer Academic Publishers, Boston, London, Dordrecht.
- Champsaur, P. and Laroque, G 1982, "Strategic behavior in decentralized planning procedures", *Econometrica*, 50 (2), 325 - 344.
- Chander P. 2003, "The γ -core and Coalition Formation", *CORE Discussion Paper* 2003/46; revised version: January 2005. Forthcoming in *International Journal of Game Theory*.
- 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.
- Chander, P., Tulkens, H., Van Ypersele, J-P. and Willems, S. 2002, "The Kyoto Protocol: An Economic and Game Theoretic Interpretation", chapter 6 (pp.98-117) in Kriström, B., Dasgupta P. and Löfgren K.-

- G. (eds), *Economic Theory for the Environment : Essays in Honor of Karl-Göran Mäler*, Edward Elgar, Cheltenham.
- Coase R.M. "The Problem of Social Cost", *Journal of Law and Economics*, 1960 (3), 1-44.
- Drèze, J. and de la Vallée Poussin, D. 1971, "A Tâtonnement Process for Public Goods", *Review of Economic Studies* 38, 133-150.
- Eyckmans, J. and Finus, M. 2004, "An almost ideal sharing scheme for coalition games with externalities", *FEEM Working Paper* No. 155.04.
- Eyckmans, J. and Finus, M. 2006a, "Coalition Formation in a Global Warming Game: How The Design of Protocols Affects The Success of Environmental Treaty-Making", *Natural Resource Modeling* 19 (3), 323-358.
- Eyckmans, J. and Finus, M. 2006b, "New Roads to international Environmental Agreements: The Case of Global Warming", *Environmental Economics and Policy Studies* 7, 391-414.
- Eyckmans, J. and Tulkens, H. 2003, "Simulating coalitionally stable burden sharing agreements for the climate change problem", *Resource and Energy Economics* 25, 299-327.
- Finus, M. 2001, *Game Theory and international Environmental Cooperation*, Edward Elgar, Cheltenham.
- Finus, M and Runshagen, B. 2003 "Endogenous Coalition Formation in Global Pollution Control: A Partition Function Approach », ch. 6, pp. 199-243 in Carraro, C. (ed.), *Endogenous Formation of Economic Coalitions*, Edward Elgar, Cheltenham.
- Groves, T. and Ledyard, J. 1977, "Optimal allocation of public goods: a solution to the 'free rider' problem", *Econometrica*, 45(4), 783-810.
- Helm, C. 2001, "On the existence of a cooperative solution for a coalitional game with externalities", *International Journal of Game Theory*, 30, 141-147.
- Henry, C., 1979, "On the free rider problem in the M.D.P. procedure", *Review of Economic Studies*, 46(2) (Symposium on Incentive Compatibility), 293-303, 1979.

- Mäler, K.G. 1989, "The Acid Rain Game", chapter 12 (pp. 231-252) in H. Folmer et E. Van Ierland (eds), *Valuation Methods and Policy Making in Environmental Economics*, Elsevier, Amsterdam.
- Maskin, E. 2003, "Bargaining, Coalitions and Externalities", presidential address to the Econometric Society European Meeting, Stockholm, mimeo (August).
- Myerson, R. 1991, *Game Theory: Analysis of Conflict*, Harvard university Press, Cambridge, Mass.
- Osborne, M. and Rubinstein, A. 1994, *A Course in Game Theory*, The MIT Press, Cambridge, Mass.
- Ray, D. and Vohra, R. 1997, "Equilibrium binding agreements", *Journal of Economic Theory* 73: 30-78.
- Ray, D. and Vohra, R. 1999, "A theory of endogenous coalition structures", *Games and Economic Behavior* 26: 286-336.
- Ray, D. and Vohra, R. 2001, "Coalitional power and public goods", *Journal of Political Economy* 109 (6): 1355-1384.
- Roberts, D.J. (1979), "Incentives in Planning Procedures for the Provision of Public Goods", *Review of Economic Studies* XLVI (2), 283-292.
- Samuelson P.A. 1954, "The Pure Theory of Public Expenditure", *Review of Economics and Statistics* 36, 387-389.
- Scarf, H. 1971, "On the existence of a cooperative solution for a general class of N-person games", *Journal of Economic Theory* 3, 169-181.
- Shapley, L. and Shubik, M. 1969, "On the core of an economic system with externalities", *American Economic Review* LIX, 678-684.
- Thrall and Lucas 1963, "n-Person Games in Partition Function Form", *Naval Research Logistics Quarterly* 10, 281-298.
- Tulkens, H. 1979, "An Economic Model of International Negotiations Relating to Transfrontier Pollution", chapter 16 (pp. 199-212) in K. Krippendorff, (ed.) *Communication and Control in Society*, Gordon and Breach Science Publishers, New York. Reprinted as chapter 5 (pp.107-122) in P. Chander, J. Drèze, C.K.Lovell and J. Mintz (eds),

Public Goods, Environmental Externalities and Fiscal Competition, Springer, Boston 2006.

- Tulkens, H. 1998, "Cooperation vs. free riding in international environmental affairs: two approaches", chapter 2 (pp. 330-44) in N. Hanley and H. Folmer (eds), *Game Theory and the Environment*, Elgar, Cheltenham.
- Tulkens, H. and Schoumaker, F. 1975, "Stability Analysis of an Effluent Charge and the 'Polluters Pay' Principle", *Journal of Public Economics* 4, 245-269.
- Van Steenberghe, V. 2004, "Core-stable and equitable allocations of greenhouse gas emission permits", *CORE Discussion Paper* 2004/75.
- Varian, H.R. 1990, *Intermediate Microeconomics: A Modern Approach*, second edition, W.W.Norton & Company, New York and London.



Polluter: $u_r(p_r, y_r)$, $y_r \leq \underline{y}_r$ (> 0 : initial endowment)
 with $\partial u_r / \partial p_r \geq 0$, $\partial u_r / \partial y_r > 0$,

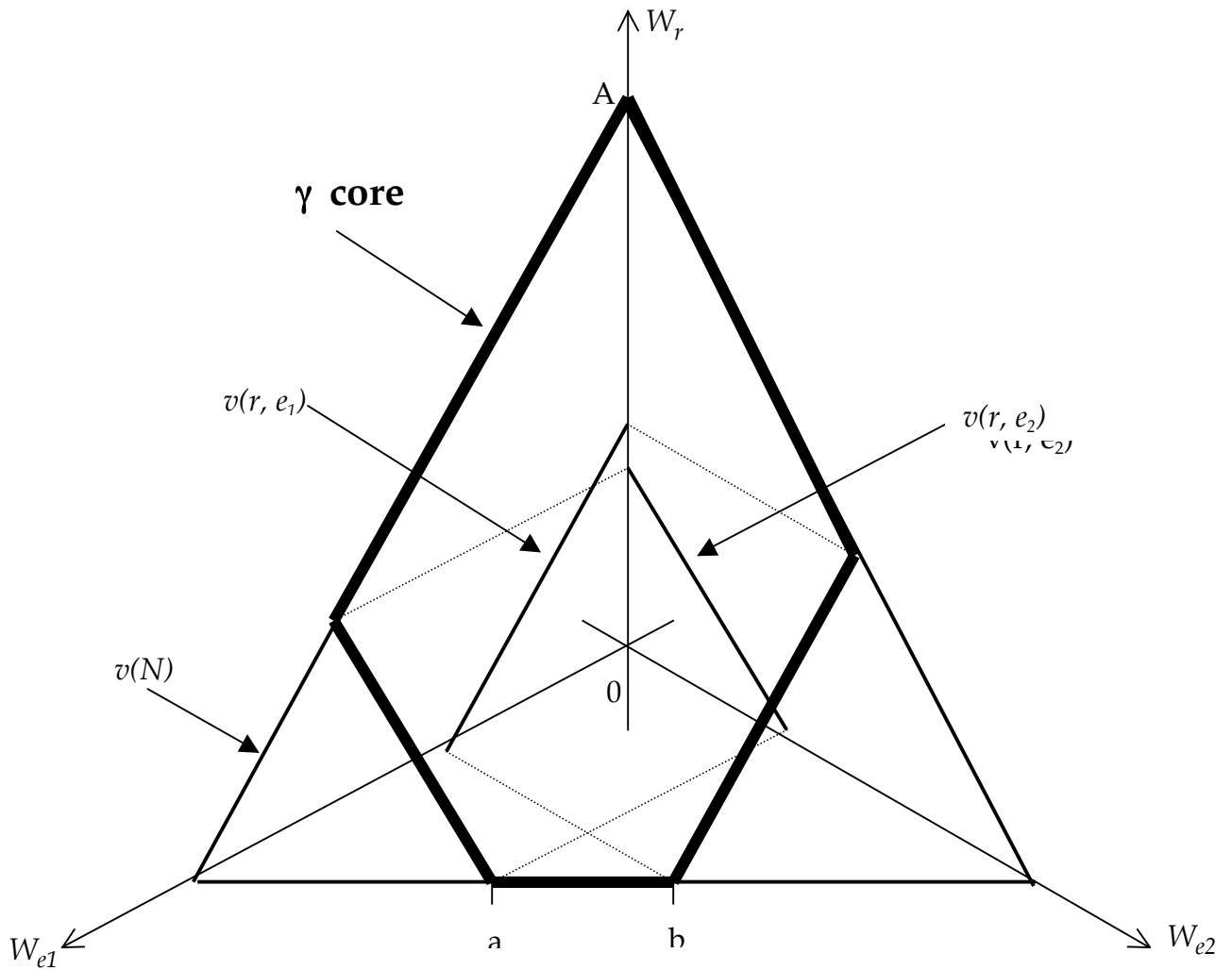
Pollutee: $u_e(p_r, y_e)$, $y_e \leq \underline{y}_e$ (> 0 : initial endowment)
 with $\partial u_e / \partial p_r < 0$, $\partial u_e / \partial y_e > 0$.

M: *Nash equilibrium*

A - B: *core* (with respect to M)

A: *CT solution* (where r receives from e a transfer KL)

Figure 8.1 - A one polluter (r) - one pollutee (e) economy
 Source: TULKENS and SCHOUMAKER 1975



The game is defined by $N = \{r, e_1, e_2\}$ and the characteristic function $v(\cdot)$.

Note that $v(e_1, e_2) = 0$.

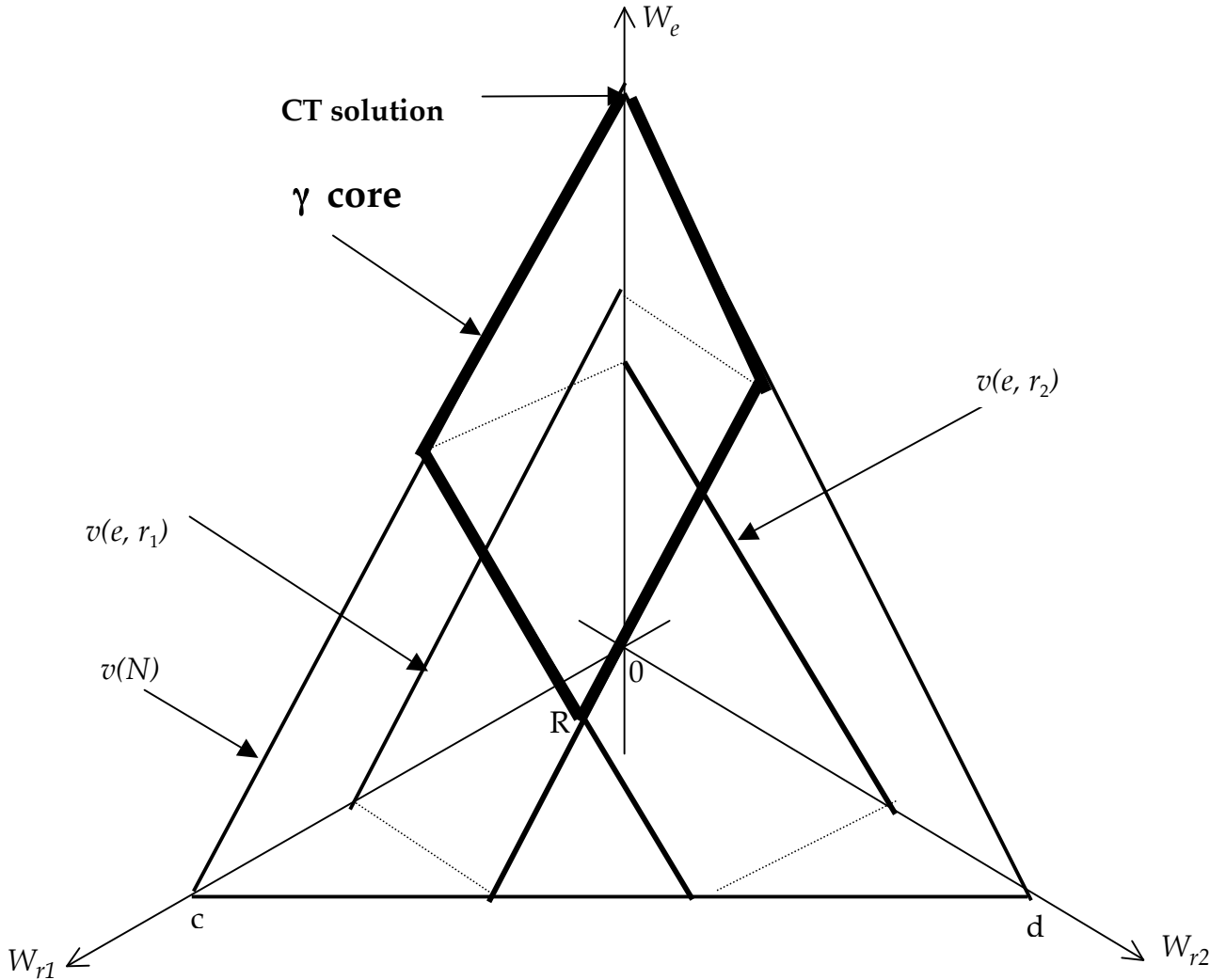
The origin is the welfare levels of players at the Nash equilibrium

The CT solution is one point along the segment $[a, b]$.

There, all the bargaining gain accrues to the pollutees.

That point A belongs to the core illustrates that the (single) polluter can reap all of the bargaining gain

**Figure 8.3 : The γ core in payoffs space
 for any one polluter (r) and two pollutees (e_1, e_2) game game**



The game is defined by $N = \{e, r_1, r_2\}$ and the characteristic function $v(\cdot)$.

Note that $v(r_1, r_2) = 0$.

The origin is the welfare levels of players at the Nash equilibrium

The CT solution is one point on the W_e axis.

There, all the bargaining gain (or ecological surplus) accrues to the (single) pollutee.

All other core solutions give some of the gain to the polluters, down to R

In this example, a solution where the two polluters would reap all of the bargaining gain (*i.e.* a point along $[c, d]$) does not belong to the core.

In general, the stronger (weaker) the coalitions between the polluter and one pollutee, the less (more) the pair of polluters can obtain from the bargaining gain.

Figure 8.4 : The γ core in payoffs space
 for any one polluter (e) and two polluters (r_1, r_2) same