Internal *vs.* Core Coalitional Stability in the Environmental Externality Game: A Reconciliation

by

Henry Tulkens

CORE, Université catholique de Louvain, Belgium

CORE Discussion Paper 2014/58[1]

First version: September 2014

This version: March 2015

Abstract

In a game with positive externalities, such as *e.g.* the standard environmental externality game used in the analysis of international environmental agreements, the solutions having the property of coalitional *internal stability*, when they exist, are compared in this paper with the solutions with the property of *γ-core stability*. Key instruments for that comparison are the notions of stable imputations, on the one hand, and on the other, of partial agreement Nash equilibria relative to a coalition as they result from unacceptable, *i.e.* unstable imputations. The relation between internal and core stable solutions is claimed to be one of compatibility, the former concept complementing the latter in the games where internally stable solutions exist. But this class of games is more restricted than the one for which only γ-core solutions exist. The argument is first presented graphically, then analytically. The relations here exhibited between core and internal forms of stability arouse some concluding thoughts on efficiency, coalitional stability, and on motivations in sharing the surplus generated by cooperation in international environmental issues.

JEL codes: C7, H4, H87, Q5

Keywords: Environmental externalities, game theory, coalitions, core, balancedness, internal stability.

[1] The content of this paper was first presented at the Thursday Seminar of the Fondazione ENI Enrico Mattei (FEEM) in Venice on May 29, 2014. The research was done while being a visiting professor at Università Ca' Foscari Venezia as well as a FEEM Research Associate. It was subsequently pursued at Core and presented there at its Welfare Economics Workshop. The present version was prepared for presentation at the 20th Coalition Theory Network Workshop to be held in Venice in March 2015. The basic intuitions that triggered off this paper were substantially reinforced thanks to insightful suggestions received at Core from my colleagues Thierry Bréchet, Claude d'Aspremont and Maurice Queyranne. I owe a great deal to them. Thanks are also due, for their comments on the earlier version, to Scott Barrett, Emanuele Massetti, Hans Peter Weikard, Carmen Marchiori and François Maniquet. The continuing support from the Chair Lhoist Berghmans "Entreprise, Économie, Environnement" at CORE, UCL of the research backing this work is gratefully acknowledged.

Internal *vs.* Core Coalitional Stability in the Environmental Externality Game: A Reconciliation

Content

Introduction 3

1. The environmental externality game and the coalitional stability issue 3

1.1 The game 3

1.2 The coalitional stability issue 4

2. Graphical representation of coalitional payoffs 5

2.1 Payoffs of coalitions in absence of externality: a reminder 6

2.2 Representing "sensitivity to the coalitional externality" and "outside option payoff" 7

2.3 The PANE wrt S and the P^{γ}-coalitional correspondence 9

2.4 Free riding *vs.* cooperation 10

3. The source of internal coalitional instability: non acceptability of the outside option payoff and the ensuing defection 10

4. Multiple reciprocal externalities and the ISI set 11

4.1 Representing multiple reciprocal externalities and internally stable imputations 11

4.2 Feasibility and the ISI set 12

5. The ISI set and the γ-core 14

6. Internal and core coalitional stability are compatible 15

7. Concluding remarks 18

7.1 On the notion of coalitional stability. 18

7.2 On efficiency, stability and cooperation 19

Appendix: Balancedness of the IS-constrained environmental externality game 21

References 22

Introduction

Two alternative concepts of coalitional stability are dealt with in the game theoretically inspired literature that deals with international environmental agreements: « internal stability » (IS) on the one hand, which was introduced in the field by Carraro and Siniscalco 1992 as well as Barrett 1992, who borrowed it from oligopoly theory[2]. « Core stability » (CS) on the other hand, introduced in the field by Chander and Tulkens (1994, 1995, 1997) who applied, and extended in a specific direction, a concept from classical cooperative game theory[3].

These two concepts were contrasted in only two occasions at the theoretical level[4], without the comparison leading to sufficient respective characterizations to show their interrelation – if any. With a numerical simulation model[5], Bréchet, Gerard and Tulkens 2011 proposed a first such comparison. It confirmed theoretical intuitions, namely the possibility of exhibiting a core stable solution on the one hand, and on the other hand, the frequent impossibility to find an internally stable solution for the grand coalition. But a logical reason for this difference is lacking in that paper.

I claim to offer such an explanation in the present note, showing that, far from being contradictory, the two concepts do belong to a same family, complementing one another. Nevertheless, they are based on quite different views of what determines coalitional stability as well as cooperation. I elaborate and contrast these views in the concluding remarks.

1. The environmental externality game and the coalitional stability issue

1.1 The game

What I call here the standard environmental externality game (henceforth EEG), is stated in coalitional function form originally in Chander and Tulkens 1997. It consists of a set N of players (countries) $i = 1,\ldots,n$, and a coalitional function[6], w^{γ}, that specifies the worth $w^{\gamma}(S)$ of every coalition $S \subseteq N$ as it results from the strategies chosen by the coalition members.

[2] To be found in d'Aspremont and Jaskold Gabszewicz 1986.
[3] The core concept finds its origin in Gillies 1959. For a textbook presentation, see *e.g.* Mas Colell, Green and Whinston 1995.
[4] Namely Tulkens 1998 and Chander and Tulkens 2009, with a discussion in Thoron 2009, p.191.
[5] Namely the CWS model, derived from Nordhaus and Yang's 1996 RICE model.
[6] Called at the time "characteristic" function. I follow here the more recent terminology used, say, by Aumann 1987.

More explicitly, letting the individual and coalitional strategies be denoted by (e_i, x_i), with $0 \le e_i \le e^\circ$, $i = 1,\ldots,n$, (pollutant emissions from production, with some exogenously given upper bound e°) and $0 \le x_i$ (private good consumption in country i, made possible from production, within limits specific to each coalition), the payoffs are concave functions of the form $u_i = x_i - v_i(\sum_{j\in N} e_j)$, with $v_i' > 0$ where the second term (a convex function) represents disutility from aggregate emissions. The game is "with externalities" because the payoff of each player is a function of the other players' strategies in addition to his own, the externality is "environmental" because it affects all players[7], and it is with transferable utility because of the quasi linearity of the payoff functions. These assumptions allow one to specify for each coalition S the coalitional function $w^{\gamma}(S)$ as the sum of the payoffs of the coalition's members.

However, in addition to accounting for the worth of a coalition, the γ-coalitional function can also specify (albeit only implicitly in the original formulation recalled here above) the payoffs that are obtained by the players who are *not* members of S. Indeed, the function is actually defined by using an equilibrium concept assumed to prevail when S forms, namely the "partial agreement Nash equilibrium with respect to S " (PANE wrt S). This equilibrium consists of a strategy profile denoted $\tilde{e}^S = (\tilde{e}_1^S,\ldots,\tilde{e}_n^S)$ that involves *all* players[8], those in the coalition as well as the others: players in the coalition choose strategies that maximize their joint payoffs, whereas the players not in S choose individually, as singletons, their strategy as best reply to the action of all other agents.

1.2 The coalitional stability issue

As a "strategically stable solution" for the game so sketched out, Chander and Tulkens 1995, 1997 propose to use the classical game theoretical concept of core of a game, appropriately modified to fit the specifics of the coalitional function w^{γ}. Thus, the γ-core of the environmental externality game is defined as the set of payoff vectors such that no coalition $S \subseteq N$ can achieve higher payoff levels for its members[9]. The authors exhibit a specific strategy that yields payoffs in the core[10]. By construction, a core solution dominates any

[7] This property of the EEG is shared with what is sometimes called the « public goods game », a terminology that I avoid because there is no production of a public good in the above model but well production of a *private* good. However, the externality (emissions) generated by the process does, like a public good, affect all agents of the economy.

[8] A "PANE wrt S " is a Nash equilibrium of the game between the coalition S that forms (acting as if it were a single player) and the other players acting as singletons. A formal definition of the concept, its properties and the way it generates the γ-coalitional function are given with full details in sections 4.2 and 4.3 of Chander and Tulkens 1997.

[9] A formal definition is given in Section 5 below.

[10] More generally, Helm 2001 shows for the game the qualitative property of balancedness, which implies a non empty core. No uniqueness property is being sought for.

outcome that would involve coalitions smaller than the "grand" coalition of all players and therefore the core property applies only to $S = N$. A first numerical test of the existence result on a numerical integrated assessment model (CWS) has been made in Eyckmans and Tulkens 2003. It has lead to a positive answer, which was confirmed in Bréchet, Gerard and Tulkens 2011.

Parallel to this development, "internal stability of coalitions" in the sense of the theory of cartels initiated by d'Aspremont and Gabszewicz has been called for by Barrett, Carraro and Siniscalco for considering whether in the environmental externality game, any coalition and possibly the grand one, N, would have that property. For any coalition $S \subseteq N$ internal stability is defined by the fact that for each one of its members, the payoff is higher when being a member of it than when staying outside[11], [12]. Analytical existence results of outcomes of the game with internally stable coalitions are obtained only under the assumption of symmetric players, and for coalitions with only a few members ($n \leq 3$). Early numerical tests, also with CWS, were presented in Eyckmans and Finus 2009. They confirmed non existence of internal stability results for the grand coalition in an environmental externality game with $n > 3$.

Over the years, the amount of environmental economics literature devoted to internal stability has been much larger[13] than the one exploring core stability[14]. Yet, on the other hand, the pure game theoretical literature has been rather sulky with respect to the internal stability concept. Are the two concepts contradictory? Are there logical reasons for preferring one to the other? This paper is built on the intuition that they are complementary, in a sense to be made precise below. But the theoretical analysis will show that the class of games for which internal stability of the grand coalition holds is a restricted one, compared to the class where core stability can be shown to hold.

2. Graphical representation of coalitional payoffs

In the hope of gaining in accessibility, I introduce the argument graphically in the case of a three players coalitional game, and extend it subsequently to the $n \geq 3$ players case. To that

---

[11] A formal definition is given in Section 4 below.

[12] The somehow parallel concept of external stability is not considered in this paper, because I wish to focus on stability of N, for which external stability is irrelevant.

[13] A non exhaustive list appears in Finus and Caparros 2015.

[14] Main examples being Van Steenberghe 2003 and 2005, as well as the various dynamic versions of the model reviewed in the introductory part of Germain, Tulkens and Magnus 2010. Cores of voting games are used in Currarini and Tulkens 2003.

effect, a well known graphical technique[15] is used, which has already been applied to the EEG in Chander and Tulkens 2009 (pp. 176-177)[16]. In the present paper it is extended so as to visualize the presence of the externality and derive thereby the insights announced in the introduction.

2.1 Payoffs of coalitions in absence of externality: a reminder

For the grand coalition of a usual 3-players game with coalitional function traditionally denoted $v(S)$, $S \subseteq N$, normalized to 0, the set of accessible payoffs $\{u_1, u_2, u_3\}$ such that [17] $\sum_{i=1}^{3} u_i = v(\{1,2,3\})$ *i.e.* the Pareto efficient surface, is representable by the points that make up the triangle ABC in the three-dimensional diagram of Fig. 1. This set is also called the set of *imputations* of the game. In the case of an Edgeworth market game, the 0-normalization reflects the convention that the origin corresponds to the payoffs of the players at their initial endowments.

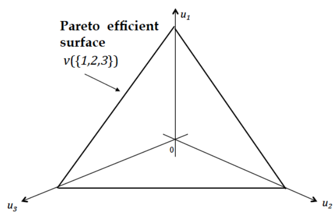


Fig. 1 The set of imputations

In such a game, if coalition $\{1,2\}$ forms, one represents in the two-dimensional space on the right of the diagram the set of accessible payoffs $\{u_1, u_2\}$ such that $\sum_{i=1}^{2} u_i = v(\{1,2\})$, which is what this coalition can achieve and is represented on Fig.2 by the line drawn in the plane (u_1, u_2).

---

[15] In chapter 18, Appendix A pp. 673 & ff., Mas Colell, Green and Whinston 1995 give excellent graphical representations of 3 players games.

[16] It was used there with the different intention to illustrate and characterize the γ-core solution. Within that set, the particular solution is exhibited that results from the use of the specific transfers formula proposed by Chander and Tulkens in their 1995 and 1997 papers.

[17] In this paper, the set of players in a coalition S supposed to form will most often be denoted by the list of numbers that identify the players, within braces: *e.g.* $\{1,2,3\}$ denotes the coalition whose members are players 1, 2, and 3. Player 1, 2, or 3 when considered as singletons, are denoted similarly {1}, {2}, {3}.

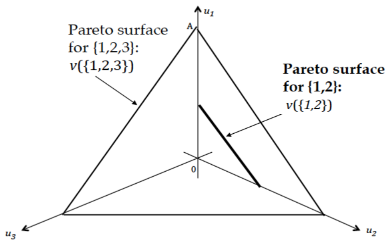


Fig. 2 What coalition {*1,2*} can do

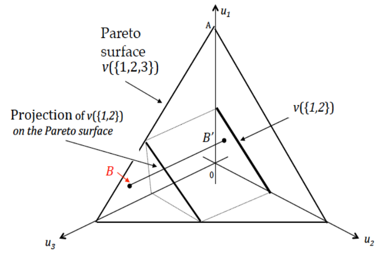


Fig.3 "B": a "blocked" imputation

Projecting the line *v*({*1,2*}) on the surface *v*({*1,2,3*}) yields on the Pareto efficient surface a similar line that is of key interest to appreciate the stability of coalition {*1,2,3*}. Indeed, this line permits to locate on the *v({1,2,3}*) surface the subset of its points that constitute imputations in the three person game which are not acceptable by coalition {*1,2*}. For instance in Fig. 3, the imputation « *B* », a Pareto efficient point for coalition {*1,2,3*}, is rejected (or « blocked ») by coalition {*1,2*} because to point *B* there corresponds point « *B′* », which is the *projection back of B* in the (u_1, u_2) space, and « *B′* » lies below *v*({*1,2*}), that is, below what coalition {*1,2*} can do by itself.

Such rejection of *B*, as well as (for the same reason) of all imputations in the triangle that *B* belongs to, is usually interpreted by asserting that that none of these imputations can be an outcome of the game, because they are blocked by a subset of the players.

I propose below another interpretation of a situation like *B*, in the special and distinct context of a game with externalities.

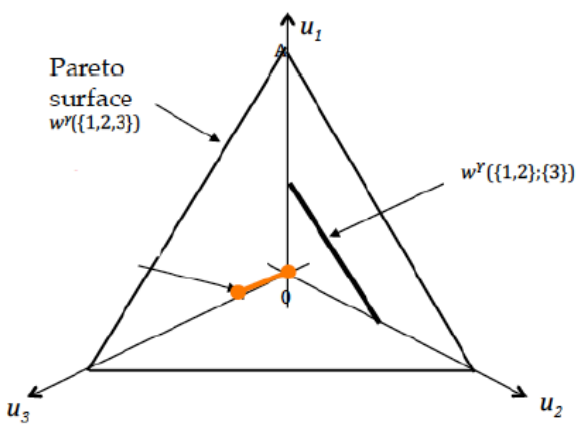
2.2 Representing "sensitivity to the coalitional externality" and "outside option payoff".

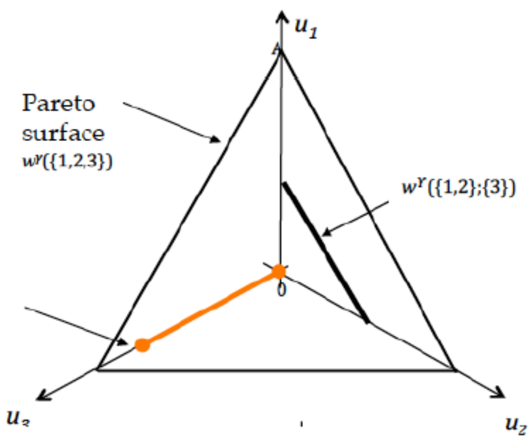
Notice that in the description just made of the formation of coalition {*1,2*}, what happens with player 3 is simply ignored. That ignorance is innocuous if the actions of the members of that coalition do not affect player 3, which is the case in games where there are no externalities such as the standard market games I have referred to. However, if an externality does occur between the players, and between all of them as in the EEG defined above, then a graphical representation of their mutual interactions can be obtained as follows. First, when each player acts individually and chooses his most preferred strategy, the outcome is obviously a Nash equilibrium of the game defined in Section 1.1 above. In a three dimensional diagram of the kind used so far, let the origin conventionally correspond to the payoff levels at that equilibrium.

Next, consider coalitions. For instance, when players 1 and 2 choose a joint action, this has an effect on player 3, favorable to 3 if the externality is of the « positive » type, and of a magnitude due to the specific fact that the external effect now *results from the jointness* of the strategies of the coalition members. Graphically, in the (u_1, u_2, u_3) space, the windfall benefit that player 3 receives from $\{1,2\}$'s action *as a coalition*[18] can be measured in terms of a segment along the u_3 axis (see Figs. 4 & 5 for two alternative cases). The amount of this benefit is determined by the subjective characteristics of player 3's payoff function. I shall call it his *sensitivity to the coalitional externality*[19].

If one now adds the consideration that while benefitting from the externality, player 3 can of course also act on his own, according to his preferences, one can interpret the extreme point of the segment just identified as the outcome in payoff terms of a simultaneous occurrence of two things: (a) a utility maximizing individual strategic choice and (b) the reception of a positive externality. I therefore shall denote this point on the u_3 axis by the expression $w^{\circ}(\{3\};\{1,2\})$ – see Fig. 4 and Fig 5 for the alternative possible cases of sensitivity to the coalitional externality just shown.

Parallel to that, it will be useful to adopt from now on, for what I have called the Pareto surface for $\{1,2\}$, the notation $w^{\gamma}(\{1,2\};\{3\})$ so as to describe the behavior of coalition $\{1,2\}$ in a way that takes into account the existence and the behavior of player 3 *outside* of it. Note that both $w^{\circ}(\,.\,;\,.\,)$ and $w^{\gamma}(\,.\,;\,.\,)$ are each just one number.





Two typical cases of player 3's sensitivity to the externality generated by coalition $\{1,2\}$

Fig. 4 Low sensitivity

Fig. 5 High sensitivity

---

[18] If {1,2} does not form, the externality described here does not occur: all players' payoffs go back to 0, *i.e.* the Nash equilibrium of the game, where only individual externalities are exerted.

[19] At an earlier stage I thought of calling "the size of the externality » the phenomenon here described — an expression which is not without relevance, as the sequel will show. Yet the term sensitivity finally chosen points better to the subjective component of it.

As to terminology, the expression of « outside option payoff » attached to player 3 at the point $w°(\{3\};\{1,2\})$ in Figs 4 and 5 is self explanatory if one thinks in the context of a three players game in which coalition $\{1,2\}$ is formed *and* player 3 is still in the game, but outside of the coalition.

2.3 The PANE wrt *S* and the P^{γ}-coalitional correspondence

At a more general level, this terminology leads one to describe the situation of the three players – actually, a conceivable outcome of the game – in terms that recognize explicitly their respective positions, a situation which is exactly the one that has been named above a « Partial Agreement Nash Equilibrium with respect to a coalition *S* » — in the case of Figs. 4 and 5, a P.A.N.E. wrt $\{1,2\}$. This is expressed by the two values $w^{\gamma}(\{1,2\};\{3\})$ and $w°(\{3\};\{1,2\})$ which are, respectively, the sum of the payoffs of $\{1,2\}$ and the payoff of player 3 at that equilibrium.

But this amounts to substitute, for the single valued coalitional function $w^{\gamma}(S)$ introduced in Section 1.1, a multi valued *coalitional correspondence* $P_{\gamma}(S)$ which in the case of Figs. 4 and 5 associates with $S = \{1,2\}$ the pair $[w^{\gamma}(\{1,2\};\{3\})\ ,\ w°(\{3\};\{1,2\})]$ just described.

Stated in general terms, the correspondence $P^{\gamma}(.)$ associates
- with every coalition $S \subset N$:
  (i) the expression $w^{\gamma}(S\ ;\ \{j\}\ j \notin S)$, that is, the scalar value of the sum of the payoffs of the coalition members at a PANE, given the presence of the non members as singletons, and
  (ii) for each $j \in N \setminus S$, the expressions $w°(\{j\}\ j \notin S\ ;\ \mathrm{S},\ \{k\}\ \mathrm{k} \in N \setminus S \cup \{j\})$, that is, the vector of the scalar values of the individual payoffs of the non members of *S* considered as singletons at that same PANE, given the presence of *S* and of the other singletons within *N*;
- with coalition $S = N$ the scalar value $w^{\gamma}(\{1,2,3\})$.

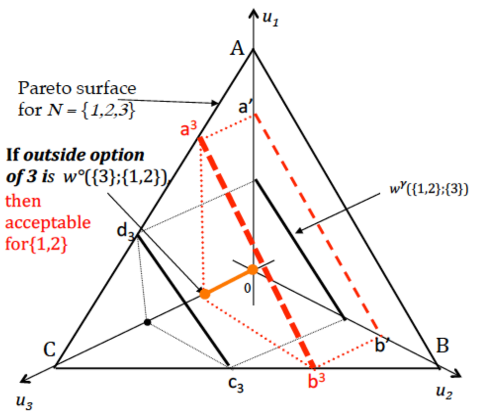
Thus, all payoffs in this definition are those induced by the strategy profile that defines the PANE wrt *S* assumed to prevail when *S* forms. This correspondence will be used from now on in the developments to follow. Strictly speaking, our game originally said to be in coalitional function form should now be called "in coalitional correspondence form"[20].

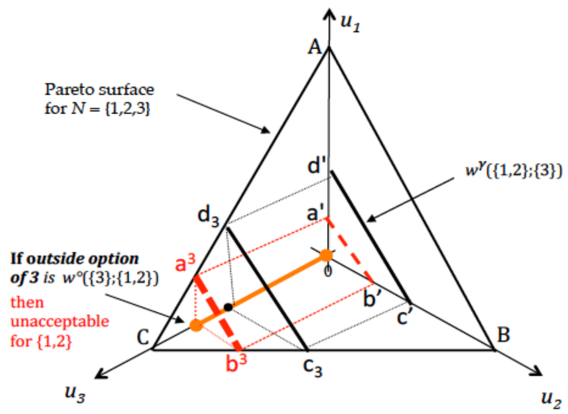
[20] Weikard 2009 (Definition 1, p.578), uses the same concept, calling it « cartel partition function ». More recently, Eyckmans and Finus 2012 similarly conduct their analysis by means of a « partition function » of the same coalitional structure as the one of the partial agreement Nash equilibria with respect to a coalition used in γ-core theory.

2.4 Free riding *vs.* cooperation

In our graphical example, the position of player 3 at the PANE wrt ({*1,2*}) can be seen from two alternative points of view: either from the point of view of a player outside of coalition {*1,2*}, who enjoys the benefits generated by {*1,2*}'s joint action without being involved in its realization – a typical free riding position; or from the point of view of the same player being a member of the coalition {*1,2,3*} and cooperating with it, but considering to break away, the best result that he can hope for is $w°(\{3\};\{1,2\})$. While the former point of view is one of coalition *formation*, the latter is rather one of *stability* of *N*, *i.e.* cooperation within *N*. In this paper I concentrate on the latter, being motivated by the search for what sustains cooperation in the alternative perspectives of internal *vs* core stability concepts.

3. The source of internal coalitional instability: non acceptability of the outside option payoff and the ensuing defection





Acceptability *vs.* non acceptability of player 3's outside option payoff within the {*1,2,3*} coalition depending upon 3's sensitivity to the externality.

Fig. 6 Low sensitivity

Fig. 7 High sensitivity

Internal stability *vs.* instability of coalition *N* can be illustrated in the following way. On both Figs 6 and 7, the point representing the outside option of player 3, $w°(\{3\};\{1,2\})$,is projected on the Pareto surface for *N* as an entire dashed line, namely a^3b^3. The difference in sensitivity to the externality implies that this line lies either above (Fig.6) or below (Fig. 7) the line d_3c_3 which is the projection of the joint payoff $w^{\gamma}(\{1,2\};\{3\})$.

The two parallel diagrams show that a request by player 3, within {*1,2,3*}, for an imputation that would correspond to a point lying "above" the line d_3c_3, may be acceptable for the two other players in Fig. 6, whereas such a claim would be unacceptable in the case represented by Fig. 7. The explanation is most obvious in the latter case: in Fig. 7, if player 3 demands an payoff that puts him on the dashed line, the "back projection" a'b' of this dashed line on the space (u_1,u_2) shows, by comparison with the coalition's payoff $w^{\gamma}(\{1,2\};\{3\})$ in that same space, that coalition $\{1,2\}$ can do better for its members without having player 3 along. It therefore prefers the PANE wrt $\{1,2\}$ to any of the grand coalition imputations on the dashed line. The demand a^3b^3 of player 3 makes the grand coalition $\{1,2,3\}$ unstable, and eventually breaking down.

Alternatively, in the case of Fig. 6, a similar back projection shows that player 3's request for a payoff that puts him on an imputation line that would grant him his outside option payoff leaves players 1 and 2 with ample possibilities of payoffs that are better than those they could achieve without player 3, namely all those represented by points in the area Aa^3b^3B. Accepting an imputation that compensates player 3 for not exerting his outside option thus does not endanger the stability of the grand coalition in this case. In addition, it allows $\{1,2\}$ to reach payoff levels for themselves that are inaccessible without player 3, as shown by the respective positions in the space (u_1,u_2) of the projection a'b' of the dashed line a^3b^3 and of the line of the joint payoff $w^{\gamma}(\{1,2\};\{3\})$.

It is now meaningful to call the dashed line a^3b^3 an "acceptability line". This graphical representation of player 3's outside option payoff confronted with the set of imputations of the game suggests that recognizing it in the negotiation process (whereby an imputation is sought for N) amounts to restricting to the area $a^3b^3c_3d_3$ the set of imputations accessible in the game, at least in the case of Fig.6. In more pedestrian words, acceptance by the grand coalition of a player's request for compensation not to free ride "bites" on the Pareto surface of the game. When that bite is big enough, such an area does not exist, the Pareto surface is not accessible anymore, and cooperation with that player within N breaks down[21].

4. Multiple reciprocal externalities and the ISI set

4.1 Representing multiple reciprocal externalities and internally stable imputations

Consider now an externality generated by coalition $\{1,3\}$ on player 2. Repeating the above reasoning, the degree of player 2's sensitivity to it is represented by the level of the payoff

[21] This argument is not without connection with the one put forward by BARRETT 1994 where he suggests that cooperation in the sense of internal stability is more likely to occur when there is little to gain from cooperation. To the extent that one correlates what I call here the importance of "sensitivity to the externality" (or its "size") with the importance of the gains to be obtained from cooperation, the idea is of the same vein. When the externality, and more precisely the sensitivity to it, is a "large" one, internal stability of a coalition is unlikely because the request for the free riding payoff is likely to be "large" too, and therefore not acceptable by the coalition members. Anticipating a bit on the topic of Section 5, let me point out that in the core stability concept, such dependency on the sensitivity to the externality does not play any role.

$w°(\{2\};\{1,3\})$. Considering the two externalities occurring simultaneously, one gets Fig. 8. Introducing the third player leads to Fig. 9 where all three players generate externalities on one another[22], having the sensitivities indicated by $w°(\{1\};\{2,3\})$, $w°(\{2\};\{1,3\})$ and $w°(\{3\};\{1,2\})$, respectively.

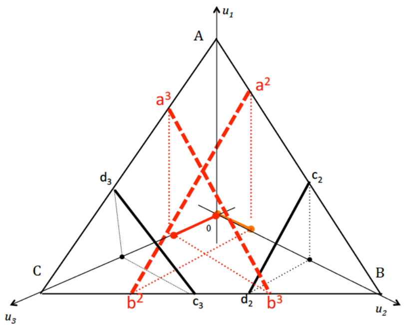


Fig. 8 Two coalitional externalities
with their respective acceptability lines

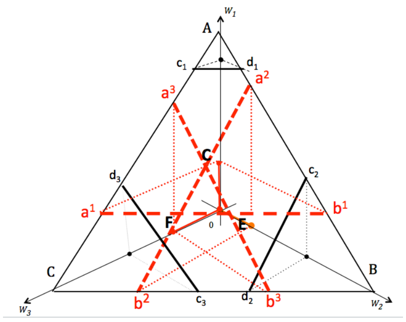


Fig. 9 An Internally Stable Imputations set
with three coalitional externalities

Thus, a set of points is generated on the Pareto surface that forms the triangle **CEF**. Each side of this triangle is obtained by the procedure followed thus far: it thus consists of the intersection of the three areas $a^1b^1c_1d_1$, $a^2b^2c_2d_2$ and $a^3b^3c_3d_3$. As each point of this set is an imputation of the game, that grants all players, as members of the grand coalition, at least their free rider option payoff, I propose to call it the set of **I**nternally **S**table **I**mputations (henceforth the ISI set, for short).

If the point is an interior one to the intersection, all players get actually more than their outside option payoff. If the point is located on a side, such as the **CE** side for example, it grants one player, player 3 in the example, just his outside option payoff. Similarly, point **C** is an imputation that grants just their outside option payoffs to both players 2 and 3, whereas the u_1 coordinate of this point shows that player 1 gets a payoff larger than what his outside option would secure him; in fact, he gets all of what remains of the gain from cooperation in *N*, if 1 and 2 receive just their outside option payoffs.

4.2 Feasibility and the ISI set

In general terms, the above amounts to state:

[22] Hence called « reciprocal ». The case of unilateral externalities where polluters and pollutees are distinct economic agents is illustrated in the diagrams of Chander and Tulkens 2009 (pp. 176-177) mentioned above.

Definition 1: For any game in the EEG class, in coalitional correspondence form $[N, P^{\gamma}]$, the **set of internally stable imputations** (the ISI set) is the set of payoff vectors $x^\circ = (x_1^o,...,x_n^o) \in R_+^n$ such that

$$(a)\ \forall i \in N,\ \ x_i \geq w^\circ(\{i\}; N \setminus \{i\})$$

$$(b) \qquad \sum_{i \in N} x_i = w^\gamma(N).$$

Properties (a) specify that the payoff of each player in the imputation reaches at least the level of his outside option payoff. Property (b) ensures that such vectors x° are feasible imputations of the game.

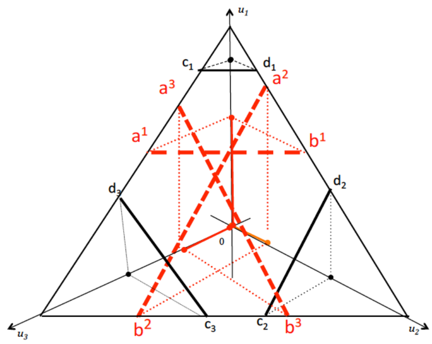


Fig. 10 A game where the ISI set is empty.

On Fig 10, the outside option payoff of *each* player *i*, considered in isolation, is acceptable for the members of $N\backslash\{i\}$ so that condition (a) is met in this case. Yet, the ISI set is empty, as revealed by the fact that the three areas $a^1b^1c_1d_1$, $a^2b^2c_2d_2$ and $a^3b^3c_3d_3$ do not intersect. The position of the dashed line a_1b_1, which is "higher" along the u_1 axis than it is in the previous figure, seems to be the cause of this emptiness. But it is not so. The cause lies in the positions of all three dashed lines, relative to one another and the non intersection reveals an impossibility of *simultaneous* acceptability of otherwise acceptable individual requests for compensation not to free ride. Formally:

Condition 1: For the ISI set to be non empty it is necessary that the game's coalitional correspondence $P^{\gamma}(S)$ has the property that

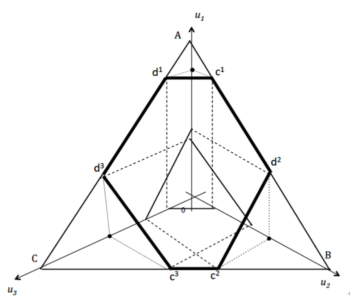
$$\sum_{i=1}^{n} w^\circ(\{i\}; N \setminus \{i\}) \leq w^\gamma(N).$$

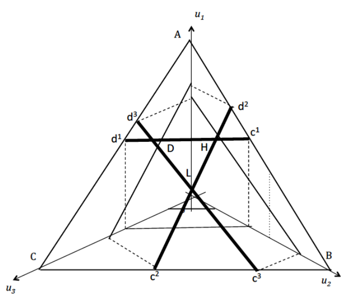
In words, the sum of the claims for outside payoffs of all players should not exceed the common payoff that the grand coalition can achieve. This feasibility, *i.e.* existence, condition implies that the concept of internal stability is applicable only to a class of games where the players' sensitivities to the externality is not "too large", not only at the level of each of the individual players but also *globally*. In other words, while the concept of an internally stable solution is well defined for classical EEGs as defined at the beginning of this paper, such a solution can only be found to prevail only for a subclass of these games and consequently of the underlying economies, namely those with "not too large" coalitional externalities.

5. The ISI set and the γ-core

The developments thus far have been made in order to formulate the concepts of internal stability theory in the terms and vocabulary of the theory of core stability, with key elements of that formulation being those of partial agreement Nash equilibrium with respect to a coalition (S) and the coalitional correspondence $P^{\gamma}(S)$ derived from it.

Using the textbook representation referred to in Section 2.1 above of coalitional functions of cooperative games, one can proceed to a similar graphical representation of the γ-core of the EEG. This is done in Fig. 11 where the core is a subset of the Pareto surface depicted by the area $c_1d_1c_2d_2c_3d_3$[23]. Actually, as the reader surely has realized, this form of the γ-core was already present from Fig. 9 on. However, the γ-core may have other forms, such as the area DHL in Fig 12 for instance, due to alternative characteristics of the coalitional correspondence: here, stronger coalitional powers are at play[24].





Shapes of the γ-core for alternative specifications
of the environmental externality game where $n = 3$.

Fig. 11 : An "anchored" core   Fig 12 An "interior" core

[23] This representation appears to be the same as the one of the core of a usual market game, although our coalitional function w^{γ} is different. But the difference is immaterial for the graphical representation because the γ-coalitional function used in the Chander and Tulkens papers referred to has a single valued image. It measures only the sum of the payoffs of the coalition members, as does the coalitional function of a market game, and ignores the payoffs of the other players.

[24] With the terminology of « anchored » and « interior » cores in the legends of Figs 11 and 12 I am following Gilles 2010.

On that basis, putting together Figs. 9 and 11, leads to Fig. 13, that makes it tempting to state that the ISI set, when non empty, is a subset of the γ-core[25]. However, the reverse inclusion can also occur, as shown in Figs. 12 and 14, again, due to alternative characteristics of the coalitional correspondence. Thus, the two sets can overlap, and considerably so. At the other extreme, the question may be raised whether the ISI set (when non empty) and the γ-core could be disjoint sets.

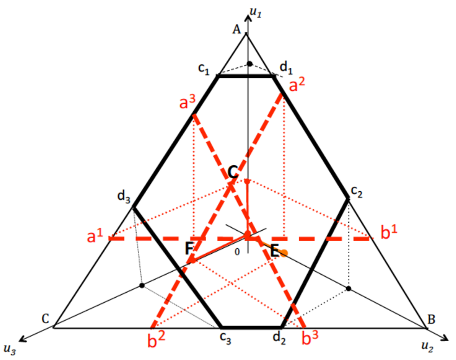


Fig. 13: A game where the ISI set (CEF) is contained in the γ-core ($c_1d_1c_2d_2c_3d_3$)

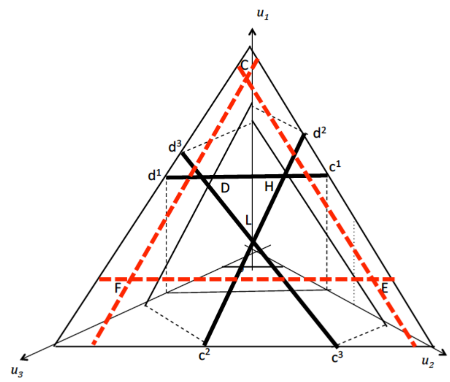


Fig 14: A game where the ISI set (CEF) contains the γ-core (DHL)

However, and more constructively, one can turn the question of the relation between the two sets by asking whether for the same game conditions can be identified under which the *intersection* of the ISI set and the γ-core would be non empty. A positive answer would establish conditions of *compatibility* between the two concepts, in the sense of the logical possibility of finding solutions of an EEG having *both* properties of internal and core coalitional stability.

Handling this question is probably not achievable with graphical techniques, but it can be found in a fairly straightforward way by turning to the mathematical concept by means of which the existence theory of the core is established, namely balancedness of the game.

6. Internal and core coalitional stability are compatible

Let Y_{ISI} and Y_{CSI} be, respectively, the sets of internally stable and γ-core stable imputations of any EEG with coalitional correspondence $P^{\gamma}(S)$. It is claimed that, under certain

[25] As was asserted in the first version of this paper, using an excessively restrictive notion of internal stability for games with $n > 3$, that Claude d'Aspremont convinced me to abandon.

conditions,

$$Y_{ISI} \cap Y_{CSI} \neq \varnothing. \qquad (1)$$

This section is entirely devoted to establish this claim, which covers the graphical cases of Figs. 13 and 14, as well as many other conceivable ones. To that effect, recall first:

Definition 2: For any game in the EEG class, in coalitional correspondence form $[N\ P^{\gamma}]$, the set of **γ-core imputations** (CSI) is the set of payoff vectors $y = (y_1,\ldots,y_n) \in R_+^n$ that satisfy the inequalities

$$(c)\ \forall S \subset N,\ \sum_{i\in S} y_i \geq w^{\gamma}(S\ ;\{j\}_{j\in N\setminus S})$$

$$(d)\qquad \sum_{i\in N} y_i = w^{\gamma}(N).$$

Theorem 1 (Bondareva 1963, Shapley 1967) **:** A coalitional game has a non empty γ-core if and only if it is a balanced game.

Theorem 2 (Helm 2001) **:** Any game in the EEG class, in coalitional correspondence form $[N, P^{\gamma}]$, is a balanced game.

Using these two known results, claim (1) can be established by verifying under which conditions the balancedness of an EEG is preserved when what defines internal stability is introduced in Definition 2 of the γ-core. That is, if conditions (a) of Definition 1 are added as new constraints to the system of inequalities of Definition 2. This results in a modified form of the EEG, call it the *IS-constrained EEG*, for which one can have:

Definition 3: For a IS-constrained EEG in coalitional correspondence form $[N, P^{\gamma}]$, the set of **IS-CS imputations** is the set of payoff vectors $z = (z_1,\ldots,z_n) \in R_+^n$ that satisfy the inequalities

$$(e)\ \forall S \subset N \text{ where } |S| \geq 2,$$

$$\sum_{i\in S} z_i \geq w^{\gamma}(S\ ;\{j\}_{j\in N\setminus S})$$

$$(f)\ \forall S \subset N \text{ where } |S| = 1 \text{ and } S \supset \{i\}$$

$$z_i \geq w^{\circ}(\{i\}; N\setminus\{i\})$$

$$(g)\qquad \sum_{i\in N} z_i = w^{\gamma}(N).$$

The novelty with respect to the γ-core is the presence of the n constraints (*f*), which restrict the set of admissible imputations to those that ensures every player i his outside option payoff from N. However, recalling condition 1 on the feasibility of an ISI set, the above definition of the set of IS-CS imputations is meaningful only if the EEGs to which it is applied do satisfy that condition 1.

Subject to this last proviso, if the IS-constrained EEG is balanced, then by applying the Bondareva-Shapley theorem one may assert that the IC-CS imputations set is nonempty, and the compatibility claim stated above is established.

Balancedness, in turn, is established by exhibiting a strategy profile for the IS-constrained EEG that satisfies the condition:

$$\sum_{S \subset C} \delta^S w^\gamma(S; \{j\}_{j \in N \setminus S}) \leq w^\gamma(N) \tag{2}$$

where C is any balanced collection of coalitions belonging to a family $\boldsymbol{X}$ of such collections, and $\delta^S \in [0,1]$, $S \in C$, are balancing weights that characterize these collections by the condition that

$$\sum_{\substack{S \in C \\ S \supset \{i\}}} \delta^S = 1, \ \ i = 1, \ldots, n.$$

The argument of Bondareva and Shapley consists in showing that, by duality, the expression (2) implies the existence of a primal solution to the mathematical program implicit in the (e)-(f)-(g) inequalities – a primal solution that is precisely a γ-core imputation of the game.

There thus remains now to exhibit a strategy profile satisfying (2). In Section 1.1 above, a strategy profile in the EEG was just denoted as a vector $\tilde{e}^S = (\tilde{e}_1^S, \ldots, \tilde{e}_n^S)$, without precise specification of what a strategy consists of in the underlying economic model. This must now be made explicit, as follows.

For each economic agent represented by a player in the EEG, the emission variable e_i enters as an "input"[26] in a specific production function $g_i(e_i)$, increasing and concave, as well as in a damage function $v_i(\sum_{j \in N} e_j)$, increasing and convex. With these two components and letting $x_i = g_i(e_i)$, the concave, quasi-linear utility function $u_i = x_i - v_i(\sum_{j \in N} e_j)$ mentioned

[26] Emissions being assimilated with pollutant emitting energy use.

in Section 1.1 becomes $u_i(e_1,...,e_n) \underset{def}{=} g_i(e_i) - v_i(\sum_{j\in N} e_j)$ that was used throughout this paper as player i 's payoff function with $(e_1,...,e_n)$ as a strategy profile. For each coalition S, its worth, *i.e.* the sum of the members' payoffs, $\sum_{i\in S} u_i = \sum_{i\in S}[x_i - v_i(\sum_{j\in N} e_j)]$, is similarly expressed in terms of strategy profiles by letting $\sum_{i\in S} x_i = \sum_{i\in S} g_i(e_i)$. As to the coalitional correspondence P^{γ} (S), defined in Section 2.3 over the set of all PANE wrt a coalition and denoted $\tilde{e}^S = (\tilde{e}_1^S,...,\tilde{e}_n^S)$, its multi-components image reads in strategic form as:

$$\text{for each } S,\ w^{\gamma}(S;\{j\}_{j\in N\setminus S}) = \sum_{i\in S}(g_i(\tilde{e}_i^S) - v_i(\sum_{j\in N}(\tilde{e}_j^S)) \tag{3}$$

and

$$\text{for each } j\in N\setminus S,\ w°(\{j\}\ j\notin S\ ;\ S,\ \{k\}\ k\in N\setminus S\cup\{j\}) = g_j(\tilde{e}_j^S) - v_j(\sum_{k\in N}\tilde{e}_k^S). \tag{4}$$

Consider now the following strategy profile:

$$\tilde{e}(C) = (\tilde{e}_1(C),...,\tilde{e}_n(C)) \tag{5}$$

where for each $i = 1,...,n$,

$$\tilde{e}_i(C) = \sum_{\substack{S\in C \\ S\supset\{i\}}} \delta^S \tilde{e}_i(S), \tag{6}$$

that is, a convex combination of the strategies adopted by player i at the PANE wrt the coalitions S of which i is a member, and which are elements of the balanced collection C .

In view of the concavity of the functions $g_i(e_i)$ and the convexity of the functions $v_i(\sum_{j\in N} e_j)$, one may verify (as suggested in the Appendix) that (2) is satisfied when specified in strategic terms as in (3) and (4) evaluated as in (5) and (6).

On that basis, one may conclude:

Proposition : For the IS-constrained EEG in coalitional correspondence form [N, P^{γ}], the set of **IS-CS imputations** (*i.e.* the IS-constrained γ-core) is non empty.

7. Concluding remarks

7.1 On the notion of coalitional stability.

The reconciliation announced in the title of this paper is essentially one of compatibility. By expressing the concept of internal stability in the classical language of coalitional (*i.e.*

cooperative) games, the two concepts appear to be formally related, and by no means contradictory.

Compared to other solution concepts in cooperative game theory, internal stability for games with externalities is of limited scope to the extent that its existence depends on a model specific characteristic of the players' payoff functions, namely their degree of sensitivity to the externality: while high sensitivity entails emptiness of the ISI set, it may be the case that existence holds with lower sensitivity. In the case of the γ-core concept, existence of coalitional stability rests only on the general property of concavity of the payoff functions, and is independent of the importance of the externality.

Finally, while the two concepts rest on the idea of sharing the "ecological surplus"[27], that is, the surplus generated by the move from the no agreement Nash equilibrium to an efficient Pareto outcome in an environmental externality framework, they organize that sharing on quite different grounds. Internal coalitional stability is based on the bargaining ability of the players to extract a part of the global surplus achieved in common within the grand coalition. By contrast, core coalitional stability rests on the comparison, made by each coalition, between what it gets from the common surplus and what it could get with its own resources.

7.2 On efficiency, stability and cooperation.

In the vocabulary used by many authors of the game theoretically inspired literature evoked in the introduction, stability of a coalition is almost always taken as a synonym for cooperation within that coalition. There is also floating around the expression of "fully cooperative solution", which is used to denote an efficient (Pareto optimal) solution. The analysis that lead to the present results gives an occasion to remind us of proper correspondences between concepts and vocabulary.

To call an efficient (*i.e.* Pareto optimal) solution "the fully cooperative solution" is really a misnomer. While it is correct that such a solution can only be reached, in the environmental externality game, by formation of the grand coalition (because of the public good nature of this kind of externality), it is an abuse of words to call it "cooperative". Indeed, the cooperative virtue of a solution does not derive from the fact that the grand coalition has to form to reach efficiency, but well from the fact that the solution be *accepted* by all coalitions of players (in the case of core stability) and/or by all individual players (in the case of internal stability). As the pictures in the preceding pages show, there are an infinity of efficient

[27] A happy expression coined by Bjorn Tuypens.

solutions (all points of the Pareto surface), but only a subset of them are cooperative in the core theoretic sense, as well as in the IS sense. On that basis, I propose to keep the expression of "fully cooperative" exclusively for solutions that are at the same time efficient, core stable, and internally stable — thus, in the IS-CS set. Let us call "cooperative" the solutions in the γ-core, "free riding free" those in the ISI set, and "efficient" those that are simply Pareto efficient.

Considering now the two forms of coalitional stability, internal vs. core, when looked at in their exact technical formulation, they appear not to proceed at all from a same motivation towards cooperation. Internal stability is obtained by granting the players the payoffs they could secure *by not cooperating* (their outside option)! These payoffs are of the nature of carrots to keep them in the coalition they consider leaving[28]. Core stability is obtained, instead, by comparing what deviating coalitions could get using heir own resources, with a solution that results from the coalitions' resources *being pooled with those of all the other players in the game*. Arguably, this is co-operation in a stricter form of the term. As well as more fruitful, as far as existence is concerned.

Finally, other motivations for cooperation are at the source of other game theoretic solution concepts not considered here, such as the nucleolus, the Shapley value, the bargaining set and the von Neumann-Morgenstern solution. Identifying such motivations within the framework of a specific game such as those in the class of our environmental externality game, and comparing them with the motivations that are found here to be at the source of core and internal forms of stability would enrich our understanding of what cooperation can mean.

[28] At an admittedly higher level of motivation, Weikard 2009 finds to be « a natural assumption to grant each member a *right* (my emphasis) to a position no worse than his outside option » (p.580), based on a rights-egalitarian sharing theory where « individual claims have priority and it is a collective responsibility to meet them ».

Appendix: Balancedness of the IS-constrained environmental externality game

Balancedness of the IS-constrained EEG $[N,\ P^{\gamma}]$ is to be established by showing that the inequalities (2) hold when expressed in the strategic form (3)-(4) and evaluated at the partial agreement Nash equilibria (PANE) wrt balanced collections of coalitions as specified in (5) and (6). Following Helm 2001, the proof is in three steps:

(1) By concavity of the production functions, one has

$$g_i\Big(\sum_{\substack{S\in C\\ S\supset\{i\}}} \delta^S \tilde{e}_i^S\Big) \geq \sum_{\substack{S\in C\\ S\supset\{i\}}} \delta^S g_i(\tilde{e}_i^S) \quad i=1,\ldots,n.$$

(2) By convexity of the damage functions, one can show that

$$v_i\Big(\sum_{j\in N} \delta^S \tilde{e}_j^S\Big) \leq \sum_{\substack{S\in C\\ S\supset\{i\}}} \delta^S v_i\Big(\sum_{j\in N} \tilde{e}_j^S\Big) \quad i=1,\ldots,n\,.$$

(3) By cohesiveness of the game and the previous steps,

$$\sum_{i\in N} g_i(\vec{e}_i(C) - \sum_{i\in N} v_i\Big(\sum_{j\in N} \vec{e}_j(C)\Big)$$

$$\leq \sum_{i\in N}\Big[g_i(e_i^*) - v_i\Big(\sum_{j\in N} e_j^*\Big)\Big] = w^{\gamma}(N)$$

where $(e_1^*,\ldots,e_n^*)$ denotes the efficient strategy profile.

In none of these inequalities does the substitution of the RHSs of the n constraints (f) for the RHSs of the corresponding n constraints (c) (*i.e.* those for $|S|=1$ and $S\supset\{i\}$) endanger the concavities or convexities invoked, since the payoff functions $w^{\circ}(\{i\};\ N\setminus\{i\})$, $i= 1,\ldots,n$, are themselves concave functions of the strategy profiles $(e_1,\ldots,e_n)$.

References

AUMANN, R. 1987, "Game Theory", pp. 460–482 in by J. Eatwell, M. Milgate, and P. Newman, eds, *The New Palgrave: A Dictionary of Economics*, (Volume 2 reprinted in 1989 as *Game Theory*), Macmillan, London.

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

BONDAREVA, O. N 1963, "Some applications of linear programming methods to the theory of cooperative games (In Russian)". *Problemy Kybernetiki* **10**: 119–139 English translation in an document issued by the Econometric Research Program, Princeton University, 1968.

BRÉCHET, T., GERARD, F. and TULKENS, H. 2011, "Climate coalitions: a theoretical and computational appraisal", *The Energy Journal* 32 (1) 49-75.

CARRARO, C. and D. SINISCALCO, 1993, "Strategies for the international protection of the environment", *Journal of Public Economics*, 52, 309-328.

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. Early version issued as FEEM *Nota di Lavoro* 69/1994.

CHANDER, P. and TULKENS, H. 2009, "Cooperation, stability and self-enforcement in international environmental agreements: a conceptual discussion", chapter 8 (pp.165-186) in R. Guesnerie and H. Tulkens, eds, *The Design of Climate Policy*, CESifo Seminar Series, The MIT Press, Boston.

CHANDER, P., DRÈZE, J., LOVELL, C.K. and MINTZ, J. (eds) 2006, *Public goods, environmental externalities and fiscal competition: 22 essays by Henry Tulkens*, Springer, New York.

CURRARINI, S. and TULKENS, H 2003, "Stable international agreements on transfrontier pollution with ratification constraints", chapter 1 (pp.9-36) in C. Carraro and V. Fragnelli (eds), *Game Practice and the Environment*, FEEM/Elgar Series in Environment, Chentelham.

D'ASPREMONT, C. and JASKOLD GABSZEWICZ, J. 1986, "On the stability of collusion", chapter 8 (pp. 243-264) of J.E. Stiglitz and G.F. Mathewson (eds), *New developments in the analysis of market structure*, The MIT Press, Cambridge (Mass.).

EYCKMANS, J. and FINUS, M. 2009, "Transfer schemes and institutional changes for sustainable global climate treaties", chapter 6 (pp.103-135) in R. Guesnerie and H. Tulkens, eds, *The Design of Climate Policy*, CESifo Seminar Series, The MIT Press, Boston.

EYCKMANS, J. FINUS, M. and MALLOZZI, L. 2012, « A New Class of Welfare Maximizing Stable Sharing Rules for Partition Function Games with Externalities", *Bath Economics Research Papers 6/12*, University of Bath.

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. and CAPARROS, A. 2015, eds, *Game Theory and International Environmental Cooperation, Essential Readings*, Elgar, Cheltenham.

GERMAIN, M., TULKENS, H. and MAGNUS, A. 2010, "Dynamic core-theoretic cooperation in a two-dimensional international environmental model", *Mathematical Social Sciences* 59, 208-226.

GILLES, R.P. 2010, *The cooperative game theory of networks and hierarchies,* Theory and Decision Library C 44, Springer-Verlag, Berlin Heidelberg.

GILLIES, D.B. 1959, “Solutions to general non-zero-sum games”, pp. 47–85 in R.D. Luce and A. W. Tucker, eds. *Contributions to the Theory of Games IV* (*Annals of Mathematics Studies Series 40*), Princeton University Press, Princeton.

MAS COLELL, A., WHINSTON, M. and GREEN, J. 1995, *Microeconomic theory*, New York, Oxford University Press.

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

HELM, C. 2001, “On the existence of a cooperative solution for a coalitional game with externalities”, *International Journal of Game Theory* 30, 141-146

SHAPLEY 1967, “On Balanced Sets and Cores”, *Naval Research Logistics Quarterly*, **14,** 453–460.

THORON, S. 2009, “Heterogeneity of countries in negotiations of international environmental agreements: A joint discussion of the Buchner-Carraro, Eyckmans-Finus, and Chander-Tulkens chapters”, Chapter 9 (pp. 186-198) in R. Guesnerie and H. Tulkens, eds, *The Design of Climate Policy,* CESifo Seminar Series, The MIT Press, Boston.

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. Translation in French published as chapter 2 (pp.47-72) in G. Rotillon, ed. *Régulation environnementale: jeux, coalitions, contrats*, Economica, Paris 2002.

VAN STEENBERGHE, V. 2005, “Core-stable and equitable allocations of greenhouse gas emission permits”, *Core Discussion Paper* 2004/75.

VAN STEENBERGHE, V. 2003, Designing Markets for Emission Permits With Applications to Climate Change, *Doctoral Dissertation,* CORE and Département des Sciences économiques, Université catholique de Louvain.

WEIKARD, H. P. 2009, “Cartel stability under an optimal sharing rule”, *The Manchester School* 77 (5), 575-593.