

# About Heterogeneity of Countries in Negotiations of International Environmental Agreements Discussion

Sylvie Thoron\*

July 19, 2006

## Abstract

The theory of coalition formation which has been developed in a symmetric framework, explains that the problem of participation in the negotiations on climate change can be attributed to the existence of positive externalities and the incentive to free ride. The three papers on negotiations in this volume present the other effects which arise when asymmetries between countries are taken into account. Here, the instability of the grand coalition and the emergence of climate-blocks are discussed. The question is whether the latter are observed in reality and is this harmful?

JEL Classifications: C68, C70, C71, C72, H23, H41, Q25.

Key words: climate change, emissions trading, flexibility mechanism, country-specific targets, climate-blocs, agreement, coalition formation, negotiation, membership, stability.

---

\*GREQAM and University of Toulon, Institute for Advanced Study, Princeton.

# 1 Introduction

An important characteristic of the Kyoto protocol is that the different partners are treated very differently. For the first time, an International Environmental Agreement (IEA) sets not only a global target and a time table but also very precise country-specific targets. In previous IEA the objective was to reach a consensual common target as, for example, in the case of the Protocol on the Reduction of Sulphur Emissions by at Least 30 percent. The Montreal Protocol on Substances that Deplete the Ozone Layer also fixes a uniform target for industrialized countries but introduced a differentiation in the time table with a 10-year grace period for developing countries. In the case of the Kyoto Protocol the differentiation is much more marked. The developing countries are not committed to any binding target for the first period (until 2012) and nothing is specified for the following period. The global target of a decrease of greenhouse gazes by 5.5% is supposed to be achieved by assigning reductions to the different countries which vary from - 8% to + 10%. Furthermore, in the same protocol, the flexibility mechanisms, the European Union Bubble, the principles of Joint Implementation, Emissions Trading and the Clean Development Mechanism allow the exploitation of differences and complementarities between countries. Indeed, the Kyoto protocol is clearly the result of a negotiation between partners whose differences, far from being ignored in order to reach a consensus on a common target, have occupied a central position in the discussions.

It is not surprising, in this context, that the theorists who wanted to analyze the IEA in general and the Climate Change negotiations in particular, were not satisfied with a theory of coalition formation which limited its analysis to a symmetric framework. Although I do not want to advocate the latter choice in this introduction, I will try to give a historical explanation. The integration of externalities and the exploitation of efficiency gains are two important economic justifications for cooperation. Chander and Tulkens explain in this volume how these two arguments have been developed in different literatures. The problem of externalities has been analyzed by economists in the framework of Samuelson's public good theory (1954) or as developments of the Coase theorem (1961). In parallel and exactly during the same period, the cooperative game theory literature was analyzing superadditive games (Shapley (1953) introduced the Shapley value, Gillies (1959) developed the Core). This second literature has always been criticized because it focused on the problem of how to exploit the efficiency gains inside a coalition, while leaving to one side the incentive problem generated by the persistent externalities between coalitions. Thrall and Lucas (1963) proposed a framework to deal with both problems simultaneously but the cooperative game theory literature did not take this aspect up. Interest in this was revived in the 90s, with the development of the theory of coalition formation.

It is worth noting that both literatures then focused on disparities between actors. The Coase Theorem explained how polluters and pollutees could reach an efficient outcome by bargaining. The aim of the solution concepts proposed by cooperative game theory was to solve the problem of sharing a coalition's worth

between heterogenous players. Why then, did the theory of coalition formation not deal with the problem of heterogeneity? Because the original aim of this theory was to explain how the existence of these externalities or spillovers could lead to inefficient outcomes. The Coase Theorem claimed the emergence of an efficient outcome, while for cooperative game theory efficiency is assumed. This new literature focused on the sources of inefficiencies. Non cooperative game theory focused on the problem of compliance to agreements and the theory of coalition formation on the problem of participation in agreements. I will not address the problem of compliance here. After all, to quote Abram Chayes and Antonia Chayes (1991):

« International lawyers and others familiar with the operations of international treaties take for granted that most states comply with most of their treaty obligations most of the time »

Conversely, participation is often a serious problem for IEA and this has been illustrated by the negotiations on climate change. However, now we know that the incentives are not conducive to the formation of the grand coalition, even in a symmetric framework, several questions remain to be answered. Is the situation better or worse off in a non symmetric framework? Did the theory of coalition formation propose solutions to the participation problem? If the answer is in the affirmative, are these solutions relevant for the heterogeneous case? Finally, are there solutions specific to heterogeneous cases? These are the kind of questions addressed in the three papers on international environmental negotiations of this volume.

## 2 Sources of heterogeneity

In this volume, Eyckmans and Finus on one hand and Buchner and Carraro on the other, use modified versions of the same RICE model or Regional Integrated model of Climate and the Economy. This model, proposed by Nordhaus and Yang in 1996, is a "regional, dynamic, general equilibrium model of the economy which integrates economic activity with the sources, emissions, and consequences of greenhouse-gas emissions, and consequences of greenhouse-gas emissions and climate change". It divides the global economy into six different regions: USA (United States), JPN (Japan), EU (European Union), CHN (China), FSU (Former Soviet Union) and ROW (Rest of the World). The novelty of the RICE model in comparison with previous models of global warming was to allow nations to adopt different strategies.

In the CLIMNEG World Simulation Model, Eyckmans and Finus introduce differences in discount rates. Discount rates for developing regions are higher than those for developed countries. The countries also differ in energy efficiency. USA, JPN, EU have steep marginal abatement costs, while CHN and ROW have flat marginal abatement costs. This means that energy efficient regions face higher costs when cutting back emissions. The countries are also more or less vulnerable or more or less sensitive to climate change. Damage functions are particularly steep in EU and ROW, to a lesser extent in USA and JPN and

relatively flat in FSU and CHN.

The FEEM-RICE model used by Buchner and Carraro incorporates an endogenous technical change which yields endogenous growth and improves the emission/output ratio. The Kyoto protocol constitutes the starting point: the EU, Japan and Russia are committed to complying with their Kyoto targets and there is a market for pollution permits. The authors present different scenarios and explain in each how the countries can move away from this situation. In each scenario two coalitions are formed. Each coalition can create a market for pollution permits. On each market, the countries draw benefit from their complementarities. Big polluters with stringent targets can buy pollution permits from small or "not yet so big" polluters with mild targets. Despite their mild targets, less energy efficient regions have an incentive to reduce their emissions. By selling their own emission permits to more energy efficient regions, they generate funds which can be used to invest in new technologies.

### 3 Transfers and Stability

Eyckmans and Finus (EF) pose the question as to the links between the stability of agreements and transfers among partners. As has already been discussed in the introduction, in most of the literature in economics and cooperative game theory dealing with the formation of agreements, efficiency is a basic criterion. With this as a principle, the outcome should be calculated to satisfy efficiency in a first step. The different partners' contributions are chosen to maximize the sum of their welfares. As a consequence, in the framework described above for example, it may be that energy inefficient regions facing smaller abatement costs will have to make bigger reductions in pollution. Then, in a second step, different normative criteria can be used to justify transfers between partners. This is necessary because the different partners will not sign an agreement they consider to be *unfair*. The difficulty is that the definition of fairness in this context is not unique. In particular, this depends on which of the countries' characteristics are to be considered. There is a whole literature on this delicate question to which Eyckmans and Finus refer.

They define a series of transfer schemes. Each of these guarantees to each partner at least the welfare level it would have had if no agreement were signed. They differ in the way the costs and benefits from partnership are shared: a normative criterion is used to define a country-specific coefficient which is applied to the difference between the sum of welfares before and after the agreement implementation. In one of these schemes, a solution originally proposed by Chander and Tulkens (1997) (from now on the CT-solution), each country's contribution to pollution abatement is proportional to the cost that it incurs as a result of the damage from climate change. As a consequence, those who care more about climate change or are more vulnerable to it, must contribute more. From a normative point of view, this is debatable. It depends on what kind of *good* the climate is. Given the objective which is to determine a contribution for each country, let us consider a parallel with private goods. Suppose that

climate could be considered as a luxury good: in this case the preference for environment in general and climate stability in particular increases with wealth. It may then be justified to ask richer countries which have a higher preference for it, to contribute more. However, climate is more often considered as an inferior good<sup>1</sup>: those who have the highest demand for it are those who are not wealthy enough to be able to protect themselves against climate change. In this case it is more difficult to justify, from a normative point of view, the idea that developing countries which are more vulnerable to climate change should contribute more.

However, the CT-solution is also attractive because it can be interpreted in a positive framework. In a purely positive approach, the stability of an agreement is not guaranteed by its fairness but rather by the fact that each partner has an incentive to sign, given the alternative and considering its own interest. This depends on each countries' anticipations about what would be the ex-partners' reaction if the country decided to leave the agreement. The literature on coalition formation explains how the assumption about the reactions of the ex-partners after a deviation determines the outcome. Suppose that each partner anticipates that a deviation would provoke a complete dismantling of the agreement. Call this anticipation Assumption I. Now suppose that, after a deviation, a country anticipates that its partners will still go ahead with the agreement. These countries will just adapt their contributions and will no longer take the outsider's welfare into account. Call this Assumption II. When externalities are positive, as in the case of an abatement game, the dismantling of the original coalition is a stronger threat than a simple adjustment of the coalition's abatement (cf Hart and Kurz (1983), Yi (1996), Thoron (2000)). As a consequence, stability under Assumption II is more difficult to attain than stability under Assumption I.

Eyckmans and Finus propose to test the claim that transfers can increase coalition stability. They consider stability under Assumption II. Each transfer scheme guarantees every partner at least its welfare level in the situation without any agreement. Hence, when transfers are added, the coalitions are certainly stable under Assumption I. For example, Chander and Tulkens (1997) prove that the CT-solution belongs to the  $\gamma$ -core which is also defined under Assumption I. However, there is no direct reason why these coalitions should still be stable under Assumption II, unless the normative criterion used to share the surplus also plays a role in stability, which is the case of the CT-solution. Indeed, the countries who care more about climate change are also those who are least likely to withdraw. If they contribute more, those who care less contribute less and this favours the attainment of stability.

However, this relationship is the exception rather than the rule. Other than by simple coincidence, there is no link between an allocation which satisfies a normative criterion and an allocation which satisfies a stability criterion. I will go even further, there is in fact a certain contradiction between the normative

---

<sup>1</sup>See for example the interpretation given by Schelling (2006) in the chapter What makes greenhouse sense? of his book Strategies of Commitment.

and the positive approaches. On one hand, the normative approach starts from an efficient outcome and organizes transfers on the basis of a normative criterion. On the other hand, the positive approach describes the outcome of a negotiation, whether or not it is efficient, and considers that the allocation of payoffs is endogenous to this negotiation process. In the first case the efficiency is a basic criterion, in the second case it is neither a starting point, nor necessarily an outcome.

I explained above the problem posed by the definition of stability in normal form games of coalition formation. In the literature on coalition formation, other types of models are proposed. Bloch (1996) and Ray and Vohra (2001) used extensive form games to represent the negotiation process. They proved the inefficiency of the outcome when there are positive externalities. In order to focus on the problem of heterogeneity, Maskin (2003) uses these extensive form games to define a Shapley value for games with externalities. Here again, Maskin's conclusion is that this new value does not satisfy the efficiency axiom. The other common feature of these positive models is that the payoffs are generated during the negotiation, with the only objective of stability. For example in an extensive form game, when a player makes a proposal to another player, it is just sufficient to convince the later to become a partner.

The conclusion is not that normative criteria can or cannot be used as arguments during a negotiation. Two problems prevent us from giving a clear answer to this question. First, it is theoretically difficult to combine the normative and positive approaches. Maybe the reason is that the basis of positive models, the maximization of individual payoffs, is too restrictive. Second, we lack empirical evidence other than simple declarations that normative criteria have some relevance for the outcome of a negotiation<sup>2</sup>.

## 4 Membership rule and climate-blocks

When the internal/external stability criterion is applied in a symmetric framework, it has been proved that an equilibrium always exists (d'Aspremont et al. (1983)). The agreement generated at the equilibrium can be restricted, in the extreme case it corresponds to the trivial situation in which the agreement is signed by a unique player (!), but the equilibrium exists. It has also been proved that this equilibrium is robust against deviations by coalitions (Thoron (1998)). This is no longer true when players are not symmetric. To illustrate this point, consider the following example: it may be that FSU wants to join coalition (EU, JPN) and that USA wants to join (EU, JPN, FSU). Both coalitions are externally unstable. However, it may be that FSU does not want to leave (EU, JPN, FSU) but would want to leave (EU, JPN, FSU, USA). In comparison with the symmetric case, the fact that one country wants to join the coalition no longer means that the extended coalition is internally stable. As a consequence,

---

<sup>2</sup>See a very interesting article by Kauppi and Widgren (2004) about the difference between declared and real arguments underlying the european budget allocation.

the equilibrium may not exist, and the membership may cycle. Eyckmans and Finus propose to reduce this instability by using an exclusive membership rule.

The difficulty, in the asymmetric case, is that the different countries have different preferences for the type of measures which should be applied to prevent climate change. Indeed, economic theory has taught us that two kind of problems threaten the efficient provision of a public good: the incentive to free ride and the difficulty with choosing the public good for partners who have different preferences for it. Whilst the recent theory of coalition formation has focused on the first problem, which can be studied in a symmetric framework, the second problem, which only occurs when there are asymmetries, is the main object of an older literature on coalition formation: the theory of jurisdiction formation and the theory of clubs. When no agreement is possible with an open membership rule, this means that preferences are too different for all countries to reach a common agreement. Then, as a consequence of what would be the equivalent of a Tiebout mechanism, different agreements may emerge. The countries which manage to reach an agreement will have similar preferences or complementary characteristics. However, the literature on jurisdiction formation which tried to model the Tiebout mechanism showed that the existence of an equilibrium cannot be proved in this framework. Furthermore, the conclusion which can be drawn from the theory of clubs is that this problem of non-existence can be solved by using an exclusive membership rule.

My interpretation of Buchner and Carraro's paper in this volume, is that the post Kyoto period may be characterized by the formation of sub-global climate blocs to exploit complementarities of countries on a market for pollution permits. Indeed, the existence of this kind of market creates new incentives. On one hand, a country which did not sign or ratify the Kyoto protocol cannot benefit from the existing market. On the other hand other markets can be created. The larger the difference between countries and their complementarities, the higher is the incentive to form a new coalition and create a new market.

However, several coalitions can exist simultaneously, even in the symmetric case. When the assumption that only one coalition can be formed is dropped, the driving forces are the same as in the restricted framework. Because the game is superadditive the countries have an incentive to cooperate but they also have an incentive to free ride, which generates the non-participation problem. Now the different coalitions free ride on each other. Because the externalities are positive, within a given structure the smaller coalitions obtain considerable benefits from the existence of the bigger ones. In Ray and Vohra's model (2001), the coalitions form sequentially. When the number of negotiating partners is intermediate, two coalitions are formed. The biggest coalition forms first, the second coalition forms afterwards, in order to free ride on the previous one. When the externalities are negative the outcome is very different. In this case, Yi (1996) showed that the grand coalition is always an equilibrium.

When describing the formation of climate-blocs, Buchner and Carraro draw a parallel with the formation of the World Trade Organization. Indeed, coordination and information problems can make the formation of the grand coalition, in one step, difficult. This is why, in the case of the organization of free trade,

countries started by signing binary agreements. However, it seems to me that there is a fundamental difference between this last case and negotiations on climate change. Thanks to negative externalities, the incentives made for a move in the right direction to reach the final outcome of the formation of the World Trade Organization. This will not work in the same way when considering negotiations on global warming. In this case, because the externalities are positive, the emergence of several agreements does not necessarily mean that we are moving in the right direction. It may mean that we have reached a sub-optimal equilibrium in which one coalition free rides on the other.

An optimistic point of view would be that there are other driving forces which could create negative externalities between blocks and that markets for permits might constitute a source of these externalities. Another force is the emergence of social norms. An argument developed by Finnemore and Sikkink (1998) is that at a certain "tipping point" in a norm's evolution, a "norm cascade" takes place, and then states join the coalition in large numbers because of pressure from other states and non-state actors. In this case again, the disparities between countries can help the process since the countries which are more committed can, in the end, convince the others to follow their lead.

## 5 Conclusion

The theory of coalition formation has been developed essentially in a symmetric framework. It explains that the problem of participation inherent to the negotiations on climate change, is generated by the existence of positive externalities and an incentive to free ride. A conclusion would then be that the countries involved are unable to reach the social optimum because the incentive to free ride is too strong. As a result they may form sub-groups, the climate-blocks, and reach a sub-optimal situation in which one coalition free rides on the other. However, the forces which could change the externalities might help to offset these incentives.

The asymmetries between potential partners introduce a new difficulty: the differences between the countries' evaluations of climate change costs and abatement benefits. The impossibility to reach a consensus among these asymmetric parties is another explanation for the emergence of climate blocks. However, the same asymmetries can also be considered as an advantage when they take the form of complementarities between countries. In this case they provide an explanation for the formation of these climate-blocks. However, it is difficult to disentangle the two explanations: free riding or efficient exploitation of similarities or complementarities.

In conclusion, even if the outcome turns out to involve differential treatment of the countries and, in the extreme case, the formation of several blocks, I would still recommend a global framework for negotiations. In the case of negotiations on climate change, all the countries felt themselves obliged, at least initially, to negotiate together under the guidance of the UN. These lengthy negotiations provided conditions for a more creative approach. Countries had to find other

ways to differentiate themselves from each other, within a single agreement. In the framework of the Kyoto protocol, they came up with the country-specific targets and the flexibility mechanisms. The advantage of this process is that it is a better guarantee that, when countries want to differentiate themselves from each other they do so not to free ride but to increase efficiency.

## 6 References

Bloch F. (1995), "Endogenous Structures of Association in Oligopolies", *Rand Journal of Economics*, 26, 537-556.

Bloch F. (1996), "Sequential Formation of Coalitions with Externalities and Fixed Payoff Division", *Games and Economic Behavior*, 14, 90-123.

Chayes, A. and Chayes, A. H. (1991), "Compliance Without Enforcement: State Regulatory Behavior Under Regulatory Treaties", *Negotiation Journal*, 7: 311-31.

Maskin E. (2003), "Bargaining, coalitions and externalities", presidential address to the Econometric Society European Meeting, Stockholm, mimeo (August).

Hart and Kurz, (1983), "Endogenous Formation of Coalitions", *Econometrica*, Vol. 51, No 4, 1047-1064.

Kauppi H. and M. Widgrén (2004), "What determines EU decision making? Needs, power or both?", *Economic Policy*, July 2004, pp. 221-266.

Lucas, William F.; Maceli, John C. (1978), "Discrete Partition Function Games", *Game Theory and Political Science*; edited by Peter C. Ordeshook; New York; New York University Press; 191-213; #1748.

Ray D. and R. Vohra (1999), "A Theory of Endogenous Coalition Structure", *Games and Economic Behavior*, 26, 286-336.

Ray D. and R. Vohra (2001), "Coalitional Power of Public Goods", *Journal of Political Economy*, Vol 109, n 6.

Roth A. E. (1998), *The Shapley value: Essays in honor of Lloyd S. Shapley*, Cambridge University Press, Cambridge.

Shapley, (1953), "A Value for n-Person Games", in *Contributions to the Theory of Games*, Vol. II, ed. By Kuhn and Tucker. Princeton; Princeton University Press, 307-317.

Schelling T. C., (2006), *Strategies of Commitment and other essays*, Harvard University Press.

Thoron, (1998), "Formation of a Coalition Proof Stable Cartel", *Canadian Journal of Economics*, February 1998, Vol 1.

Thoron, (2000), "Conjectures about the Reaction of a Frustrated Coalition", working paper GREQAM n°31A00, 2000.

Thrall, RM and Lucas, WF (1963), "n-Person Games in Partition Function Form", *Naval Research Logistics Quarterly* 10 , 281-298.

Yi, (1997), "Stable Coalition Structure with externalities", *Games and Economic Behavior*, 20, 201-237.