



Lebanese American University Repository (LAUR)

Post-print version/Author Accepted Manuscript

Publication metadata

Title: International Environmental Agreements: Doomed to Fail or Destined to Succeed? A Review of the Literature

Author(s): Walid Marrouch & Amrita Ray Chaudhuri

Journal: International Review of Environmental and Resource Economics

DOI/Link: <http://dx.doi.org/10.1561/101.00000078>

How to cite this post-print from LAUR:

Marrouch, W., & Ray Chaudhuri, A. (2016). International Environmental Agreements: Doomed to Fail or Destined to Succeed? A Review of the Literature. *International Review of Environmental and Resource Economics*, DOI, [10.1561/101.00000078](http://dx.doi.org/10.1561/101.00000078), <http://hdl.handle.net/10725/7314>

© Year 2016

Electronic version of an article published as *International Review of Environmental and Resource Economics*, 9 (3-4), 2016, 245-319, <http://dx.doi.org/10.1561/101.00000078> © 2016 World Scientific Publishing Company, <https://ideas.repec.org/a/now/jirere/101.00000078.html>

This Open Access post-print is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives (CC-BY-NC-ND 4.0)



This paper is posted at LAU Repository

For more information, please contact: archives@lau.edu.lb

International environmental agreements: Doomed to fail or destined to succeed? A review of the literature

Walid Marrouch

Department of Economics, Lebanese American University

P.O.Box 13-5053, Chouran Beirut 1102-2801 - Lebanon

Email: walid.marrouch@lau.edu.lb

Amrita Ray Chaudhuri

Department of Economics, University of Winnipeg; CentER & TILEC, Tilburg University

515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9

Email: a.raychaudhuri@uwinnipeg.ca

Abstract: We survey the economics literature on International Environmental Agreements (IEAs). We classify the extant literature into the following categories: pure IEA games without linkages, which focus on a single externality stemming from global pollution; IEA games with side-payments and issue linkages, which consider carrot or stick mechanisms to lure cooperators or punish defectors; political economy models, which examine intra- and inter-governmental issues that affect international negotiations; design issues of IEAs; behavioral factors, which affect the willingness of countries to cooperate; and additional externalities resulting from open economy issues and adaptation, which interact with the global pollution externality and affect the free-riding incentives of countries. Based on this classification of the literature, we attempt to identify conditions under which an international environmental agreement would be sustainable and effective.

Keywords: international environmental agreements; coalition formation; transboundary pollution; climate change

Acknowledgement: We would like to thank Rita Antar for excellent research assistance. We appreciate insightful comments by two anonymous reviewers and the editors.

1 Introduction

International environmental issues, such as climate change due to the anthropogenic emissions of greenhouse gases, the destruction of the ozone layer due to the emissions of ozone depleting gases, the destruction of trans-oceanic fisheries, or the eutrophication of international lakes, are often characterized by the existence of diffused and multilateral externalities and the absence of a supra-national institution to enforce policy. Given the global public good nature of the problem, abating/mitigating transboundary pollution requires international cooperation in the form of international environmental agreements (IEAs). Two important features that must be satisfied for an IEA to be sustainable are the following: first, the agreement must be profitable, i.e. each country must gain by joining the IEA; second, the agreement must be self-enforcing, i.e. each country must have an incentive to join the agreement when it acts in its own self-interest, rather than remaining outside the IEA. While this paper reviews the broader economics literature on IEAs, many of the insights found in this literature are applicable to current global problems including climate change.

Over 1000 international environmental agreements have been signed, the majority of them since the United Nations Conference on the Human Environment (Stockholm Conference) in 1972, which is now seen as a turning point in global environmental governance. Global environmental problems that countries came together to tackle with varying degrees of success include climate change and the depletion of the ozone layer. International cooperation to tackle climate change was epitomized by the Kyoto Protocol (signed in 1997), which came into force in 2005 and expired in 2012. One of the main weaknesses of the protocol was the fact that some of the world's leading emitters including China, and India, were not required to commit to abatement targets, while others like the US did not ratify the protocol. This has exacerbated the problem of carbon leakage whereby production of carbon-intensive goods shifts to non-participating countries, mainly developing ones (Olmstead and Stavins, 2012). Moreover, in the newly signed Paris Agreement (2015) under the United Nations Framework Convention on Climate Change (UNFCCC), the current pledges of emission abatement by countries may not be sufficient to achieve the United Nations' target for limiting temperature increases to two degrees Celsius. Compared to the climate change problem, tackling the problem of ozone layer depletion by mitigating the emissions of chlorofluorocarbons (CFCs) has proved to be more successful, and resulted in the signing of the Montreal (1987) protocol and its subsequent amendments. Empirical evidence suggests that while some IEAs have been effective at reducing pollution damage, others have not. Hence the relevance of our question: Is there hope of sustaining meaningful environmental agreements, and if so, under what conditions?

It turns out that it has been just as difficult to identify conditions conducive to global cooperation in theory as it has been in practice. This has sparked a large research interest over the past 25 years that has led to the appearance of a wide body of literature. In this essay, we review the related literature on self-enforcing IEAs, and the various instruments that have been proposed to attain meaningful cooperation. A number of existing surveys provide comprehensive summaries of the game theoretic models analyzing the formation and stability of self-enforcing IEAs, including Finus (2001, 2003, 2008), Wagner (2001), Barrett (2003), Yi (2003), and Benchekroun and Long (2012). While we briefly present the canonical game-theoretic models, our focus is on updating the reader on the latest game theoretical

applications within this context. We also contribute by extending the scope of our survey beyond game theoretic models to include the latest empirical, behavioral, and institutional research on IEAs.

Specifically, we classify the literature into three main strands. Section 2 focuses on “pure IEA” games where signatories’ commitment to mitigate emissions generate positive externalities that benefit non-signatories, reducing individual countries’ incentives to participate in the IEA. Section 3 focuses on side-payments and linkages between IEA games and related issues, which may qualify some of the canonical results found in the pure games. Section 4 explores a number of areas of recent developments in the literature. These include the political economy and behavioral aspects of negotiations, and additional externalities arising from modeling an open economy context and allowing for adaptation to climate change. Finally, Section 5 provides concluding remarks.

2 Pure IEA games

In the first part of the survey, we focus on pure IEA games and provide a detailed classification of the extant literature. Pure IEA games refer to games where countries cooperate on a single issue, that is, the mitigation of the global pollutant, but without relying on transfers, linkages, or side issues. These IEA games are considered to be positive externality games since collaborative effort on mitigating global pollution generates benefits to free-riders in terms of reduced pollution damage. We first classify the literature into non-cooperative games, cooperative games, and games with farsighted stability. We, then, further classify non-cooperative games into games where countries are *ex ante* identical, *ex ante* different, and where uncertainty plays a role. Within each of these three categories, we examine static membership games, repeated games, and dynamic games. In general, these models yield a single coalition that is self-enforcing in equilibrium.¹

2.1 Non-cooperative games

The literature on non-cooperative games is divided into two categories. The first uses a two-stage model where, in the first stage, countries decide on their membership status, while in the second stage, countries decide on their emissions or abatement strategies. Within the second stage, we distinguish between papers that allow coalition and non-coalition members to simultaneously choose their strategies, and those papers that allow the coalition to act as a Stackelberg leader vis-à-vis the non-members. The second category uses dynamic membership decisions that can change every period.

2.1.1 Symmetric players

In this section we review models of countries that exhibit symmetry in their benefit and damage functions.

¹While most of the IEA models have this feature, there exist a few exceptions that allow for multiple coalitions to co-exist in equilibrium. These will be mentioned as we proceed.

Static membership games (full information) *Simultaneous move games*

Membership models consider a set, N , consisting of n countries. Country i emits $e_i \geq 0$ units of the pollutant with the global emissions denoted by $E = \sum_{i=1}^n e_i$. Country i 's welfare is given by the reduced-form payoff function:

$$W_i = B_i(e_i) - D_i(E) \quad \forall i \in N, \quad (1)$$

where $B_i(e_i)$ represents the benefit to country i from its own emissions. Typically, papers assume that the benefit is an increasing function of e_i , which could be linear or concave. The function $D_i(E)$ represents the damage to country i from global pollution E . Typically, papers assume the damage to be an increasing function of E , which could be either linear or convex. Papers that consider symmetric countries assume $B_i(e_i) = B(e_i) \forall i$ and $D_i(E) = D(E) \forall i$. The curvature of the benefit and damage functions play a role in the resulting equilibrium of the emissions game. For example, assuming that both functions are linear in emissions results in discrete choices (either pollute or abate). However, using a linear-quadratic framework results in emissions being a continuous function of the parameters of the model.

Carraro and Siniscalco's (1991, 1993) canonical model considers a two-stage open membership game. In the first stage, countries decide whether or not to join the coalition and form an international environmental agreement. More specifically, in the first stage, let the set $S \subset N$ of countries sign an agreement and become members of the coalition S , while the set $N \setminus S$ of countries do not and thus become non-members. We denote the size of coalition S by s . In the second stage, countries decide on their equilibrium emissions where the welfare of each member country, at the equilibrium, is given by W_s^* , while the welfare of a non-member is given by W_{ns}^* .

The game is solved by backward induction. In the second stage, the coalition and the non-members choose their emissions simultaneously. Each player chooses its emissions strategy taking as given all other players' emissions strategies. When a coalition of size s is formed the number of players is reduced to $n - s - 1$. In the first stage, countries decide whether or not to join the coalition. This choice depends on the stability criterion used to explain the self-enforcement of the coalition in a world with no supra-national government. In fact, most papers in the IEA literature use the now well-known cartel stability criterion first introduced by d'Aspremont *et al.* (1983) where a coalition is said to be stable if it is both internally and externally stable. Formally, internal stability implies that $W_s^*(s) \geq W_{ns}^*(s-1)$ for all $i \in S$, i.e. there is no incentive for any one member to leave the coalition and reduce its size to $(s-1)$, where $W_s^*(s)$ and $W_{ns}^*(s-1)$ are derived from (1). External stability implies that $W_{ns}^*(s) \geq W_s^*(s+1)$ for all $i \in N \setminus S$, i.e. there are no further incentives for any non-member to join the coalition and increase its size to $(s+1)$.

In papers assuming that the coalition and non-members move simultaneously in the second stage of the game, the stable size of the self-enforcing coalition depends on the curvatures of the benefit and damage functions. When benefits are quadratic and damages are linear, the best-response functions in emissions are orthogonal. In this case, the largest stable size is $s^* = 3$, as shown in Carraro and Siniscalco (1991) and Finus (2001, 2003). When benefits and damages are quadratic, the best-response functions are downward-sloping as in the standard Cournot oligopoly setup. In this case, the largest stable size is $s^* = 2$ (Carraro and Siniscalco, 1991; Rubio and Casino, 2001; Finus, 2003). Finally, with linear benefits

and quadratic damages, it is found that the best-response functions are downward-sloping and the largest stable size is $s^* = 1$ (Finus, 2003). Thus, the general conclusion has been that a so-called ‘puzzle of small coalitions’ arises under the scenario where the members and non-members move simultaneously (also referred to as Cournot games).²

Finus and Rundshagen (2001) consider various versions of the first stage while keeping the second stage similar to the canonical model described above. In the first stage, they introduce and compare different coalition formation processes. While the canonical model considers the scenario where all countries decide whether to join simultaneously and open membership, i.e. all players can freely join a coalition if they wish, Finus and Rundshagen (2001) also consider models of exclusive membership where non-members need members’ consent in order to join, with different degrees of unanimity required to form coalitions. The authors also consider games where countries sequentially decide whether to join a coalition. In these various contexts, they find that multiple coalitions may coexist in equilibrium.

De Cara and Rotillon (2003) keep the first stage similar to the canonical model (internal and external stability), but introduce the possibility of multiple pollutants in the payoff functions in the second stage. They compare the cases where some pollutants are neglected in the negotiations, and where all pollutants are included in the agreement (comprehensive agreement). They find that the latter better resist free-riding incentives.

Another natural question that follows is whether larger coalitions can be stable if the coalition acts as a Stackelberg leader instead of moving simultaneously with the non-members. This feature reflects the fact that IEAs often announce their targets before the non-members decide their strategies.

Sequential move games

Some papers model the open membership game by considering the scenario where the coalition acts as a Stackelberg leader vis-à-vis the non-members in the second stage of the game. That is, within the second stage countries move sequentially. One of the first papers using this approach is the seminal paper by Barrett (1994a). In contrast to the papers mentioned earlier, Barrett (1994a) uses abatement instead of emissions as the choice variable of each country in the second stage of the game. The setup is similar to the canonical model described above, except that country i abates $q_i \geq 0$ units of the pollutant with the global abatement denoted by $Q = \sum_{i=1}^n q_i$. Country i ’s welfare is given by the reduced-form payoff function:

$$W_i = A_i(Q) - C_i(q_i) \quad \forall i \in N, \quad (2)$$

where $A_i(Q)$ represents the benefit to country i from global abatement. The function $C_i(q_i)$ represents the private cost of abatement of country i . In Barrett (1994a), $A_i(Q) = A(Q) \forall i \in N$ and $C_i(q_i) = C(q_i) \forall i \in N$. As in the case with emissions as the choice variable,

²We note that in contrast to the theoretical predictions of models deriving the small coalition puzzle, the Kyoto Protocol (1997) resulted in a stable coalition of 37 countries and not in a stable coalition of 2 or 3 countries. Similarly, the Paris Agreement (2015) has been signed by 196 countries and not by 2 or 3 countries. Such discrepancies may be partially explained by the fact that the theoretical models typically assume that countries signing the IEA also ratify it simultaneously and implement mitigation levels consistent with the joint welfare maximization effort of the IEA, which may not apply to IEAs in reality. For example, the Kyoto Protocol included large emitters like China and India who did not have to commit to abatement levels, while the Paris Agreement has not yet been ratified yet.

the curvatures of the benefit and cost function determine whether or not the equilibrium solution is discrete or continuous.

The author first considers the case where $A_i(Q)$ and $C_i(q_i)$ are both quadratic functions. He finds that when the marginal cost of abatement, $C'(q_i)$, is small (large), and the marginal benefit of abatement, $A'(Q)$, is large (small), non-cooperative abatement by individual countries is substantial (little). In the latter case, countries abate little even under full cooperation. In both these cases, the gains to cooperation are relatively small. When both $C'(q_i)$ and $A'(Q)$ are small, the gains to cooperation are also small. However, when both marginal values are large and of similar magnitude, global abatement under full cooperation is much larger than non-cooperation, which implies that the gains to cooperation are large. However, as Barrett (1994a) shows, using numerical simulations, when the gains to cooperation are large only a small subset of countries become members. It follows that when a large number of countries become members, the gains to cooperation are small, i.e. a ‘shallow coalition’ is formed.

Moreover, when other functional forms for the benefits and costs are considered, Barrett (1994a) shows that the puzzle of small coalitions persists. For example, when $A_i(Q)$ is linear and $C_i(q_i)$ is convex, the self-enforcing IEA consists of three countries when $n > 3$. Also, when both functions are linear, a self-enforcing IEA consisting of two or more countries does not exist.

Other papers considering the Stackelberg framework include Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006), which retain a similar framework to Barrett (1994a) but use emissions instead of abatement as the choice variable of countries. The framework is similar to the canonical model as described by (1) except that countries do not move simultaneously. Diamantoudi and Sartzetakis (2006), using quadratic benefit and damage functions, find that when the set of parameters are restricted to ensure that equilibrium emissions levels are strictly positive, the largest stable size is $s^* = 4$. Rubio and Ulph (2006) extend this analysis by explicitly considering corner solutions, i.e. when equilibrium emissions equal zero. They find that stable coalitions of size larger than four are possible in line with the results first proposed by Barrett (1994a). Overall, these papers establish the duality between the abatement and emissions games, which yield qualitatively similar results in terms of the stable coalition size.

Fimus (2003) summarizes the results that are obtained under the Stackelberg framework with emissions as the choice variable, using different functional forms. Under quadratic benefits and linear damages, the best-response functions are orthogonal and the largest stable size is $s^* = 3$. Under quadratic (or linear) benefits and quadratic damages, the best-response functions are downward sloping and the stable size is $s^* \in [2, N]$. Thus, under the Stackelberg framework it is possible to obtain larger coalitions than under the Cournot framework since the first-mover advantage given to the coalition members reduces the cost of cooperation. Nevertheless, as shown by Barrett (1994a) and others, the larger the coalition the smaller the gains to cooperation.

Dynamic membership games A number of papers extend the canonical IEA model by considering a stock pollutant instead of a flow pollutant. The global flow of emissions, E , accumulates over time into a stock of pollution, Z . This is an important feature to consider

in the climate change problem given that greenhouse gases are accumulative pollutants.

Rubio and Casino (2005) consider the standard two-stage open membership model where each country decides noncooperatively whether or not to join an IEA in the first stage, and signatories and non-signatories determine their emissions simultaneously in the second stage. The novel feature of the model is the consideration of an infinite horizon dynamic setting defined in a continuous time framework. The stock dynamics, at time t , are given by the following transition equation:

$$\dot{Z}(t) = E(t) - \rho Z(t), \quad \rho \in (0, 1), \quad (3)$$

where ρ denotes the natural sequestration rate of pollution. Open-loop emissions strategies refers to the case where each country commits to a time-path of emissions, at the beginning of the second stage of the game, that maximizes the present value of its stream of welfare given the emission path of others. Using open-loop strategies and considering a benefit function that is quadratic in emissions and a damage function that is quadratic in Z , the authors find that a coalition of size two is the unique internally and externally stable IEA. However, the authors find that when a minimum participation clause is added to the IEA (i.e. the agreement is implemented only if a minimal number of countries sign it), it is possible to generate large stable coalitions of size equal to the minimum. As such, it possible to have the grand coalition of all countries as a stable outcome. This is the case because under the minimum participation clause if one country deviates the IEA is dissolved. This effectively breaks the assumption that when one country unilaterally deviates, the others continue to cooperate, as is implicitly assumed in the internal and external stability concept.

An important contribution to the dynamic IEA games literature is made by Rubio and Ulph (2007) who allow for the possibility of IEA membership changing endogenously over time (i.e. dynamic or variable membership). This is in contrast to the standard two-stage game where once membership is decided in the first stage, it does not change for the remainder of the game (i.e. fixed membership), as in the static version of the canonical model (e.g. Carraro and Siniscalco, 1993) or its dynamic version (e.g. Rubio and Casino, 2005).

Rubio and Ulph (2007) consider a discrete time infinite horizon model. The stock dynamics, in period t , are given by the following transition equation:

$$Z_{t+1} = E_t - \rho Z_t, \quad \rho \in (0, 1). \quad (4)$$

The authors use a benefit function that is linear in emissions and a damage function that is quadratic in Z . Each period t has two stages. In the first stage, countries decide on their membership using expected welfare based on a Random Assignment Rule for determining which countries become signatories. The probability of any country being a signatory in that period is the stable size, s^* (as determined by the internal and external stability criteria), in that period divided by the total number of countries, n . Thus, at the onset of each period, the probability of being a signatory is the same for all countries, but it can vary over time. In the second stage of each period t , all countries (signatories and non-signatories) decide their emission strategies simultaneously. Within this setting, they find that in the transition to the steady state, the stable size of the IEA decreases. Moreover, in the steady state, s^* is more likely to be small the greater is the marginal damage from pollution. These results are

reminiscent of the pessimistic result reported by Barrett (1994a), for the Stackelberg game, that when the gains from cooperation are large, it is more likely that s^* is small.

Nkuiya (2012) uses a model with linear benefit and quadratic damages, similar to that of Rubio and Ulph (2007), but allows for the discretization of a continuum of time into discrete commitment periods for the IEA of equal length h . The author allows the length, h , of each time period over which the IEA is effective to vary exogenously. The author finds that when h is very large, only small stable coalitions can be sustained. However, below a threshold, as h is decreased the size of the stable coalition increases.

Biancardi (2010) uses a fixed membership two-stage IEA model in continuous time, with abatement as the choice variable for countries instead of emissions. Global emissions are exogenously given at a constant rate and their accumulation over time depends on abatement choices.³ Each country is assumed to minimize its costs of abatement plus the damage, which is linear in the stock of pollution. Using both open loop and closed loop strategies,⁴ the author finds that only coalitions of size 2 and 3 can be stable under the internal and external stability criteria.

Breton *et al.* (2010) consider an IEA model with a quadratic benefit function in emissions and linear damages in the pollution stock. Countries are assumed to use closed loop emissions strategies. First, the authors consider a fixed membership dynamic IEA setup but allow signatories to punish defectors at a cost. They find that an increase in the punishment reduces the steady state stock of pollution and yields greater cooperation. Second, they consider a novel membership game that is different from either the fixed or the dynamic membership game of Rubio and Ulph (2007). This game uses discrete-time replicator dynamics to determine the evolution of the proportion of signatory countries over time. In every period, each country compares the welfare it can get by cooperating and by defecting and chooses whichever yields a greater payoff. The rate at which countries switch their membership decision is modeled as the ratio of the total discounted welfare of a signatory country to a weighted average of the welfare of a signatory and that of a defector. The authors find that fewer initial signatories are required for convergence to a stable IEA when the initial level of pollution is large.

For a more detailed survey of dynamic IEA models, please refer to Calvo and Rubio (2012).

Repeated games (compliance games) Another approach to analyze the problem of cooperation on abatement considers the infinite repetition of the static IEA game. However, in this sub-literature, discrete time is used because in a continuous time framework a one period gain from deviation is very small given that defection is instantaneously detected. As

³In models that use abatement instead of emissions as the choice variable within a dynamic context, the non-negativity of the stock of pollution may not always be satisfied. This may be the reason why most dynamic models consider emissions rather than abatement as the choice variable.

⁴Closed loop strategies refers to the case where each country's choice variable is a function of the state variable at each instant in time. Since each player's actions cause the state variable to evolve, the continuation of the game may be regarded as a subgame of the entire game. Using closed loop strategies ensures that a subgame perfect equilibrium is obtained.

such, the continuous time dynamic game framework, presented earlier, could be seen as the limiting case of repeated games as the length of each period tends to zero. It is important to note that the setup of repeated games is different from the membership games that have been described thus far because the focus is on compliance with an existing agreement rather than coalition formation. Another distinction is related to the fact that membership games lack the appropriate temporal dimension needed to model reciprocity as manifested by cooperating countries punishing defectors.

Barrett (1999) considers a reduced form setup where the defecting country's payoff is linearly increasing in the number of cooperating countries because of the positive abatement externality caused by the coalition. Countries face a binary choice, either to cooperate, i.e. to abate, or to defect, i.e. no abatement. The payoff of the cooperators includes a cost of cooperation but is also linearly increasing in the number of those cooperators. The author restricts the payoff parameter values such that the unique Nash equilibrium of the one shot game is full defection by all countries (i.e. a Prisoners' dilemma setup). Barrett (1999) considers three types of punishment strategies. The first is the grim strategy, also known as the trigger strategy, whereby countries commit to playing the non-cooperative strategy forever if any country defects. This implies that, in contrast with empirical observations, all countries always play the cooperative strategy for a sufficiently low discount rate in equilibrium (referred to as the Folk Theorem in repeated games). This strategy is shown not to be collectively rational since cooperating countries have an incentive to re-negotiate the agreement given that punishing the defector is costly. The second is the tit-for-tat strategy whereby cooperators punish a defector by playing the non-cooperative strategy in the subsequent period and revert to cooperation once the defector returns to compliance. This strategy is not individually rational because a single defection results in an infinite series of defections such that countries realize they can do better by deviating from the tit-for-tat strategy. The third is the getting-even strategy, which is similar to the tit-for-tat strategy except the fact that the punishment period is adjusted to be proportional to the defection. Barrett (1999)'s main contribution is to show that such a strategy is both individually and collectively rational (i.e. the equilibrium is re-negotiation proof) and can therefore sustain full cooperation while, at the same time, explain periods of non cooperation in line with empirical observations. However, even when considering such a strategy, the gains from cooperation are low for the grand coalition. This result extends the pessimistic result of shallow coalitions in the context of static membership games (Barrett, 1994a) to the context of compliance games.⁵

Asheim *et al.* (2006) extend the model of Barrett (1999) by allowing for regional agreements. They consider N identical countries divided into two regions. The authors analyze and contrast compliance under a global agreement and two regional ones. They consider a penance scheme which is similar to the getting-even strategy in Barrett (1999). Under the global penance strategy, non-signatories are allowed to play 'defect' given any history of the game. However, if a signatory plays 'defect' in one period, the other signatories punish the

⁵Barrett (1994a) considers more detailed functional forms for countries' payoffs when modeling the compliance game, but reaches the same conclusion on the relationship between coalition size and gains from cooperation. Finus (2001) develops a more general framework for compliance games in a two-country setting, and in chapter 14 provides a general definition of weakly re-negotiation proof equilibria (WRPE) in the N-country case. His results are consistent with those of Barrett (1994a, 1999).

defector by playing ‘defect’ instead of ‘cooperate’ in the subsequent period. Similarly, under the regional penance strategy, non-signatories are allowed to play ‘defect’ given any history of the game. However, if a signatory to one of the regional agreements plays ‘defect’ in one period, only the other signatories within the same region punish it by playing ‘defect’ in the subsequent period. The authors find that allowing for a single global agreement achieves a small stable coalition size. The reason is that the larger is the number of participating countries, the greater is the benefit from cooperation, which in turn makes re-negotiation attractive, thus undermining the incentives of the coalition to punish defectors. However, by splitting the world into two regions, each regional agreement involves a smaller number of participating countries, which facilitates sustaining the re-negotiation proof agreement. This leads to the main result of the paper indicating that two regional agreements will encompass a larger number of participating countries than a single global agreement and will Pareto dominate the latter outcome.⁶ Further, Froyen and Hovi (2008) extend this idea to show that the grand coalition can be stable as a weakly renegotiation-proof equilibrium within a binary choice setting with the possibility of only a subset of signatories punishing deviation. Asheim and Holtmark (2009) further extend the result to the case where abatement is a continuous choice variable rather than a binary choice. In the stage game, they consider linear benefits from abatement and quadratic cost of abatement. They determine the conditions under which full cooperation can be implemented as a weakly renegotiation-proof equilibrium. These include a sufficiently high discount factor and a short detection lag of deviation. Under these conditions, they show that it is possible to achieve a large and deep coalition in contrast to Barrett (1999).

2.1.2 Asymmetric players

Although introducing asymmetric players into the IEA model significantly increases the complexity of the analysis, a number of attempts have been made to model asymmetry because in reality countries are heterogenous in terms of benefits and damages associated with emissions. It is noteworthy that asymmetry gives rise to the possibility of transfers among countries. As such, most of the literature that considers asymmetric countries also tackles the issue of transfers, which are discussed in detail in Section 3. In this section, we focus on the results for pure IEA games without transfers.

Static membership games (full information) Most asymmetric IEA models use the standard two stage coalition formation game and adopt a simultaneous move setup for the second stage emissions game. Hoel (1992) is among the first to model asymmetry in IEA games. He considers a payoff function similar to the one given by (1) with quadratic benefits from emissions and linear damages that are country specific. Countries are ranked from the lowest to the highest in terms of their marginal damage. The author considers a different setup for the rules of the IEA game than in the canonical model. He studies coalition formation where signatories have to emit uniformly and another scenario where the emissions level is chosen by the median country. Using the standard internal and external stability

⁶Although not explicitly modeled in their paper, Asheim *et al.* (2006) note that there may be other factors that facilitate regional rather than global agreements such as, geographic proximity, cultural affinity, and institutional similarities, which enhance trust and reduce uncertainty.

conditions, his main finding is that the maximum stable size is equal to two and therefore the puzzle of small coalitions persists even when countries are heterogeneous. Barrett (1997b) uses two groups of countries with different characteristics and finds similar results to Hoel (1992). In the absence of transfer payments across countries, it is not surprising to find that the puzzle of small coalition persists since countries with lower pollution damage and higher abatement costs have less incentives to participate in IEAs than others.

It is important to note that the size of a coalition is a more meaningful concept when it comes to homogeneous countries. In the presence of heterogeneity even a small coalition may be effective at mitigating emissions if it involves the largest emitters. Moreover, the relatively few papers that consider asymmetric countries, as discussed in this subsection, typically rely on very specific assumptions in the second stage of the membership game, since the standard sequential or simultaneous move structures are difficult to apply in these cases. Therefore, their results are not always readily comparable to the pure IEA games, which derive the small coalition puzzle.

Barrett (2001) considers a simplified version of the canonical model where countries' payoffs are linearly increasing in the number of cooperating countries. He considers binary choices to either emit or abate. The payoff of countries that choose to abate includes a fixed cost of abatement. The author restricts the payoff parameter values such that the unique Nash equilibrium is where all countries choose to pollute. He considers two types of countries where one type has a higher marginal benefit from the coalition size. Signatories abate while non-signatories emit by assumption. He finds that large stable coalitions are possible and consist of only one type if countries are sufficiently asymmetric, and of both types if they are weakly asymmetric. Barrett (2001) allows for multiple coalition structures to be stable in equilibrium. Kolstad (2010) builds on Barrett (2001) by considering two types of equilibrium refinements to select the stable coalition and by allowing countries to differ by size as well as damage. The first is an efficiency refinement which selects the coalition that maximizes the joint payoff of signatories. The second is an equity refinement which selects the coalition structure that equates the payoffs of the two types of countries in the coalition. He finds that heterogeneity increases the stable size but reduces global welfare. The efficiency criterion leads to a unique stable coalition of only the low per capita damage countries, while the equity criterion results in a unique coalition made up of a combination of high and low damage countries. Fuentes-Alberto and Rubio (2010) generalize the model of Barrett (2001) by considering continuous instead of discrete choice of emissions. The authors consider a simultaneous open membership game with a single agreement. Each non-signatory country minimizes the sum of the abatement cost, which is quadratic in abatement of emissions, and the pollution damage, which is linear in emissions. Signatory countries minimize the joint cost of the coalition. Regardless of whether the asymmetry is in the marginal abatement cost or marginal damages, the largest stable size is equal to three. Biancardi and Villani (2010) also find similar results using a cost minimization problem with abatement as the choice variable where two sets of countries differ in their pollution damage. Tol (2001), using an integrated assessment model of climate change called the Climate Framework for Uncertainty, Negotiation and Distribution (FUND), also finds that stable coalitions are small and consist of similar regions if no transfers are allowed. These pessimistic results are in line with the puzzle of small coalitions that exist with symmetric countries. This suggests that heterogeneity without transfers in simultaneous games does not stabilize larger coalitions.

Yet, when asymmetry in both abatement cost and marginal damages coexist, Pavlova and de Zeeuw (2012) find large but shallow stable coalitions. Related empirical evidence is provided by Almer and Winkler (2010) who find countries expecting high damages from climate change (high compliance cost) were more (less) likely to sign the Kyoto protocol.

Botteon and Carraro (2001) consider a computational model of five different world regions. They compare two cooperative bargaining rules regarding the distribution of gains from cooperation across different coalition members. They find that the stable coalition size depends on the bargaining rule that is applied.

Finus and Rundshagen (2003) consider a similar setting as Finus and Rundshagen (2001) (described in the previous subsection) except that heterogeneity is allowed among countries. In an example with $n = 12$, they find that up to five coalitions can coexist, with the largest being of size three. Finus and Rundshagen (2003) conclude that stability is greater under exclusive than under open membership and is increasing in the requirements to join, such as the degree of the required unanimity. Qualitatively similar results are confirmed by Eyckmans and Finus (2006) using an integrated assessment model, the CLIMNEG World Simulation Model. Finus (2008) also examines similar issues i.e. comparing open versus exclusive membership and single versus multiple coalition using another integrated assessment model, the STACO model.⁷ Osmani and Tol (2010) extend the model of Barrett (1994a) to allow for asymmetric countries and two self-enforcing IEAs and show that when the number of countries is few and environmental damage is high, two IEAs are more effective than one at reducing emissions and increasing welfare. More recently, Hagen and Eisenack (2015) show that, in the two-stage IEA membership game with asymmetric countries, the possibility of multiple coalitions increases the number of cooperating countries and total abatement under constant marginal benefits from abatement. However, if the marginal benefits are decreasing, the number of coalitions does not affect total emissions.

Dynamic membership games Biancardi and Villani (2014) is one of the only papers to study the stability of IEAs using a dynamic pollution abatement game given heterogeneous countries. Heterogeneity considers two types of environmental awareness: high awareness (developed country) and low awareness (developing country). The authors model a standard two-stage IEA membership game where players decide whether or not to join the agreement (coalition) in the first stage and in the second stage they play a differential game in abatement (between signatories and non-signatories) subject to pollution stock dynamics. The authors consider an open-loop Nash equilibrium framework where the membership decision is decided at the outset of the game. After solving for the abatement equilibrium outcomes, the objective functions are evaluated at equilibrium and used to examine the internal and external stability of the coalition. The authors then use numerical simulations to compute the stable size of the coalition. Their main finding is that a large stable coalition -inclusive of all developing and developed countries- cannot be sustained including the grand coalition; instead they find values of countries' environmental awareness for which exclusive coalitions composed of the set of all developed or developing countries exist or small global coalitions. It may be useful to further analyze dynamic games with heterogeneous countries using alter-

⁷A more detailed description of integrated assessment models is available later in this and the next sections.

native setups such as models with emissions as the choice variable and closed-loop strategies.

Repeated games (compliance games) One of the complexities introduced when asymmetry is considered is that there might not be a consensus on what constitutes a first-best emissions or abatement target. Usually, in static asymmetric IEA models, it is assumed that the abatement target reached by the grand coalition is optimal. However, Finus and Rundshagen (1998) who model a repeated game with asymmetric countries find that sub-coalitions may be more efficient than global coalitions. They consider a reduced form game in emissions that is repeated over an infinite horizon. The authors consider heterogeneous quadratic benefits and damages in emissions. They study the difference between a uniform quota and a uniform emissions tax where the policy is set by the most polluting country. They find that in the single shot game the grand coalition is not stable under either of the two policies. However, in the repeated game, when the discount factor is equal to one, the quota regime does better than the tax regime. When the discount factor is less than one, the quota regime is also better but sub-coalitions are more efficient, similar to the result obtained by Asheim *et al.* (2006).

As discussed in the previous subsection, a number of attempts to extend the canonical model have been made. While this subsection summarizes the main results that are derived when the symmetry assumption is relaxed, the next subsection summarizes those that are obtained when the full information assumption is relaxed.

2.1.3 Games with uncertainty

Static membership games A natural extension of the pure IEA literature is to consider the implications of uncertainty because the benefits from abatement or the damages from emissions leading to climate change are as of yet not fully known. It is well established that, in a single country setup or where countries are not strategic, acquiring more information is always beneficial to the countries. However, when countries behave strategically, such an assertion may no longer hold. With a few exceptions, most of the papers in the sub-literature on uncertainty are ones where the authors use simultaneous move games to model the second stage of the standard two-stage membership framework. To our knowledge, the IEA literature on compliance (i.e. repeated games) does not examine the role of uncertainty.

One of the first papers to analyze this issue is by Na and Shin (1998). The authors explore the implications of relaxing the full information assumption in a three country version of the canonical IEA model where countries move simultaneously in the second stage. They introduce uncertainty about the distribution of the benefits of abatement among countries by considering the following payoff function:

$$W_i = \theta_i Q - \frac{q_i^2}{2} \quad \forall i \in N, \quad (5)$$

where Q and q_i represent total and individual abatement efforts respectively. The parameter θ_i is a random variable where $\theta_i \in \{z_1, z_2, z_3\}$. Using the standard internal and external stability criteria, the authors compare two scenarios that differ in the timing of the IEA negotiation. In the first scenario, the IEA is negotiated *ex ante*, that is, before each country knows its realization of θ_i (no learning). In the second scenario, the IEA is negotiated *ex*

post, that is, after the uncertainty is resolved (complete learning). Within this setup, the authors show analytically that the grand coalition is always stable in the no learning scenario, whereas it is never stable in the complete learning scenario. The intuition underlying this basic result is that *ex ante*, the countries are homogenous because they maximize the same expected payoff function, that is, $E[W_i]$, where W_i is as defined in (5). However, *ex post* the countries are heterogenous since each country knows its individual realization of θ_i . The country with the least benefits from abatement has the least incentive to participate in the IEA, causing the grand coalition to be unstable.

Kolstad and Ulph (2011) present a model similar to that of Na and Shin (1998) except that they consider $n > 3$ countries and the choice variable is emissions rather than abatement. The benefits and damages are linear in emissions, and the random component of the damage function can take on a low value with probability p and a higher value with probability $(1 - p)$. In contrast to Na and Shin (1998), they find that the grand coalition may not be stable under no learning. They also introduce a scenario of partial learning where the countries learn about the realization of their individual marginal damage after the IEA membership has been negotiated but before the emission decisions are made. They find that, with partial learning, the stable coalition size may be higher than that with complete learning, but less than with no learning.

Another approach to modeling uncertainty is to consider systematic uncertainty, rather than individual uncertainty, where the realization of the random component of the damage function is the same for all countries as is the case in Kolstad (2007). The author uses a similar setting as Kolstad and Ulph (2011) and finds that complete learning increases the stable coalition size relative to no learning. Also, he finds that with partial learning, the puzzle of large but shallow coalitions found in the IEA literature with full information persists in the presence of systematic uncertainty.

Finus and Pintassilgo (2013) build on the work of Na and Shin (1998), Kolstad (2007) and Kolstad and Ulph (2011) by comparing, within a single framework, uncertainty about both the distribution and level of benefits of abatement. They contrast the cases of no learning, partial learning, and complete learning. They consider a payoff function with concave benefits and convex costs of abatement. Their main conclusion is that uncertainty is detrimental in most relevant cases in contrast with the results of Na and Shin (1998) and others.

A notably different approach to modeling uncertainty is to consider a catastrophe that causes a discrete jump in damages from emissions. Barrett (2013) introduces such a catastrophe within the standard two-stage IEA model where the choice variable is abatement. The author considers the payoff function given by (2) if abatement is sufficiently high, and by $W_i = A_i(Q) - X - C_i(q_i) \quad \forall i \in N$, if abatement is below a given threshold. In the setup, $A_i(Q)$ is linear in total abatement and $C_i(q_i)$ is quadratic in individual abatement. The main finding is that as long as the potential benefits of avoiding the catastrophe are sufficiently large, allowing for the possibility of a catastrophe helps to increase IEA participation. However, if the threshold of abatement below which the catastrophe occurs is uncertain, then coordination is less likely.

While most of the papers considering uncertainty use simultaneous move games in the second stage of the two-stage IEA membership game, Boucher and Bramoullé (2010) use sequential decisions in the second stage. They extend the Barrett (1994a) model to allow for

uncertainty and, while most related papers consider risk neutral countries, they are among the first to introduce risk aversion. Using linear functional forms for benefits and damages in (1) and benefits and costs in (2), they compare the games where emissions and abatement are the choice variables respectively. They consider systematic uncertainty in the benefits from abatement and in the damages from emissions with no learning. One of their main findings is that the duality between the emissions and abatement IEA games no longer holds in the presence of risk aversion. When considering the abatement game, increases in either uncertainty or risk aversion reduce the abatement level of the signatories, while participation in the IEA is increased in the presence of uncertainty compared to the full information case. Taken together these results reinforce those of Barrett (1994a) on large but shallow coalitions. All these results are reversed when the emissions game is considered whereby a small but deep coalition is more likely when uncertainty and risk aversion increase.

Finus *et al.* (2014) use a similar framework to the emissions game in Boucher and Bramoullé (2010) except that in the second stage they consider simultaneous decisions. Their contribution is to model both learning and risk aversion within a single framework. They show that learning can help international cooperation when risk aversion is below a certain threshold, whereas above the threshold the opposite is true. They also show that if risk aversion is sufficiently high, it leads to more cooperation in the absence of any learning.

Overall, the literature on IEAs in the presence of uncertainty clearly indicates that the role of information is a function of the type of uncertainty. If uncertainty is in the form of a potential catastrophe, more information can foster cooperation on abatement. However, in the case of systematic uncertainty about the damages from emissions or benefits from abatement, more information may have a perverse effect.

Dynamic membership games Ulph (2004) extends the Barrett (1994a) model to allow for systematic uncertainty and stock dynamics in pollution. Similar to Boucher and Bramoullé (2010), but with risk neutral countries, the author uses linear functional forms for benefits and damages from emissions in (1) with $n > 3$ countries. The author considers simultaneous decisions in the second stage of the IEA membership game. Uncertainty is such that the damages to benefits ratio from emissions can take on one of two values with equal probability. In order to model the stock dynamics, he uses a simplified version of the model by Rubio and Ulph (2007) with two periods instead of an infinite horizon. The stock of pollution in the second period is given by the cumulative emissions of both periods. He compares four scenarios. The first is with no learning and with fixed membership where countries commit to their membership decision for both periods. The second is with fixed membership and learning, where countries learn about the realization of the damage between the first and second periods. The third is with no learning and variable membership, where an IEA is negotiated at the beginning of each period and membership is determined by the Random Assignment Rule as in Rubio and Ulph (2007). The fourth is with learning and variable membership. The author shows that learning is beneficial both in terms of participation in the IEA and global welfare under fixed membership as compared to no learning. Under variable membership, although learning increases the number of participants it nonetheless decreases global welfare.

Nkuiya *et al.* (2015) model a potential catastrophe as in Barrett (2013) and endogenize

the probability of the catastrophe. The authors consider a two-period dynamic IEA model similar to Ulph (2004) with a stock of pollution equal to cumulative emissions subject to some natural decay. In contrast to Ulph (2004), the authors consider quadratic benefits and damages from emissions. Moreover, uncertainty about the damage function is only present in the second period in contrast to Ulph (2004) where uncertainty is present in both periods. Crucially, the probability of occurrence of the catastrophe in the second period is increasing in the emissions of the first period. The main finding is that the largest stable coalition size is equal to two under exogenous uncertainty and no learning. However, in the presence of endogenous uncertainty, it is possible to have large stable coalitions including the grand coalition. Moreover, for a given coalition size, the pollution stock in both periods under endogenous uncertainty is always lower than under exogenous uncertainty. Comparing fixed and variable membership games, the authors show that full cooperation can be sustained for both periods under fixed membership but only for the first period under variable membership. Thus, this paper provides conditions under which the puzzle of small coalitions, found in IEA models with simultaneous move emissions games, is solved by endogenizing the uncertainty about damages.

The above papers focus on uncertainty about countries' own damage from emissions or benefit from abatement. A couple of papers, Espinola-Arredondo and Munoz-Garcia (2011) and Jakob and Lessmann (2012), consider the case where countries have less information about other countries' damages than their own. They show that each country may have an incentive to signal its private information in order to influence the IEA negotiations, which may improve global welfare. Since this is a policy relevant issue, further research is warranted in this area, given that the existing signaling models are restricted to a two-player setting and as such do not address the coalition formation process in detail.

2.2 Cooperative games

Another strand of the IEA membership literature follows the cooperative approach to game theory and applies the core concept of stability to examine coalition formation. A cooperative game with externalities $[N, v(S)]$ is a game with a set of N countries that is defined by its characteristic function $v(S)$, where $S \subseteq N$ and

$$v(S) = \max_{e_i, i \in S} \sum_S W_i(e_i, E), \quad (6)$$

with $W_i(e_i, E)$ being given by (1). The core of the game $[N, v(S)]$ consists of a set of imputations of the grand coalition such that these imputations yield higher payoffs for each country than what the country would get under any sub-coalition S where the payoff is given by $v(S)$.

Chander and Tulkens (1995, 1997, 2006) compare two different versions of the core, the α -core and the γ -core, which differ in their assumption regarding the behavior of the non-signatories. The α -core assumes that non-signatories minimize the maximized welfare of the sub-coalition S which weakens the sub-coalition, whereas the γ -core assumes that each non-signatory maximizes its own welfare taking as given the strategies of the coalition and other non-signatories. While Chander and Tulkens (1995, 1997) note that the previous cooperative game theory literature, including Shapley and Shubik (1969), Scarf (1971), and

Mäler (1989), relied on the α -core concept in the hope of generating a non-empty core, the behavioral assumption on the non-signatories in the α -core is rather strict. Thus, Chander and Tulkens' main contribution is to relax this assumption by introducing the concept of γ -core. The underlying concept of equilibrium consistent with the γ -core is referred to in the literature as the Partial Agreement Nash Equilibrium (PANE) with respect to a coalition S . In fact, the PANE concept bridges the cooperative and non-cooperative strands of the IEA literature, since it is also used as the underlying assumption when defining the behavior of signatories and non-signatories in non-cooperative membership games. It is implicitly assumed that defection by one country leads to the breakup of the entire coalition. Chander (2007) not only justifies the definitional assumption of the γ -core, but also shows that the γ -core imputations can be supported as equilibrium outcomes of a repeated non-cooperative game. Thus, the γ -core is also a non-cooperative solution concept.

The main contribution of this strand of literature is to determine the conditions under which the γ -core is non-empty. Since these conditions involve transfer payments, we will examine the results of these papers in Section 3, which discusses the IEA literature with side-payments and transfers.

2.3 Farsighted games

Thus far we have presented two polar cases for the stability concept in IEA games. On the one hand, we have the internal and external stability concept as represented by d'Aspremont *et al.* (1983), where defection by one country does not affect the membership status of the remaining countries. This concept is often referred to as myopic stability. On the other hand, we have the standard assumption made in the cooperative game theory literature stating that defection by one country leads to the breakup of the entire coalition (e.g. Chander and Tulkens, 1995, 1997). The concept of farsighted stability lies between these two polar cases. This concept also relies on the concept of internal and external stability. However, farsighted stability considers the case where each country, when deciding to join or leave the coalition, takes into account the fact that its own decision affects the membership decisions of other countries.

Diamantoudi and Sartzetakis (2002) consider the standard canonical two-stage IEA game identical to Diamantoudi and Sartzetakis (2006) except for the stability concept. Their main contribution is to formally define the concept of farsighted stability as applicable to the pure IEA game and to show that larger stable coalitions are possible than under myopic stability as found in Diamantoudi and Sartzetakis (2006). The intuition behind this result is that under farsighted stability the incentive to free-ride is reduced since each country cannot rely on the others to cooperate if it deviates unilaterally. This bridges the cooperative and non-cooperative strands of the literature in two ways. First, the reaction of signatories to defection is not as extreme as in the polar cases (no reaction versus coalitional shattering). Second, the analysis is extended to the case of coalitional defection, which is common in the cooperative literature but not common in the non-cooperative one. In this latter case, the authors find that the grand coalition can be supported as a coalitionally farsighted stable set. Lise and Tol (2004) find similar results when they apply the concept of farsighted stability and coalitional stability into a multi-regional computable general equilibrium model (WIAGEM). Biancardi and Villani (2011) find similar results using a static IEA membership

model where the choice variable is abatement rather than emissions. Biancardi (2010), using a dynamic model with a stock pollutant where countries play a differential game in the second stage, also confirms the results of Diamantoudi and Sartzetakis (2002, 2006) on the different implications of myopic and farsighted stability. Both Biancardi (2010) and Biancardi and Villani (2011) show that the higher the environmental awareness of countries the more likely that large coalitions are formed under farsighted stability.

De Zeeuw (2008) builds on Diamantoudi and Sartzetakis (2002) by analyzing farsighted stability within a dynamic setup. He begins by considering a static model where countries minimize the sum of the cost of abatement and pollution damage, which are both assumed to be quadratic. In this case, large coalitions are found to be farsightedly stable, in line with Diamantoudi and Sartzetakis (2002). He then introduces the dynamic context by assuming that emissions in one period are increasing in emissions of the previous period and decreasing in the abatement effort of the previous period. He shows that in the dynamic context if deviations are not detected immediately, large coalitions are only possible when the benefits from cooperation are low. The intuition is that since countries' abatement in a given period reduce future periods' emissions regardless of future cooperation, the threat to each country of being a non-signatory in a future world with a small coalition is reduced. Therefore, the incentive not to free-ride, which is strong under the static farsighted stability concept (Diamantoudi and Sartzetakis, 2002), is weakened in the dynamic case. In this way, de Zeeuw (2008) generalizes Barrett's (1994a) results on large but shallow coalitions that arise within a static context under myopic stability to a richer framework that includes dynamics and farsighted stability.

Another paper to implement farsighted stability is Benchekroun and Ray Chaudhuri (2015). Their focus is to determine the impact of cleaner technologies (reduction in the emissions-output ratio) on the stable coalition size. In the three-country case, they show that cleaner technologies may destabilize an otherwise farsightedly stable grand coalition, whereas with more than three countries the impact of cleaner technologies on the stable IEA size is ambiguous. These results are obtained both in a static context and a dynamic context with stock pollutant.

Thus far we have discussed three main types of stability criteria that are frequently used in the IEA literature, d'Aspremont *et al.* (1983) internal and external stability, the gamma core stability concept, and the farsighted stability concept. We note that there are a few other stability concepts that have been discussed by game theorists, which may be relevant to IEA analysis, as discussed in Thoron (1998) and Yi (2003). Whereas the d'Aspremont *et al.* (1983) stability concept considers only unilateral deviations by players, Thoron (1998) examines alternate stability criteria where players may deviate multilaterally. The author models the coalition formation game as a non-cooperative game in which players can make a binary choice of whether to cooperate or not. The author compares two refinements: the Strong Nash Equilibrium (SNE) and the Coalition-Proof Nash Equilibrium (CPNE). She shows that while a SNE coalition may not exist, a CPNE stable coalition exists and is unique.

Yi (2003) discusses further concepts of stability, such as the use of the partition function approach, linked closely to the works of Bloch (1996) and Ray and Vohra (1997), which allows for the possibility of externalities across coalitions since it allows the change in coalition structures to change the value of all coalitions, a feature that the characteristic function approach

does not take into account. The author also discusses the broader literature on coalition formation including the simultaneous move single coalition formation game (d'Aspremont et al., 1983), the simultaneous move open membership game with multiple coalitions (Yi and Shin, 2000), the simultaneous move exclusive membership game (Hart and Kurz, 1983), and the simultaneous move unanimity game (Von Neumann and Morgenstern, 1944).

2.4 Unresolved issues in pure IEA games

While the pure IEA games constitute the fundamental building block of the IEA formation literature, there are a few unresolved issues within this strand. For example, there is no consensus on a number of key features of the canonical IEA membership game. First, the literature is divided between simultaneous and sequential move games when it comes to the second stage of the membership game. These different assumptions can lead to quite different results. While under simultaneous move games the puzzle of small coalitions persists, under sequential games the puzzle of large but shallow coalitions prevails. Second, the literature is divided as to which stability criterion to apply i.e. the internal-external stability as used in the cartel literature, the gamma core concept from the cooperative games literature, and the farsighted stability concept, which acts as a bridge between these two polar cases. Third, there is no consensus on how to model dynamic games, i.e. fixed versus dynamic membership. It would be helpful to conduct empirical and experimental research in order to discern which of these modeling techniques are more applicable and under what conditions.

Another matter of concern for future research relates to the difficulty of accounting for heterogeneity across countries in terms of benefits and damages. The few existing papers that model asymmetric countries show that the intuitions from symmetric games do not readily extend to the asymmetric case. Further these models often rely on very specific assumptions in the second stage of the membership game, since the standard sequential or simultaneous move structures are difficult to apply in these cases. One possible way of modeling asymmetry satisfactorily might be to apply evolutionary game theory. There are currently very few papers that attempt this. Moreover, it is not easy to extend the canonical game into an evolutionary game framework where the identity of countries matters. Therefore, there is scope to conduct further research on this front.

A further limitation of using the internal and external stability criteria is that it is not useful for characterizing stable coalition structures in a model of some generality with heterogeneous countries (see the concluding section of Yi, 2003). When any coalition of size $s < n$ is stable, this approach does not specify which countries will form the coalition and which will get to free-ride. It also does not account for the existence of multiple equilibria that may arise with more than three countries, and therefore, cannot address the question of how to choose between these multiple equilibria. Even in the case of symmetric countries, the internal-external stability literature restricts attention to pure strategy asymmetric equilibria, without justifying why pure strategy asymmetric equilibria are more relevant than mixed strategy equilibria.

A further concern is that while there are a number of papers modeling uncertainty, few papers have modeled asymmetric information problems within membership games. Some recent attempts have been made in this direction, but there is scope for further research.

Moreover, without transfers, it is not surprising that the grand coalition may not form if

countries are heterogeneous. For instance, in a game with heterogeneous countries, a country that emits a lot but does not suffer much from pollution damage, cannot be induced to reduce its emissions unless it is compensated for doing so. Thus, in the presence of heterogeneity, transfers are a necessary mechanism for the formation of the grand coalition. Transfers and other issue linkages may resolve the puzzles of small and large but shallow coalitions, yet these features are not included in the pure IEA games. In Section 3, we therefore turn to papers that include these features.

3 IEA games with side-payments and issue linkage

As mentioned in the introduction, pure IEA games belong to the family of games where the coalition generates positive externalities for the non-signatories. This set of games as described above typically leads to dismal results in terms of cooperation with some exceptions such as in the case of repeated games or farsighted stability. Another approach to examining whether more cooperation is sustainable is to design games where the coalition generates some negative externalities for the non-signatories in addition to the standard positive externalities in terms of reduced emissions. This approach effectively creates a hybrid model of club games, where the coalition generates only negative externalities for the non-signatories, and pure IEA games, where the coalition generates only positive externalities for the non-signatories⁸

The extant literature has considered a number of different approaches of introducing negative externalities into the IEA game. We first consider the set of studies which take the pure IEA game and introduce the possibility of side/transfer payments. By not signing the IEA, non-signatories essentially lose the transfer payment, which constitutes a form of a negative externality generated by the coalition (i.e. the environmental club). These papers define side or transfer payments broadly by considering a form of transferable utility. Second, we consider cases where there exist multiple externalities among countries, for example through reputation, trade effects, or R&D spillover, in contrast to the pure IEA games, which assume that the only externality between countries is in the form of transboundary pollution (or its abatement). In these cases, the environmental club generates negative externalities for non-signatories by linking another source of externality to that of abating transboundary pollution, for example by linking trade sanctions or loss of goodwill to defection from IEAs.⁹

3.1 IEA games with side-payments

The literature on IEAs with side-payments considers both non-cooperative and cooperative games. We start by reviewing the former strand.

⁸For a discussion of positive and negative externality games within the context of coalition formation please refer to Yi (2003).

⁹As noted in Section 2, one way to achieve this is through compliance games where signatories have some punishment strategies against defectors. The difference between the pure compliance games and issue linkage games is that in the former the signatories typically harm themselves when punishing defection, whereas in the latter the punishment harms the defector and at the same time benefits the signatories.

3.1.1 *Non-cooperative games*

Recall the model of Carraro and Siniscalco (1993) as presented in Section 2. The authors extend their analysis of the pure IEA membership model with simultaneous emission choices in the second stage to include transfers among countries. The authors consider a chicken game instead of the prisoners' dilemma setup typically used in the pure IEA games. In the chicken game, some countries cooperate while others do not in equilibrium. They ask whether cooperating countries can transfer payments to non-cooperating countries to expand the coalition. The transfers are assumed to be self-financing, i.e. the total transfer must be lower than the gain arising from cooperation inside the coalition. They conclude that, despite being self-financing, such transfers could destabilize the existing coalition and as such, to be effective, they must be accompanied by some form of commitment from the cooperating countries. They also consider the case where some non-cooperating countries can transfer payments to other non-cooperating countries to convince them to cooperate. Carraro *et al.* (2006) applies these different transfer schemes to the CLIMNEG integrated assessment model.

Petrakis and Xepapadeas (1996) extend the model of Carraro and Siniscalco (1993) by considering country heterogeneity in damages with quadratic benefits from emissions and linear damages. They consider two types of countries: more environmentally conscious and less environmentally conscious. If the asymmetry is sufficiently large then the more environmentally conscious countries are willing to pay a side payment to the less environmentally conscious countries to convince the latter to reduce their emissions. This leads to their main result that heterogeneity in damages creates the possibility for transfers to stabilize large coalitions. The authors recognize that any transfer scheme entails the potential of moral hazard on the part of the less environmentally conscious countries. They propose a mechanism where transfers are conditional on total emissions, which allows environmentally conscious countries to convince the less conscious to emit at the globally efficient level. In such a case, the grand coalition can be a stable outcome. While Carraro and Siniscalco (1993) and Petrakis and Xepapadeas (1996) rely on some form of commitment to make the transfers effective, Botteon and Carraro (1997) extend the analysis to show that, even without commitment, transfers can be effective in a scenario with five heterogeneous countries. Botteon and Carraro (2001), discussed in Section 2, also consider a model of five heterogeneous countries. They compare two cooperative burden sharing rules, the Shapley value and the Nash bargaining solution. They find higher cooperation under the former rule with transfer payments.

Another paper to consider the standard two stage open membership game is Biancardi and Villani (2010), as mentioned in Section 2. They consider two kinds of countries characterized by different damages from emissions and the choice variable is the abatement level. The authors find that without transfers only small coalitions arise in equilibrium. However, when a transfer scheme in line with the Kaldor-Hicks criterion is implemented, they find that the grand coalition may also be sustained as a Nash equilibrium. When such transfers are implemented, they are Pareto improving for all countries. As a result, when a country deviates from the first-best, all countries are made worse off since the disagreement outcome is a smaller coalition. Therefore, such transfers ensure the stability of the grand coalition.

As discussed in Section 2, Barrett (2001) considers a simplified version of the canonical

model with two types of asymmetric countries and the possibility of having side-payments. When the asymmetries in benefits from forming a coalition are strong, he finds that allowing countries with the higher benefit to provide side-payments to countries with low benefits, increases the stable size of the coalition and the participation from both types, as compared to the case without side-payments. Moreover, the larger coalition generates higher global welfare. Taken together, these results indicate that the provision of side-payments can overcome the twin problems of small coalitions or large but shallow coalitions provided that asymmetry among countries is large enough. Fuentes-Alberto and Rubio (2010), after generalizing the model of Barrett (2001) by considering continuous instead of discrete choice of emissions, also find that large and deep coalitions are stable if the heterogeneity across countries is in terms of pollution damage but not if it is in terms of the abatement cost. These two papers suggest that cooperation from low damage countries can be bought by countries with high damages if (i) asymmetry is large enough, and (ii) self-financing transfers are possible.

McGinty (2007) introduces tradable emissions permits into the standard two-stage IEA model where the permits are used to transfer payments between countries. The author considers simultaneous abatement decisions in the second stage of the game with quadratic benefits and cost functions. He considers a coalition to be stable when the aggregate payoff to members is larger than the sum of the payoffs to individual countries from defection. Within this context, he designs a transfer scheme based on the benefit-cost ratio, which distributes the remaining surplus after each coalition member receives a payoff equivalent to what each coalition member would get as a non-signatory. He then applies this rule to a scenario with 20 heterogeneous countries and demonstrates that the puzzle of large but shallow coalitions (see Barrett, 1994a) can be broken in such a case. Dellink (2011) also designs a permit allocation scheme across countries which stabilizes larger coalitions within the integrated assessment STACO model.

While most papers model transfers among countries directly as opposed to modeling the presence of an arbitrator in the form of a supra-national authority, Batabyal (2000) explicitly models the role of such an authority in designing and implementing the transfers among countries. The author considers a scenario where the authority makes separate transfers to governments and firms located in a given country, and it does not observe the abatement efforts of firms perfectly. The author compares two possibilities, one where the contract is signed *ex ante* by the authority, i.e. before the uncertainty is resolved, and one where it is signed *ex post*. In the *ex ante* case, in the first stage of the game, the authority signs the contract with the governments and the firms. In the second stage, the firms observe their abatement technology and send a signal to their government. In the third stage, the firms choose their abatement level. In the fourth stage, the government observes the emissions and sends a report to the authority. In the last stage, the transfers are implemented. In the *ex post* case, the structure is similar except that the uncertainty is resolved before the contract is signed by the players. While the authority would prefer *ex ante* contracting to *ex post*, the authors shows that *ex post* contracting is more likely to be renegotiation-proof. Also, these contracts typically do not yield the first best outcome given the asymmetric information and the possibility of collusion between the firms and their governments.

While most papers in this stream of the literature consider game theoretic models, Tol (2001) consider an integrated assessment model without transferable utility. He uses the

Climate Framework for Uncertainty, Negotiation and Distribution (FUND) model, which captures changes in population, economic activity, emissions, and atmospheric composition. He finds that monetary side-payments do not achieve large emissions reductions or generate full cooperation. Large stable coalitions, if found, neither include the most polluting countries nor the countries most damaged by climate change. Another paper to examine transfers empirically is by Finus *et al.* (2006) who consider the STACO model with 12 regions. They find that transfer schemes are beneficial for stabilizing otherwise unstable coalitions. By contrast, Bosetti *et al.* (2013) use the WITCH model to show that only the grand coalition can achieve sufficient GHGs mitigation but that it is internally unstable under various combinations of parameter values, and despite allowing for transfer payments.

A more recent meta analysis of computational IEA models is presented by Lessmann *et al.* (2015) who compare the effect of several different transfer schemes on IEA stability across five different integrated assessment models: RICE, CWS, STACO, MICA, and WITCH. While MICA considers 11 regions, WITCH considers 13. The models also vary along several other dimensions, such as, the time horizon, discount rate, and carbon price growth rate. The authors find a robust result across models indicating that, without transfers, stable coalitions are generally small and achieve little in terms of welfare improvement. These results are in line with the dismal game theoretic results discussed in Section 2. Comparing across different transfer schemes, another robust result is that only transfer schemes which redistribute payoffs within a coalition such that it becomes internally stable (referred to as the potential internal stability (PIS) criterion, Carraro *et al.*, 2006) are effective.

3.1.2 *Cooperative games*

The issue of whether transfers can increase the likelihood of cooperation between countries is also explored in the cooperative games literature. As noted by Tulkens (1998), the cooperative literature introduced transfer schemes which are linked to the equilibrium abatement levels of different countries rather than being lump-sum. Such transfer schemes have since been introduced in both non-cooperative and cooperative IEA games.

Chander and Tulkens (1995, 1997) design a transfer rule which lies within the γ -core (a non-empty core, as described in Subsection 2.2), i.e. the grand coalition is stable when this rule is applied. The rule stipulates that the transfer can be negative or positive for a given country. Given country heterogeneity, the transfer has two components. The first component is related to the cost of abatement such that it is negative (positive) for countries whose cost of abatement under full cooperation is smaller (larger) than under non-cooperation. This is equivalent to the following: The first component is negative (positive) for countries whose benefits from emissions under full cooperation are larger (smaller) than under non-cooperation. The second component, which represents how the transfers are financed, is always negative and depends on each country's marginal environmental damage relative to the aggregate damage. Therefore, under this rule, a country with high damage and low benefits from emissions pays others, while a country with low damage and high benefits from emissions receives transfers in equilibrium. These conclusions hold for the welfare maximization problem, defined in (6), as shown in Chander and Tulkens (1997), and for the dual cost (of abatement plus damage) minimization problem as shown in Chander and Tulkens (1995). Helm (2001) generalizes the main result of Chander and Tulkens (1997) by

proving that the γ -core is non-empty under standard convexity assumptions on the payoff function. Moreover, recall that in the non-cooperative strand of the literature using the internal and external stability criteria of d'Aspremont *et al.* (1983), a coalition is stable when there are no incentives for unilateral deviations. By contrast, the resulting positive result in the cooperative game literature, that is the stability of the grand coalition, is robust to deviations by one country or by a group of countries (see Tulkens, 1998).

Germain *et al.* (1997) introduce a transfer scheme similar to that of Chander and Tulkens (1995, 1997) into a differential game where countries behave strategically in the presence of a stock pollutant. Effectively, they develop a hybrid model with elements of cooperative game theory and differential games. They modify the transfer scheme such that it is consistent with the dynamic setup and ensures that the grand coalition is stable and coalitionally rational. While Germain *et al.* (1997) uses open-loop strategies, Germain *et al.* (2003) design the corresponding transfer scheme for the closed-loop case.

A number of papers apply the γ -core concept to computational general equilibrium models. Eyckmans and Tulkens (2003) introduce the transfer scheme of Germain *et al.* (1997) into the CLIMNEG world simulation (CWS) model, which is based on the RICE model. They characterize the parameters of the transfer scheme necessary to stabilize the grand coalition within the CWS model. Bréchet *et al.* (2011) also use the CWS model, but their focus is to compare the γ -core stability concept to the internal and external stability concept of d'Aspremont *et al.* (1983). While the CWS model considers 6 regions, the STACO model is based on 12 regions. More recently, Yang (2015) applies the γ -core to an integrated assessment model, the RICE model (developed by Nordhaus and Yang, 1996), which is a computational general equilibrium model of climate change with up to six regions: the United States (USA), European Union (EU), other high-income countries (OHI), China (CHN), Eastern European countries and former Soviet Union (EEC), and the rest of world (ROW). The welfare function of each region is different to represent country heterogeneity across regions. The author asks whether the shrinking core theorem of Arrow-Debreu carries over to this setting. This theorem states that as the number of homogenous agents increases the core shrinks. In contrast to the Arrow-Debreu setting, Yang (2015) considers a finite number of heterogenous agents (regions) in the presence of negative externalities between the regions arising from the accumulation of a pollution stock. The results, from numerical simulations, show that the core shrinks as the number of regions increases. This indicates that an increase in heterogeneity makes the grand coalition less likely to be a stable outcome.

In sum, the literature on transfers presents mixed results depending on the design of the transfer scheme and the degree of heterogeneity across countries. On the one hand, transfers can be seen as a natural way to increase cooperation by buying-in cooperative behavior. On the other hand, if transfers are required to be self-financing they may destabilize existing coalitions rather than expanding them. Although the previous papers reviewed here define transfers broadly either as monetary or transferable utility, in practice transfers can be implemented in various ways such as by reducing trade barriers or transferring technologies. An alternative stream of the literature examines these avenues of transfers under the framework of issue linkages, a term first coined by Cesar and de Zeeuw (1994, 1996). These papers link other sources of externality to that of abating transboundary pollution found in the pure IEA games.

3.2 IEA games with interconnection

The papers we discussed thus far model an emissions/abatement only game, or a single subject game. The literature on interconnected games, however, focuses on the concept of issue linkage to achieve stability. Folmer *et al.* (1993) and Folmer and van Mouche (1994) are among the first papers to formalize the linkages between several sources of externalities among countries within the context of IEA games. The authors specify two types of interconnected games. The first is direct sum games, which are one shot games. They consider multiple isolated one shot games, for example, a prisoners' dilemma abatement game and a club game. The authors define the direct sum game, also known as the trade-off sum game, by assigning weights on each player's payoffs from the isolated games and by summing the weighted payoffs across the isolated games. As a result, the strategies in the trade-off sum game are the weighted best response functions of the isolated games. This allows free-riding to be mitigated since, for each player, the non-excludable benefits of abatement are augmented by the excludable benefits stemming from becoming a member of the club. The second is tensor games, which are repeated games, whereby, the direct sum game is the stage game of the tensor game. The authors extend the existence theorem of Nash equilibrium to the interconnected context and prove that interconnection augments the strategy space available for punishing defection from cooperation. They show that for tensor games, interconnecting two isolated games e.g. environment and trade, makes cooperation more likely. This provides an alternative to side-payments to promote cooperation. Folmer and van Mouche (1994) argue that for interconnection to be effective in promoting cooperation via the exchange of concessions, countries need to have reversed interests on different issues of approximately the same size. Kroeze-Gil and Folmer (1998) extend the analysis to the cooperative game setting and find conditions under which the core of the interconnected game is non-empty. They illustrate the positive result by constructing an example where the core of the isolated game on abatement is empty but the core of the interconnected game is non-empty.

As noted by Finus (2003), issue linkage may be analyzed either under compliance models with repeated games (e.g. Folmer *et al.*, 1993; Folmer and van Mouche, 1994) or under membership models as in Carraro and Marchiori (2004) who endogenize the choice of countries on whether to engage in issue linkage. Carraro and Marchiori (2004) consider two isolated games: a public good provision game (abatement) and a club good game (trade or R&D). The authors consider three stages. First, countries decide whether to link negotiations over the two games. This is determined using the unanimity voting rule, whereby consensus by all countries is required. If countries decide to link the issues in the first stage, they decide whether to sign the linked agreement in the second stage. If not, they decide whether to sign separate agreements on the two issues. Third, they choose their equilibrium strategies regarding abatement and the club good. They find that countries are more likely to link issues if each country gets more from the linked agreement than by free-riding on the agreement, that is if the benefits from the public good provision resulting from the larger coalition are sufficiently large. Linking the two agreements has a potential drawback, that is it may cause the agreement on the club good to become smaller (e.g. fewer countries willing to share technologies or join free trade areas). The increased payoff for each country from the linked agreement must compensate any loss resulting from the smaller agreement on the club good.

Since the concept of interconnected games was brought into the IEA context, three strands of the literature have evolved focusing on different club games that are linked to an IEA membership game. Barrett (1994b) and others propose to link environment discussions with trade negotiations. Carraro and Siniscalco (1995) and others propose linking IEAs with Research and Development (R&D) cooperation. Hoel and Schneider (1997) and others attempt to model issue linkage by introducing reputation effects. All these proposals rely on the idea of linking the environmental game, which is a public good provision game where signatories cannot exclude non-signatories from enjoying the benefits of improved global environmental quality, to a stable club good game, where the non-signatories can be excluded from the benefits resulting from the club membership.

3.3 IEA games with trade effects

Barrett (1997a) presents one of the first papers to analyze the use of trade sanctions by signatories against non-signatories of an IEA as a threat to induce cooperation. The author presents an environment-trade game with N countries with one representative firm in each country with fixed location. Firms produce a homogenous tradable good with segmented markets at the country level. He considers only traded goods that are linked to environmental problems. The game has four stages. In the first stage, countries decide whether to become members of an IEA. In the second and third stages, signatories and non-signatories choose their abatement standards sequentially. In the fourth stage, firms choose their output (emissions) levels. His main finding is that linkage of an IEA to trade restrictions can increase participation, as was the case with the Montreal protocol on CFCs, provided that the sanctions are accompanied by a minimum participation clause. Dong and Zhao (2009) extend the work of Barrett (1997a) by endogenizing the tariff level chosen by countries, instead of focusing on either free trade or a complete trade ban. They find that in equilibrium the tariff is positive and it is not clear that allowing for trade linkages increases cooperation. Finus and Rundshagen (2000) qualify the results of Barrett (1997a) by considering the pollution-haven-hypothesis within the context of issue linkage. They consider a three country model with trade tariffs and endogenous location of polluting firms. They find instead that linking trade to environmental negotiations can potentially reduce both welfare and participation when countries behave strategically to attract firms.

Kemfert (2004) and Kemfert *et al.* (2004) apply the concept of issue linkages through trade on the WAGEM model, which is an intertemporal computational general equilibrium regional trade model for the global economy. The model consists of 11 regions that are linked via bilateral trade flows. They find that trade effects on abatement costs and emissions can either increase or decrease incentives to reduce emissions and to cooperate on abatement.

Conconi and Perroni (2002) introduce a cooperative game theoretic framework that models the trade-environment link explicitly. They propose a multidimensional core concept and apply it to a generic trade-environment game with three identical countries. Their results suggest that linking trade and environmental decisions into one super-agreement can have positive effects in terms of better cooperation only if environmental problems are relatively small in terms of welfare costs and benefits when compared to the costs and benefits of trade

policies.¹⁰

Egger *et al.* (2011) provide empirical evidence on the impact of trade and investment liberalization on a country's propensity to sign an IEA. They use data on 353 IEAs, including IEAs on environmental issues other than climate change, that have been ratified between 1960 and 2006. They use a dynamic panel data model covering 105 countries and show that both trade and investment liberalization increase participation in IEAs. These results are in line with (Neumayer 2002b) who finds that trade openness is an important determinant of signing IEAs.

3.4 IEA games with R&D

As discussed earlier, one type of issue linkage studies R&D effects on the cost of abatement. Carraro and Siniscalco (1997) use an IEA model with R&D effects on abatement technology and find that this linkage may not increase participation, while Katsoulacos (1997) and Ruis and de Zeeuw (2010) find the opposite. Botteon and Carraro (1998) consider R&D linkage with 5 heterogeneous players/countries to study the impact of R&D linkage on the stability of IEAs and on profitability of participants when linkage is used as a strategy. Using calibration and two different cost sharing rules among coalition members (Nash and Shapley), they find that the stable coalition is usually larger under linked negotiations, but it may not be optimal. The optimal coalition is found to be smaller than the stable one. And due to heterogeneity, countries may disagree on the coalition they find optimal, such that no equilibrium may exist. More recently, Biancardi and Villani (2015) extend the dynamic model of Biancardi and Villani (2014) (discussed in Section 2) to include R&D linkages and they find this leads to increased cooperation compared to the pure IEA game.

As discussed in the previous subsection, Kemfert (2004), using a computational general equilibrium model finds that participation in IEAs increases if the R&D cost sharing and the IEA are linked because more innovation leads to cheaper ways to reduce emissions. At the same time, there may be positive spillover effects on the non-signatories from the innovation undertaken by the signatories. Buchner *et al.* (2005) apply the basic model of Carraro and Siniscalco (1997) to the ETC-RICE integrated assessment model with six regions, where technical change can lead to increased economic growth via increased returns to scale and reduces the emissions-output ratio. They find that linking R&D cooperation to the IEA is not effective in inducing the US to join the IEA since the linkage is based on a non-credible threat. This is because countries like the EU, Japan and Russia cooperate with the US on technological innovation despite free-riding on the part of the US in terms of abating emissions.

An alternative approach of the role played by R&D in mitigating emissions is explored

¹⁰Some related papers, which do not explicitly model coalition formation, include the following. Diamantoudi *et al.* (2015) generalize the Shapley value applied in single issue cooperative games to multiple issue games that exhibit externalities within each issue and across issues e.g. environment and trade games. They propose a way to extend the Shapley axioms to a linked setting and thereby propose a surplus sharing solution that respects Pareto efficiency and a set of desirable properties. Another contribution is found in Leroux and Marrouch (2012) who propose a linked bargaining solution that ties-in trade and environmental surplus sharing games. This solution is based on the weighted average of relative bargaining powers of players on each separate issue.

by Barrett (2006), who proposes a two-treaty framework. The first treaty is for promoting cooperation on R&D, while the second is on the adoption of the cleaner technology arising from the R&D. He finds that such a bundle of treaties is not more successful at achieving large coalitions and reduced emissions than a standard IEA on abatement, unless the cleaner technology exhibits increasing returns (network externalities). Hoel and de Zeeuw (2010) extend the Barrett (2006) model by allowing the adoption cost of the technology to vary with the level of R&D and show that this leads to larger stable coalitions and higher welfare.

3.5 IEA games with reputation effects

Hoel and Schneider (1997) and Cabon-Dhersin and Ramani (2006) attempt to model issue linkage by introducing reputation effects. These attempts to solve the small coalitions puzzle in IEAs in the non-cooperative game context suggest that reputation costs resulting from defection can be costly enough to act as deterrents. As such larger coalitions become plausible.

Hoel and Schneider (1997) extend the basic IEA membership setup of Carraro and Siniscalco (1993) by introducing a non-environmental cost to each country of not joining the IEA. One interpretation of such a cost is a loss of goodwill or reputation in international forums. This cost is country specific and depends on the number of countries in the coalition. Allowing for this cost increases the size of the stable coalition. Cabon-Dhersin and Ramani (2006), instead of modeling the cost of non-participation, introduces a non-environmental benefit to each country of joining the IEA. This benefit is increasing in the size of the IEA. They find that larger coalitions are possible, including the grand coalition. They also find that the higher the non-environmental benefit of joining the IEA, the greater the abatement and global welfare. This result is in contrast to both the puzzle of small coalitions (e.g. Carraro and Siniscalco, 1993) and the puzzle of large but shallow coalitions (e.g. Barrett, 1994a).

3.6 Unresolved issues in IEA games with linkages and side-payments

Although, theoretically, it is argued in the inter-connected games literature that issue linkages are helpful in sustaining cooperation, when linkages are examined within a practical context the results are mixed. For example, the effects of linking trade restrictions and R&D treaties to IEAs have been shown to be stabilizing or destabilizing depending on a number of factors such as the degree of heterogeneity and the functional forms of benefits and damages from emissions. The same also holds for the effectiveness of transfer payments. In order to resolve some of these issues it may be beneficial to conduct a meta-analysis to identify the conditions under which transfers and linkages can be effective in promoting participation into IEAs. Also, some empirical or experimental evidence can help to disentangle some of the conditions under which transfers and linkages can be effective.

We note that neither the Kyoto Protocol (1997) nor the Paris Agreement (2015) makes linkages between abatement and trade tariffs. The WTO and UNFCCC negotiations have thus far been independent of each other. At the same time, several economists have raised the question about whether games with linkages have the potential to generate optimal outcomes (see Barrett, 2001; Brewer, 2003, 2004; Frankel, 2005; Stiglitz, 2006; Nordhaus,

2015; and others). Thus, it becomes relevant to examine issue linkage as a possible route for the future design of IEAs. Existing models of issue linkage often ignore some important aspects of the coalition formation problem. Thus, there is scope for further research in this direction.

4 Novel approaches to the IEA formation problem

In this section, we highlight some areas of research that have recently gained traction, and provide directions to potential routes that the IEA literature may follow. These include studies that examine the design of IEAs from a political economy perspective, behavioral issues from a theoretical and an experimental perspective, open economy externalities, and externalities arising from allowing countries to adapt to climate change.

4.1 Design of IEAs and political economy aspects

When we think about the political economy aspects of IEAs, a number of inter-governmental factors play a role, for example, as noted by Carraro and Siniscalco (1998), transfers, issue linkage, and threats, which have been discussed in detail in Section 3. In this subsection, we focus on a few other factors focusing mainly on the design of the IEAs. In much of the existing IEA literature, surveyed in Sections 2 and 3, it is assumed that governments maximize their countries' social welfare when deciding to join an IEA. In reality, governments often have to balance different objectives from winning votes from elections to satisfying demands from lobby and pressure groups and budgeting prioritization issues. As noted by Wangler *et al.* (2013), these intra-governmental aspects are important for the design of IEAs. While most of the extant IEA models focus on the negotiation stage as discussed by Barrett (1998), there are several stages before and after the negotiation stage, which influence the final outcome, such as pre-negotiations, ratification, implementation, and renegotiations. In this subsection, we review the studies that attempt to open the black box of IEA design by looking into intra-governmental and inter-governmental issues at these various design stages.

4.1.1 Design issues of IEAs

Ecchia and Mariotti (1998) focus on the role played by a potential supra-national authority in IEA negotiations. They consider both prisoners' dilemma and chicken games and show that in a one shot game, with three players, it is difficult to achieve a Pareto efficient outcome. Their main message is that a supra-national authority may facilitate in reaching the full cooperation (Pareto superior) outcome by allowing countries to communicate with each other. This is the case even when the supra-national authority has limited powers, e.g. no ability to punish defectors.

Barrett (2002) is one of the first papers to model the pre-negotiations stage by allowing countries to vote, under the veil of ignorance, i.e. without knowing if they will be a coalition member or not, on whether they want a sub-coalition, referred to as incomplete focal treaty, or the grand coalition, referred to as the consensus treaty. That is, either the countries can choose to form the grand coalition first and agree to modest abatement efforts, which will

maximize the welfare of all countries, or they can choose to form the stable sub-coalition where the coalition members make larger abatement efforts. He considers a repeated abatement game framework with symmetric players and no discounting, and finds conditions under which the consensus treaty is better than the incomplete, focal treaty in terms of welfare. Finus and Maus (2008) extend the results of Barrett (2002) from repeated games to membership games. Unlike Barrett (2002), they allow countries to choose between the incomplete, focal treaty and all other possible forms of consensus treaties including the grand coalition, i.e. in their model countries are effectively choosing how modest their abatement levels will be, which is equivalent to choosing how large the coalition resulting from the consensus treaty will be. Along the same lines, Barrett and Stavins (2003), when discussing how to improve the UNFCCC framework, propose that emissions reduction targets should be modest in the short-run and only increase in stringency over time. The consensus treaties considered in Barrett (2002) and Finus and Maus (2008) and the various alternative architectures proposed by Barrett and Stavins (2003) do not achieve the first-best outcome. A more recent paper discussing second-best agreements is by Courtois and Haeringer (2012) who consider an IEA membership game with symmetric countries and abatement as choice variable. However, contrary to most of the membership literature where the coalition maximizes joint welfare, an IEA is assumed to come into force when countries commit to a binding environmental target, which does not necessarily maximize the joint welfare of coalition members. They also consider a minimum participation clause and find that the IEA that is internally and externally stable is exactly equal to the size stipulated by the minimum participation clause and achieves the pre-announced abatement target. In the game, the target and the minimum participation clause are both negotiated. They find that the set of self-enforcing IEAs of different size and abatement levels include the first-best and several second-best agreements.

In fact, Courtois and Haeringer (2012) build on a strand of the literature that focuses on the Minimum Participation Clause (MPC) in IEA negotiations. As discussed in Section 2, Carraro and Siniscalco (1993) were among the first to model the MPC. Other papers to model the MPC with symmetric countries and complete information include Rutz (2001), Courtois and Haeringer (2005), and Carraro *et al.* (2009), while Weikard *et al.* (2015) allow for asymmetric countries. Most of these papers find that the stable IEA will be of the same size as the MPC, which may include all countries resulting in a stable grand coalition. When the MPC is chosen endogenously, Weikard *et al.* (2015) show that the grand coalition may not emerge since the bargaining power of countries is asymmetric. Some papers consider uncertainty when modeling the MPC, e.g. Black *et al.* (1993) and Harstad (2006). While Black *et al.* (1993) show that it may not be optimal to implement the MPC containing all countries in a symmetric country setup, Harstad (2006) show that a stable equilibrium MPC may not evolve endogenously when countries are asymmetric (see Wangler *et al.*, 2013 for further details on the MPC literature).

Another approach to the design of IEAs is proposed by Karp and Zhao (2010) who study how emission permit trading among countries that join an IEA affect the level of abatement and participation. While permit trading helps to achieve the efficient abatement level, it may reduce participation in the IEA. They consider two policy designs. The first is an escape clause, which prohibits trade in permits across countries. If a country exceeds its emissions target, it must pay a fine. Thus, combining emission trading with an escape clause, results in a successful IEA, which achieves efficient abatement and high participation. Second, they

consider a safety valve by which a country that exceeds its target abatement level can get permits, which can be resold to other countries. They find that safety valves are not effective in increasing participation.

Another design issue relates to the scope of the IEA, in particular whether it should cover abatement only or both abatement and R&D efforts on abatement. Goeschl and Perino (forthcoming) consider international agreements on emission abatement only without any related agreement on R&D in green technologies. This is sometimes referred to as an incomplete contract in the related literature. The authors examine non-cooperative investment in R&D of cleaner technologies when an IEA on emission reduction is impending.¹¹ The authors assume that there exist international intellectual property rights that are perfectly enforceable, which enable the patent holder to charge a license fee for the abatement technology after signatories and non-signatories of the IEA have decided on their abatement commitment. The patent holder is able to charge a higher fee when the abatement level committed to by the IEA is high. This provides a perverse incentive to countries, leading to fewer signatories joining in equilibrium and to weaker commitments on abatement levels, which in turn reduces the demand for the abatement technology. This ultimately results in the holdup problem, i.e. lower investment in R&D in abatement technologies. Endres and Rundshagen (2012) rather than focusing on countries' R&D incentives, analyzes firms' incentives to diffuse and adopt cleaner technologies in a framework in which governments negotiate an IEA on emission reduction. They show that unlike the closed-economy setup, where emission taxes rather than emission standards provide greater incentives for firms to innovate, in an open economy setup, the opposite may hold.

In a similar context of commitment problems as in Goeschl and Perino (forthcoming), Battaglini and Harstad (2016) extend the analysis to endogenize the length of commitment by signatories, and compare the cases of incomplete contract and complete contract, the latter being an agreement on both abatement and R&D. The authors find that the puzzle of small coalitions persists in the case of complete contracts. However, in the case of incomplete contracts, unilateral deviations reduce the length of the agreement, resulting in lower R&D levels. These perverse effects act as a credible threat, which discourages free-riding. Thus, the authors conclude that incomplete contracts may help in sustaining meaningful cooperation. Thus, their results provide a cautionary tale against increasing the scope of the IEA to include R&D cooperation, which has been often advocated by policymakers involved in climate negotiations. Harstad (2012) also contributes to this debate on mechanism design, as follows. The author shows that countries have similar incentives to free ride on R&D cooperation as on emission reduction. This is the case because each country, by reducing current investment, strategically induces others to pollute less and invest more in the future. The author shows that these problems can be overcome by negotiating a long lasting IEA with small quotas of emission reduction. Better negotiated outcomes can be achieved if the IEA is tougher to satisfy towards the end of its duration and for countries with lower investment cost in R&D, and if the IEA is renegotiable.¹²

Another mechanism is proposed by Martimort and Sand-Zantman (forthcoming). The

¹¹Their framework is similar to Barrett (2006), except that it does not allow for international cooperation on R&D.

¹²These findings are in line with the suggestions made by Olmstead and Stavins (2012) for the post-Kyoto phase for tackling climate change.

authors consider a scenario where countries set up a green fund at the international level. Countries that emit at levels close to their business-as-usual (BAU) levels contribute to the fund, while those that abate significantly away from their BAU receive an abatement subsidy. They propose that the contributions to the fund be limited so as not to discourage countries from participating in the IEA. They propose a menu that offers the choice between a fixed contribution and a per-unit subsidy together with another contribution. Using numerical simulations, they show that efficient countries choose the incentive option while inefficient countries choose to make fixed contributions to the green fund, and that the resulting welfare is higher than the BAU level but lower than the first-best outcome.

4.1.2 Political economy aspects of IEAs

An important intra-governmental aspect that can affect the success of an IEA is strategic voting. Buchholz *et al.* (2005) show, within a symmetric two-country framework, that the median voter will choose a less environmentally friendly government if voters know that the country is inside an IEA than if the country behaves non-cooperatively. This behavior is explained by the desire to obtain a larger Nash Bargaining surplus, which is the sharing rule applied to the IEA. Thus, the formation of an IEA may lead to lower environmental quality and welfare than the non-cooperative equilibrium. This pessimistic result indicates that even if stability is assured, the grand coalition may be Pareto dominated by the non-cooperative outcome in the presence of environmentally conscious voters who behave strategically.

Another relevant intra-governmental aspect is lobbying. Haffoudhi (2005) considers the standard two-stage IEA membership model with symmetric countries that is augmented with a pre-IEA lobbying stage. The government's payoff function is given by the difference between benefits from global abatement and the country specific abatement cost plus the contribution from the local special interest group. There are two types of special interest groups: The environmentalists, who seek to maximize the difference between the benefits from abatement and the damage from emissions, and the industrialists, who seek to minimize the abatement cost. For both types, the lobbying contribution to the government is equal to the surplus generated by their lobbying. The author finds that when lobbies are located in both signatory and non-signatory countries, the puzzle of small coalitions is maintained. However, when lobbies are located in signatory countries only, the grand coalition is stable. Unlike Haffoudhi (2005), Altamirano-Cabrera *et al.* (2007) assume that signatory countries do not obtain any foreign lobbying revenues and find that the grand coalition is no longer stable.

Using panel data from 170 countries, Fredriksson *et al.* (2007) provide empirical evidence in support of the claim that environmental lobbying is effective in facilitating a country's participation in an IEA. Other empirical papers examine different factors which affect the likelihood of countries signing IEAs. While Neumayer (2002a) provides evidence that countries are more likely to sign an IEA if they have a democratic regime, Roberts *et al.* (2004) and Von Stein (2008) examine related factors, in particular, showing that NGOs and strong institutions play an important role.

4.2 Behavioral economics and IEA games

As noted by Cramton and Stoft (2012), in the standard two-stage open membership IEA game, in the first stage, it is assumed that countries who choose to become signatories to the IEA not only sign the climate treaty but also ratify it at the same time, thereby committing to the prescribed future emissions levels. In practice this is often not the case. For example, the US signed the Kyoto (1997) protocol but did not ratify it. A recent set of studies has tried to model this process explicitly by applying the concept of focal points (Schelling, 1960). These papers recognize that the stage at which countries decide whether to sign the treaty requires negotiations since these countries need to mutually agree on the design to be adopted by the IEA, for example, how to allocate a global cap on emissions or to choose a uniform global carbon price. An interpretation of these papers is that they present a description of a stage game preceding the standard two-stage membership game. In this stage game, there are different possible policy designs to choose from, which results in a coordination problem that a focal point can help to solve. According to Weitzman (2014), "a focal point of a n-party coordination game is some salient feature that reduces the dimensionality of the problem and simplifies the negotiations by limiting bargaining to some manageable subset, hopefully of one dimension". A natural candidate for a focal point is a single universal carbon price (see Weitzman, 2015a, 2015b). The existence of the focal point increases the number of countries signing the treaty and thereby increases the number of potential players who will subsequently ratify it. The lack of a focal point has, thus, been argued to be one of the reasons behind the failure of cooperation on climate. At the same, these papers acknowledge that in addition to the existence of a focal point, for the IEA to be successfully implemented, the treaty design must "embody countervailing force against narrow self interest by automatically incentivizing all negotiating parties to internalize the externality" (Weitzman, 2015b). For example, Cramton and Stoft (2012) suggest the creation of a green fund, which makes transfers from rich to poor countries, while Nordhaus (2015) suggests using tariffs. The lack of a supra-national authority to impose a common global climate policy necessitates a multilateral bargaining framework, which entails high transactions costs among countries. As noted by Weitzman (2015a), a focal point provides a way of circumventing these transaction costs.

In contrast to the previous papers, which discuss in detail the application of a focal point, Nordhaus (2015) assumes the existence of a focal point and then focuses on the standard two-stage IEA membership game. He proposes a way of implementing a global climate treaty by converting the climate change problem from a prisoners' dilemma into a club good problem. While the idea is similar to issue linkage, which links a public good provision game to a club good game, the approach to implementation in Nordhaus (2015) is different because he relies on the forming the club around a focal point. He proposes that the focal point consist of a carbon price that would be uniformly implemented across the club members. Countries would have the choice to use different policy instruments, such as carbon tax or cap-and-trade, to reach the uniform carbon price. The club members would charge a positive tariff on imports from non-club members. He applies this concept to the Coalition-DICE model, which is an integrated assessment model consisting of 15 regions, based on the DICE-RICE model, and is specifically designed to find stable climate coalitions. Nordhaus (2015) finds that as long as the carbon price, which is typically set equal to the global social cost of

carbon, is not too high (not higher than \$50 per ton), small tariffs against non-club members are sufficient to foster high participation with substantial gains from cooperation in contrast to the large but shallow coalition paradox (e.g. Barrett, 1994a). Since it is in countries' self-interest to join the club, the resulting treaty is self-enforcing. It is worth noting that the model of Nordhaus (2015) considers the stability concept based on the Coalition-Proof Nash Equilibrium as opposed to much of the previous literature on non-cooperative IEA games that consider only unilateral deviations.

Another behavioral aspect of IEA negotiations that has been recently discussed is that of the tipping point. Heal and Kunreuther (2012) present the idea as follows: In a game with strategic complementarity, if a sufficiently large subset of influential countries choose to cooperate, i.e. the tipping point in terms of the number of cooperating countries is reached, they may induce others to do the same. The policy implication is that if large emitters like the US and China adopt unilateral or bilateral measures to combat climate change, they should be encouraged since they may be triggering a move towards such a tipping point. As noted by Nyborg (2015), under the standard assumptions on preferences used in the IEA literature, which typically results in a prisoners' dilemma framework, it is not likely that the tipping point concept can be explained. However, when reciprocity is allowed for in the behavior of countries, Nyborg (2015) shows that the tipping point concept may be explained since non-signatory countries make an effort to reward the countries that have already signed the treaty.

Another set of papers examines coalition formation from an experimental perspective. These papers set up laboratory experiments to determine conditions under which the behavior of players is more conducive to cooperation.

Bosetti *et al.* (2015) consider an investment game where countries choose between a high return but dirty technology and a low return but clean technology. There exists a tipping point such that if the joint investment in the dirty technology is lower than this point, countries will get the relevant payoff with certainty, whereas if it exceeds the tipping point, the payoff will be random. The authors consider a coalition formation game with technological cooperation where the payoff from the clean technology is increasing in the number of signatories. Next, the authors design an experiment in a neutral laboratory setting where the players are not made aware of the environmental dimension of the game. Multiple groups of 7 players are given some endowments and asked to invest in the two possible technologies. They consider different treatments with different levels of benefits from the clean technology. They find that subjects cooperate the most when the gap between the clean and dirty technology, in terms of costs, is the smallest. When the spillovers from R&D are small, such that the coalition members retain most of the benefits from the clean technology, it is the coalition members who undertake most of the cost of implementing the clean technology. When the positive spillovers from the clean technology are large, the participation in the coalition is reduced but at the same time, the non-signatories adopt the clean technology and thus avoid exceeding the climate tipping point.

McEvoy (2010) tests the prediction of the standard two-stage IEA membership game that each country should be among the first to defect from a given coalition. The experiment considers two treatments with varying minimum participation constraints for forming a coalition. There are 13 rounds of play with two sets of players, randomly matched into groups of 10 players in each round. Each participant is given the choice about whether to produce a

unit of a neutrally defined public good and an agreement is set to hold if a sufficient number of participants choose to produce. He finds that only when the minimum participation clause is sufficiently high, then the laboratory results match the theoretical predictions about the timing of defection. Using a similar experimental design as McEvoy (2010), McEvoy *et al.* (2011) examine whether allowing coalition members to fund a third party enforcer to prevent defection leads to higher participation and more public good provision, as predicted by theory. However, they find that this result does not hold in the laboratory.

Cherry and McEvoy (2015) consider an experimental IEA game where the baseline is an N-country prisoners' dilemma game. The experiment considers five treatments, with and without uncertainty with various participation clauses. There are 20 rounds of play with 20 players in each, randomly matched into groups of 5 players in each round. Each round consists of two stages. In the first stage, players decide to join an agreement. In each treatment, the experimenters decide on the minimum number of players that must join for the agreement to enter into force. In the second stage, players are given a monetary endowment and asked to contribute to the public good. The members have to make a refundable deposit payment, while the non-members do not. If the contribution to the public good falls below a threshold, the members lose their deposit. Their main finding is that refundable deposits increase participation and reduce emissions significantly even when uncertainty about damages is introduced. McEvoy *et al.* (forthcoming) consider another experiment to test the effectiveness of minimum participation clauses in IEAs. There are 20 rounds of play with 18 players in each, randomly matched into groups of 6 players in each round. The players are asked to first choose the number of countries required in the minimum participation clause and then to choose between cooperation and non-cooperation. They consider different treatments where the payoffs are varied such that in one case the efficient minimum participation clause includes all countries and in the second it does not. They find that when the minimum participation clause includes all countries, it is very effective in fostering abatement and cooperation. If the participation clause includes a subset of countries only, it is not as effective due to players' concerns with equity issues.

Dannenberg *et al.* (2014) design a public good provision game in a laboratory setting and compare two cases: First, they allow countries to voluntarily contribute to the public good using an endogenously determined minimum contribution rule. Second, they force all players to contribute an exogenously given amount. They find that although participation is higher in the first case, the resulting coalition is less effective, given the voluntary nature of the contributions, than when they all countries are forced to contribute. These results are in line with the theoretical predictions discussed in Section 2 on the puzzle of large but shallow coalitions. Dannenberg (2016) tests the theoretical predictions about small but meaningful coalitions versus large but shallow coalitions in public good provision games. In line with these predictions, the author finds that full cooperation is not sustainable unless there are punishment mechanisms in place against deviators (refer to the discussion in Section 3). Only in the presence of such punishments, full cooperation is better than partial cooperation.

McGinty *et al.* (2012) design an experiment based on the canonical two-stage IEA membership model but with asymmetric players. They compare different surplus -resulting from cooperation- sharing rules: the Shapley value, the Nash bargaining solution, the equal split, and a rule which accounts for players' payoffs as singletons. They find that the latter rule yields the most participation in terms of coalition size and public good provision.

4.3 IEA games in an open economy context

As noted in Section 2, the early IEA models in an open economy context focus on issue linkage where trade sanctions can be used to punish defection from cooperation on the environment. There are a few recent papers which model IEAs in an open economy context but without the focus on issue linkages. That is, these papers do not model free trade as a club good, rather the focus is on comparing autarky to the case where free trade across countries is exogenously given. For example, Eichner and Pethig (2014a) propose an IEA model in the presence of capital mobility across countries to examine the pollution haven effect. As long as the pollution is not too damaging, emissions taxes are strategic substitutes across countries and result in small stable IEAs. However, the grand coalition may be stable if pollution is sufficiently damaging such that emissions taxes become strategic complement across countries. Eichner and Pethig (2013) introduce free trade into the standard open membership IEA model and compare this case to autarky. They consider a model with perfectly competitive firms which produce fossil fuel and a consumer good both of which are produced and consumed in each country and traded internationally. The fossil fuel is the polluting good; it generates some utility and also acts as an input in the production of the consumer good. They consider integrated world markets for both goods. The game unfolds as follows: In the first stage, countries decide whether to become members of an IEA. In the second and third stages, signatories and non-signatories choose their emissions standards sequentially. In the fourth stage, producers choose their output and input levels. Under autarky, they retrieve the small coalitions puzzle. Under free trade, they retrieve the results of large but shallow coalitions. Eichner and Pethig (2014b) extend the model of Eichner and Pethig (2013) by allowing countries to simultaneously use two policy instruments: a fossil tax in addition to the emissions standard. They show that signatories, by choosing a higher fossil fuel tax, are able to shift the burden of mitigation onto non-signatories. The resulting stable IEA is smaller than the case without the fossil fuel tax under free trade, but the IEA achieves greater emissions reduction and higher welfare levels.

Baksi and Ray Chaudhuri (2014) also analyze the stability of IEAs in an open economy context. The production of the tradable good causes transboundary pollution, and IEA signatories and non-signatories choose pollution taxes simultaneously. They consider segmented markets with imperfect competition. They find that the grand coalition is unstable under autarky, and that free trade may stabilize the grand coalition. As markets become more competitive, it becomes more likely that the global IEA is stable. Further, they also introduce a border tax adjustment (BTA), where countries with a higher pollution tax impose a tariff on imports from countries with a lower pollution tax and the tariff rate reflects the difference in pollution tax rates across a pair of trading countries. With symmetric countries, they find that the imposition of a BTA destabilizes an otherwise stable grand coalition such that the BTA cannot be used as a threat against defecting countries to sustain cooperation. At the same time, if the grand coalition is unstable under free trade, imposing a BTA may improve welfare.

Bencheikroun and Yildiz (2011) analyze the sustainability of cooperation between two identical countries, when setting their emission standards, within an infinitely repeated game framework where cooperation is sustained through the use of trigger strategies. They find that such cooperation is sustainable under autarky but not under free trade. Baksi and

Ray Chaudhuri (2016) extend the analysis in Benchekroun and Yildiz (2011) to allow for heterogeneity in pollution damage across countries. They consider an oligopoly model of trade between two regions, "North" and "South", where production generates transboundary pollution, and the regions differ in terms of their pollution damage parameter. In particular, the damage parameter is assumed to be higher in the North. Each country imposes a pollution tax on its domestic firms, where the tax rate can be chosen either cooperatively or non-cooperatively. While the North has a higher incentive to cooperate than the South, Baksi and Ray Chaudhuri (2016) find that (i) an increase in the degree of heterogeneity between the two countries in terms of their pollution damage parameter reduces the likelihood that cooperation is sustained between them, and (ii) trade liberalization increases the likelihood that cooperation is sustained between the two countries. Further, they find that imposing a BTA increases the likelihood of sustainable cooperation between the countries if and only if they are sufficiently heterogeneous in terms of their pollution damage parameter.

Labriet *et al.* (2015) capture how the changes in energy requirements when countries adhere to an IEA on climate change affect other countries' macroeconomic outcomes, especially countries that trade with signatories either through relocation of energy intensive industries or through reduced export volumes of energy intensive products to the signatories. In order to analyze these issues, they couple two models: TIAM-WORLD, which explicitly models technology choices and focuses on the energy dimension of an economy, and GEMINI-E3 model, which is a computational general equilibrium model capturing the basic macroeconomic features in a long-term growth model. They find that an agreement involving only the most energy intensive sectors in developing countries and all industries in developed countries increases participation. Since such an agreement does not affect consumers in developing countries, it makes developing countries more willing to participate. At the same, since energy intensive industries in developing countries are constrained, developed countries are more willing to participate.

4.4 IEA games with adaptation

Another nascent strand of the literature considers the effect of adaptation to climate change on the incentives of countries to participate in IEAs. One of the first papers to model a membership IEA model in the presence of adaptation is Marrouch and Ray Chaudhuri (2011) who focus on the derivation of the stable coalition size where the coalition and non-signatories choose emissions sequentially in the second stage of the IEA membership game. They assume simultaneous decisions about adaptation and mitigation. They consider a payoff function that includes quadratic benefits from individual emissions and quadratic costs from adaptation, which is a private country decision. Following the standard IEA literature, global emissions damage all countries equally, given their public bad nature as opposed to the private good nature of the adaptation effort. These modeling features follow Tulkens and van Steenberghe (2009) who model the trade-off between mitigation and adaptation in a non-cooperative setup. Benchekroun *et al.* (2011) use a similar setup, but with all countries choosing emissions simultaneously, to study the effect of an improvement in adaptation technologies on free-riding incentives. They show that incentives to free-ride decrease as more efficient adaptation technologies are implemented. This is the case because the presence of adaptation renders the best-response in terms of emissions less responsive (flatter). This

effect reduces the gains from free-riding. As a result, larger coalitions may be stable. At the same time, they show that the gains from cooperation may also increase in response to improved adaptation technologies. As such, large and meaningful coalitions may arise in contrast to the puzzle of large but shallow coalitions found in pure IEA games (e.g. Barrett, 1994a).

Lazkano *et al.* (forthcoming) extends the model of Benckroun *et al.* (2011) by considering cost asymmetry in adaptation across countries, specifically between developed and developing countries. They identify two channels through which adaptation affects free-riding incentives: carbon leakage and cost asymmetry in adaptation. They find that the presence of adaptation is not always a destabilizing factor in coalition formation. They find that reducing the adaptation cost difference between countries leads to higher global emissions when damages are linear. Furthermore, such a reduction in cost difference can eliminate carbon leakage between countries and reduces free-riding incentives as long as adaptation is not too effective across countries.

De Bruin *et al.* (2011) introduce adaptation into the AD-STACO integrated assessment model. This model allows for coalition formation within a setup with 12 heterogeneous regions. They consider two different levels of adaptation (those pertaining to the non-cooperative equilibrium and the cooperative equilibrium). They assume sequential decisions about adaptation and mitigation, and a damage function that is linear in emissions. They find that, under the grand coalition, more adaptation leads to less incentives to free-ride but reduces the equilibrium welfare level.

4.5 Directions for future research

As mentioned in the survey, Section 4 discusses novel approaches to modeling IEAs. Therefore, there are few existing models tackling political economy, mechanism design, behavioral and open economy aspects of IEAs. It would be useful to conduct further research on each of these topics.

Regarding mechanism design aspects of IEAs, further extensions of the two-stage IEA membership game are needed to tackle the various stages of pre-negotiations, ratification, implementation, and renegotiations of IEAs. Existing papers modeling the minimum participation clause are mostly restricted to models with symmetric countries under perfect information. Thus, it is important to explore the effectiveness of MPCs further under imperfect information and in models with asymmetric players. Moreover, existing papers usually consider a global carbon price as the focal point. It would be useful for future models to explore alternative focal points such as the widely discussed target of not exceeding an average increase of global temperature above 2°C. Regarding political economy aspects of IEAs, strategic voting issues should be explored in settings with more than two countries. The existing models of lobbying rely on very specific assumptions regarding who is lobbying and how the rents from lobbying are distributed. Therefore, there is scope to generalize these models in future research, with particular focus on asymmetric countries.

Regarding open economy aspects of IEAs, very few models account for market segmentation, asymmetric countries, and imperfect competition. Further research on trade restrictions other than sanctions, such as border tax adjustments, would also be useful. Finally, regarding the presence of the adaptation option, it is important to explore the issue of timing of

adaptation, whether ex ante or ex post relative to the pollution damage, within the context of IEA membership games.

5 Conclusion

This review provides a classification of the IEA literature examining the problem of trans-boundary pollution into the following categories: (i) pure IEA games without linkages, which focus on a single externality stemming from global pollution, (ii) IEA games with side-payments and issue linkages, which consider carrot or stick mechanisms to lure cooperators or punish defectors, (iii) design of IEAs and political economy aspects, which examine intra- and inter-governmental issues that affect international negotiations, (iv) behavioral factors, which affect the willingness of countries to cooperate on international agreements, (v) additional externalities resulting from open economy issues, which interact with the global pollution externality, and (vi) adaptation to climate change, which affects the free-riding incentives of countries. Based on this classification of the literature, the review attempts to identify conditions under which international environmental agreements would be sustainable and effective.

Section 2 on pure IEA games gives some pessimistic results. In the non-cooperative setup with symmetric players, Cournot games result in the puzzle of small coalitions, while Stackelberg games result in the puzzle of large but shallow coalitions. These results are broadly consistent across static and dynamic models with a few exceptions such as when minimum participation clauses are considered, or when the length of commitment is shortened. With repeated games, the results are mixed, with more optimistic results being achieved when the discount rate is low, the detection lag for defection is short, and multiple regional agreements are allowed to coexist. With asymmetric players, the literature also provides mixed results. In models with uncertainty, the results depend on the presence of learning. When countries face uncertainty about their individual and heterogeneous damages, the veil of ignorance is better for cooperation, whereas with systematic uncertainty facing all countries, learning can promote cooperation. Moreover, introducing the possibility of a catastrophe can increase cooperation, while the effect of risk aversion is ambiguous. When the assumption about myopic internal and external stability is replaced with farsightedness, the literature demonstrates that the puzzle of small coalition no longer persists. Once the dynamics of emissions accumulation are considered, however, the puzzle of large but shallow coalitions may persist, even under farsighted stability.

Topics for future empirical research related to Section 2 of our review include the following. It would be useful to conduct future empirical research to identify which of the following features of the pure IEA games are applicable and under what conditions, such as simultaneous or sequential move games for the second stage of the membership game, the stability criterion which is applicable, and fixed versus dynamic membership decisions. More specifically, some of the testable hypotheses that have been derived in the theoretical literature are as follows. First, while under simultaneous move games the puzzle of small coalitions persists, under sequential games the puzzle of large but shallow coalitions prevails. Second, these twin puzzles persist regardless of whether countries use trigger strategies or are heterogeneous in terms of their benefits and damage functions. Large emitters that do

not suffer much from pollution damage will not ratify the IEA unless they are compensated for reducing their emissions. Third, the twin puzzles of small and large but shallow coalitions may not hold when countries are farsighted in their decision to deviate from cooperation. Finally, when damages are uncertain, if the uncertainty is idiosyncratic then learning leads to smaller coalitions, whereas if the uncertainty is systematic, then learning leads to larger coalitions.

Section 3, on IEA games with side-payments and issue linkages, provides important insights on factors that affect the success of international agreements. In non-cooperative games, one conclusion is that transfers may increase the stable coalition size when asymmetry across country is large enough and self-financing transfers are possible. More generally, the results on the effectiveness of transfers are mixed since, on the one hand, non-signatories could be lured to join an IEA, but on the other, existing signatories may have an incentive to leave the coalition if they have to pay sufficiently high transfers. The cooperative literature focuses on designing transfer schemes, which are found to stabilize the grand coalition. At an abstract level, the literature indicates that interconnecting a public good game and a club good game is effective in promoting more successful IEAs. However, when the club good is modeled in more detail, the results are ambiguous. For example, while the games linking IEA with reputation effects generate positive results, those linking IEAs with R&D or trade games generate mixed results.

Topics for future empirical research related to Section 3 of our review include the following. A number of testable hypotheses can be drawn from the strand of literature on IEAs with transfer payments and linkages. Although, theoretically, it is argued in the inter-connected games literature that issue linkages, such as using trade sanctions against defectors from IEAs, are helpful in sustaining cooperation, when linkages are examined within a practical context the results are mixed. First, transfer payments are more likely to be effective the more asymmetric the countries in terms of their benefits and damage functions. Second, if existent signatories have to pay large transfers to increase the size of the coalition they might have an incentive to leave the agreement. Third, countries with high damages and low benefits from emissions pay others, while countries with low damages and high benefits from emissions receive transfers. Fourth, trade restrictions are less likely to be effective when firms can relocate across countries with different environmental regulations. Finally, a two-treaty framework, with the first treaty promoting cooperation on R&D and the second adopting cleaner technologies, is more likely to be effective when the adoption cost of the technology varies with the level of R&D, and when the cleaner technology exhibits increasing returns.

Section 4 begins by examining IEA design issues and finds that the following factors, if included in the design, may lead to more successful IEAs: minimum participation clauses, a supra-national authority, requiring countries to commit to modest rather than significant emission reductions, an agreement on only emission reduction rather than an agreement on both emission reduction and R&D, and the existence of a focal point in terms of some policy target. Intra-governmental issues may also impact the success of IEAs. While strategic voting may have a detrimental impact on the effectiveness of IEAs, lobbying has an ambiguous effect depending on the nature of the lobby applying the pressure on the government and whether the lobbying occurs in signatory or non-signatory countries. Section 4 also examines behavioral aspects in environmental negotiations. For example, reciprocity may allow for the existence of a tipping point whereby if a few countries initiate an IEA, other countries

subsequently join. Moreover, Section 4 summarizes the results from experimental economics related to IEAs. In the laboratory, the puzzle of small coalitions and the puzzle of large but shallow coalitions are generally confirmed. Participation clauses, if sufficiently high, and refundable deposits are shown to be effective. Requiring signatories to pay a third party for detecting deviation is not effective. Section 4 also summarizes the results on the effects of trade liberalization on IEAs, which are ambiguous and depend on trade patterns, market structure, and whether international markets are integrated or segmented. Finally, Section 4 discusses the effects of adaptation on IEAs. Allowing for adaptation to climate change and reducing the asymmetry of adaptation costs across countries reduce the free-riding incentives in negotiations.

Testable hypotheses related to Section 4 of our review include the following. First, the stable coalition size is typically equal to the size stipulated by the minimum participation clause and achieves the pre-announced abatement target. Second, the existence of a focal point during negotiations, or a supra-national authority, yields a larger stable IEA. Furthermore, it is important to analyze other mechanism design features, which have not received much attention in the IEA literature. These include flexibility mechanisms like the ones implemented in the Kyoto Protocol (1997) such as emissions trading, the clean development mechanism, and joint implementation. Third, according to the tipping point literature, if a small number of large emitters sign an IEA more countries would join the IEA thereafter. Fourth, when countries are sufficiently symmetric, freer trade induces more cooperation, whereas, border tax adjustments are only effective if countries are sufficiently asymmetric. Finally, if adaptation occurs simultaneously with emissions or ex post, then this may increase the stable size of the IEA.

The common thread linking the different topics included in Section 4 is that they encompass some of the more recent developments in the IEA literature. They point to directions for further analysis that may provide new insights, in addition to the ones we summarized in this review, to answer the question regarding whether international environmental agreements are doomed to fail or destined to succeed.

References

- Almer, C. & Winkler, R. (2010). Strategic behavior in IEAs: when and why countries joined the Kyoto Protocol. Discussion Paper 10-14, Bern, Switzerland: University of Bern.
- Altamirano-Cabrera, J-C., Weikard, H-P., & Haffoudhi, H. (2007). Influence of political pressure groups on the stability of international climate agreements. NCCR Climate Working Paper 2007-03. Bern, Switzerland: University of Bern-EPFL-Paul Scherrer Institute.
- Asheim, G. B., Froyn, C. B., Hovi, J., & Menz, F. C. (2006). Regional versus global cooperation for climate control. *Journal of Environmental Economics and Management*, 51(1), 93-109.
- Asheim, G. B., Holtsmark, B. (2009). Renegotiation-proof climate agreements with full participation: Conditions for pareto-efficiency. *Environmental and Resource Economics*, 43(4), 519-533.
- d'Aspremont, C. A., Jacquemin, J., Gabszewicz, J., & Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics*, 16, 17-25.
- Baksi, S. & Ray Chaudhuri, A. (2014). Trade, border tax adjustments, and the stability of international environmental agreements. Presented at World Congress of Environmental and Resource Economists (Istanbul, 2014).
- Baksi, S. & Ray Chaudhuri, A. (2016). International trade and environmental cooperation across heterogeneous countries. Mimeo, University of Winnipeg.
- Barrett, S. (1994a). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46, 878-894.
- Barrett, S. (1994b). Trade restrictions in international environmental agreements. Centre for Social and Economic Research on the Global Environment.
- Barrett, S. (1997a). The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics*, 19(4), 345-361.
- Barrett, S. (1997b). Heterogeneous international agreements. In *International Environmental Negotiations: Strategic Policy Issues*, ed. Carraro, C. (Edward Elgar), 9-25.
- Barrett, S. (1998). On the theory and diplomacy of environmental treaty-making. *Environmental and Resource Economics*, 11, 317-333.
- Barrett, S. (1999). A theory of full international cooperation. *Journal of Theoretical Politics*, 11(4), 519-541.
- Barrett, S. (2001). International cooperation for sale. *European Economic Review*, 45(10), 1835-1850.
- Barrett, S. (2002). Consensus treaties. *Journal of Institutional and Theoretical Economics*, 158(4), 529-547.
- Barrett, S. (2003). *Environment and Statecraft: The Strategy of Environmental Treaty-Making: The Strategy of Environmental Treaty-Making*. Oxford University Press.
- Barrett, S. (2006). Climate treaties and "breakthrough" technologies. *American Economic Review Papers & Proceedings*, 96(2), 22-25.
- Barrett, S. (2008). Climate treaties and the imperative of enforcement. *Oxford Review of Economic Policy*, 24 (2), 239-258.
- Barrett, S. (2013). Climate treaties and approaching catastrophes. *Journal of Environmental Economics and Management*, 66(2), 235-250.

- Barrett, S., & Stavins, R. (2003). Increasing participation and compliance in international climate change agreements. *International Environmental Agreements*, 3(4), 349-376.
- Batabyal, A. A. (2000). On the design of international environmental agreements for identical and heterogeneous developing countries. *Oxford Economic Papers*, 52(3), 560-583.
- Battaglini, M., & Harstad, B. (2016). Participation and duration of environmental agreements. *Journal of Political Economy*, 124(1), 160-204.
- Benckekroun, H., & Long, N. V. (2012). Collaborative environmental management: A review of the literature. *International Game Theory Review*, 14(04).
- Benckekroun, H., Marrouch, W., & Ray Chaudhuri, A. (2011). Adaptation effectiveness and free-riding incentives in international environmental agreements (No. 2011-120).
- Benckekroun, H., & Ray Chaudhuri, A. (2015). Cleaner technologies and the stability of international environmental agreements. *Journal of Public Economic Theory*, 17(6), 887-915.
- Benckekroun, H., & Yildiz, H. M. (2011). Free trade, autarky and the sustainability of an international environmental agreement. *The BE Journal of Economic Analysis & Policy*, 11(1).
- Biancardi, M. E. (2010). International environmental agreement: A dynamical model of emissions reduction. In *Nonlinear Dynamics in Economics, Finance and Social Sciences*, eds. Bischi, G. I., Chiarella, C., Gardini, L. (Springer), 73-93.
- Biancardi, M., & Villani, G. (2010). International environmental agreements with asymmetric countries. *Computational Economics*, 36(1), 69-92.
- Biancardi, M., & Villani, G. (2011). Largest consistent set in international environmental agreements. *Computational Economics*, 38(3), 407-423.
- Biancardi, M., & Villani, G. (2014). International environmental agreements with developed and developing countries in a dynamic approach. *Natural Resource Modeling*, 27(3), 338-359.
- Biancardi, M., & Villani, G. (2015). The effects of R&D investments in international environmental agreements with asymmetric countries. *Chaos, Solitons & Fractals*, 79, 30-39.
- Black, J., Levi, M. D., & De Meza, D. (1993). Creating a good atmosphere: minimum participation for tackling the 'greenhouse effect'. *Economica*, 60, 281-293.
- Bloch, F. (1996). Sequential formation of coalitions in games with externalities and fixed payoff division. *Games and Economic Behavior*, 14(1), 90-123.
- Bosetti, V., Carraro, E., de Cian, E., Massetti, & M. Tavoni (2013). Incentives and stability of international climate coalitions: An integrated assessment. *Energy Policy*, 55, 44-56.
- Bosetti, V., Heugues, A., Tavoni, (2015). Luring others into climate action: Coalition formation games with threshold and spillover effects, *Nota di Lavoro 21.2015*, Milan, Italy: Fondazione Eni Enrico Mattei.
- Botteon, M., & Carraro, C. (1997). Burden-sharing and coalition stability in environmental negotiations with asymmetric countries. In *International Environmental Negotiations: Strategic Policy Issues*, ed. Carraro C. (Edward Elgard), 26-55.
- Botteon, M., & Carraro, C. (1998). Strategies for environmental negotiations: issue linkage with heterogeneous countries. In *Game Theory and the Environment*, eds. Hanley, N., Folmer, H. (Edward Elgard), 181-218.

- Botteon, M., & Carraro, C. (2001). Environmental coalitions with heterogeneous countries: burden-sharing and carbon leakage. In *Environmental Policy, International Agreements, and International Trade*, ed. Ulph, A. (Oxford University Press), 26-55.
- Boucher, V., & Bramoullé, Y. (2010). Providing global public goods under uncertainty. *Journal of Public Economics*, 94(9), 591-603.
- Bréchet, T., F. Gerard & H. Tulkens (2011). Efficiency vs. stability in climate coalitions: a conceptual and computational appraisal. *Energy Journal* 32(1), 49-76.
- Breton, M., Sbragia, L., & Zaccour, G. (2010). A dynamic model for international environmental agreements. *Environmental and Resource Economics*, 45(1), 25-48.
- Brewer, T. L. (2003). The trade regime and the climate regime: institutional evolution and adaptation. *Climate Policy*, 3(4), 329-341.
- Brewer, T. L. (2004). The WTO and the Kyoto Protocol: interaction issues. *Climate Policy*, 4(1), 3-12.
- de Bruin, K., Weikard, H.-P. and Dellink, R. (2011). The role of proactive adaptation in international climate change mitigation agreements, CERRE Working Paper 2011:9, Centre for Environmental and Resource Economics (CERRE), Umea University, Sweden.
- Buchholz, W., Haupt, A., & Peters, W. (2005). International environmental agreements and strategic voting. *Scandinavian Journal of Economics*, 107, 175-195.
- Buchner, B., Carraro, C., Cersosimo, I., & Marchiori, C. (2005). Back to Kyoto? US participation and the linkage between R&D and climate cooperation. In *The Coupling of Climate and Economic Dynamics*, eds, Haurie, A., Viguier, L. (Springer), 173-204.
- Cabon-Dhersin, M. L., & Ramani, S. (2006). Can social externalities solve the small coalitions puzzle in international environmental agreements. *Economics Bulletin*, 17(4), 1-8.
- Calvo, E., & Rubio, S. J. (2012). Dynamic Models of International Environmental Agreements: A Differential Game Approach. *International Review of Environmental and Resource Economics*, 6, 289-339.
- De Cara, S., & Rotillon, G. (2003). Multi greenhouse gas international agreements, mimeo.
- Carraro, C., Eyckmans, J., & Finus, M. (2006). Optimal transfers and participation decisions in international environmental agreements. *The Review of International Organizations*, 1(4), 379-396.
- Carraro, C., & Marchiori, C. (2004). Endogenous strategic issue linkage in international negotiations. In *Game Practice and the Environment*, eds, Carraro, C., Fragnelli, V. (Edward Elgar), 65-86.
- Carraro, C., Marchiori, C., & Orefice, S. (2009). Endogenous minimum participation in international environmental treaties. *Environmental and Resource Economics*, 42(3), 411-425.
- Carraro, C. & Siniscalco, D. (1991). Strategies for the international protection of the environment. Discussion Paper 568, Centre for Economic Policy Research, London
- Carraro, C., & Siniscalco, D. (1993). Strategies for the international protection of the environment. *Journal of public Economics*, 52(3), 309-328.
- Carraro, C., & Siniscalco, D. (1995). R&D cooperation and the stability of international environmental agreements (No. 1154). CEPR Discussion Papers.
- Carraro, C., & Siniscalco, D. (1997). R&D cooperation and the stability of international environmental agreements. In *International Environmental Negotiations*, ed. Carraro, C.

(Edward Elgar).

Carraro, C., & Siniscalco, D. (1998). International institutions and environmental policy: International environmental agreements: Incentives and political economy. *European economic review*, 42(3), 561-572.

Cesar, H., & de Zeeuw, A. (1994). Issue linkage in global environmental problems, paper presented at the FEEM-IMOP workshop on 'Market Structure and Environmental Policy', Athens, 14-16 January 1994.

Cesar, H., & de Zeeuw, A. (1996). Issue linkage in global environmental problems. In *Economic Policy for the Environment and Natural Resources*, ed. Xepapadeas, A. (E. Elgar, Cheltenham).

Chander, P. (2007). The gamma-core and coalition formation. *International Journal of Game Theory*, 35(4), 539-556.

Chander, P., & Tulkens, H. (1995). A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. *International tax and public finance*, 2(2), 279-293.

Chander, P., & Tulkens, H. (1997). The core of an economy with multilateral environmental externalities. *International journal of game theory*, 36, 379-401.

Chander, P., & Tulkens, H. (2006). Cooperation, stability and self-enforcement in international environmental agreements: a conceptual discussion. CORE Discussion Paper.

Cherry, L. & McEvoy, D. (2015). Effective climate agreements under uncertainty. Mimeo.

Conconi, P., & Perroni, C. (2002). Issue linkage and issue tie-in in multilateral negotiations. *Journal of International Economics*, 57(2), 423-447.

Courtois, P., & Haeringer, G. (2005). The making of international environmental agreements. Working Paper 652.05. Barcelona, Spain: Universitat Autònoma de Barcelona.

Courtois, P., & Haeringer, G. (2012). Environmental cooperation: ratifying second-best agreements. *Public Choice*, 151(3-4), 565-584.

Cramton, P., & Stoft, S. (2012). Global climate games: How pricing and a green fund foster cooperation. *Economics of Energy & Environmental Policy*, 1(2), 125-136.

Dannenber, A. (2016). Non-binding agreements in public goods experiments. *Oxford Economic Papers*, 68(1), 279-300.

Dannenber, A., Lange, A., & Sturm, B. (2010). On the formation of coalitions to provide public goods: experimental evidence from the lab. NBER Working Paper 15967.

Dannenber, A., Lange, A., & Sturm, B. (2014). Participation and commitment in voluntary coalitions to provide public goods. *Economica*, 81(322), 257-275.

Dellink, R. (2011). Drivers of stability of climate coalitions in the STACO model. *Climate Change Economics*, 02(02), 105-128.

Diamantoudi, E., Macho-Stadler, I., Pérez-Castrillo, D., & Xue, L. (2015). Sharing the surplus in games with externalities within and across issues. *Economic Theory*, 60, 315-343.

Diamantoudi, E. and Sartzetakis, E. S. (2002) International environmental agreements: The role of foresight, *Economics Working Papers 2002-10*, School of Economics and Management, University of Aarhus.

Diamantoudi, E., & Sartzetakis, E. S. (2006). Stable international environmental agreements: An analytical approach. *Journal of public economic theory*, 8(2), 247-263.

Dong, D., & Zhao, X. (2009). International Environmental Agreement Formation and Trade. *Asia-Pacific Journal of Accounting & Economics*, 16(3), 339-356.

- Ecchia, G., & Mariotti, M. (1998). Coalition formation in international environmental agreements and the role of institutions. *European Economic Review*, 42(3), 573-582.
- Egger, P., Jeßberger, C., & Larch, M. (2011). Trade and investment liberalization as determinants of multilateral environmental agreement membership. *International Tax and Public Finance*, 18(6), 605-633.
- Eichner, T., & Pethig, R. (2013). Self-enforcing environmental agreements and international trade. *Journal of Public Economics*, 102, 37-50.
- Eichner, T., & Pethig, R. (2014a). Self-enforcing environmental agreements and capital mobility. *Regional Science and Urban Economics*, 48, 120-132.
- Eichner, T., & Pethig, R. (2014b). Self-enforcing environmental agreements, trade, and demand-and supply-side mitigation policy. *Journal of the Association of Environmental and Resource Economists*, 1(3), 419-450.
- Endres, A., & B. Rundshagen (2012). Incentives to diffuse advanced abatement technology under the formation of international environmental agreements. *Environmental and Resource Economics*, 56, 177-210.
- Espinola-Arredondo, A., & Munoz-Garcia, F. (2011). Free-riding in international environmental agreements: A signaling approach to non-enforceable treaties. *Journal of Theoretical Politics*, 23(1), 111-134.
- Eyckmans, J. & Finus, M. (2006). Coalition formation in a global warming game: how the design of protocols affects the success of environmental treaty-making. *Natural Resource Modeling*, 19, 323-358.
- Eyckmans, J. & H. Tulkens (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.
- Finus, M. (2002). Game theory and international environmental cooperation: any practical application?. In *Controlling Global Warming: Perspectives from Economics, Game Theory, and Public Choice*, eds. Christoph Bohringer, C., Finus, M., Vogt, C. (Edward Elgar), 9-104.
- Finus, M. (2003) Stability and design of international environmental agreements: The case of transboundary pollution. In *The International Yearbook of Environmental and Resource Economics 2003-04*, eds. Folmer, H., Tietenberg, T. (Edward Elgard), 82-158.
- Finus, M. (2008). Game theoretic research on the design of international environmental agreements: Insights, critical remarks, and future challenges. *International Review of Environmental and Resource Economics*, 2, 29-67.
- Finus, M., & Maus, S. (2008). Modesty may pay!. *Journal of Public Economic Theory*, 10(5), 801-826.
- Finus, M., van Ierland, E., & Dellink, R. (2006). Stability of climate coalitions in a cartel formation game. *Economics of Governance*, 7, 271-291.
- Finus, M., & Pintassilgo, P. (2013). The role of uncertainty and learning for the success of international climate agreements. *Journal of Public Economics*, 103, 29-43.
- Finus, M., Pintassilgo, P., & Ulph, A. (2014). International environmental agreements with uncertainty, learning and risk aversion (No. 4589). CESifo Working Paper.
- Finus, M., & Rundshagen, B. (1998). Toward a positive theory of coalition formation and endogenous instrumental choice in global pollution control. *Public Choice*, 96(1-2), 145-186.

Finus, M., & Rundshagen, B. (2000). Strategic links between environmental and trade policies if plant location is endogenous. *Diskussionsbeitrag Nr. 283 Fernuniv., Fachbereich Wirtschaftswiss.*

Finus, M., & Rundshagen, B. (2001). Endogenous coalition formation in global pollution control. *Nota di Lavoro 43.2001*, Milan, Italy: Fondazione Eni Enrico Mattei.

Finus, M., & Rundshagen, B. (2003). How the rules of coalition formation affect stability of international environmental agreements. *Nota di Lavoro 62.2003*, Milan, Italy: Fondazione Eni Enrico Mattei.

Finus, M. & Rundshagen, B. (2009). Membership rules and stability of coalition structures in positive externality games. *Social Choice Welfare*, 32(3), 389-406.

Folmer, H., van Mouche, P., & Ragland, S. (1993). Interconnected games and international environmental problems. *Environmental and Resource Economics*, 3(4), 313-335.

Folmer, H., & van Mouche, P. (1994). Interconnected games and international environmental problems, II. *Annals of Operations Research*, 54(1), 97-117.

Frankel, J. (2005). Climate and trade: Links between the Kyoto Protocol and WTO. *Environment: Science and Policy for Sustainable Development*, 47(7), 8-19.

Fredriksson, P. G., Neumayer, E., & Ujhelyi, G. (2007). Kyoto Protocol cooperation: Does government corruption facilitate environmental lobbying?. *Public Choice*, 133(1-2), 231-251.

Froyen C.B., Hovi J., (2008). A climate agreement with full participation. *Economics Letters*, 99, 317-319.

Fuentes-Albero, C., & Rubio, S. J. (2010). Can international environmental cooperation be bought?. *European Journal of Operational Research*, 202(1), 255-264.

Germain, M., Toint, P.L., & Tulkens, H., (1997). Financial transfers to ensure international optimality in stock pollutant abatement. In *Sustainability and Firms: Technological Change and Changing Regulatory Environment*, eds. Duchin, F., Faucheux, S., Gaudy, J., Nicolai, I. (Edward Elgar).

Germain, M., Toint, P., Tulkens, H., & de Zeeuw, A. (2003). Transfers to sustain dynamic core-theoretic cooperation in international stock pollutant control. *Journal of Economic Dynamics and Control*, 28(1), 79-99.

Goeschl, T., & Perino, G. (forthcoming). The climate policy hold-up: Green technologies, intellectual property rights, and the abatement incentives of international agreements. *Scandinavian Journal of Economics*.

Haffoudhi, H. (2005). The logic of two-level games with endogenous lobbying: Case of IEAs. *Cahiers de la Maison des Sciences Économiques 2005-54*. Paris, France: University of Paris I.

Hagen, A., & Eisenack, K. (2015). International environmental agreements with asymmetric countries: Climate clubs vs. global cooperation, *Nota di Lavoro 58.2015*, Milan, Italy: Fondazione Eni Enrico Mattei.

Harstad, B. (2006). Flexible integration? Mandatory and minimum participation rules. *Scandinavian Journal of Economics*, 108(4), 683-702.

Harstad, B. (2012). Climate contracts: A game of emissions, investments, negotiations, and renegotiations. *Review of Economic Studies*, 79(4), 1527-1557.

Hart, S., & Kurz, M. (1983). Endogenous formation of coalitions. *Econometrica*, 51(4), 1047-1064.

Heal, G., & Kunreuther, H. (2012). Tipping climate negotiations. In *Climate Change and Common Sense: Essays in Honor of Tom Schelling*, eds Hahn, R., Ulph, A. (Oxford University Press), 50-60.

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

Hoel, M. (1992) International environmental conventions: the case of uniform reductions of emissions. *Environmental and Resource Economics*, 2, 141-159.

Hoel, M., & Schneider, K. (1997). Incentives to participate in an international environmental agreement. *Environmental and Resource economics*, 9(2), 153-170.

Hoel, M., & de Zeeuw, A. (2010). Can a focus on breakthrough technologies improve the performance of international environmental agreements? *Environmental and Resource Economics*, 47(3), 395-406.

Jakob, M., & Lessmann, K. (2012). Signaling in international environmental agreements: the case of early and delayed action. *International Environmental Agreements: Politics, Law and Economics*, 12(4), 309-325.

Karp, L., & Zhao, J. (2010). International environmental agreements: Emissions trade, safety valves and escape clauses. *Revue Économique*, 61(1), 153-182.

Katsoulacos, Y. (1997). R&D spillovers, R&D cooperation, innovation and international environmental agreements. *International Environmental Agreements: Strategic Policy Issues*.

Kemfert, C. (2004). Climate coalitions and international trade: assessment of cooperation incentives by issue linkage. *Energy Policy*, 32(4), 455-465.

Kemfert, C., Lise, W., & Tol, R. S. (2004). Games of climate change with international trade. *Environmental and Resource Economics*, 28(2), 209-232.

Kolstad, C. (2003). Uncertainty in self-enforcing international environmental agreements. Manuscript, Donald Bren School of Environmental Science and Management.

Kolstad, C. D. (2007). Systematic uncertainty in self-enforcing international environmental agreements. *Journal of Environmental Economics and Management*, 53(1), 68-79.

Kolstad, C. D. (2010). Equity, heterogeneity and international environmental agreements. *The BE Journal of Economic Analysis & Policy*, 10(2).

Kolstad, C., & Ulph, A. (2008). Learning and international environmental agreements. *Climatic Change*, 89(1-2), 125-141.

Kolstad, C. D., & Ulph, A. (2011). Uncertainty, learning and heterogeneity in international environmental agreements. *Environmental and Resource Economics*, 50(3), 389-403.

Kroeze-Gil, J., & Folmer, H. (1998). Linking environmental and non-Environmental problems in an international setting: the interconnected games approach. In *Game Theory and the Environment*, eds. Hanley, N., Folmer, H. (Edward Elgar), 165-180.

Labriet, M., L. Drouet, M. Vielle, R. Loulou, A. Kanudia, A. Haurie, (2015). Assessment of the effectiveness of global climate policies using coupled bottom-up and top-down models, *Nota di Lavoro 23.2015*, Milan, Italy: Fondazione Eni Enrico Mattei.

Lazkano, I., Marrouch, W., & Nkuiya, B. (forthcoming). Adaptation to climate change: How does heterogeneity in adaptation costs affect climate coalitions? *Environment and Development Economics*.

Leroux, J., & Marrouch, W. (2012). Bargaining with linked disagreement points. *Strategic Behavior and the Environment*, 2, 259-277.

- Lessmann, K., Kornek, U., Bosetti, V., Dellink, R., Emmerling, J., Eyckmans, Nagashima, M., Weikard, H-S., & Yang, Z. (2015). The stability and effectiveness of climate coalitions. *Environmental and Resource Economics*, 62, 811-836.
- Lise, W., & Tol, R. S. (2004). Attainability of international environmental agreements as a social situation. *International Environmental Agreements*, 4(3), 253-277.
- Mäler, K. G. (1989). The acid rain game. *Studies in Environmental Science*, 36, 231-252.
- Marrouch, W., & Ray Chaudhuri, A. (2011). International environmental agreements in the presence of adaptation. *Nota di Lavoro 35.2011*, Milan, Italy: Fondazione Eni Enrico Mattei.
- Martimort, D., & Sand-Zantman, W. (forthcoming). A mechanism design approach to climate agreements. *The Journal of the European Economic Association*.
- McEvoy, D. (2010). Not it: opting out of voluntary coalitions that provide a public good. *Public Choice*, 142, 9-23.
- McEvoy, D. M., Cherry, T. L., & Stranlund, J. K. (forthcoming). Endogenous minimum participation in international environmental agreements: An experimental analysis. *Environmental and Resource Economics*, 1-16.
- McEvoy, D., Murphy, J., Stranlund, J. K. & Spraggon J. (2011). The problem of maintaining compliance within stable coalitions: experimental evidence. *Oxford Economic Papers*, 63, 475-498.
- McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers*, 59, 45-62.
- McGinty, M., Milam, G., & Gelves, A. (2012). Coalition stability in public goods provision: testing an optimal allocation rule. *Environmental and Resource Economics*, 52(3), 327-345.
- Na, S. L., & Shin, H. S. (1998). International environmental agreements under uncertainty. *Oxford Economic Papers*, 50(2), 173-185.
- Neumayer, E. (2002a). Weak versus strong sustainability: exploring the limits of two opposing paradigms. (Cheltenham: Edward Elgar), Northampton, USA.
- Neumayer, E. (2002b). Does trade openness promote multilateral environmental cooperation?. *World Economy*, 25, 815-832.
- Nkuiya, B., Marrouch, W., & Bahel, E. (2015). International environmental agreements under endogenous uncertainty. *Journal of Public Economic Theory*, 17(5), 752-772.
- Nkuiya, B. (2012). The effects of the length of the period of commitment on the size of stable international environmental agreements. *Dynamic Games and Applications*, 2, 411-430.
- Nordhaus, W. (2015). Climate clubs: Overcoming free-Riding in international climate policy. *American Economic Review*, 105, 1339-1370.
- Nordhaus, W.D. & Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *American Economic Review*, 86, 741-765.
- Nyborg, K. (2015). Reciprocal climate negotiators. IZA Discussion Papers No. 8866.
- Olmstead, S.M. & Stavins, R.N. (2012). Three key elements of a post-2012 international climate policy architecture. *Review of Environmental Economics and Policy*, 6, 65-85.
- Osmani, D., & Tol, R. S. (2010). The case of two self-enforcing international agreements for environmental protection with asymmetric countries. *Computational Economics*, 36(2), 93-119.

- Pavlova, Y., & de Zeeuw, A. (2013). Asymmetries in international environmental agreements. *Environment and Development Economics*, 18(01), 51-68.
- Petrakis, E., & Xepapadeas, A. (1996). Environmental consciousness and moral hazard in international agreements to protect the environment. *Journal of Public Economics*, 60(1), 95-110.
- Ray, D., & Vohra, R. (1997). Equilibrium binding agreements. *Journal of Economic theory*, 73(1), 30-78.
- Roberts, J.T. and Parks, B.C. & Vásquez, A.A. (2004). Who ratifies environmental treaties and why? Institutionalism, structuralism and participation by 192 nations in 22 treaties. *Global Environmental Politics*, 4(3), 22-64.
- Rubio, S. J., & Casino, B. (2005). Self-enforcing international environmental agreements with a stock pollutant. *Spanish Economic Review*, 7(2), 89-109.
- Rubio, J. S., & Casino, B. (2001) International cooperation in pollution control. Working Papers. Serie AD 2001-21, Instituto Valenciano de Investigaciones Económicas, S.A. (Ivie)
- Rubio, S. J., & Ulph, A. (2006). Self-enforcing international environmental agreements revisited. *Oxford economic papers*, 58(2), 233-263.
- Rubio, S. J., & Ulph, A. (2007). An infinite-horizon model of dynamic membership of international environmental agreements. *Journal of Environmental Economics and Management*, 54(3), 296-310.
- Ruis, A., & de Zeeuw, A. (2010). International cooperation to combat climate change. *Public Finance and Management*, 10(2), 379-404.
- Rutz, S. (2001). Minimum participation rules and the effectiveness of multilateral environmental agreements. Working paper 01/22. Zurich, Switzerland: Centre for Economic Research-SFIT.
- Scarf, H. E. (1971). On the existence of a cooperative solution for a general class of N-person games. *Journal of Economic Theory*, 3(2), 169-181.
- Schelling, T.C. (1960). *The strategy of conflict*. Harvard University Press.
- Shapley, L. S., & Shubik, M. (1969). On the core of an economic system with externalities. *American Economic Review*, 59, 678-684.
- Stiglitz, J. (2006). A new agenda for global warming. *The Economists' Voice*, 3(7).
- Thoron, S. (1998). Formation of a coalition-proof stable cartel. *Canadian Journal of Economics*, 31(1), 63-76.
- Tol, R. (2001). Climate coalitions in an integrated assessment model. *Computational Economics*, 18, 159-172.
- Tulkens, H. (1998) Cooperation versus free riding in international environmental affairs: Two approaches. In *Game Theory and the Environment*, eds. Hanley, N., Folmer, H. (Edward Elgar), 330-344.
- Tulkens, H., & Steenberghe, V. V. (2009). Mitigation, adaptation, suffering: in search of the right mix in the face of climate change. *Nota di lavoro 79.2009*, Milan, Italy: Fondazione Eni Enrico Mattei.
- Ulph, A. (2004). Stable international environmental agreements with a stock pollutant, uncertainty and learning. *Journal of Risk and Uncertainty*, 29(1), 53-73.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton university press: Princeton.

- Von Stein, J. (2008). The international law and politics of climate change. *Journal of Conflict Resolution*, 52, 243-268.
- Wagner, U. J. (2001). The design of stable international environmental agreements: Economic theory and political economy. *Journal of Economic Surveys*, 15(3), 377-411.
- Wangler, L., Altamirano-Cabrera, J. C., & Weikard, H. P. (2013). The political economy of international environmental agreements: a survey. *International Environmental Agreements: Politics, Law and Economics*, 13(3), 387-403.
- Weikard, H. P. (2009) Cartel stability and optimal sharing rule. *The Manchester School* 77, 575–593.
- Weikard, H. P., Wangler, L., & Freytag, A. (2015). Minimum participation rules with heterogeneous countries. *Environmental and Resource Economics*, 62(4), 711-727.
- Weitzman, M. L. (2014) Can negotiating a uniform carbon price help to internalize the global warming externality? *Journal of the Association of Environmental and Resource Economists* 1(1/2), 29-49.
- Weitzman, M. L. (2015a). Voting on prices vs. voting on quantities in a world climate assembly. NBER Working Paper No. 20925.
- Weitzman, M. (2015b). Internalizing the climate externality: Can a uniform price commitment help?. *Economics of Energy & Environmental Policy*, 4(2), 37-50.
- Yang, Z. (2015). Likelihood of environmental coalitions and the number of coalition members: evidences from an IAM model. *Annals of Operations Research*, 1-20.
- Yi, S. S. (2003). Endogenous formation of economic coalitions: a survey on the partition function approach. *The Endogenous Formation of Economic Coalitions*, Carlo Carraro, ed., Edward Elgar: London.
- Yi, S. S., & Shin, H. (2000). Endogenous formation of research coalitions with spillovers. *International Journal of Industrial Organization*, 18(2), 229-256.
- de Zeeuw, A. (2008). Dynamic effects on the stability of international environmental agreements. *Journal of Environmental Economics and Management*, 55(2), 163-174.