Self-Enforcing International Environmental Agreements with Endogenous Uncertainty

by

Joshua Murphy

An essay submitted to the Department of Economics

in partial fulfillment of the requirements for

the degree of Master of Arts

Queens University

Kingston, Ontario, Canada

April 2010

Abstract

In this paper I study the effect of uncertainty on the incentives countries face to participate in an international environmental agreement using a two-stage single-coalition model. One important innovation is that uncertainty is modelled endogenously, which provides additional opportunities for strategic interaction among nations. For example, this may lead signatory countries to undertake emission reductions in excess of those that are socially optimal, a behavior that reflects risk-reducing considerations. Although recent studies have indicated that persistent and endogenous uncertainty may be beneficial in a strategic setting, I find that uncertainty may only affect participation in an agreement negatively. This conclusion is likely robust to altering the assumption that countries are symmetric. I also find that the level of participation varies inversely with the ratio of marginal damage parameters to marginal benefit parameters, consistent with a core theoretical result.

Acknowledgments

First, I would like to express my deepest gratitude towards my supervisor, Professor Devon Garvie, for providing encouragement and guidance over the past several years. More than anyone else, Professor Garvie has contributed to my development as an economist. She is also responsible for sparking my interest in the increasingly important topic of international environmental agreements, which is the subject matter of this essay. I am also appreciative of the help I received from Professor Robin Boadway in seeing my essay through to completion. Professor Boadway's support on this project was very reassuring, and I owe him a million thanks for his timely comments and thoughtful advice. In addition, I must thank Rachel Ferreira for taking care of many administrative matters related to this essay.

Outside of the Queen's Economics Department, my family, and especially my partner, Monika Rahman, deserve special credit for providing me with helpful advice and unconditional support.

Contents

List of Figures

List of Tables

1 Introduction

Global environmental problems, such as climate change, are plagued by uncertainty.[1] Uncertainty stems from the extraordinary complexity of both our natural ecosystems and tightly integrated, dynamic economies, so that it is often difficult to assess how our actions translate into environmental change, and how these changes, in turn, will impact economic activity and societal well-being. While the presence of considerable uncertainty has been generally acknowledged for some time now, only recently have economists begun to analyze its strategic implications in a multi-decision-maker setting. This development is appropriate because the existence of multiple decision-makers is an inherent feature of global environmental problems. In the present study, I investigate the effect of persistent uncertainty on the incentives countries face to voluntarily participate in an international environmental agreement. The particular nuance of this paper is that uncertainty is modelled endogenously, which provides additional opportunities for strategic interaction among nations.

The most obvious example of an international environmental problem involving pervasive uncertainty is that of climate change. Recent estimates from leading climate scientists predict, for example, a 90% chance of warming by century's end ranging from just 1.1°C to as much as 6.4°C (IPCC 2007, p. 13). Although this level of uncertainty regarding the relationship between global mean temperature and the concentration of greenhouse gases in the atmosphere is disconcerting, more uncertain still are estimates of the economic damages resulting from climate change, which depend upon, among other things, how global mean temperature affects climate, how climate affects the global economy, and to what extent future generations will be able to adapt to these changes. To give some indicative figures, the recent and much-debated *Stern Review*, conducted by a team of economists led by Sir Nicholas Stern, estimated that the total cost of ignoring climate change would amount to a loss of

---

[1] In this essay I make no distinction between uncertainty (in the Knightian sense) and risk.

Table 1: Global mean temperature increase for various greenhouse gas concentrations

Stabilization level	Temperature increase relative to pre-industrial (°C)		
(ppm CO_2 equivalent)	IPCC TAR 2001	Hadley Center	Eleven Studies
400	0.8 - 2.4	1.3 - 2.8	0.6 - 4.9
450	1.0 - 3.1	1.7 - 3.7	0.8 - 6.4
500	1.3 - 3.8	2.0 - 4.5	1.0 - 7.9
550	1.5 - 4.4	2.4 - 5.3	1.2 - 9.1
650	1.8 - 5.5	2.9 - 6.6	1.5 - 11.4
750	2.2 - 6.4	3.4 - 7.7	1.7 - 13.3
1000	2.8 - 8.3	4.4 - 5.9	2.2 - 17.1

Table shows equilibrium temperature projections using 5% - 95% climate sensitivity ranges. See Table 1.1 on p. 12 of Stern (2007) for references.

welfare equivalent to a permanent reduction in yearly global per-capita consumption of between 5% and 20% (Stern 2007, p. 144).

Stern's estimates were among the first to incorporate the possibility of catastrophic climate impacts arising from the acknowledged yet highly uncertain possibility of abrupt, non-linear changes in the climate system.[2] In recognition of the idea that these devastating impacts become more likely as global mean temperature increases, Stern, in his modelling exercise, allowed the probability with which they occur to depend indirectly upon the concentration of greenhouse gases in the atmosphere (Stern 2007, p. 153). This points to the endogeneity of uncertainty in the climate problem, a feature that has also been recognized in various forms by, among others, Chichilnisky and Heal (1993), Tsur and Amos (1996), and Fisher and Narain (2003), and has recently been emphasized by Weitzman (2007), who argues for a re-characterization of climate policy as insurance against worst-case catastrophes. It can be seen from Table 1, which gives current scientific predictions of the equilibrium effect of various levels of greenhouse gas concentrations in the atmosphere on global mean temperature, that the variance of predicted temperature increases with the atmospheric concentration of

---

[2] Specific catastrophic impacts are discussed in Stern 2007, Chapter 3.

greenhouse gases. Since economic damages and its variance are expected to increase with global mean temperature,[3] more emissions of greenhouse gases will tend to lead to greater and more uncertain damages.

This last observation has been incorporated into analyses of non-cooperative international emissions games with risk-averse countries by Endres and Ohl (2003) and by Bramoulle and Treich (2009). Both studies conclude that persistent uncertainty may have positive social consequences, in that Nash equilibrium emissions are lower with uncertainty relative to the case of certainty, reflecting risk-reducing considerations.[4] Moreover, global welfare may be positively affected by the presence of uncertainty if damages avoided due to reduced emissions outweigh the cost of bearing the associated risk. In other words, uncertainty may be beneficial in a strategic setting. It is this possibility that I am interested in exploring further in the present study, particularly because theory (as well as anecdotal evidence) indicates that meaningful international cooperation on the climate issue will be difficult to achieve and maintain.

This paper contributes to two branches of literature. First, and most directly, it contributes to the large and growing body of literature on international environmental agreements (IEAs), which studies the endogenous formation of cooperative coalitions among sovereign nations. Uncertainty is mostly ignored in this literature aside from a handful of studies contrasting the effects of different assumptions regarding the timing of the resolution of exogenous uncertainty (learning) on the prospects for agreement. I am not concerned with the effects of learning in the present study. Instead, I provide the first analysis of an international environmental agreement under endogenous uncertainty. Second, there is a small body of literature, some of which is experimental, studying the effect of collective and endogenous uncertainty on the behavior of agents in social dilemmas. Included in this literature are the two studies just men-

---

[3] See Table 13.2 on p. 195 of Stern (2007).

[4] In Endres and Ohl (2003), emissions choices are discrete and expected payoffs take the form of a prisoner's dilemma, so that uncertainty must be large enough or countries must be sufficiently risk-averse before they are willing to switch their strategies from *pollute* to *abate*.

tioned, and also studies such as those by Walker and Gardner (1992), who consider the probabilistic destruction of a common property resource, and Ihori and McGuire (2007), who consider odds improving self-protection as a public good.[5] This branch of literature focuses on non-cooperative Nash equilibria in situations where agreement among the parties provides a means to escape typically sub-optimal outcomes. Thus, introducing coalition formation into such models is a logical extension.

The remainder of this paper is organized as follows. Section 2 introduces important aspects of the literature on IEAs and evaluates alternative modelling choices. Some relevant literature on uncertainty in a strategic setting is also discussed. In Section 3, a model of an international emissions game is extended to include endogenous uncertainty and benchmark outcomes of the extended model are characterized. Some parameter restrictions that apply in later sections are also derived. I describe in detail the standard, two-stage single coalition model of an international agreement in the opening segment of Section 4, and then apply this approach to the model developed in Section 3 to determine participation in an IEA with endogenous uncertainty. A brief discussion follows in Section 5, where issues such as the robustness of results are considered, and Section 6 concludes. Lengthy proofs may be found in the appendix.

2 Literature Review

In this section I discuss relevant literature on IEAs[6] in order to introduce and evaluate available modelling choices strongly affecting the subsequent analysis. In particular, I devote attention to outlining the dominant approach to studying IEAs and to assessing its merits and drawbacks. I then review literature focusing on uncertainty in environmental problems with multiple decision-makers to provide further context for

---

[5] Self-protection is a term associated with individual insurance that applies to actions that reduce the probability of a loss. See Ehrlich and Becker (1972) for a precise definition.

[6] More comprehensive surveys of the IEA literature can be found in Wagner (2001) and Barrett (2005). Finus (2001) gives a book-length treatment.

the ideas explored in this paper, as well as to look more closely at the available options for modelling uncertainty endogenously. It should be noted, however, that the subject matter of this paper is related to several branches of literature not discussed in detail here, including that on common property resources, the private provision of public goods, coalition formation, and cartel stability.

2.1 International Environmental Agreements

International environmental agreements (IEAs) are often conceptualized as attempts to solve international common property resource dilemmas. Alternatively (but equivalently), they can be thought of as means to remedy the under-provision of international public goods. In either case, the underlying problem involves multiple, interdependent decision-makers (countries) operating in situations involving reciprocal externalities.[7] In this setting, theory tells us that when policy levels are set non-cooperatively (unilateralism), purely self-interested countries will over-exploit a shared environmental resource or devote too little attention to protecting it (Hardin 1968, Dasgupta 1982). To escape Pareto-inefficient environmental outcomes, sovereign countries must negotiate an agreement (Barrett, 1990).

Most economic analyses of IEAs study the endogenous formation of coalitions within a non-cooperative, game-theoretic framework. Issues of primary concern are the size of coalitions that form at equilibrium and the extent to which these coalitions are able to resolve environmental problems. Aside from a few notable exceptions, a rather unfortunate and disappointing prediction regarding the potential for meaningful international cooperation has emerged: IEAs will achieve relatively little when the gains from cooperation (as measured by aggregate welfare under the fully coopera-

---

[7] When countries share a common property resource, their use of the resource imposes a *negative* externality on all other countries sharing the resource. An example is an openly accessible fishing ground or rain-forest. By contrast an example of an international public good is a clean, stable climate. It can be 'supplied' by a country that reduces its emissions of harmful pollutants such as greenhouse gases and chlorofluorocarbons. Abatement activities in one country impose a *positive* externality upon all other countries benefiting from the climate.

tive outcome less that under unilateralism) are large. The majority of literature has reached this conclusion using a static, two-stage coalition model, consisting first of a membership game followed by an emissions game.[8] Early studies such as Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994) defined what have become the standard assumptions of the game. These include, among many others: (i) that there is only one agreement, and countries simultaneously and individually decide whether or not to sign it in the membership game; (ii) in the emissions game, signatories to the IEA are bound by the agreement to set their emissions levels to maximize their joint payoff/welfare, while each non-signatory acts as a singleton, setting its emissions level to maximize its own payoff/welfare;[9] (iii) signatories and non-signatories set their emissions levels in a Nash-Cournot fashion; (iv) countries interact in a partial equilibrium framework where payoff functions consist solely of benefits from emitting and damage costs from the aggregate level of pollution; (v) payoffs are known with certainty; (vi) countries are identical; and (vii) the only agreements of interest are those that are stable—that is, those in which no country has an incentive to change its membership decision, given the membership decisions of all other countries.

Finus (2001) confirms the results of numerous previous studies by showing that, under these assumptions, a stable (in the sense described above) IEA can consist of only very few signatories for a wide range of payoff functions. But at the same time, in the real world, there exist numerous IEAs involving high levels of participation.[10] This fact has led to an outpouring of literature aimed at altering the assumptions of the standard model so that the resulting approach can rationalize greater participation. In light of this more recent literature, some of which I will discuss shortly, it is my view

---

[8] The result has been confirmed in the more realistic repeated-game (Barrett, 1994) and dynamic settings (Rubio and Casino 2005, Rubio and Ulph 2007).

[9] The model can also be formulated in terms of emission reductions, in which case the second stage is an abatement game rather than an emissions game. As one would expect, models cast in emission reductions space and those cast in emissions space produce results that do not differ in any important respect.

[10] The Montreal Protocol on Substances that Deplete the Ozone Layer is one such example. As of 2003, it had been signed by 183 countries (Barrett, 2003).

that the standard model is still useful in that it provides a well-understood, sharply focused framework within which to investigate the effects of various extensions, while at the same time maintaining an acceptable trade-off between realism and analytical convenience. For these reasons I apply it in the present study.

I do not maintain that all of the assumptions in the standard model are ideal. Two assumptions in particular may actually seem egregious. First, while the standard model allows countries to free-ride on an agreement through non-participation, an assumption is that participants comply with their obligations. Of course, participants may neglect to do so; they may also free-ride through non-compliance. However, if non-participation can be deterred, compliance can be enforced, which means that non-participation is the binding constraint on international cooperation (Barrett, 1998b). Thus, simplifying things by avoiding the issue of non-compliance seems reasonable provided that countries' participation decisions are made voluntarily. Second, the assumption that countries are identical cannot be defended as reflective of reality. And it is true that dropping this assumption can dramatically affect participation in the agreement (Barrett, 2001). However, the assumption of identical countries does allow results to be derived analytically rather than by simulation, so that dropping it can be costly depending on whether or not differences among countries are likely to interact with the particular extension under consideration (in this case, uncertainty) to qualitatively affect the results. If not, or at least to start out with, it makes sense to sharpen one's focus by assuming symmetry.

While some assumptions of the standard model can be justified by the amount of tractability they purchase, others are imperfect but reasonable in that they capture reality in a powerful and analytically simple way. However, they tend to be less subtle and lead to more pessimistic results than one might prefer. The assumption that signatories are bound by the agreement to set their emission levels to maximize their joint payoff falls into this category. An agreement with these terms is referred to in the literature as strongly collectively rational (SCR). In understanding why the SCR assumption is made, it is best to consider an agreement that requires signatories to restrict their emissions to fully cooperative levels if and only if all countries sign the agreement, and otherwise to revert to their fully non-cooperative emission levels (Nash-reversion). The mere threat of triggering Nash-reversion can induce all countries to become part of a stable IEA (Chander and Tulkens 1995, 1997). However, in the real world it is possible for countries to renegotiate an agreement. Once one recognizes this, it is easy to see that such an agreement would never be put forth because the threat upon which it is based is not credible. Suppose that a portion of the countries did not sign the agreement. Following the membership stage, the signatories would then have the option of implementing the treaty, which would effectively involve implementing unilateralism, or renegotiating the terms of the agreement amongst themselves. In most cases they would prefer renegotiation (Barrett, 1998a). A SCR agreement does not suffer from this problem: once membership decisions are given, the signatories can do no better than to carry out the terms of the agreement in the emissions game.[11] Unlike the agreement involving Nash-reversion, an SCR agreement is 'collectively rational' (in the sense that signatories can do no better as a group than by following the agreement) both on *and* off the equilibrium path.

SCR agreements are compelling because they are credible in the sense that they are not vulnerable to renegotiation. But one could argue that such agreements are also too rigid. This criticism is rooted in the fact that the SCR assumption provides strong incentives for countries to free-ride through non-participation while simultaneously restricting the ability of signatories to deter it. As was just discussed, the SCR assumption rules out the possibility that signatories will carry out the terms of a self-damaging agreement. Yet in a partial equilibrium framework in which countries

---

[11] The SCR assumption is compatible with the concept of a strongly renegotiation-proof (SRP) equilibrium employed in the repeated-game counterpart to the one-shot game considered here. Farrell and Maskin (1989) provide a definition of a SRP equilibrium.

interact on emissions alone, signatories will find it difficult to deter free-riding unless they are willing to suffer at least a little bit themselves. It does not seem beyond belief, and indeed it is supported by behavioral evidence, that countries may undertake self-damaging actions if there is sufficiently good reason for doing so, such as arriving at a situation that is overall socially desirable (Brekke and Johansson-Stenman, 2008). At the same time, the SCR assumption encourages free-riding by requiring signatories to undertake deep cuts in their emissions. This makes non-participation attractive. Here, again, the SCR assumption may be too strong. For instance, Downs et al. (1996) argue that although compliance with international agreements ostensibly seeking to limit some collectively dysfunctional behavior (such as polluting the environment) has in general been observed to be good, this may simply be a symptom of weak agreements. That is, the depth of cooperation among the signatories may be shallow rather than deep, where the depth of cooperation is defined as 'shallow' if signatories choose policy levels close to those implied by unilateralism, and 'deep' if these policy levels are close to those that maximize the signatories' joint payoff. Furthermore, since it is almost certainly the case that countries make the connection between an agreement requiring deep cuts in emissions and the result of low participation (because deep cuts heighten free-rider incentives), one might also expect that countries have an interest in limiting the ambition of their agreements in order to boost participation, perhaps out of a concern for fairness or perhaps because it is simply more efficient to share the burden of environmental protection more evenly.

Both Barrett (2002) and Finus and Maus (2008) have shown that there does indeed exist a trade-off between the depth and breadth of cooperation.[12] As cooperation among the signatories is made more shallow, IEAs with higher levels of participation become stable. Even full participation can be stable for a non-trivial agreement

---

[12] Finus and Maus assume that signatories consider only a fraction of the coalition's aggregate damages when setting their emissions both on and off the equilibrium path. As this fraction, which the authors refer to as a 'modesty parameter', decreases, the emissions reductions required by the agreement become less ambitious.

(i.e. requiring reductions beyond those implied by unilateralism). Nevertheless, the conclusion that an agreement will achieve relatively little when the gains from cooperation are large still obtains. So while departing from the SCR assumption has lead to a theory of IEAs that can explain varying levels of participation, the fundamental conclusion regarding the potential for meaningful international cooperation has remained. In this paper I address the question of whether endogenous uncertainty alters this conclusion—does it improve upon the trade-off between the depth and breadth of cooperation? By assuming that countries negotiate a SCR agreement, the depth of cooperation is held constant (it is assumed to be full) and higher participation indicates a less severe trade-off. Thus, the SCR assumption seems reasonable here provided that it is recognized that participation is not the focus *per se*.

Another important aspect of the standard model falling into the category of 'intuitively appealing but analytically blunt' is the particular definition of stability that it assumes. In the standard model an agreement is stable if (1) no signatory would want to exit given the membership decisions of all other countries (internal stability), and (2) no non-signatory would want to join the agreement, again given the membership decisions of all other countries (external stability). I follow de Zeeuw (2008) in referring to stability in this sense as myopic stability. Myopic stability was the first formal concept to capture the notion that any agreement among sovereign nations must be self-enforcing (Barrett, 1994). The term self-enforcing is used because, unlike local environmental problems that can be addressed through existing governing bodies and regulatory institutions, in the international arena there can be no appeal to a higher authority to enforce cooperative behavior among the parties involved. Therefore, the terms of the agreement must be such that incentives to participate in the agreement and to comply with its terms are implicit in the agreement itself. By using myopic stability to determine the equilibrium to the membership game, we can be sure that each country is participating voluntarily. However, as is evident in the concept's name, myopic stability implies that countries are short-sighted: they assume that other countries' membership decisions will not change when they decide to either opt into or exit from the agreement.[13] If a country decides to exit from the agreement, it assumes that the coalition will remain otherwise intact. But cooperation among the remaining coalition members may unravel; the resulting agreement may not be an equilibrium (i.e. it may not be stable) and thus not a valid point of comparison. The implication is that a country should consider the dynamic consequences of its exit from the agreement on the membership decisions of the remaining signatories.

Diamantoudi and Sartzetakis (2002) remedy this inconsistency by recursively defining a stability concept that imposes no exogenous restrictions on the behavior of the remaining signatories. They draw on Chwe (1994) in concluding that a far-sighted country will recognize that, if it leaves the agreement, it may trigger other countries to leave until a new stable situation emerges. Thus the country ought to compare its welfare in the initial situation to that at the end of this process (i.e. the equilibrium that eventually occurs) when deciding whether or not to exit from the agreement. If countries behave in this way and no country has an incentive to exit, then the agreement satisfies what Diamantoudi and Sartzetakis call far-sighted stability. They show that far-sightedness leads to multiple equilibria in the membership game; both high and low levels of participation in a SCR agreement can be supported.[14] This result suggests that the problem of environmental protection is one of coordination. However, de Zeeuw (2008) shows that far-sightedness may not lead to such optimistic conclusions in a more realistic dynamic framework. He finds

---

[13] The concept of stability implies that governments can re-optimize their membership decisions and that a stable agreement is reached through a process of adjustment. Of course, in a stage-game this re-optimization and adjustment can only be hypothetical. So while using the language 'exiting from the agreement' rather than 'not signing the agreement to begin with' is technically incorrect, it can be justified on the grounds that the stage-game is just a simplification of a game in which countries have plenty of opportunities to revise their membership decisions.

[14] Other studies based on the concept of far-sighted stability include Eyckmans (2001), Biancardi and Liddo (2008), and Osmani and Tol (2009). These studies also conclude that far-sightedness leads to multiple stable agreements including those with both high and low participation.

that large coalitions can be stable only when the gains to cooperation are very small. Osmani and Tol (2008) also report pessimistic results regarding far-sightedness in a dynamic context. They show that large coalitions cannot be sustained for very long, even in a far-sighted world. Given current theory then, it is not clear whether adopting a far-sighted concept of stability over a myopic one is worth the additional complexity in terms of explanatory power. In the present study I work with the more simple concept of myopic stability.

Apart from dropping the SCR assumption and considering alternative stability concepts, the IEA literature has altered other aspects of the standard model in order to improve realism and arrive at more optimistic results. For instance, Barrett (1994), Rubio and Casino (2001), Diamantoudi and Sartzetakis (2006), and Rubio and Ulph (2006) take the view that membership in an IEA acts as a commitment device, and, consequently, they model decisions in the emissions game as occurring sequentially. In some cases, Stackleburg leadership by the signatories can lead to higher levels of participation in a SCR agreement. Other authors have shown that cooperation may be easier when the sphere of interaction among countries is broadened to include non-environmental issues. In such a case reputation effects become important (Hoel and Schneider 1997; Jeppensen and Anderson 1998; Cabon-Dhersin and Ramani 2006; Breton et al. 2010), and it is possible to make cooperation in other areas contingent upon environmental cooperation through 'issue-linkage' (Barrett 1997b; Botteon and Carraro 1998; Le Breton and Soubeyran 1997; Katsoulacos 1997; Carraro and Siniscalco 1997, 1998; Mohr and Thomas 1998; Finus 2000). More fundamentally, Lange and Vogt (2003) demonstrate that multiple stable coalitions, including the grand coalition, can be supported if countries have a sufficiently strong preference for equity. Self-financed transfers may also help if some countries can commit to cooperation, or if countries are strongly asymmetric (Carraro and Siniscalco 1993; Barrett 1997a; Hoel and Schneider 1997; Eychmans and Finus 2003; Carraro et al. 2006). Less successful extensions include consideration of a stock pollutant (Rubio and Casino, 2005), the ‘ancillary’ benefits of environmental protection (Finus and Rubbelke, 2008), and a minimum participation clause (Black et al. 1993; Barrett 1998b; Carraro et al. 2004). Finally, the most recent studies, such as McGinty (2010) and Breton et al. (2010), apply replicator dynamics from evolutionary game theory to model the process through which membership decisions become stable (loosely defined) over time. The evolutionary approach applies if governments are not explicitly forward-looking and thus seems to be compatible with the concept of myopic stability.

Overall, the theory of IEAs provides interesting insight into the problem of environmental cooperation. On a broad level it tells us that meaningful action cannot be expected unless additional instruments can be used to stabilize large coalitions of countries cooperating ‘deeply.’ This conclusion is obtained in the standard model. It has also been obtained in richer, more elaborate models based on less restrictive assumptions. In this paper I ask: can the fact that uncertainty is endogenous alter this general conclusion? I have argued above that the standard model provides a simple and valid framework within which to provide an initial answer to this question.

I should mention, however, that there is one potentially important caveat to this argument. Overwhelming evidence from the field and from the lab indicates that, when it comes to problems of collective action, a large proportion of individuals are willing to cooperate conditional upon the cooperation of others (Ostrom, 2000). Assuming that the same applies at the country level, this evidence suggests that the rational, self-interested model of behavior used in the present study (and universally in the IEA literature) is insufficient in capturing an essential aspect of actual behavior. In that case, any question regarding the effect of various factors on the incentives that purely self-interested countries face to participate in an IEA may be misguided. Instead, the focus ought to be shifted towards determining and creating the conditions under which conditional cooperation is likely to take root and grow. However, my aim in this paper is not to rewrite the theory of IEAs, but rather to extend a well-established model to better account for an important feature of the climate problem, and to determine whether the model's predictions still hold.

2.2 Environmental Uncertainty in a Strategic Setting

In what follows I restrict attention to analyses of environmental problems involving the interaction of two or more agents and in the presence of uncertainty.[15] I begin by discussing literature in which uncertainty is exogenous and learning is the main focus. This branch of literature relates most closely to the present study since uncertainty is considered in combination with coalition formation. I then move on to consider the various ways in which environmental uncertainty has been modelled endogenously in games with more than one player.

The question of whether the resolution of exogenous uncertainty through learning is beneficial or detrimental to achieving efficiency in a strategic setting was first addressed in a string of papers that includes Ulph and Ulph (1994), Ulph and Maddison (1997), Ulph (1997), Ulph and Ulph (1997), Kolstad (2005), and Baker (2005). These studies consider the impact of learning on non-cooperative emissions levels and welfare under risk-neutrality using a two country, two period model with a stock pollutant where damage from the accumulation of emissions occurs at the end of the second period. Under learning, the value of an uncertain damage parameter for each country is revealed between the first and second periods. Without learning, the parameters are revealed after second period emissions levels have been set, so that both first and second period emissions decisions must be taken on expected damages. An interesting finding is that the value of information (aggregate welfare under learning less that under no learning) may be negative if the realized damage parameters are

---

[15] There exists a vast literature devoted to environmental policy under uncertainty, mostly in relation to climate change, that assumes a single world decision-maker. See, for example, Manne and Richels (1992), Nordhaus (1994), Kolstad (1996), and Stern (2007).

sufficiently negatively correlated across countries, so that learning uncovers large differences in damage parameters. In other words, learning can be bad in a strategic setting if it reveals winners and losers who would be better off collectively by setting their emissions behind a veil of uncertainty.

Studies by Na and Shin (1998), Fujita (2004), Ulph (2004), Dellink et al. (2005), Kolstad (2007), Kolstad and Ulph (2008a), and Kolstad and Ulph (2008b) ask whether learning can also be harmful to participation in an IEA. All of these studies assume that countries are risk-neutral and adopt the standard, two-stage single coalition model of an IEA requiring agreements to be SCR and myopically stable, but their assumptions regarding uncertainty are idiosyncratic. Together, they provide mixed evidence for the effect of learning, which can occur before the membership game (full learning), after the membership game but before the emissions game (partial learning), or after the emissions game (no learning), on the prospects for international cooperation. Finus and Pintassilgo (2009) work with a more general formulation of uncertainty and are able to go some distance towards resolving the issue by providing the conditions under which learning can reduce participation in an IEA. They distinguish between uncertainty regarding the value of a common damage parameter (level uncertainty) and uncertainty regarding the distribution of different damage parameters (distributional uncertainty). In the former case, countries face the same damage parameter once uncertainty is resolved, but prior to learning cannot be sure of how damaging each unit of emissions will be. In the latter case, nature takes a set of different damage parameters with as many elements as there are players and randomly assigns, without replacement, a damage parameter to each country. Here, countries know how damaging each unit of emissions will be but are unsure as to how these damages will be distributed. Finus and Pintassilgo show that it is only in a pure version of the latter case that learning unambiguously affects participation in an IEA negatively. And, even then, the negative effect of learning can be made positive through use of transfers and an appropriately designed hedging strategy.

Learning is de-emphasized in the present study relative to this earlier literature. Rather than contrasting different learning scenarios I assume that learning does not take place, so that countries must make their membership and emissions decisions prior to the resolution of uncertainty. This particular learning scenario seems most relevant in the context of climate change, the leading example considered in this paper. Moreover, I ignore distributional uncertainty and assume that all countries realize the same damage parameter, again on the grounds that the distribution of impacts is not the key source of uncertainty affecting international climate policy; countries are already aware of their relative vulnerabilities.[16]

With regard to the modelling of endogenous uncertainty, some initial guidance can be found in Endres and Ohl (2003) and Bramoulle and Treich (2009), both of which study the effect of uncertainty on the Nash equilibrium emission levels and welfare of a non-cooperative game of international emissions. These studies assume that countries are risk-averse and that the level but also, crucially, the variance of economic damage is increasing in aggregate emissions. Under these assumptions countries have an extra incentive to abate emissions: in addition to reducing damages in any given state of the world, limiting one's emissions also favourably alters the probabilities over which these states occur, thereby reducing expected welfare losses due to risk. Following this approach to modelling uncertainty endogenously would preclude the analytical derivation of results, however, since it requires the assumption of risk-aversion.[17]

An alternative approach that can be applied with risk-neutral agents is based on the notion of self-protection. Ehrlich and Becker (1972) define self-protection as actions taken by an individual to reduce the probability of the occurrence of an

---

[16] See Tol et al. (2004) and Buys et al. (2007). Tol et al. (2004) suggest determinants of vulnerability to climate change impacts while Buys et al. (2007) explicitly rank countries along various dimensions of vulnerability.

[17] Boucher and Bramoulle (2007) are forced to resort to simulation in their analysis of a SCR, myopically stable IEA with risk-averse countries under exogenous uncertainty.

undesirable state. In the present study I adapt this concept for a group of agents facing a common risk, as in Ihori and McGuire (2007) and Schmidt et al. (2007). With this approach, "self-protecting" actions (reducing emissions in this case) undertaken by one country have positive social consequences: they reduce the probability of an undesirable state (e.g. a high damage scenario) occurring in other countries as well.

This particular method of modelling uncertainty endogenously can be contrasted with that taken by Walker and Gardner (1992), who consider, in an experimental setting, a repeated game of common property resource use with an endogenous continuation probability. In their experiment, higher aggregate exploitation of the resource in one period translates into a higher probability that the resource is destroyed before the next period. With this approach, when a bad state occurs (i.e. when the resource is destroyed), countries, in aggregate, suffer a discrete, exogenous loss equal to the entire value of the resource. In other words, the loss suffered in a bad state is fixed. This approach would capture well the possibility of a true climate catastrophe—an event that becomes more likely as aggregate emissions increase but whose impact, once it occurs, is largely independent of aggregate emissions. However, it does not allow for more severe instances of gradual climate change where economic damage is still proportional to the aggregate level of emissions. An example of such an intermediate outcome would be if positive natural feedbacks from global warming (e.g. the large-scale release of greenhouse gases from sinks) turned out to be strong rather than weak, so that economic damage per unit of aggregate emissions is higher than expected. In such a case the loss suffered in a bad state will depend positively on aggregate emissions. An advantage of the approach taken in the present study is that while it allows the probability distribution over all possible states of the world to depend on aggregate emissions, it does not restrict the form of damages in any particular state of the world. Thus, in principle, both catastrophic impacts and impacts resulting from gradual climate change can be accounted for.

3 Underlying Model and Benchmark Outcomes

3.1 A Model of International Emissions with Endogenous Uncertainty

Consider a world of $N \geq 3$ identical countries, each of which emits a pollutant that damages a shared environmental resource. A typical country $i \in \{1, ..., N\}$ is involved in a production activity that generates emission level $q_i \in [0, q_{max}]$ and yields a private benefit of $B(q_i) \geq 0$. Since $B(q_i)$ differs by country only insofar as emission levels are concerned, the same benefit function applies in each country. As in Hoel (1992), I assume that $B'(q_i) > 0 > B''(q_i)$ for $q_i < q_{max}$, $B(0) = 0$, and $B'(q_{max}) = 0$. The implications of last two requirements are that $B(q_i)$ measures i's benefit from emitting relative to the case of no emissions and $q_i = q_{max}$ is the optimal amount of emissions for i in the absence of environmental concerns (business-as-usual emissions). I assume the following linear-quadratic form for $B(q_i)$:

$$B(q_i) = b\left(q_i - \frac{1}{2}q_i^2\right)$$

which implies $q_{max} = 1$ since $B'(q_i) = b(1 - q_i)$. The parameter $b > 0$ measures the slope and intercept of each country's marginal benefit from emissions curve.

Pollution is a global public bad: at some time in the future each country suffers a damage cost of $D(Q, \omega) \geq 0$ from the global level of pollution, denoted Q. Individual emissions are aggregated into global emissions using the unweighted summation technology $Q \equiv \sum_{j=1}^{N} q_j$, which implies that different countries' emission are perfectly substitutable (i.e. a unit of pollution contributes equally to the global stock regardless of where it is emitted). Since damages resulting from the accumulation of pollution are subject to uncertainty, the damage cost depends upon the state of the world, $\omega \in \Omega$, where Ω is the set of all possible states of the world. I assume that

$D_Q > 0$ and $D_{QQ} \geq 0$, so that the damage cost is increasing at a non-decreasing rate in aggregate emissions in any particular state.[18] These assumptions are standard in the IEA literature but, for reasons mentioned above, are not ideally suited to capture catastrophic impacts resulting from discrete, abrupt environmental change. On the other hand, they are well-suited to capture uncertain impacts resulting from more severe instances of gradual climate change. Damages are assumed to take the following linear form:

$$D(Q, \omega) = d(\omega)Q$$

where $d(\omega) \geq 0$ is a constant that measures the private marginal damage cost of emissions in state ω. Here, uncertainty relates to the particular value for a parameter of a more general damage function that does not differ across states. Numerous studies of environmental problems under uncertainty adopt the same approach. For example, in Bramoulle and Treich (2009), there is uncertainty with regard to a damage parameter that enters multiplicatively into the damage function, just as is the case here. Similarly, Stern (2007) assumes that differences in damages across states can be captured by differences in the exponent of the damage function.[19]

Notice also that, in this model, uncertainty applies to the level of a common marginal damage parameter rather than the distribution of a particular set of damage parameters, since all countries realize the same $d(\omega)$. This implies that countries are both *ex-ante* and *ex-post* identical with regard to the economic impacts of environmental change. Thus, risk is perfectly correlated across countries and so there are no opportunities to enter into risk-sharing arrangements.[20]

Countries can only speculate about the true state of the world when choosing their irreversible emissions levels and thus expected utility is relevant when making these

---

[18] Subscripts denote partial derivatives.

[19] This applies to economic impacts resulting from gradual climate change. Damage resulting from catastrophic events are modelled differently. See Dietz et al. (2007) p. 141-3.

[20] Heal and Kristrom (2002) defend the view that risk-sharing opportunities are limited due to the high correlation of climate risks across countries.

choices.[21] I assume risk-neutrality, which implies that expected utility is linear in payoffs, so for what follows I shall equate i's expected utility with its expected payoff. Assuming that countries do not discount the future, i's payoff in state ω is given by its benefits from emitting less the damages it incurs later from global emissions:

$$\pi_i = b\left(q_i - \frac{1}{2}q_i^2\right) - d(\omega)Q$$

The expected payoff to country i is given by $\bar{\pi}_i = E_{\omega\in\Omega}\pi_i(q_i, Q, \omega)$.

In keeping with previous analyses of IEAs under uncertainty, I adopt a simple specification of uncertainty by assuming that the marginal damage parameter can be either high or low (see, for example, Ulph 2004, Boucher and Bramoulle 2007, and Kolstad 2007). That is, I assume $\Omega = \{H, L\}$, $d(H) = d_H$ and $d(L) = d_L$, where $d_H > d_L \geq 0$. Under these assumptions, i's expected payoff can be expressed as follows:

$$\bar{\pi}_i = b\left(q_i - \frac{1}{2}q_i^2\right) - \Big(p_H(Q)d_H + (1 - p_H(Q))d_L\Big)Q$$

where $p_H(Q) \in [0, 1]$ is the probability that $\omega = H$, or, in other words, the probability that the private marginal damage cost of emissions is high rather than low. This probability depends on Q, signifying that uncertainty is endogenous. The restriction $p'_H > 0$ applies because the risk of a high damage scenario is increasing in aggregate emissions. For simplicity, I let $p''_H(Q) = 0$ in assuming the following linear form for $p_H(Q)$:

$$p_H(Q) = \frac{\theta Q}{N}$$

where $\theta \in (0, 1]$ is the probability that the private marginal damage cost of emissions will be high if all countries continue to emit at business-as-usual levels, since $p_H(N) = \theta$. Note that $p_H(0) = 0$; if countries eliminate their emissions then the high damage

---

[21] This assumption regarding the time at which information becomes available corresponds to what Kolstad and Ulph (2008b) refer to as 'No Learning.'

scenario will not occur.

Given the specific functional form for $p_H(Q)$, the expected payoff function forming the basis for all subsequent analysis can be obtained. It is given by:

$$\bar{\pi}_i = b\left(q_i - \frac{1}{2}q_i^2\right) - \bar{d}(Q)Q \tag{1}$$

where $\bar{d}(Q) = \left(\frac{\theta Q}{N}\right) d_H + \left(1 - \frac{\theta Q}{N}\right) d_L$ is the expected marginal private damage cost of emissions. The expected marginal social damage cost of emissions is given by $N\bar{d}(Q)$.

3.2 Benchmark Outcomes and the Gains to Cooperation

In this section I demonstrate the need for an IEA by showing that non-cooperative Nash equilibrium emission levels are above those that are socially optimal and thus countries may gain from international cooperation.

3.2.1 Social Optimum

Socially optimal emission levels can be found by choosing each country's emission level in order to maximize the sum of expected payoffs:

$$\max_{q_1,\ldots,q_N \geq 0} \quad \sum_{i=1}^{N} \bar{\pi}_i = \sum_{i=1}^{N} \left[b\left(q_i - \frac{1}{2}q_i^2\right) - \bar{d}(Q)Q \right] \quad s.t. \quad q_i \leq 1 \quad \forall i = 1, \ldots, N$$

where $Q = q_i + Q_{-i}$ and Q_{-i} is the total amount of pollution emitted by all countries other than country i. Since countries are symmetric and the marginal benefit of emitting in each country is declining, efficiency requires that each country emit the same level of pollution. Thus, $q_i = q_c$, the socially optimal emission level, for all $i \in \{1, \ldots, N\}$, so that the previous optimization problem reduces to:

$$\max_{q_c \geq 0} \quad N\bar{\pi}_c = N\left[b\left(q_c - \frac{1}{2}q_c^2\right) - \bar{d}(Q)Q \right] \quad s.t. \quad q_c \leq 1$$

where $Q = Nq_c$. At an interior solution $(0 < q_c < 1)$ the social planner determines q_c by equating the private marginal benefit from emitting to the marginal expected social cost (MESC) of emissions:

$$b(1 - q_c) = N\bar{d}(Q) + Q\frac{\partial \bar{d}(Q)}{\partial q_c} \tag{2}$$

As can be seen from the RHS of (2), the MESC of emissions consists of two terms, indicating that when a country emits an additional unit of emissions its effect can be broken down into two components. First, the additional unit causes damage in all N countries in either state of the world, holding the probability of realizing d_H constant. Thus, the marginal benefit of emitting must be weighed against the expected marginal social damage cost of emissions, which is given by $N\bar{d}(Q)$. Moreover, an additional unit of emissions results in an increase in the expected marginal social damage cost of emissions by making the realization of d_H instead of d_L more likely. This increase applies to each unit of global emissions, Q. Multiplying gives the second term on the RHS of (2). These two terms together constitute the MESC of emissions. Given $Q = Nq_c$, the MESC of emissions can be expressed as a function of q_c:

$$MESC(q_c) = Nd_L + 2N\theta(d_H - d_L)q_c \tag{3}$$

which is non-negative and strictly increasing in q_c. This implies that business-as-usual emissions $(q_c = 1)$ can never be socially optimal, since $B'(1) = 0$. At an interior solution, the socially optimal level of emissions is given by:

$$q_c = \frac{b - Nd_L}{b + 2N\theta(d_H - d_L)}$$

which occurs when $b > Nd_L$. When $b \leq Nd_L$ the socially optimal level of emissions is $q_c = 0$, a corner solution. This type of solution, which implies that there exists a substitute for the production activity responsible for generating emissions, may arise in this model because I am considering the case of non-essential emissions. Emissions are non-essential when the marginal benefit from emitting is positive but finite for zero emissions (Rubio and Ulph, 2006). These conditions hold in the present study, since $0 < b < \infty$. With non-essential emissions it makes sense to eliminate emissions completely if marginal environmental damage for zero emissions is high enough. In the social optimum this will be so if $b \leq Nd_L$.

3.2.2 Nash Equilibrium

The fully non-cooperative, symmetric Nash equilibrium emission level can be found by solving the following optimization problem for country i:

$$\max_{q_i \geq 0} \bar{\pi}_i = b\left(q_i - \frac{1}{2}q_i^2\right) - \bar{d}(Q)Q \quad s.t. \quad q_i \leq 1$$

where $Q = q_i + Q_{-i}$ and Q_{-i} is total amount of pollution emitted by all countries other than country i. At an interior solution ($0 < q_i < 1$), i determines its emission level by equating the marginal benefit of emitting (labelled B' in Figure 1) to the marginal expected private cost (MEPC) of emissions:

$$b(1 - q_i) = \bar{d}(Q) + Q\frac{\partial \bar{d}(Q)}{\partial q_i} \tag{4}$$

The MEPC of emissions differs from the MESC of emissions in two important respects. First, it only partially accounts for the damage cost that an additional unit of emissions causes in either state of the world, holding the probability of the high damage scenario constant. The first term on the RHS of (4) is the expected marginal private (rather than social) damage cost of emissions, $\bar{d}(Q)$. Second, while the MEPC accounts for the fact that emitting an additional unit heightens the risk of a high damage scenario, it only values this impact insofar as a single country's payoff is affected.

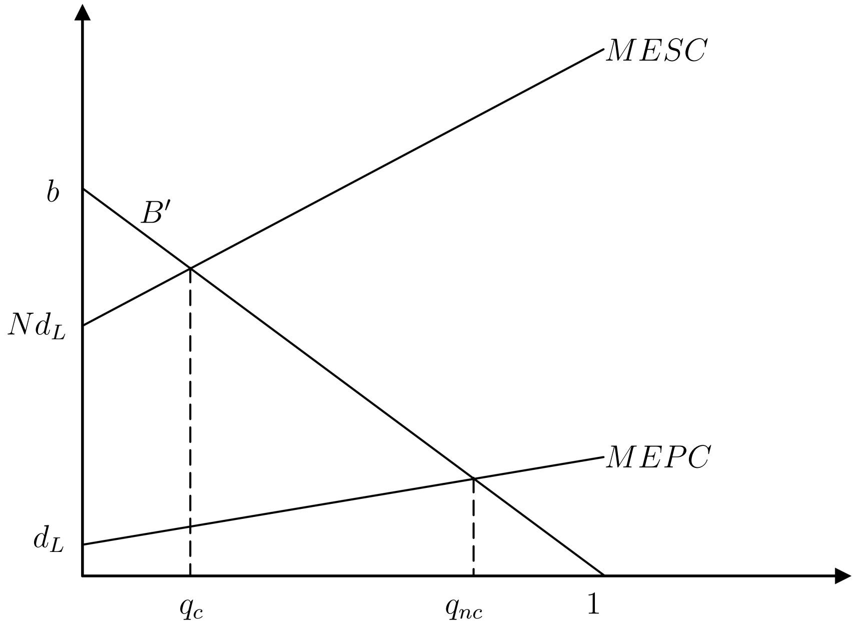


Figure 1: Benchmark outcomes for $b > Nd_L$

In a symmetric equilibrium $q_i = q_{nc}$, the non-cooperative emission level, for all i, and the MEPC of emissions is given by:

$$MEPC(q_{nc}) = d_L + 2\theta(d_H - d_L)q_{nc} \tag{5}$$

From comparison of (3) and (5) it is evident that $MEPC(q) = \frac{1}{N}MESC(q)$. Figure 1 shows that this discrepancy is to blame for higher than socially optimal emissions levels in the non-cooperative Nash equilibrium, provided $b > d_L$. At an interior solution the Nash equilibrium level of emissions is given by:

$$q_{nc} = \frac{b - d_L}{b + 2\theta(d_H - d_L)}$$

which occurs when $b > d_L$. When $b \leq d_L$, the optimal unilateral policy is for a country to abate all of its emissions by setting $q_{nc} = 0$. As was the case in the social optimum, business-as-usual emissions cannot be optimal here either, since $q_{nc} = 1$

never holds.

3.2.3 Gains to Cooperation

Lemma 1 summarizes results pertaining to the benchmark outcomes.

Lemma 1. *For every value of b, the fully cooperative (socially optimal) and fully non-cooperative (Nash equilibrium) emission levels of the international emissions game in which a typical country has expected payoff function (1) are as follows:*

(i) $b > Nd_L$:

$$q_c = \frac{b - Nd_L}{b + 2N\theta(d_H - d_L)}, \quad q_{nc} = \frac{b - d_L}{b + 2\theta(d_H - d_L)}$$

(ii) $d_L < b \leq Nd_L$:

$$q_c = 0, \quad q_{nc} = \frac{b - d_L}{b + 2\theta(d_H - d_L)}$$

(iii) $b \leq d_L$:

$$q_c = 0, \quad q_{nc} = 0$$

There is no need for countries to negotiate an IEA in order to limit their emissions when $b \leq d_L$, since, by Lemma 1 (iii), emission levels resulting from unilateral policy and those that are socially optimal coincide. Therefore, I maintain the following assumption for the remainder of this paper:

Assumption 1. $b > d_L$

When $b > d_L$, $q_c < q_{nc}$ and countries have reason to negotiate an agreement; there are gains to cooperation.

4 The Self-Enforcing IEA

4.1 Two-Stage Single-Coalition Model of an IEA

Following established literature, I represent an IEA with a single-coalition model in which countries are identical. Only one coalition is permitted to form, which may seem restrictive.[22] However, for global environmental problems, restricting the coalition structure can be justified, as for this type of problem IEAs are usually unique and launched by the United Nations. The model has two stages. In the first stage (the Membership Game), countries decide whether or not to sign an IEA, and in the second stage (the Emissions Game), emission levels are determined. Uncertainty is resolved and damages are incurred at the end of the second stage, after emission levels have been chosen and implemented. The model is solved backwards, reflecting the idea that countries look forward to the outcome of the Emissions Game when making their membership decisions. I describe each stage below, in reverse order.

The outcome of the Membership Game is a number of signatories $n \in \{0, ..., N\}$ to the agreement. n is fixed during the Emissions Game. Since the agreement is SCR by assumption, its only clause is that each of the n signatories (a typical signatory being denoted by s) sets its emission level to maximize the coalition's joint payoff, taking as given the emission levels of the non-signatories. Effectively, the signatories act as one unit. By contrast, the $N-n$ non-signatories (a typical non-signatory being denoted by f, for free-rider) behave selfishly. Each non-signatory sets its emission level to maximize its own expected payoff, taking as given the emission levels of all other countries. Equilibrium of the Emissions Game is thus a Nash equilibrium where the signatories are considered as one player, or a 'Partial Agreement Nash

---

[22] Carraro (2000) reviews literature on the endogenous formation of coalitions in which the coalition structure is not artificially restricted. A general result, which applies to games in which coalition formation imposes a positive externality on non-members (as is the case here, since the signatories agree to limit their emissions) is that multiple coalitions tend to form at equilibrium. Osmani and Tol (2006) analyze the possibility of two SCR, self-enforcing IEAs.

Equilibrium' (Tulkens and Chander, 1997). Solving the Emissions Game determines $q_s(n)$ and $q_f(n)$, the equilibrium emission levels of a typical signatory and of a typical non-signatory, respectively, both as functions of n. Equilibrium expected payoffs as functions of n, $\bar{\pi}_s(n)$ and $\bar{\pi}_f(n)$, can then be defined. These functions inform countries' membership decisions in the first stage.

There are two fundamentally different conceptions of the Membership Game that are nonetheless equivalent in terms of the resulting equilibrium. One can think of the Membership Game as a simultaneous-move game in which each country must announce, individually, one of two strategies—to *sign* or to *not sign* the agreement—and a Nash equilibrium results. In a Nash equilibrium each country is playing its best response to the strategies of the others. At an equilibrium in which n countries sign the agreement, it must be true that, given the strategies of the other countries, a signatory could not do any better by switching its strategy from *sign* to *not sign*, so that $\bar{\pi}_s(n) \geq \bar{\pi}_f(n-1)$. Also, again given the strategies of all other countries, a non-signatory could not do any better by switching its strategy from *not sign* to *sign*, or $\bar{\pi}_f(n) \geq \bar{\pi}_s(n+1)$. Any Nash equilibrium of the announcement game described above must simultaneously satisfy these two conditions.

Alternatively, one can assume that a coalition of size n exists and ask whether it is myopically stable, a concept with origins in the literature on cartel stability (d'Aspremont et al., 1983). If the coalition is not myopically stable then it will either shrink or expand, depending on whether there is an incentive for a signatory to exit from the agreement or for a non-signatory to join it. A key assumption underlying the concept of myopic stability is that governments are able to re-optimize their membership decisions and a stable situation is reached through a process of adjustment. Clearly this is a valid assumption in the international arena, since even international law permits countries to withdraw from an agreement and membership is typically open to all countries. However, when thinking in terms of stable coalitions it is best to regard the Membership Game as a modelling tool ensuring that participation in an agreement is voluntary, rather than as a stage occurring prior to the Emissions Game in which countries announce their membership strategies. For all $2 \leq n \leq N$, the following definition of myopic stability applies:

Definition 1. *An IEA with n signatories is myopically stable if $\bar{\pi}_s(n) \geq \bar{\pi}_f(n-1)$ (internal stability condition is satisfied) and $\bar{\pi}_f(n) \geq \bar{\pi}_s(n+1)$ (external stability condition is satisfied).*

These are precisely the conditions defining a Nash equilibrium in the game described above. The internal stability condition ensures that each signatory would prefer to remain a signatory rather than exiting from the agreement, given the membership decisions of all other countries. Complementary to this condition is the external stability condition, which ensures that each non-signatory would prefer to remain outside of the agreement rather than becoming a signatory, again given the membership decisions of all other countries.

Given the n satisfying myopic stability as defined in Definition 1, equilibrium emission levels and welfare—of a typical signatory, of a typical non-signatory, and in aggregate—can be determined as functions of the underlying parameters of the model, and inferences regarding the amount of cooperation achieved by the agreement can be drawn.

4.2 Myopically Stable SCR Agreements with Endogenous Uncertainty

4.2.1 Emissions Game

Suppose that, as the outcome of the Membership Game, there are n signatories and $N-n$ non-signatories. A typical non-signatory $j \in \{1, ..., N-n\}$ chooses emission level q_{fj} to maximize its expected payoff taking the emissions levels of all other countries as given:

$$\max_{q_{fj}\geq 0} \bar{\pi}_{fj} = b\left(q_{fj} - \frac{1}{2}q_{fj}^2\right) - \left[\left(\frac{\theta Q}{N}\right)d_H + \left(1 - \frac{\theta Q}{N}\right)d_L\right]Q \quad s.t. \quad q_{fj} \leq 1$$

where $Q = q_{fj} + Q_{-fj}$ and Q_{-fj} is total amount of pollution emitted by all countries other than non-signatory j. The Lagrangian for this problem is given by:

$$\mathcal{L}_f = b\left(q_{fj} - \frac{1}{2}q_{fj}^2\right) - \left[\left(\frac{\theta Q}{N}\right)d_H + \left(1 - \frac{\theta Q}{N}\right)d_L\right]Q + \lambda_{fj}\left(1 - q_{fj}\right)$$

and the Kuhn-Tucker conditions (KTCs) are:

$$\begin{aligned} \frac{\partial \mathcal{L}_f}{\partial q_f} &= b(1 - q_f) - \frac{1}{N}2\theta(d_H - d_L)Q - d_L - \lambda_f \leq 0 && (6)\\ q_f &\geq 0, \quad q_f\frac{\partial \mathcal{L}_f}{\partial q_f} = 0, \\ \frac{\partial \mathcal{L}_f}{\partial \lambda_f} &= 1 - q_f \geq 0 && (7)\\ \lambda_f &\geq 0, \quad \lambda_f\frac{\partial \mathcal{L}_f}{\partial \lambda_f} = 0 \end{aligned}$$

where symmetry, which implies $q_{fj} = q_f$ and $\lambda_{fj} = \lambda_f$ for all $j \in \{1, ..., N - n\}$, was imposed after differentiation.

The n signatories are assumed to coordinate by choosing their emission levels to maximize their collective expected payoff, taking the emission levels of the non-signatories as given. They face the following optimization problem:

$$\max_{q_{s1},\ldots,q_{sn}\geq 0} \quad \Pi_s \equiv \sum_{i=1}^{n} \bar{\pi}_{si} = \sum_{i=1}^{n}\left[b\left(q_{si} - \frac{1}{2}q_{si}^2\right) - \left[\left(\frac{\theta Q}{N}\right)d_H + \left(1 - \frac{\theta Q}{N}\right)d_L\right]Q\right]$$
$$s.t. \quad q_{si} \leq 1 \quad \forall \quad i = 1, ..., n$$

where $Q = q_{si} + Q_{-si}$ and Q_{-si} is the total amount of pollution emitted by all countries other than signatory i. Since signatories are identical, $q_{si} = q_s$ for all $i \in \{1, ..., n\}$,[23] and the previous optimization problem reduces to:

$$\max_{q_s \geq 0} \quad n\bar{\pi}_s = n\left[b\left(q_s - \frac{1}{2}q_s^2\right) - \left[\left(\frac{\theta Q}{N}\right)d_H + \left(1 - \frac{\theta Q}{N}\right)d_L\right]Q\right] \quad s.t. \quad q_s \leq 1$$

where $Q = nq_s + (N-n)q_f$. The Lagrange function for this problem is given by the following equation:

$$\mathcal{L}_s = n\left[b\left(q_s - \frac{1}{2}q_s^2\right) - \left[\left(\frac{\theta Q}{N}\right)d_H + \left(1 - \frac{\theta Q}{N}\right)d_L\right]Q + \lambda_s(1 - q_s)\right]$$

and the associated KTCs, which apply for all $n \in \{1, ..., N\}$, are:

$$\frac{\partial \mathcal{L}_s}{\partial q_s} = n\left[b(1 - q_s) - \frac{n}{N}2\theta(d_H - d_L)Q - nd_L - \lambda_s\right] \leq 0 \tag{8}$$

$$q_s \geq 0, \quad q_s\frac{\partial \mathcal{L}_s}{\partial q_s} = 0,$$

$$\frac{\partial \mathcal{L}_s}{\partial \lambda_s} = 1 - q_s \geq 0 \tag{9}$$

$$\lambda_s \geq 0, \quad \lambda_s\frac{\partial \mathcal{L}_s}{\partial \lambda_s} = 0$$

Determining the equilibrium to the Emissions Game involves solving KTCs (6)-(9) simultaneously for $q_f(n)$ and $q_s(n)$ using $Q = nq_s + (N-n)q_f$ when $n \in \{1, ..., N-1\}$. When $n = 0$, finding the equilibrium to the Emissions Game requires solving KTCs (6) and (7) for $q_f(0)$ only, since $q_s(0)$ is undefined and irrelevant. In such a case no coalition forms. On the other hand, when $n = N$, equilibrium is defined by KTCs (8) and (9) only, since $q_f(N)$ is not a quantity of interest. In such a case all countries are signatories.

---

[23] The assumption that each signatory chooses emission level q_s can be justified on the grounds that equal burden-sharing within the coalition emerges as the equilibrium of a cooperative game in which countries bargain over individual obligations. This is true for various burden-sharing rules (e.g. Nash's bargaining rule or assigning each country its Shapley value) provided that countries are identical. See Barrett (1997a) and Botteon and Carraro (1997).

Lemma 2. *Given KTCs (6)-(9) and* $n \in \{1, ..., N-1\}$*, the unique equilibrium strategy of a typical signatory and of a typical non-signatory in the Emissions Game as functions of the number of signatories in the first stage, n, is as follows:*

(i) Interior Solution $(0 < q_s < 1 \text{ and } 0 < q_f < 1)$

$$q_s(n) = 1 - \frac{nN\left(2\theta d_H + (1-2\theta)d_L\right)}{Nb + 2\theta\left(d_H - d_L\right)\left(n^2 + N - n\right)} \tag{10}$$

$$q_f(n) = 1 - \frac{N(2\theta d_H + (1-2\theta)d_L)}{Nb + 2\theta(d_H - d_L)\left(n^2 + N - n\right)} \tag{11}$$

so that $q_s > 0$ *iff*

$$g(n, b, \chi) \equiv N(b - nd_L) - 2\theta(d_H - d_L)(n-1)(N-n) > 0$$

(ii) Signatory Corner Solution $(q_s = 0 \text{ and } 0 < q_f < 1)$

$$q_s(n) = 0 \tag{12}$$

$$q_f(n) = \frac{N(b - d_L)}{Nb + 2\theta(d_H - d_L)(N-n)} \tag{13}$$

where $q_s = 0$ *iff*

$$g(n, b, \chi) \leq 0$$

Proof: See Appendix

In Lemma 2 and for the remainder of this paper, I use the symbol χ as shorthand for the set of parameters (d_H, d_L, θ, N).

Turning briefly to the possibilities for n not covered by Lemma 2, when $n = N$, the emission level of a typical signatory is given by equation (10) if $g(n, b, \chi) > 0$ and by equation (12) otherwise (see Lemma 1). In this case the coalition consists of all countries cooperating with one another. Thus, all reciprocal externalities are fully accounted for and emission levels coincide with those that are socially optimal. When $n = 0$, the equilibrium emission level of a typical non-signatory is given by equation (11). In this case all countries act as singletons and emission levels are the same as those in the non-cooperative Nash equilibrium.

Two observations with regard to Lemma 2 can be made. First, at an interior solution, each signatory reduces its emissions by n times the amount that each non-signatory reduces its emissions (recall that $q_{max} = 1$ in every country). This occurs because the agreement binds each signatory to account for the negative effect of its emissions on the other $n - 1$ signatories' expected payoffs, in addition to the effect on its own expected payoff. Each non-signatory is not so inclined. The second observation, which is closely related to the first, is that if either group of countries are to reduce their emission levels to zero, it will be the signatories.[24]

I now want to determine more precisely for what parameter values the two different types of solutions occur. For the purposes of stability analysis I am particularly interested in obtaining, for every possible parameter combination, the critical value(s) of n at which one type of solution yields to another, if such values exist. This knowledge will allow me to define expected payoffs as functions of n for any given b and values for the parameters in χ.

My approach is to use the function implicitly defined by $g(n, b, \chi) = 0$, which I denote by $b(n, \chi)$, to map the type of solution into (b, n) space. Given values for the parameters in χ, this allows me to determine, for each value of b, the type of

---

[24] This would not necessarily be true if the signatories were to act as Stackleburg leader with respect to the non-signatories in the Emissions Game. The Stackleburg assumption is usually justified on the grounds that membership in an agreement acts as a commitment device. Rubio and Ulph (2006) show that, in a model with declining marginal benefits and increasing marginal damage costs (as is the case here), signatories may use their first-mover advantage to corner non-signatories into eliminating their emissions entirely by selecting high emission levels themselves. High levels of participation can be sustained through this mechanism. However, a non-signatory corner solution occurs only when marginal environmental damage is high relative to the marginal benefit of emitting. In this case, if non-signatories were to challenge the commitment of the signatories to their reduction targets by flaunting the rules of the Stackleburg Emissions Game, instead emitting high levels of pollution, then the signatories would have a strong incentive to renegotiate their agreement. The Stackleburg assumption reflects a presumed degree of commitment, but does not imply absolute commitment.

solution presenting at every n, and, therefore, the critical values of n at which one type of solution gives way to another. Since it is true in general that the mapping of solution types into (b, n) space is affected by the parameters in χ, I then vary these parameters and distinguish between different cases. This analysis of the parameter space is developed in the following pages and the final results are illustrated in Figure 2. The next lemma fixes n and the parameters in χ and provides the value of b separating the interior solution from the signatory corner solution.

Lemma 3. *Provided $n \in \{1, ..., N\}$, the equilibrium to the Emissions Game is given by the interior solution for $b > b(n, \chi)$, and by the signatory corner solution otherwise.*

Since $g(n, b, \chi) > 0 \Leftrightarrow b > b(n, \chi)$, Lemma 3 follows directly from Lemmas 1 and 2, which, taken together, indicate that the equilibrium to the Emissions game is given by the interior solution when $g(n, b, \chi) > 0$ for all $n \in \{1, ..., N\}$. Essentially, Lemma 3 states that signatories in a coalition of given size prefer to eliminate their emissions entirely only for b sufficiently low.

Given Lemma 3, mapping the type of solution into (b, n) space is simply a matter of characterizing the function $b(n, \chi)$, a task to which I now turn. Regardless of the values for the parameters in χ, the endpoints of $b(n, \chi)$ are given by $b(1, \chi) = d_L$ and $b(N, \chi) = Nd_L$. Assuming that n may be any positive real number,[25] $b(n, \chi)$ is initially increasing and is concave in n, and achieves a unique maximum at $n = \hat{n}$, where $\hat{n} < N$ if $Nd_L < 2\theta(d_H - d_L)(N - 1)$. Moreover, for all $b \leq b(\hat{n}, \chi)$, it is possible to define n_1 and n_2, where $n_1 \leq n_2$, as the two positive real roots of $b = b(n, \chi)$. Given this behavior of the function $b(n, \chi)$, $\hat{n} \geq N$ represents a fundamentally different case than $\hat{n} < N$, since in the latter case the critical value n_2 becomes relevant over some

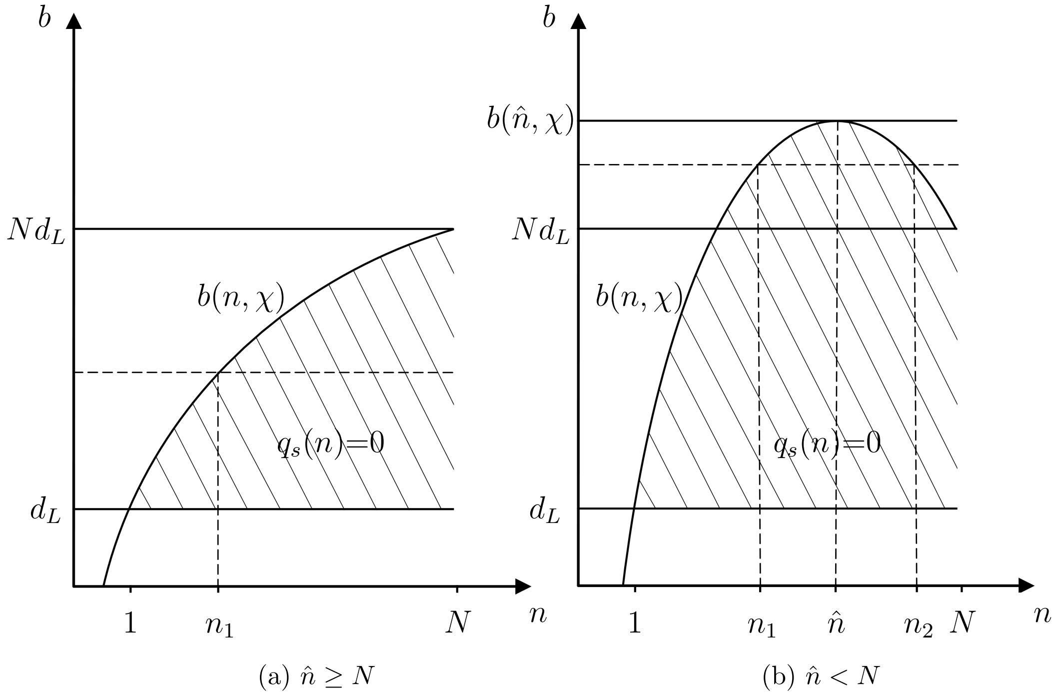
---

[25] It does no harm to assume that n can be any positive real number in order to study the behavior of the function $b(n, \chi)$. Once I know the properties of $b(n, \chi)$, I can characterize the values of this function with respect to n, but with n restricted to an integer in the interval $[0, N]$. My approach is to ignore the fact that n must be an integer, however, since accounting for this would not lead to any significant changes to what follows. Proposition 2, for example, which provides the main findings of the present study, would be unaffected.

range of b. Thus, a single condition on the parameters in χ (i.e. that implied by $\hat{n} < N$) can be used to distinguish between two qualitatively different mappings of solution types into (b, n) space. For each of these qualitatively different mappings, Proposition 1 provides the critical values of n yielding one type of solution or another by first fixing b and the parameters in χ and then considering the type of solution that presents at each n.

Proposition 1. *Given Lemma 3, (a) if $\hat{n} \geq N$ and (i) $d_L < b \leq Nd_L$, then the unique equilibrium of the Emissions Game is given by the interior solution for all $n \in [0, n_1)$, and by the signatory corner solution for all $n \in [n_1, N]$; (ii) $b > Nd_L$, then the equilibrium is the interior solution for all n. (b) If $\hat{n} < N$ and (i) $d_L < b \leq Nd_L$, then the equilibrium is again the interior solution for all $n \in [0, n_1)$, and the signatory corner solution for all $n \in [n_1, N]$; (ii) $Nd_L < b \leq b(\hat{n}, \chi)$, then the equilibrium is given by an interior solution for all $n \notin [n_1, n_2]$, and by the signatory corner solution for all $n \in [n_1, n_2]$, where $0 < n_1 \leq n_2 < N$; (iii) $b > b(\hat{n}, \chi)$, then the equilibrium is given by the interior solution for all n.*

Proposition 1, which is illustrated in Figure 2, states that if b is low enough (i.e. $d_L < b < Nd_L$), there will be a critical coalition size defined by n_1 at and above which the signatories prefer to cut their emissions to zero. However, if b is high enough (i.e. $b > Nd_L$ when $\hat{n} \geq N$, and $b > b(\hat{n}, \chi)$ when $\hat{n} < N$), no such critical coalition size will exist and the signatories will choose positive emission levels for all n. When $\hat{n} < N$, there exists an intermediate range of b, given by $Nd_L < b \leq b(\hat{n}, \chi)$, for which the signatories prefer to eliminate their emissions completely once the coalition reaches a critical size, again defined by n_1, but when the coalition grows to include more than n_2 countries the signatories prefer to emit positive levels of pollution once more. Interestingly, in this case, for some coalition sizes the signatories prefer to cut their emissions to levels below those that are socially optimal, since $b > Nd_L$ implies socially optimal emission levels are positive (see Lemma 1).



(a) $\hat{n} \geq N$

(b) $\hat{n} < N$

Figure 2: Parameter Space

Notice that these three different solution patterns can be explained with reference to the functions $B'(0)$ and MECC(0), where $B'(0) = b$ is a signatory's marginal benefit from emitting at $q_s = 0$ and MECC(0) is the marginal expected coalitional damage cost of emissions on the first unit of pollution emitted by a signatory. The MECC of emissions at $q_s = 0$ can be deduced from KTC (8) and is given by:

$$MECC(0) = n\left[\left(\frac{\theta Q_f}{N}\right)d_H + \left(1 - \frac{\theta Q_f}{N}\right)d_L\right] + n\left[\frac{1}{N}\theta(d_H - d_L)Q_f\right] \tag{14}$$

where $Q_f = (N-n)q_f$ is the total amount of pollution emitted by the non-signatories. There are two opposing effects on MECC(0) as the coalition size n increases. First, given Q_f, an increase in n causes MECC(0) to rise since the presence of an additional coalition member leads each signatory to take more fully into account the expected damage arising from its first unit of emissions. However, Q_f responds to a change in n, and it turns out that an increase in n leads to a decrease in Q_f. While the level of pollution emitted per non-signatory q_f, which is given by equation (13) in equilibrium

(q_f at the signatory corner solution), increases with the size of the coalition, the number of non-signatories, $N - n$, falls, such that the net effect is a decrease in Q_f. The strength of these two opposing effects, both of which are affected by each one of the parameter values, determines whether the MECC(0) curve cuts the line $B'(0) = b$ once, twice, or not at all, and, therefore, the type of solution pattern as n increases. For example, when the parameters are such that $\hat{n} < N$, MECC(0) is initially increasing in the coalition size but eventually begins to decline as n increases. In this case, there will exist at least one value of b such that MECC(0) cuts $B'(0) = b$ twice, once on the way up at n_1, and once on the way down, at n_2, so that the signatories prefer to eliminate their emissions for all coalitions of size $n \in [n_1, n_2]$, but otherwise prefer to emit positive levels of pollution.

I now move on to analysis of the Membership Game. Using the equilibrium emission levels in Lemma 2, I am able to find the equilibrium expected payoffs at each type of solution. Combining this information with Proposition 1 allows me to determine, for any combination of parameter values, the equilibrium expected payoff to a signatory and to a non-signatory, both as functions of n. Finding the stable coalition size is then simply a matter of applying the definition of myopic stability found in Definition 1.

4.2.2 Membership Game

In this section I conduct stability analysis and consider how the level of participation in an IEA varies with the parameters of the model. I derive results for the special case of $d_L = 0$, but claim that these results hold more generally. When $d_L = 0$, the condition $\hat{n} < N \Leftrightarrow Nd_L < 2\theta(d_H - d_L)(N - 1)$ must hold. Thus, Proposition 1(b) applies and, if $b > b(\hat{n}, \chi)$, there will be an interior solution for all n. If, on the other hand, $b < b(\hat{n}, \chi)$, then the signatories eliminate their emissions entirely for all coalitions of size $n \in [n_1, n_2]$, but will otherwise emit positive levels of pollution.

Setting $d_L = 0$, the expected payoff to a typical signatory and to a typical non-signatory in equilibrium can be obtained by substitution using the emission levels given in Lemma 2.

Lemma 4. *Given Lemma 2 and $d_L = 0$, for any $n \in \{1, ..., N\}$, the equilibrium expected payoff received by a typical signatory and by a typical non-signatory in the Emissions Game are as follows:*

(i) Interior Solution ($0 < q_s < 1$ and $0 < q_f < 1$)

$$\bar{\pi}_s(n) = \frac{1}{2}b - \frac{\theta b d_H N^2 \left(Nb + 2\theta d_H n^2\right)}{\left(Nb + 2\theta d_H (n^2 + N - n)\right)^2}$$
$$\bar{\pi}_f(n) = \frac{1}{2}b - \frac{\theta b d_H N^2 \left(Nb + 2\theta d_H\right)}{\left(Nb + 2\theta d_H (n^2 + N - n)\right)^2}$$

(ii) Signatory Corner Solution ($q_s = 0$ and $0 < q_f < 1$)

$$\bar{\pi}_s(n) = -\frac{\theta b^2 d_H N (N-n)^2}{\left(Nb + 2\theta d_H (N - n)\right)^2}$$
$$\bar{\pi}_f(n) = \frac{1}{2}b - \frac{\theta b d_H (N-n)^2 \left(Nb + 2\theta d_H\right)}{\left(Nb + 2\theta d_H (N - n)\right)^2}$$

Proposition 1 determines for which $n \in \{1, ..., N\}$ each set of expected payoffs apply.

From Lemma 4 it is clear that, for a coalition of given size, a non-signatory receives a higher expected payoff than a signatory at an interior solution for all $n > 1$. This result arises from the assumptions that all countries are identical with respect to their benefit and damage cost functions, and that each country's damage cost depends only on the global level of emissions. Under these assumptions, all countries bear the same damage cost but those emitting higher levels of pollution obtain larger benefits. Provided $n > 1$, each signatory emits less than a non-signatory at an interior solution (see Lemma 2), and, therefore, each non-signatory receives a higher expected payoff. The same logic tells us that non-signatories must receive higher expected payoffs than signatories at the signatory corner solution as well, since signatories do not emit any pollution while non-signatories emit positive levels.

While it is true that a non-signatory receives a higher payoff than a signatory for a coalition of given size, Definition 1 indicates that stability depends on a comparison between the expected payoff of a signatory at one coalition size and the expected payoff of a non-signatory at another. For instance, a coalition of size n is internally myopically stable if $\bar{\pi}_s(n) \geq \bar{\pi}_f(n-1)$. Thus, to understand why a particular coalition may be stable is to understand the mechanism preventing the exit of a signatory or the entrance of a non-signatory. The following two lemmas shed light on this mechanism.

Lemma 5. *For all $n \in \{1, ..., N\}$, the total amount of pollution emitted is strictly decreasing in the size of the coalition.*
Proof: See Appendix

Lemma 6. *For all $n \in \{2, ..., N\}$, $q_f(n-1) - q_s(n) > 0$.*
Proof: See Appendix

Suppose that a signatory is contemplating exit from the agreement (or not signing it to begin with). Lemma 5 implies that, by exiting, the signatory would expect to incur additional damages in the future, since the total amount of pollution must increase if the coalition size decreases. On the other hand, the signatory would expand its emissions, since, by Lemma 6, $q_s(n) - q_f(n-1) > 0$, and would therefore gain emission benefits. If the rise in expected damage cost outweighs the increase in emission benefits, the signatory will remain within the coalition.

Alternatively, one can think in terms of the mechanism preventing a non-signatory from joining the agreement. Lemmas 5 and 6 imply that by opting into the agreement (or signing it to begin with) a non-signatory saves on expected damage costs but must give up emission benefits. Next, I determine the coalition size at which these implicit punishments and rewards give rise to a stable situation, as defined in Definition 1.

Proposition 2. *The unique, myopically stable IEA consists of two countries when the parameters satisfy* $Nb \geq 2\theta d_H \left(N - 4 + 2\left(N^2 - 3N + 3\right)^{1/2}\right)$. *Otherwise, participation in a myopically stable IEA will be limited to just one country.*

Proof: See Appendix

Proposition 2, which relies on analytical arguments as well as numerical simulations, implies that any coalition consisting of three or more countries cannot be myopically stable. The reason for this is that the internal myopic stability condition is not satisfied for coalitions of size $n \geq 3$, so that by exiting from the agreement a signatory gains more in emission benefits than it expects to lose in increased damage costs. The same may or may not be true for a coalition of size $n = 2$, which will be internally myopically stable provided that the condition on the parameters given in Proposition 2 is satisfied. Since whether or not this condition is satisfied determines the size of the myopically stable IEA, it can be used to study the relationship between the parameters of the model and the stable coalition size.

Corollary 1. *The myopically stable coalition size is non-increasing in* θ, *and* d_H, *and non-decreasing in* b.

Corollary 1 is consistent with a common finding of previous literature, which is that participation varies inversely with the ratio of marginal damage cost parameters (d_H and θ) to marginal benefit parameters (b).

5 Discussion

The results of the previous section indicated that no more than two countries would be willing to sign an SCR IEA in the presence of endogenous uncertainty. One question this paper means to address is whether persistent uncertainty can be beneficial to international cooperation. When uncertainty is modelled endogenously and agents are risk-neutral, what qualifies as an increase in uncertainty is an open question.[26] Clearly, when $d_H = d_L$ there is no uncertainty as to the marginal damage cost of emissions. In this case it can be shown that either two or three countries would be willing to sign an SCR IEA (see Finus 2001, Chapter 13). Thus, it appears as though the presence of uncertainty can only harm the prospects for effective international collaboration. An increase in d_H holding d_L constant may also plausibly be regarded as an increase in uncertainty, and, by Corollary 1, this too can only reduce participation in an IEA. I now turn to assessing the generality of this result by considering to what extent the present analysis captures essential features of the climate problem—the leading example used to guide modelling decisions throughout this paper. Possible extensions present themselves along the way.

According to experts, the five key features of climate change are uncertainty, strategic interactions, asymmetric impacts, long time horizons, and potential irreversibilities (IPCC, 1995). The present modelling exercise was designed to capture the first two features (uncertainty and strategic interactions) so that they could be analyzed jointly. On the other hand, the last three features were either included but were not the focus (irreversibilities) or were ignored completely (asymmetric impacts and long time horizons) for the sake of simplicity.

Consider first how the inclusion of asymmetric impacts might affect the prediction that uncertainty will be harmful to participation in an IEA. Asymmetric impacts are

---

[26] Rothschild and Stiglitz (1970) provide now well-accepted definitions of increasing risk when risk is exogenous and agents are risk-averse. However, it is not obvious to me how to apply these definitions in the present study.

likely to apply for small increases in temperature (relative to pre-industrial times), but as temperatures rise impacts are expected to become more severe and increasingly correlated (Dietz et al., 2007). Thus, an appropriate method of modelling the feature of asymmetric impacts together with endogenous uncertainty would be to write country i's expected damages as follows:

$$p_H(Q)\,(d_H Q) + (1 - p_H(Q))\,(d_i Q) \tag{15}$$

where d_i varies by country and $d_H > d_i$ for all $i \in \{1, ..., N\}$. When i's expected damages are given by (15), the probability with which all countries realize the same damage parameter is increasing in aggregate emissions. In other words, damages are expected to become more correlated as emissions rise. I suspect that this relationship between aggregate emissions and expected damages in each country may work against an optimistic theoretical result, which is that asymmetric impacts in the absence of uncertainty can be exploited by self-financed transfers to expand the stable coalition size (see Barrett 2001). As shown by Carraro and Siniscalco (1993), self-financed transfers (payments from the coalition to non-signatories that are funded by the coalition's surplus) are ineffective when countries are symmetric, unless some countries can commit to cooperation. Assuming they cannot, endogenous uncertainty may restrict the potential for self-financed transfers to support meaningful cooperation to the extent that it renders countries more alike in terms of impacts.

Combining the features of uncertainty and long time horizons may also lead to a model that is more than the sum of its parts. Long time horizons raise intergenerational issues because today's countries may discount future damages when choosing how much to emit. A country's discount rate will typically depend on, among other things, its future rate of consumption growth. Since consumption and its growth are affected by the economic damage resulting from climate change, a country's discount rate ought to be modelled endogenously and will vary with aggregate emissions.[27] This holds true whether or not damages are subject to uncertainty, but, under certain assumptions, the discount rate will also be affected by the presence of uncertainty (Stern 2007, Technical Annex to Chapter 2). That is, uncertainty may interact with the feature of long time horizons through the discount rate.

Finally, the feature of potential irreversibilities is intimately related to uncertainty, since it may be desirable to reverse decisions only if better information will be available in the future. Studies assuming a single decision-maker (e.g. Kolstad 1996) emphasize the trade-off between delaying action, so as to avoid needlessly sinking resources into costly abatement capital should damage from climate change turn out to be low, and restraining emissions growth early on, in case climate change turns out to be severe and the stock of emissions cannot be quickly reduced. The model developed and analyzed in the present study is able to capture these two types of irreversibilities, since abatement costs cannot be recovered, nor can aggregate emissions be reduced, once the true state of the world is known. However, the related trade-off between earlier and later action is not present. This omission is due to the fact that I have taken a rather simplistic view of the learning process, assuming that uncertainty is resolved instantly and completely at the moment economic damages are incurred. A more realistic approach would be to assume a two period model in which some measure of learning occurs between periods, as in Ulph (2004). By allowing countries to refine their abatement policies and to reconsider their membership decisions given new information, the issue of whether to reduce emissions earlier or later arises, and may also be manifest in first-period membership choices.

---

[27] To my knowledge, all studies of IEAs with discounting have assumed an exogenous discount rate. When discount rates are exogenously given, higher discount rates (lower discount factors) have the effect of making cooperation through an IEA more difficult to achieve (Finus, 2000). Although higher discount rates dull free-rider incentives by compelling the coalition to emit more, they also limit the effectiveness of punishments because a signatory exiting from the agreement will suffer less when the remaining coalition members expand their emissions.

6 Conclusion

In this paper I have extended the standard model of an IEA to better account for uncertainty with regard to economic impacts resulting from environmental change. Not only has uncertainty been identified as an important feature of international environmental problems, scientific evidence also indicates that the probabilities over which uncertain impacts occur are affected by today's actions. One contribution of the present study was to provide the first analysis of an IEA under endogenous uncertainty, with a special focus on the climate issue. This was accomplished in the context of risk-neutral countries by modifying the concept of self-protecting actions to apply at a group level. Assuming a linear damage function in any particular state of the world, I found that the endogeneity of uncertainty provides additional opportunities for strategic behavior among countries, since the level of pollution emitted by one country affects the expected marginal damage cost of emissions in all countries.

As was just discussed, the modelling exercise conducted above should not be regarded as an all-encompassing analysis of an IEA on the climate issue, as some important features of the problem were excluded from the analysis. Since it would appear that the effect of uncertainty cannot be analyzed separately from the features of the climate problem that have been excluded, the predictions of the present analysis cannot be accepted without caution (although, at least in the case of asymmetric impacts, these predictions are unlikely to change qualitatively by the inclusion of additional features).

Nevertheless, with an appeal to the need to engage in tactical modelling, I propose two findings. First, I found that the level of participation in a myopically stable, SCR IEA varies inversely with the ratio of marginal damage parameters to marginal benefit parameters. This finding is consistent with a core theoretical result, confirming that the standard theory is robust to the particular extension considered here. Second, I found that uncertainty may only affect the prospects for meaningful international cooperation negatively. This implies that global environmental problems involving considerable uncertainties, such as climate change, may require more attention in terms of additional stabilizing measures.

However, even recognizing the need to engage in tactical modelling, the model used here ought to be extended to include the highly uncertain yet catastrophic impacts that have recently been argued to command special influence over global climate policy. Thus, the present analysis must be viewed as providing only an initial answer to the question of how uncertainty may affect the prospects for meaningful international cooperation. Future work on the under-appreciated yet important feature of uncertainty can build on the modelling framework adopted and developed in the present study to address the qualifications mentioned above.

References

Baker, E. (2005). Uncertainty and learning in a strategic environment: global climate change. *Resource and Energy Economics 27*(1), 19–40.

Barrett, S. (1990). The problem of global environmental protection. *Oxford Review of Economic Policy 6*(1), 68–79.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers 46*(0), 878–94.

Barrett, S. (1997a). Heterogeneous international environmental agreements. In C. Carraro (Ed.), *International Environmental Negotiations: Strategic Policy Issues*, pp. 9–25. Cheltenham: Edward Elgar Publishing.

Barrett, S. (1997b). The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics 19*(4), 345–361.

Barrett, S. (1998a). On the theory and diplomacy of environmental treaty-making. *Environmental & Resource Economics 11*(3), 317–333.

Barrett, S. (1998b). Political economy of the Kyoto Protocol. *Oxford Review of Economic Policy 14*(4), 20–39.

Barrett, S. (2001). International cooperation for sale. *European Economic Review 45*(0), 1835–50.

Barrett, S. (2002). Consensus treaties. *Journal of Institutional and Theoretical Economics (JITE) 158*(4), 529–547.

Barrett, S. (2003). *Environment and Statecraft*. Oxford, UK: Oxford University Press.

Barrett, S. (2005). The theory of international environmental agreements. In K. G. Mler and J. R. Vincent (Eds.), *Handbook of Environmental Economics*, Volume 3 of *Handbook of Environmental Economics*, Chapter 28, pp. 1457–1516. Elsevier.

Biancardi, M. and A. D. Liddo (2008). International environmental agreement: a dynamic model of emissions reduction. Quaderni DSEMS 13-2008, Dipartimento di Scienze Economiche, Matematiche e Statistiche, Universita' di Foggia.

Black, J., M. D. Levi, and D. de Meza (1993). Creating a good atmosphere: Minimum participation for tackling the 'greenhouse effect'. *Economica 60*(239), 281–93.

Botteon, M. and C. Carraro (1997). Burden-sharing and coalition stability in environmental negotiations with asymmetric countries. In C. Carraro (Ed.), *International Environmental Negotiations: Strategic Policy Issues*. Cheltenham: Edward Elgar Publishing.

Botteon, M. and C. Carraro (1998). Strategies for environmental negotiations: Issue linkage with heterogeneous nations. In N. Hanley and H. Folmer (Eds.), *Game Theory and the Environment*, pp. 181–203. Cheltenham: Edward Elgar Publishing.

Boucher, V. and Y. Bramoulle (2007). Risk aversion and international environmental agreements. Cahiers de recherche 0739, CIRPEE.

Bramoulle, Y. and N. Treich (2009). Can uncertainty alleviate the commons problem? *Journal of the European Economic Association 7*(5), 1042–1067.

Brekke, K. and O. Johansson-Stenman (2008). The behavioral economics of climate change. *Oxford Review of Economic Policy 24*(2), 280–297.

Breton, M., L. Sbragia, and G. Zaccour (2010). A dynamic model for international environmental agreements. *Environmental & Resource Economics 45*(1), 25–48.

Buys, P., U. Deichmann, C. Meisner, T. Ton-That, and D. Wheeler (2007). Country stakes in climate change negotiations: two dimensions of vulnerability. Policy Research Working Paper Series 4300, The World Bank.

Cabon-Dhersin, M.-L. and S. V. Ramani (2006). Can social externalities solve the small coalitions puzzle in international environmental agreements? *Economics Bulletin 17*(4), 1–8.

Carraro, C. (2000). Costs, structure and equity of international regimes for climate change mitigation. *FEEM Working Paper* (61).

Carraro, C., J. Eyckmans, and M. Finus (2006). Optimal transfers and participation decisions in international environmental agreements. Working Papers 2006.44, University of Venice “Ca’ Foscari,” Department of Economics.

Carraro, C., C. Marchiori, and S. Oreffice (2004). Endogenous minimum participation in international environmental treaties. CEPR Discussion Papers 4281.

Carraro, C. and D. Siniscalco (1993). Strategies for the international protection of the environment. *Journal of Public Economics 52*(3), 309–328.

Carraro, C. and D. Siniscalco (1997). R&D cooperation and the stability of international environmental agreements. In C. Carraro (Ed.), *International environmental negotiations: strategic policy issues*, pp. 71–96. Cheltenham: Edward Elgar Publishing.

Carraro, C. and D. Siniscalco (1998). International institutions and environmental policy: International environmental agreements: Incentives and political economy. *European Economic Review 42*(3-5), 561–572.

Chander, P. and H. Tulkens (1995). A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. *International Tax and Public Finance 2*(2), 279–293.

Chichilnisky, G. and G. Heal (1993). Global environmental risks. *Journal of Economic Perspectives 7*, 65–86.

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economic Theory 63*(2), 299–325.

Dasgupta, P. (1982). *Control of Resources.* Cambridge, MA: Harvard University Press.

d'Aspremont et al., C. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics 16*(1), 17–25.

de Zeeuw, A. (2008). Dynamic effects on the stability of international environmental agreements. *Journal of Environmental Economics and Management 55*(2), 163–174.

Dellink, R., M. Finus, and N. Olieman (2005). Coalition formation under uncertainty: The stability likelihood of an international climate agreement. Working Papers 2005.98, Fondazione Eni Enrico Mattei.

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

Diamantoudi, E. and E. S. Sartzetakis (2006). Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory 8*(2), 247–263.

Dietz, S., C. Hope, N. Stern, and D. Zenghelis (2007). Reflections on the Stern Review (1): A robust case for strong action to reduce the risks of climate change. *World Economics 8*(1), 121–168.

Downs, G. W., D. M. Rocke, and P. N. Barsoom (1996). Is the good news about compliance good news about cooperation? *International Organization 50*(03), 379–406.

Ehrlich, I. and G. S. Becker (1972). Market insurance, self-insurance, and self-protection. *The Journal of Political Economy 80*(4), 623–648.

Endres, A. and C. Ohl (2003). International environmental cooperation with risk aversion. *International Journal of Sustainable Development 6*, 378–392.

Eychmans, J. and M. Finus (2003). Coalition formation in a global warming game: how the design of protocols affects the success of environmental treaty-making. CORE Discussion Papers 2003088, Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Eyckmans, J. (2001). On the farsighted stability of the Kyoto Protocol. Energy, Transport and Environment Working Papers Series ete0103, Katholieke Universiteit Leuven, Centrum voor Economische Studin, Energy, Transport and Environment.

Farrell, J. and E. Maskin (1989). Renegotiation in repeated games. *Games and Economic Behavior 1*(4), 327–360.

Finus, M. (2000). Game theory and international environmental co-operation: A survey with an application to the Kyoto-Protocol. *FEEM Working Paper* (86).

Finus, M. (2001). *Game theory and international environmental co-operation*. Cheltenham, UK: Edward Elgar.

Finus, M. and S. Maus (2008). Modesty may pay. *Journal of Public Economic Theory 10*(5), 801–826.

Finus, M. and P. Pintassilgo (2009). The role of uncertainty and learning for the success of international climate agreements. Stirling Economics Discussion Papers 2009-16, University of Stirling, Department of Economics.

Finus, M. and D. T. Rubbelke (2008). Coalition formation and the ancillary benefits of climate policy. Working Papers 2008.62, Fondazione Eni Enrico Mattei.

Fisher, A. C. and U. Narain (2003). Global warming, endogenous risk, and irreversibility. *Environmental and Resource Economics 25*, 395–416.

Fujita, T. (2004). Design of international environmental agreements and uncertainty. *Environmental Economics and Policy Studies 6*, 103–118.

Hardin, G. (1968). The tradgedy of the commons. *Science 162*(3859), 1243–1248.

Heal, G. and B. Kristrom (2002). Uncertainty and climate change. *Environmental & Resource Economics 22*(1), 3–39.

Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental & Resource Economics 2*(2), 141–159.

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

Ihori, T. and M. C. McGuire (2007). Collective risk control and group security: The unexpected consequences of differential risk aversion. *Journal of Public Economic Theory 9*(2), 231–263.

Intergovernmental Panel on Climate Change (1995). *Working Group 3, Economic and Social Dimensions of Climate Change*. Intergovernmental Panel on Climate Change. Cambridge University Press.

Intergovernmental Panel on Climate Change (2007). *Climate Change 2007: The Physical Basis. Contribution of Working Group I to the Fourth Assessment Report of IPCC*. Intergovernmental Panel on Climate Change. Cambridge University Press.

Jeppensen, T. and P. Anderson (1998). Commitment and fairness in environmental games. In N. Hanley and H. Folmer (Eds.), *Game Theory and the Environment*, pp. 65–83. Cheltenham: Edward Elgar Publishing.

Katsoulacos, Y. (1997). R&D spillover, cooperation, subsidies, and international agreements. In C. Carraro (Ed.), *International environmental negotiations: strategic policy issues*, pp. 97–109. Cheltenham: Edward Elgar Publishing.

Kolstad, C. (2005). Piercing the veil of uncertainty in transboundary pollution agreements. *Environmental & Resource Economics 31*(1), 21–34.

Kolstad, C. D. (1996). Fundamental irreversibilities in stock externalities. *Journal of Public Economics 60*(2), 221–233.

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. and A. Ulph (2008a). Learning and international environmental agreements. *Climatic Change* (89), 125–41.

Kolstad, C. D. and A. Ulph (2008b). Uncertainty, learning, and heterogeneity in self-enforcing international environmental agreements. Paper presented at a workshop sponsored by the Frisch Center, May, in Oslo, Norway.

Lange, A. and C. Vogt (2003). Cooperation in international environmental negotiations due to a preference for equity. *Journal of Public Economics 87*(9-10), 2049–2067.

Le Breton, M. and A. Soubeyran (1997). The interaction between environmental and trade policy. In C. Carraro (Ed.), *International environmental negotiations: strategic policy issues*, pp. 126–149. Cheltenham: Edward Elgar Publishing.

Manne, A. and R. Richels (1992). *Buying Greenhouse Gas Insurance: The Economic Costs of CO2.* Cambridge, MA: MIT Press.

McGinty, M. (2010). International environmental agreements as evolutionary games. *Environmental & Resource Economics 45*(2), 251–269.

Mohr, E. and J. Thomas (1998). Pooling sovereign risks: the case of environmental treaties and international debt. *Journal of Development Economics 55*(1), 153–169.

Na, S.-l. and H. S. Shin (1998). International environmental agreements under uncertainty. *Oxford Economic Papers 50*(2), 173–185.

Nordhaus, W. (1994). *Managing the Global Commons: The Economics of Climate Change.* Cambridge, MA: MIT Press.

Osmani, D. and R. Tol (2006). The case of two self-enforcing international agreements for environmental protection. MPRA Paper 4256, University Library of Munich, Germany.

Osmani, D. and R. Tol (2009). Toward farsightedly stable international environmental agreements. *Journal of Public Economic Theory 11*(3), 455–492.

Osmani, D. and R. S. Tol (2008). Evolution in time of farsightedly stable coalitions: An application of FUND. Working Papers FNU-162, Research unit Sustainability and Global Change, Hamburg University.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives 14*(3), 137–158.

Rothschild, M. and J. E. Stiglitz (1970). Increasing risk: I. A definition. *Journal of Economic Theory 2*(3), 225–243.

Rubio, S. and B. Casino (2005). Self-enforcing international environmental agreements with a stock pollutant. *Spanish Economic Review 7*(2), 89–109.

Rubio, S. J. and B. Casino (2001). International cooperation in pollution control. Working Papers Series AD 2001-21, Instituto Valenciano de Investigaciones Economicas, S.A. (Ivie).

Rubio, S. J. and A. Ulph (2006). Self-enforcing international environmental agreements revisited. *Oxford Economic Papers 58*(2), 233–263.

Rubio, S. J. and A. Ulph (2007). An infinite-horizon model of dynamic membership of international environmental agreements. *Journal of Environmental Economics and Management 54*(3), 296–310.

Schmidt, U., J. R. Robledo, and T. Lohse (2007). Self-insurance and self-protection as public goods. Economics Working Papers 2007.16, Christian-Albrechts-University of Kiel, Department of Economics.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review.* Cambridge: Cambridge University Press.

Tol, R. S., T. E. Downing, O. J. Kuik, and J. B. Smith (2004). Distributional aspects of climate change impacts. *Global Environmental Change 14*, 259–272.

Tsur, Y. and Z. Amos (1996). Accounting for global warming risks: Resource management under event uncertainty. *Journal of Economic Dynamics and Control 20*, 1289–1305.

Tulkens, H. and P. Chander (1997). The core of an economy with multilateral environmental externalities. *International Journal of Game Theory 26*(3), 379–401.

Ulph, A. (1997). Learning about global warming? In N. Hanley and H. Folmer (Eds.), *Game Theory and the Environment*, Chapter 13, pp. 255–286. Cheltenham: Edward Elgar Publishing.

Ulph, A. (2004). Stable international environmental agreements with a stock pollutant, uncertainty and learning. *Journal of Risk and Uncertainty 29*(1), 53–73.

Ulph, A. and D. Maddison (1997). Uncertainty, learning and international environmental policy coordination. *Environmental & Resource Economics 9*(4), 451–466.

Ulph, A. and D. Ulph (1994). Who gains from learning about global warming. Discussion Paper Series In Economics And Econometrics 9407, Economics Division, School of Social Sciences, University of Southampton.

Ulph, A. and D. Ulph (1997). Global warming, irreversibility and learning. *Economic Journal 107*(442), 636–50.

Wagner, U. J. (2001). The design of stable international environmental agreements: Economic theory and political economy. *Journal of Economic Surveys 15*(3), 377–411.

Walker, J. M. and R. Gardner (1992). Probabilistic destruction of common-pool resources: Experimental evidence. *The Economic Journal 102*(414), 1149–1161.

Weitzman, M. L. (2007). A review of the Stern Review on the economics of climate change. *Journal of Economic Literature 45*, 703–724.

Appendix

Proof of Lemma 2.

I begin by restating KTCs (6)-(9) here for convenience.

$$\frac{\partial \mathcal{L}_f}{\partial q_f} = b(1-q_f) - \frac{1}{N}2\theta(d_H - d_L)(nq_s + (N-n)q_f) - d_L - \lambda_f \leq 0 \tag{6}$$

$$q_f \geq 0, \quad q_f \frac{\partial \mathcal{L}_f}{\partial q_f} = 0,$$

$$\frac{\partial \mathcal{L}_f}{\partial \lambda_f} = 1 - q_f \geq 0 \tag{7}$$

$$\lambda_f \geq 0, \quad \lambda_f \frac{\partial \mathcal{L}_f}{\partial \lambda_f} = 0$$

$$\frac{\partial \mathcal{L}_s}{\partial q_s} = n\left[b(1-q_s) - \frac{n}{N}2\theta(d_H - d_L)(nq_s + (N-n)q_f) - nd_L - \lambda_s\right] \leq 0 \tag{8}$$

$$q_s \geq 0, \quad q_s \frac{\partial \mathcal{L}_s}{\partial q_s} = 0,$$

$$\frac{\partial \mathcal{L}_s}{\partial \lambda_s} = 1 - q_s \geq 0 \tag{9}$$

$$\lambda_s \geq 0, \quad \lambda_s \frac{\partial \mathcal{L}_s}{\partial \lambda_s} = 0$$

First, I show that $q_f > 0$ by contradiction. So suppose that $q_f = 0$. By KTC (7), $q_f = 0 \Rightarrow \frac{\partial \mathcal{L}_f}{\partial \lambda_f} = 1 \Rightarrow \lambda_f = 0$. Inserting $q_f = 0$ and $\lambda_f = 0$ into KTC (6) results in the following inequality:

$$q_s \geq \frac{N(b - d_L)}{2\theta(d_H - d_L)n}$$

The RHS of the above inequality is positive by Assumption 1, so $q_f = 0 \Rightarrow q_s > 0$. In turn, $q_s > 0 \Rightarrow \frac{\partial \mathcal{L}_s}{\partial \lambda_s} = 0$ by KTC (8). Inserting $q_f = 0$ into KTC (8) yields the following equality:

$$\lambda_s = b(1-q_s) - \frac{n^2}{N}2\theta(d_H - d_L)q_s - nd_L$$

By KTC (9), $\lambda_s \geq 0$, or:

$$q_s \leq \frac{N(b - nd_L)}{b + 2\theta(d_H - d_L)n^2}$$

So $q_f = 0$ implies the following two conditions on q_s:

$$\frac{N(b - d_L)}{2\theta(d_H - d_L)n} \leq q_s \leq \frac{N(b - nd_L)}{b + 2\theta(d_H - d_L)n^2}$$

If both conditions are to hold it must be that $b + 2\theta(d_H - d_L)(n-1)n \leq 0$, which can never be true. Thus, $q_f > 0$.

Next, I show that $q_f < 1$. To see this, suppose instead that $q_f = 1$. By KTC (6) $q_f = 1 \Rightarrow \frac{\partial \mathcal{L}_f}{\partial q_f} = 0$ and the following equality must hold:

$$\lambda_f = -\frac{1}{N} 2\theta(d_H - d_L)(nq_s + (N - n)) - d_L$$

The RHS of this expression is negative. However, by KTC (7), $\lambda_f \geq 0$, so $q_f = 1$ cannot form part of an equilibrium. It must be that $q_f < 1$. Summarizing to this point, $0 < q_f < 1$; the equilibrium emission level of a non-signatory is derived from an interior solution. By KTCs (6) and (7), this means that $\lambda_f = 0$ and $\frac{\partial \mathcal{L}_f}{\partial q_f} = 0$.

Now consider possible types of solutions for the emission level of a typical signatory. Can $q_s = 1$ form part of an equilibrium? By KTC (8) $q_s = 1 \Rightarrow \frac{\partial \mathcal{L}_s}{\partial q_s} = 0$, implying that the following equality must be satisfied:

$$\lambda_s = -\frac{n}{N} 2\theta(d_H - d_L)(n + (N - n)q_f) - nd_L$$

the RHS of which is negative (recall that Lemma 2 applies only for $n \in \{1, ..., N - 1\}$). However, by KTC (9), $\lambda_s \geq 0$, so $q_s = 1$ cannot form part of an equilibrium. Therefore, $q_s < 1$ must hold in equilibrium.

Two cases can be distinguished moving forward. First, consider the interior solution with $0 < q_s < 1$ and $0 < q_f < 1$. $0 < q_s < 1 \Rightarrow \lambda_s = 0$ by KTC (9) and

$\frac{\partial \mathcal{L}_f}{\partial q_f} = 0$ by KTC (8). This permits me to find $q_s(q_f, n)$ using KTC (8). Similarly, $0 < q_f < 1 \Rightarrow \lambda_f = 0$ by KTC (7) and $\frac{\partial \mathcal{L}_f}{\partial q_f} = 0$ by KTC (6), and this permits me to find $q_f(q_s, n)$ using KTC (6). Substitution of $q_s(q_f, n)$ into $q_f(q_s, n)$ gives $q_f(n)$, which can be inserted back into $q_s(q_f, n)$, yielding $q_s(n)$. I find:

$$\begin{aligned} q_s &= 1 - \frac{nN\,(2\theta d_H + (1-2\theta)d_L)}{Nb + 2\theta(d_H - d_L)(n^2 + N - n)} \\ q_f &= 1 - \frac{N\,(2\theta d_H + (1-2\theta)d_L)}{Nb + 2\theta(d_H - d_L)(n^2 + N - n)} \end{aligned}$$

It is only necessary to derive a condition ensuring that $q_s > 0$, since I have shown above that $0 < q_f < 1$ and $q_s < 1$ hold for all possible combinations of the underlying parameters and n. Notice that q_s can also be expressed as:

$$q_s = \frac{N(b - nd_L) - 2\theta(d_H - d_L)(n-1)(N-n)}{Nb + 2\theta(d_H - d_L)(n^2 + N - n)}$$

from which it is clear that $q_s > 0$ requires that the following condition is satisfied:

$$g(n, b, \chi) \equiv N(b - nd_L) - 2\theta(d_H - d_L)(n-1)(N-n) > 0$$

Otherwise, the signatory corner solution with $q_s = 0$ and $0 < q_f < 1$ occurs. If the above condition is not satisfied, the equilibrium emission level of a typical non-signatory can be found by first recognizing that $0 < q_f < 1 \Rightarrow \lambda_f = 0$ by KTC (7) and $\frac{\partial \mathcal{L}_f}{\partial q_f} = 0$ by KTC (6), and then substituting these conditions and $q_s = 0$ into KTC (6). Simplifying, I obtain:

$$q_f = \frac{N(b - d_L)}{Nb + 2\theta(d_H - d_L)(N-n)}$$

This has been a proof of Lemma 2.

Proof of Lemma 5.

For any number of signatories n, the total amount of pollution emitted at an interior solution and at the signatory corner solution, respectively, are given by:

$$Q_{q_s>0} = \frac{N^2 b}{Nb + 2\theta d_H(n^2 + N - n)}, \quad Q_{q_s=0} = \frac{Nb(N-n)}{Nb + 2\theta d_H(N-n)}$$

As indicated by Proposition 1, it may be that increasing the number of signatories causes the type of equilibrium to change. Thus, the fact that both of $Q_{q_s>0}$ and $Q_{q_s=0}$ are strictly decreasing in n for all $n \in \{1, ..., N\}$ does not prove Lemma 5.

Suppose that the number of signatories increases from $n = m$ to $n = m + 1$. Proposition 1 indicates four different scenarios. First, $Q_{q_s>0}$ could apply at both m and $m + 1$. In this case, total emissions fall, since $Q_{q_s>0}$ is strictly increasing in n. Second, $Q_{q_s=0}$ could apply at both m and $m+1$. Total emission would fall in this case too, since $Q_{q_s>0}$ is strictly increasing in n. Third, $Q_{q_s>0}$ could apply at m and $Q_{q_s=0}$ at $m+1$, where m is largest integer less than n_1 and $m+1$ is the smallest integer no greater than n_1. In this case the equilibrium changes from the interior solution to the signatory corner solution as the coalition size increases. Fourth, $Q_{q_s=0}$ could apply at m and $Q_{q_s>0}$ at $m+1$, where m is largest integer less than n_2 and $m+1$ is the smallest integer no greater than n_2. In this case the equilibrium changes from the signatory corner solution to the interior solution as the coalition size increases. In these latter two scenarios, assuming that n_1 and n_2 are integers, it is easy to show that total emissions must fall by relying on the fact that $b = nd_L + 2\theta d_H(N - n)(n - 1)/N$ for $n \in \{n_1, n_2\}$, thus completing the proof.

Proof of Lemma 6.

Equations (10)-(13) with $d_L = 0$ provide the emission levels relevant to Lemma 6. Equations (10) and (11) give $q_s(n)$ and $q_f(n)$ at an interior solution to the Emissions

Game, while equations (12) and (13) give $q_s(n)$ and $q_f(n)$ at the signatory corner solution.

To begin, I focus on each case separately, assuming that an increase in the coalition size from $n-1$ to n does not lead to a change in the type of solution. Clearly, the difference between $q_f(n-1)$ and $q_s(n)$ at the signatory corner solution is positive, since $q_s(n) = 0$ and $0 < q_f(n-1) < 1$, regardless of n and $n-1$. It is also true that the difference between $q_f(n-1)$ and $q_s(n)$ at the interior solution is positive. When an interior solution occurs at n and $n-1$, I find:

$$q_f(n-1) - q_s(n) = \frac{2\theta d_H N(n-1)\left[Nb + 2\theta d_H\left(n^2 - 3n + N\right)\right]}{\left[Nb + 2\theta d_H\left(n^2 + N - n\right)\right]\left[Nb + 2\theta d_H\left((n-1)^2 + N - n + 1\right)\right]}$$

which is strictly positive for all $n \in \{2, ..., N\}$.

Of course, as Proposition 1 indicates, there are two additional scenarios that may present. First, the interior solution could occur at $n-1$ while the signatory corner solution could occur at n, where $n-1$ is largest integer less than n_1 and n is the smallest integer no greater than n_1. In this case it must once again be that $q_s(n) - q_f(n-1) > 0$, since $q_s(n) = 0$ and $0 < q_f(n-1) < 1$. Second, the signatory corner solution could occur at $n-1$ while the interior solution could occur at n, where $n-1$ is largest integer less than n_2 and n is the smallest integer no greater than n_2. I find that the difference between $q_f(n-1)$ and $q_s(n)$ in this case is given by:

$$q_f(n-1) - q_s(n) = \frac{2\theta d_H(n-1)\left[Nb(N+1) + 2\theta d_H(N-n+1)(N-n)\right]}{\left[Nb + 2\theta d_H\left(n^2 + N - n\right)\right]\left[Nb + 2\theta d_H(N-n+1)\right]}$$

which is strictly positive for all $n \in \{2, ..., N\}$. This has been a proof of Lemma 6.

Proof of Proposition 2.

Lemma 4 gives the expected payoff functions at each type of solution, while Proposition 1 indicates, for any combination of parameter values, the type of solution that applies at each n.

I begin by analyzing the internal stability of a coalition of size n, when the interior solution occurs at both n and $n-1$. I obtain the difference between $\bar{\pi}_s(n)$ and $\bar{\pi}_f(n-1)$ in this case, which must be non-negative if the coalition is to be internally stable, as:

$$\begin{aligned}
\bar{\pi}_s(n) - \bar{\pi}_f(n-1) &= -\frac{2N^2 b\theta^2 d_H^2 (n-1)\left[N^2 b^2 (n-3) + 4Nb\theta d_H A(n) + 4\theta^2 d_H^2 B(n)\right]}{\left[Nb + 2\theta d_H (n^2 + N - n)\right]^2 \left[Nb + 2\theta d_H (n^2 + N - 3n + 2)\right]^2} \\
A(n) &\equiv (n-1)N + n^3 - 4n^2 + 3n - 2 \\
B(n) &\equiv (n+1)N^2 + 2nN(n^2 - 2n - 1) + n^2(n-1)^2(n-3)
\end{aligned}$$

where $A(n)$ and $B(n)$ are positive for $n \geq 3$. Thus, coalitions of size n for $n \in \{3, ..., N\}$ cannot be internally stable when the interior solution occurs at both n and $n-1$.

Continuing with the case in which an interior solution occurs at both n and $n-1$, a coalition of size 2 may be internally stable. Internal stability of a coalition of size 2 requires that the following expression be non-negative, so that a signatory has no incentive to leave the agreement:

$$\bar{\pi}_s(2) - \bar{\pi}_f(1) = \frac{2N^2 b\theta^2 d_H^2 \left[N^2 b^2 - 4Nb\theta d_H (N-4) - 4\theta^2 d_H^2 \left(3N^2 - 4N - 4\right)\right]}{\left[Nb + 2\theta d_H (N+2)\right]^2 \left[N(b + 2\theta d_H)\right]^2}$$

Whether this expression is non-negative depends on the term in large square brackets in the numerator. If $Nb \geq 2\theta d_H (N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$, this term as well as the larger expression will be non-negative, and, thus, a coalition of size 2 will be internally stable.

Supposing further that the interior solution occurs at $n+1$, a coalition of size $n = 2$ is also externally stable. External stability requires the following expression to be non-positive, so that a non-signatory has no incentive to join the agreement:

$$\bar{\pi}_s(3) - \bar{\pi}_f(2) = -\frac{32N^3 b\theta^3 d_H^3 \big[b(N-1) + 2\theta d_H(N+3)\big]}{\big[Nb + 2\theta d_H(N+6)\big]^2 \big[Nb + 2\theta d_H(N+2)\big]^2}$$

This expression is clearly negative, regardless the parameter values. Thus, provided the interior solution occurs at $n-1$, n, and $n+1$, the unique, myopically stable coalition size is 2 if $Nb \geq 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$. Notice that $Nb \geq 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$ implies that an interior solution occurs at $n = 1$, $n = 2$, and $n = 3$ (since it implies $g(n, b, \chi) > 0$ for $n \in \{1, 2, 3\}$ given $d_L = 0$). Therefore, a coalition of size 2 will be myopically stable whenever this condition holds.

Through simulation, I find that a coalition of size 2 is the *unique* myopically stable IEA when $Nb \geq 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$. Defining γ as b/d_H, and $\gamma(\theta)$ as the ratio of b to d_H such that $Nb = 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$ is satisfied, Table 2 gives the myopically stable coalition size for various values of γ and θ. It is assumed that $N = 100$. Table 2 demonstrates that, regardless of θ, for $\gamma \geq \gamma(\theta)$ (i.e. $Nb \geq 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$ is satisfied), the unique myopically stable coalition consists of two countries. This result does not depend on the number of countries sharing the environmental resource, N.

For $N = 100$, Table 2 also shows that, for all $\gamma < \gamma(\theta)$ (i.e. $Nb < 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$ is satisfied), the unique myopically stable coalition size is 1. The same holds true regardless of N. The rationale behind this result is that a coalition of size 1 is always internally stable, but is externally stable only when a coalition of size 2 is *not* internally stable. A coalition of size 1 is always internally stable since it consists of a single country cooperating fully with itself in the presence of $N-1$ non-signatories acting as singletons. If the signatory country were to exit from the coalition, so that $n = 0$, it would still be acting on its own, still in the presence of $N-1$ non-signatories acting as singletons. Thus, the signatory has no

Table 2: Stable coalition size given $d_L = 0$ and $N = 100$

	θ				
$\gamma = b/d_H$	0.1	0.25	0.5	0.75	1
$\gamma < \gamma(0.1)$	1	1	1	1	1
$\gamma = \gamma(0.1)$	2	1	1	1	1
$\gamma(0.1) < \gamma < \gamma(0.25)$	2	1	1	1	1
$\gamma = \gamma(0.25)$	2	2	1	1	1
$\gamma(0.25) < \gamma < \gamma(0.5)$	2	2	1	1	1
$\gamma = \gamma(0.5)$	2	2	2	1	1
$\gamma(0.5) < \gamma < \gamma(0.75)$	2	2	2	1	1
$\gamma = \gamma(0.75)$	2	2	2	2	1
$\gamma(0.75) < \gamma < \gamma(1)$	2	2	2	2	1
$\gamma = \gamma(1)$	2	2	2	2	2
$\gamma > \gamma(1)$	2	2	2	2	2

Table shows the coalition size that is both internally and externally myopically stable for various values of $\gamma = b/d_H$ and θ, assuming $N = 100$. In the table, $\gamma(.1) = 0.586$, $\gamma(.25) = 1.47$, $\gamma(.5) = 2.93$, $\gamma(.75) = 4.40$, and $\gamma(1) = 5.86$.

incentive to leave and the coalition is internally stable. When $Nb < 2\theta d_H(N-4) + 4\theta d_H \left(N^2 - 3N + 3\right)^{1/2}$ it is also externally stable, since in this case a coalition of size 2 is not internally stable. This has been a proof of Proposition 2.