Uncertainty in Self-Enforcing International Environmental Agreements

by

Charles D. Kolstad*

Prepared for

CLIMNEG Workshop on International Environmental Agreements
Center for Economic Studies, Catholic University
Leuven, Belgium
March 14, 2003

DRAFT
March 4, 2003

ABSTRACT

This paper addresses the subject of self-enforcing international environmental agreements (IEA’s). These are voluntary agreements to control transboundary pollution among cooperating countries. The standard model of IEA’s is adapted to include uncertainty in environmental costs and benefits, as well as learning about these costs and benefits. The paper investigates the extent to which the size of the coalition changes as a result of learning and uncertainty.

* 3M Visiting Professor of Environmental Economics, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts; Donald Bren Professor of Environmental Economics and Policy, Bren School of Environmental Science and Management and Department of Economics, University of California, Santa Barbara; and University Fellow, Resources for the Future, Washington, District of Columbia. Email: ckolstad@bren.ucsb.edu. Comments from Rohini Somanathan, E. Somanathan and from seminar participants at the Delhi School of Economics, the Indian Institute of Management (Calcutta), the Indian Statistical Institute (Delhi), and the Madras School of Economics are appreciated.
I. INTRODUCTION

Environmental problems are plagued by uncertainty – uncertainty regarding physical and biological processes as well as uncertainty regarding societal costs and benefits. Certainly the public policy debate over climate change has focused on how certain we are about the nature, costs and benefits of the environmental problem. Some policymakers have used uncertainty to justify taking action now, before it is too late, and others have used uncertainty to justify delaying action. The economic literature suggests that uncertainty in and of itself should have little effect on action, provided decision-makers are risk neutral.¹

At the international level, the problem is compounded by the fact that countries are the decision-makers and they must voluntarily agree to solve the environmental problem: the problem is solved cooperatively through an international environmental agreement (IEA), coordinating pollution control. Although uncertainty (and learning) may influence the decision of individual countries, there is an issue as to how uncertainty affects the agreement process, if at all. For instance, if in fact the costs and benefits do not fall equally on all countries, it may be easier to forge an agreement before uncertainty is resolved and the identity of the winners and losers is revealed.

¹ The story is different if learning is taking place over time and action can have irreversible consequences. In this case, the environmental protection decision by individual decision makers can be biased upwards or downwards, depending on the nature of the learning process and the irreversibilities. See Kolstad (1996), among others.
More specifically, consider a set of countries contributing different quantities of pollution which in the aggregate generates a global externality. This externality affects all of the countries, though to different degrees and possibly even some positively and some negatively. Think of greenhouse gas emissions leading to global warming. The “solution” of the environmental problem is for each country to reduce emissions and for the cost of doing so to be shared in some equitable fashion. How much each country reduces emissions depends on their costs of emission control; how much each country contributes towards the cost of emission control depends on each country’s benefits from controlling the problem as well as other issues of equity. An agreement to solve the problem would contain both an actions component, prescribing the emissions control undertaken by each country, as well as a sharing component, stipulated how costs are to be shared.

In this paper we introduce uncertainty and learning into a standard model of self-enforcing international environmental agreements. We posit uncertainty regarding benefits and costs, uncertainty which may be resolved between the point at which a country commits to an international agreement and the point at which the agreeing countries decide on emission levels.

Our conclusions are somewhat ambiguous. We find that indeed uncertainty and learning can change the size of an international environmental agreement. The basic idea is that learning allows participants to condition their actions within the
coalition, thus increasing the efficiency of the coalition, decreasing its cost, and thus decreasing the incentives to defect from the coalition.²

The next section of the paper reviews literature on IEA’s and on learning and uncertainty. The subsequent section of the paper presents a standard model of self-enforcing international environmental agreements, into which uncertainty and learning is introduced. The paper closes with conclusions.

II. BACKGROUND

The literature on international environmental agreements (IEA’s) has grown over the past decade.³ Most of the literature focuses on self-enforcing agreements; ie, agreements which are structured so that they are effective and stable without recourse to a larger context of international law and interaction. Some of the earliest work (Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994) finds that such agreements are either unlikely to consist of very many participants or, the converse, if the agreements involve a large number of countries, then the gains from cooperation must be low.⁴ The basic idea is that the incentives for free-riding must be low or else most countries will choose to free-ride and not belong to the agreement. Other authors have focused less on free-riding incentives than on

---

² The “timing” of regulatory action in the presence of learning has been addressed by a number of authors, particularly in the context of climate change (Kolstad, 1994, 1996; Ulph and Ulph, 1998; Kelly and Kolstad, 1999). A key variable is the extent to which there are irreversibilities, either environmental or in terms of pollution control capital investments. These irreversibilities operate on the timing question in opposite directions with respect to the base case of timing control based purely on costs and benefits without any learning: environmental irreversibilities call for the acceleration of control whereas abatement capital irreversibilities imply the optimality of the delay of control. Several authors have analyzed this systematically in the context of climate change

³ Barrett (2003) provides a recent and comprehensive review of this literature. See also Finus (2001).

⁴ Many of the results in the literature rely on simulation models and are thus are less proofs than illustrations. Rubio and Ulph (2003) provide analytic proof of some of these early results.
burden-sharing. Chander and Tulkens (1992, 1994) provide a way of sharing the burden such that the agreement consists of all countries. Commitment is achieved through a trigger strategy by assuming that the coalition will fall apart should any single country defect. Most of the results in this literature rely on homogeneity of participating countries. Barrett (2001) shows that heterogeneity of countries can reduce the free-riding problem and thus help support larger coalitions. In much the same way as heterogeneity can support public goods provision generally (eg, Bliss and Nalebuff, 1984), heterogeneity facilitates commitment. And commitment is the big problem in self-enforcing agreements.

The results reviewed in the previous paragraph focus on forming agreements in an environment of certainty. The literature on uncertainty in the context of international agreements is more sparse and spans economics and political science. The political science literature tends to be somewhat general, though deep, whereas the economics literature tends to build on idiosyncratic models of the bargaining process.

Young (1994) adopts the concept of the “veil of uncertainty” from Brennan and Buchanan (1985), who develop it for analyzing the emergence of constitutional rules in a society. Young (1994) suggests that uncertainty can be “good,” serving to facilitate agreement on the core of international environmental agreements. To quote Brennan and Buchanan (1985, p30): “The uncertainty introduced in any choice among rules or institutions serves the salutary function of making potential agreement more rather than less likely....[An individual] will tend to agree on
arrangements that might be called ‘fair’ in the sense that patterns of outcomes generated under such arrangements will be broadly acceptable, regardless of where the participant might be located in such outcomes.” Though the authors are persuasive, neither offers an analytic version of their arguments.

One of the fundamental problems in quantitatively examining these hypotheses is that “difficulty to agree” is not an easy concept to quantify. Game theory generally focuses on equilibria, not the difficulty in attaining an equilibrium.

Several other authors contribute specific insights to this debate. Cooper (1989) is interested in understanding international macroeconomic cooperation and to do this, analyzes a quite different international forum: public health agreements. In analyzing a century of such agreements, he comes up with the opposite conclusion: “So long as costs are positive and benefits uncertain, countries are unlikely to cooperate systematically” (p. 181). He argues that with diverse views on the link between actions and ultimate outcomes, countries are unlikely to cooperate. Only when that uncertainty is reduced will cooperation occur.

Iida (1993) takes a game theoretic approach to international agreements, providing a nice review of how asymmetric information has entered into this literature, though primarily in the context of macroeconomic agreements. He distinguishes between strategic uncertainty, which is uncertainty about the types of opponents, and analytic uncertainty, which amounts to uncertainty about your own payoffs (as well as the payoffs of others). Depending how one defines “type,” this may or may not involve uncertainty about one’s own type. Although this type of
uncertainty (analytic) is the focus of his paper, Iida points out that strategic uncertainty has dominated the literature on international agreement. Iida’s interpretation of analytic uncertainty (sometimes termed *model* uncertainty) is that there are underlying characteristics of the international economic system which are unknown to all agents – these characteristics will be revealed *ex post* and will determine payoffs. Iida argues, though the use of a simple example, that analytic uncertainty will tend to retard international cooperation. Other literature is mixed (e.g., Frankel and Rockett, 1988) as to whether analytic uncertainty tends to facilitate or retard international macroeconomic agreement.

Helm (1998) comes closest to analyzing the problem of this paper. He considers the case of an international agreement on acid rain, though much of the paper is independent of the application. He repeats many of the arguments above regarding the veil of uncertainty and goes on to construct a simple two-country model of cooperation and non-cooperation. In his example, he confirms Young (1994)’s hypothesis that uncertainty is favorable for cooperation. He then goes on to investigate a repeated game and shows that generally a trigger strategy can support a cooperative equilibrium. What is not clear from the analysis is how the veil of uncertainty *facilitates* cooperation.

Na and Shin (1998) compare cooperation from both an *ex ante* (before uncertainty is resolved) and *ex post* (after uncertainty resolved) perspective. Their model is quite specific, though they do conclude that countries are unequivocally better off with *ex ante* negotiations. This is not quite the same as saying *ex ante*
negotiations are easier. Further, the result depends on their very specific assumptions about cooperation. In their model, countries have costs of abatement and benefits from collective abatement. Cooperation is defined as non-cooperative bargaining among stable coalitions. They show that \textit{ex ante}, when all countries view themselves as identical in expectation, the grand coalition (in a three country example) is stable and supports a true joint benefits maximum equilibrium. \textit{Ex post}, after uncertainty has been resolved, the grand coalition is no longer stable since one or more countries may have an incentive to defect. Since bargaining is non-cooperative among coalitions, the joint payoff is bound to be lower. Thus their result.

Recent work by Ulph (2002) has addressed the challenging complication of stock pollutants – pollutants which accumulate. This leads to an explicitly dynamic framework (though only two periods), involving repeated interaction among the countries (see also Rubio and Ulph, 2001). His results focus on the extent to which learning changes the number of signatories and the overall level of welfare. Results vary, depending on the magnitude of costs and benefits as well as the extent to which membership in an agreement can be recontracted between periods.

Ulph and Maddison (1997), following up on earlier work by Ulph and Ulph (1996), focus on the role of learning on aggregate utility. This is an extension of work done by a variety of authors on the effect of learning on current period emission control when there is a single decision-maker (e.g., Kolstad, 1996ab). Naturally, the cooperative equilibrium is analogous to the single decision maker
and consequently they find information is always valuable. In examining non-
cooperative Nash equilibria, they find more ambiguity – information may have
negative value. These results are interesting and important but somewhat
tangential to the problem being investigated here.

III. A MODEL OF AGREEMENTS

In examining self-enforcing environmental agreements, there are several
types of comparisons that can be made in the context of learning and uncertainty:
an IEA with no uncertainty, an IEA with uncertainty but no learning, and an IEA
with uncertainty and learning (ie, the uncertainty is resolved over time). To make
these comparisons, we first present a simple, static, and standard, model of a self-
enforcing environmental agreement. We then turn to introducing uncertainty and
learning into the model.

A. A Self-Enforcing IEA

We will adapt the classic model of a self-enforcing international
environmental agreement to include uncertainty and learning. Consider \( i=1, \ldots, N \)
homogeneous countries, each emitting pollution \( (q_i) \) which contributes to the global
commons \( (Q=\sum q_i) \). For simplicity, assume each country makes a discrete choice
regarding pollution: to abate \( (q_i=0) \) or to pollute \( (q_i=1) \). Each identical country’s
payoff is represented as a linear function of own emissions and aggregate emissions:

\[ \text{payoff} = a q_i + b (Q - q_i) \]

---

5 The basic model which we adapt here is used by a number of authors, though often with emissions as a continuous
variable (see Rubio and Ulph, 2003). The model here has been adapted from Ulph (2002).
\[
\Pi_i(q_i, Q_{-i}) = c q_i - b Q \\
= c q_i - b(q_i + Q_{-i}) \\
= q_i - \gamma (q_i + Q_{-i})
\] (1)

where \(Q_{-i}=\sum_{j \neq i} q_j\) and \(\gamma = b/c\) and, without loss of generality, let \(c \equiv 1\). Thus \(\gamma\) is simply the benefit-cost ratio – the ratio of the marginal environmental damage to the marginal cost of emissions control. Clearly we wish to focus on the case of \(\gamma < 1\); otherwise abatement is a dominant strategy for individual countries and cooperation is unnecessary.

We represent the formation of an IEA as a two-stage game, consisting first of a membership game followed by an emissions game. In the membership game, countries choose whether or not to join the IEA. In the emissions game, countries decide how much to emit. In the emissions game, we assume the members of the IEA decide on emissions jointly and the non-members (the fringe) decide individually: the coalition acts as a singleton and each member of the fringe act in a Nash noncooperative manner. A Nash equilibrium results. Membership of the coalition cannot change in the emissions game.

Since all countries are identical, the primary question we ask is how big will the IEA be: i.e., how many countries? Drawing on the cartel stability literature (eg, Donsimoni et al, 1986), we define two conditions for stability in the membership game, internal stability (no country has an incentive to leave the IEA) and external stability (no country has an incentive to join the IEA). These stability conditions are defined in terms of the payoff that members of the coalition and the fringe can expect, as viewed from the membership game. Let those payoffs be, respectively, \(\Pi^c(n)\) and \(\Pi^f(n)\), where \(n\) is the number of countries in the coalition/IEA. We can then define the stability conditions:
**Defn:** A coalition of size n is **internally stable** if $\Pi^c(n) \geq \Pi^f(n-1)$.

**Defn:** A coalition of size n is **externally stable** if $\Pi^f(n) \geq \Pi^c(n+1)$.

**Defn:** A coalition of size n is **stable** if it is externally and internally stable.

We will let $n^*$ denote the size of a stable coalition; of course, $n^*$ need not exist nor need it be unique.6

To solve for $n^*$, we must work backwards from the emissions game, determining payoffs as a function of n; then in the membership game, find $n^*$ which satisfies the stability conditions.7

It is useful to introduce the function $I(x)$:

**Defn:** Define $I(x)$ as the smallest integer greater than or equal to x.

The following is a well-known result:

**Proposition 1:** For the dichotomous choice homogeneous countries self-enforcing IEA with payoffs as in Eqn. (1), there is a unique stable number of countries in the IEA, $n^*$, equal to $I(1/\gamma)$.

**Proof:** In the emissions game with n members of the IEA, the dominant strategy of each member of the fringe is to pollute. The payoff to individual members of the coalition, $\Pi^c(n,\gamma)$, is given by

$$\Pi^c(n,\gamma) = \begin{cases} -\gamma(N-n) & \text{if coalition members abate} \\ \end{cases}$$

6 Hoel (1992) and others show that for a model with a continuous choice of emissions levels and with the second stage game Cournot, then at $n^* \leq 3$; with the coalition acting as a Stackelberg leader with respect to the fringe, $n^*$ can take on any value up to N. For the case of the dichotomous choice emissions levels as considered here, $n^*$ is not restricted.

7 The stability conditions characterize the incentives for defection of either the fringe or the coalition. However, defection itself is non-Nash behavior as defined since, as the definition of stability is stated, should defection occur both the new fringe and new coalition change behavior.
\[
1 - \gamma N \quad \text{if coalition members pollute} \quad (2)
\]

which implies that members of the coalition will abate if 
\[-\gamma (N-n) \geq 1 - \gamma N \iff 1 - \gamma n \leq 0 \iff n \geq 1/\gamma. \]
Thus the payoff to the coalition is given by

\[
\Pi^c(n,\gamma) = \max[0, 1 - \gamma n] - \gamma (N-n) \quad (3)
\]

where the first argument of the max function corresponds to coalition members abating; if the second term applies, members are polluting.

Turning to the membership game, it is straightforward to show that internal and external stability hold at 
\[n = I(1/\gamma) \]
and for no other \(n\).

The graphical interpretation of this proposition is straightforward. In Figure 1, the payoffs to members of the coalition (Eqn. 2) and the fringe are plotted as a function of \(n\). It is easy to see that for any \(n\) to the right of \(1/\gamma\), coalition members can do better by defecting to the fringe, with the exception of \(n = I(1/\gamma)\). At that point, defection to the fringe brings everyone into the “everyone pollutes” section of the figure, where payoffs are lowest. Similarly, any \(n\) to the left of \(1/\gamma\) is not internally stable. The discrete nature of the problem (an integer number of countries) gives us the result. If \(n\) were continuous, then stability would be elusive.

It is also easy to see that for large \(\gamma\), the coalition will be small and for small \(\gamma\), the reverse: the coalition will be large. Since \(\gamma\) is the ratio of environmental damage to abatement costs, this means that if damage is low, a coalition is likely to form. On the other hand if damage is significant, the coalition is likely to be small. This is a depressing result.

**B. Correlated Uncertainty**
We now introduce uncertainty about $\gamma$ into the model. Let there be two states of the world, H and L, which occur with probabilities $\pi$ and $1-\pi$, respectively. Let $\gamma$ take on a different value in each of these states of the world, with $\gamma_H > \gamma_L$:

$$\gamma = \begin{cases} 
\gamma_H, & \text{with probability } \pi \\
\gamma_L, & \text{with probability } 1-\pi 
\end{cases}$$  \quad (4)

In other words, the countries are uncertain about the benefit-cost ratio but that uncertainty is shared: when uncertainty is resolved, all countries will realize the same $\gamma$ – uncertainty is correlated. We consider three cases, one where actions can be fully conditioned on the state-of-the-world (no uncertainty), one where actions cannot be conditioned on states-of-the-world and uncertainty is never resolved (no learning), and one where uncertainty is resolved between the membership and emissions games (learning). We are interested in the difference in $n^*$ among these three cases: $n_{NU}^*$, $n_{NL}^*$, and $n_L^*$.

The case of full conditioning of actions on the state-of-the-world results in $n^*$ equal to $I(1/\gamma_L)$ and $I(1/\gamma_H)$, depending on whether the state-of-the-world is L or H, respectively. The expected number of members of the coalition is simply

$$n_{NU}^* = (1-\pi) \ I(1/\gamma_L) + \pi \ I(1/\gamma_H)$$  \quad (5)

The no-learning case is identical to the case considered in the previous section, except that the expected value of $\gamma$ is used. Consequently the number of members of the coalition in the “No Learning” case ($n_{NL}^*$) is as before:

$$n_{NL}^* = I(1/\Gamma)$$  \quad (6)
where $\Gamma \equiv \mathbb{E}(\gamma)$ and $\mathbb{E}$ is the expectation operator.

**Proposition 2.** If $n^{*}_{NL}$ and $n^{*}_{NU}$ are defined as in Eqn. 5 and 6, then $n^{*}_{NL} \leq n^{*}_{NU}$.

**Proof:** Follows from the convexity of the $1/x$ function. $\blacksquare$

The learning case must be solved by backwards induction. The emissions game is not one of uncertainty but rather a situation where the action of the coalition can be conditioned on the realized state-of-the-world, the realization of $\gamma$. Thus the payoff to the coalition is as in Eqn. 2 except that $\gamma$ takes on one of two values, $\gamma_L$ or $\gamma_H$. The action of the coalition depends on which of three regions $n$ is in. If $n < I(1/\gamma_H)$, then the coalition members pollute in both states-of-the-world. If $I(1/\gamma_H) < n < I(1/\gamma_L)$, then coalition members abate in state-of-the-world $H$ and pollute in state-of-the-world $L$. Finally, if $n > I(1/\gamma_L)$, then the coalition members abate in both states-of-the-world.

There are two possibilities we need to consider. One is when the $\gamma_L$ and $\gamma_H$ are so close together that there does not exist an integer which can be inserted between $I(1/\gamma_H)$ and $I(1/\gamma_L)$. In this case, trivially there is no difference between $n^{*}_L$ and $n^{*}_{NL}$, the equilibrium number of coalition members in the “Learning” and “No Learning” cases, respectively. We will assume away this case.

The basic difference between the learning and no-learning cases is most pronounced in the region for $n$ between $I(1/\gamma_H)$ and $I(1/\gamma_L)$. Here coalition members have more flexibility under learning than they had in the no-learning case. Under one state-of-the-world (L), no abatement need be undertaken. Thus their profits are higher, holding $n$ constant. One might think this would provide an incentive to increase the number of members of the coalition. Table I shows
expected profits in the Membership game for each member of the fringe and of the coalition, as a function of the number of members of the coalition, n.

Using the information in Table I, Figure 2 shows expected payoffs to the fringe and coalition members as a function of n, from the perspective of the Membership Game. Note that there are three numbered regions for n, corresponding to three possible outcomes in the emissions game. In region 1, it will be optimal for the coalition to pollute, no matter what state-of-the-world is realized. In region 3, it will be optimal to always abate. In region 2, it will be optimal to abate if the realized state-of-the-world is H, otherwise to pollute.

It is easy to show that there are only two possible equilibrium sizes of the coalition. No n in region 1 satisfies external stability: the fringe has an incentive to move into the coalition in region 2 and do better. Similarly, in region 2, internal stability fails at all points but \( I(1/\gamma_H) \): making the coalition one smaller reduces fringe and coalition payoffs by \( \pi\gamma_H \), which is less than the extra payoff the fringe enjoys, \( \pi \). Thus there is an incentive to defect to the fringe. At \( I(1/\gamma_H) \), defection from the coalition moves the defector into region 1, where the fringe and the coalition have the same payoff which is clearly lower. Thus, internal stability holds at \( I(1/\gamma_H) \), marked with a * in Figure 2. In region 3, the same logic applies: only at \( I(1/\gamma_L) \) is there a possibility that internal stability will hold and thus that a coalition member will not have an incentive to move to region 2. \( I(1/\gamma_L) \) is marked with a + in Figure 2 and it is easy to see that a country’s payoff may increase by defecting from the coalition to the fringe. The payoff is on the solid line at + when in the coalition, moving to the dashed line at the far right of region 2, should the country defect.
Which of these two possible equilibria will prevail? The smaller number of countries is always an equilibrium. The question is whether a second, higher number of countries may also be supported.

**Proposition 3:** Provided \( \pi > 0 \), then \( n^* = I(1/\gamma_h) \) is a stable coalition.

**Proof:** That internal stability holds follows from the logic of Prop. 1. From Table I, we see that external stability also holds. Any member of the fringe wishing to join the coalition will lose \( \pi \) in payoff and gain \( \pi \gamma_h < \pi \) from joining the coalition. ■

The first proposition gives sufficient conditions for there being only one equilibrium number of coalition members:

**Proposition 4.** If \( \pi > \Gamma \) then internal stability fails at \( n= I(1/\gamma_l) \), resulting in the equilibrium number of countries in the coalition of \( n^* = I(1/\gamma_h) \).

**Proof:** By the arguments in the proof of Proposition 1, \( I(1/\gamma_h) \) meets the conditions for both internal and external stability. Also by similar arguments, all other \( n \) fail stability tests with the possible exception of \( n' = I(1/\gamma_l) \). Internal stability will fail at this point if \( \Delta P \equiv \Pi'(n') - \Pi'(n' - 1) < 0 \); ie, there is an incentive to leave the coalition for the fringe at \( n' \) if \( \Delta P < 0 \). From Table I, we see that

\[
\Delta P = (1-\pi) n \gamma_l + \pi \gamma_h - 1 \tag{7}
\]

From the definition of the \( I(\cdot) \) function, we know that
\[ \frac{1}{\gamma_L} \leq n' < \frac{1}{\gamma_L} + 1 \]  (8)

which can be rearranged into

\[ \pi(\gamma_H - 1) \leq \Delta P < \Gamma - \pi \]  (9)

Thus if \( \pi > \Gamma \), \( \Delta P \) is unequivocally negative and internal stability fails at \( n' \).

The intuition behind this is easy to see from Figure 2. Suppose the coalition is at the point marked with a + in the Figure. If the coalition loses a member, payoffs for the coalition drop by as much as \( \Gamma \). In region 2, the fringe reaps an extra payoff of \( \pi \). So if \( \pi > \Gamma \), it is attractive to defect from the coalition.

The obvious next question is what conditions will assure us that \( n^* = I(1/\gamma_L) \) is an equilibrium? Unfortunately, there are no such general conditions, primarily because of the discrete nature of \( n \). Observe from Figure 2 that if the point \( I(1/\gamma_L) \), marked with a + in the Figure, is just to the right of \( 1/\gamma_L \), then even very small \( \pi \)’s will be large enough to make defection to the fringe attractive. It is the nature of the “integerization” of \( 1/\gamma_L \) that \( n \) may be very close to \( 1/\gamma_L \) or it may be nearly \( 1+1/\gamma_L \). This is quite clearly a somewhat synthetic result, since one would not expect the real world to be as sensitive to how close to an integer \( 1/\gamma_L \) is.

**Proposition 5.** If \( \pi < \Gamma \), then \( n^* = I(1/\gamma_L) \) is a stable coalition, provided \( I(1/\gamma_L) - 1/\gamma_L < 1 \) is sufficiently larger than zero. Further there always exists a small perturbation of \( \gamma_L, \gamma_L^{**} = \gamma_L - \varepsilon \), with \[ \left| \frac{1}{\gamma_L} - \frac{1}{\gamma_L^{**}} \right| < 1 \], such that \( n^{**} = I(1/\gamma_L^{**}) \) is a stable coalition, provided \( \pi < \gamma^{**} \).
Proof: The greatest value of \( I(1/\gamma_L) - 1/\gamma_L \) is just shy of 1; i.e., \( I(1/\gamma_L) \approx 1/\gamma_L + 1 \). In this case,

\[
\Delta P \approx \gamma[N-(1/\gamma_L + 1)] - [1 - \Gamma N + \pi \gamma_H/\gamma_L] = \Gamma - \pi > 0
\]  

(10)

This proves both parts of the proposition. ■

The interpretation of these results hinges on the interpretation of the relative magnitude of \( \pi \) and \( \Gamma \). The most natural interpretation of \( \pi \) is the advantage of being in the fringe in region 2 of Figure 2 – the most natural region to consider because sometimes you abate, sometimes not, depending on the state-of-the-world. The more likely the H state is, then the more likely it will be that abatement will occur in the emissions game, and thus the greater the advantage of free-ridding in the fringe.

The variable \( \Gamma \) on the other hand is the magnitude of expected environmental benefits from abatement and also the slope of the payoff function in region 3 of Figure 2. There is a loss in environmental benefits associated with leaving the coalition due to the fact that one less country is abating. That is \( \Gamma \). So we compare the loss in environmental benefits from a smaller coalition with the advantages of not having to pay to abate. Whichever is larger tends to drive the decision. When \( \pi > \Gamma \), then the cost saving advantage of being in the fringe is larger than the environmental damage benefit of being in the coalition; thus Prop. 4 applies, and uncertainty tends to dilute the coalition-building potential. The equilibrium size of the coalition is smaller than under certainty. On the other hand, if the advantages of being in the fringe are modest, then Prop. 5 applies and uncertainty will tend to allow grow the size of the coalition to grow.

Another way of interpreting these results is to start with no uncertainty over \( \gamma \) and slowly introduce uncertainty. Start with \( \pi = 0 \), which implies that \( \Gamma = \gamma_L \). A coalition of size \( I(1/\gamma_L) \) will
be stable. Now start slowly increasing $\pi$ (the incentive for defection), which has the effect of slowly increasing $\Gamma$ (the incentive for cooperation) towards $\gamma_H$. However, $\pi$ increases more rapidly than $\Gamma$. While $\pi$ remains small compared to $\Gamma$, the size of the coalition will not change. And in fact it will be larger than the case of no learning, in which case $n^*=I(1/\Gamma')$. The implication is that learning results in a larger coalition. However, as $\pi$ becomes larger, then the likelihood of the $H$ state-of-the-world becomes more difficult to ignore. Eventually the coalition at $I(1/\gamma_L)$ is no longer stable and the number of coalition members drops.

It is important to point out however, that $n^*=I(1/\gamma_H)$ is always an equilibrium, unless $\pi=0$. It is just that for sufficiently small $\pi$ relative to $\Gamma$, it is possible to support a second equilibrium number of coalition members. So another way of stating the result is that under learning, the number of coalition members may always be smaller than the size of the coalition without learning.

C. Uncorrelated Uncertainty.

We now consider the case in which individual countries may end up in different states-of-the-world: some countries may be $H$ and some may be $L$. In the membership game, all countries are identical; in the emissions game, some countries are $H$ and some are $L$, differing in their benefit-cost ratios, $\gamma$. \textit{Ex ante}, no country knows whether it will be $H$ or $L$.

Assume $n$ countries have decided, in the membership game, to constitute a coalition. After that decision is completed, the state-of-the-world for each country is revealed. Let $\phi$ be the proportion of countries in the coalition that end up being in state $H$. Clearly the proportion $1-\phi$ are in state $L$. The expectation, from the point of view of the membership game, is that $\phi = \pi$. 
To solve this problem, we work backwards from the emissions game, determining the optimal action by the coalition in the emissions game as a function of \( \varphi \) and the resulting payoff for the members of the coalition. We then move to the membership game and ask what is the expected payoff to the coalition and fringe, as a function of \( n \). We can then determine \( n^* \).

Consider first the extremes outcomes of \( \varphi = 0 \) and \( \varphi = 1 \). Using Eqn. 3, we can plot the payoff in the emissions game, as a function of \( n \). These two lines are shown in Fig. 3 (solid lines). Also shown in the Figure (dashed line) is the expected payoff, as a function of \( n \). This line is exactly the same as in Figure 2, in the case developed in the previous section.

The implication is that the outcomes in the case of uncorrelated uncertainty are exactly the same as correlated uncertainty. What is important to us is actions taken in the membership game. And in the membership game, the view of the emissions game is purely expectations. Expectations of the emissions game are the same for an uncertain state-of-the-world shared by all countries vs. an uncertain state of the world for each individual country. In expectation, the effect is the same. Thus the choice of the number of countries in the coalition will be the same in the two cases.

V. CONCLUSIONS

In this paper, we have taken the standard model of self-enforcing international environmental agreements and introduced uncertainty about costs and benefits as well as a very specific type of learning. Learning is of a very specific type and occurs between commitment to an agreement and the actual decision to emit. Clearly there are other ways of representing learning; thus this paper only scratches the surface of the topic.
We find that uncertainty (without learning) tends to increase the size of the cooperating coalition in an international environmental agreement. Learning is introduced by positing that countries commit to belong to a coalition for pollution control in a state of uncertainty but then the coalition decides on how much emissions control to undertake after learning. In comparing the two extremes of no learning and complete learning, we find that learning typically decreases the size of the pollution control coalition. This in turn reduces the potential welfare gains from the agreement. Whether this makes agreement easier or not is not known.
REFERENCES


\[ n < I(1/\gamma_H) \quad \text{if} \quad I(1/\gamma_H) \leq n < I(1/\gamma_L) \quad \text{if} \quad n \geq I(1/\gamma_L) \]

\[ \Pi^i(n) = 1 - \Gamma N \quad 1 - \Gamma N + \pi(\gamma_H n - 1) - \Gamma (N - n) \]

\[ \Pi^f(n) = 1 - \Gamma N \quad 1 - \Gamma N + \pi \gamma_H n \quad 1 - \Gamma (N - n) \]

Table I: Expected payoffs with Correlated Uncertainty, Membership Game
Figure 1: Payoffs in Emissions game

\[
\begin{align*}
\text{Payoff} & \quad \Pi^f & \Pi^c \\
1 - \gamma N & \quad \text{pollute} \quad \text{abate} \rightarrow \\
1/\gamma & \quad n
\end{align*}
\]
Figure 2: Payoff in Emissions Game, Correlated Uncertainty

Uncertainty in SE-IEA’s  

26  

4 March 2003
Figure 3: Payoffs for different fractions of coalition members in H state

\[
\text{Payoff} \\
\Pi(\varphi=1) \\
\Pi(\varphi=0) \\
E[\Pi(\varphi)]
\]

\[
\begin{align*}
1 - \gamma_L N \\
1 - E(\varphi) N \\
1 - \gamma_H N
\end{align*}
\]

\[
\begin{align*}
1/\gamma_H \\
1/\gamma_L
\end{align*}
\]