# Systematic Uncertainty in Self-Enforcing International Environmental Agreements[+]

by

Charles D. Kolstad[*]
UCSB & RfF

DRAFT
June 17, 2004

## ABSTRACT

This paper addresses the subject of self-enforcing international agreements. The emphasis is on international environmental agreements (IEA's), though the results are more general. 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 systematic uncertainty (also known as model uncertainty). Results are that systematic uncertainty by itself decreases the size of an IEA. Learning has the further effect of either increasing or decreasing the size of an IEA, depending on parameters of the problem.

---

[*] 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. This work was conducted, for the most part, while the author was the 3M Visiting Professor of Environmental Economics in the Department of Economics at the Massachusetts Institute of Technology in Cambridge, Massachusetts.

+This paper has benefited from comment from Brian Copeland, Ita Falk, Michael Finus, Don Fullerton, Andy King, Eric Orts, Rohini Somanathan, Eswaran Somanathan, Samudra Vijay, David Zilberman and from seminar participants at Dartmouth College, Harvard University, North Carolina State University, Yale University, the Delhi School of Economics, the Indian Institute of Management (Calcutta), the Indian Statistical Institute (Delhi), the Madras School of Economics, and the Universities of Central Florida, Connecticut, Wyoming, and Leuven (Belgium) and. I am particularly grateful to Alistair Ulph of the University of Manchester for many discussions early on and throughout the development of this paper. In fact, the basic model presented here is a simplification of a dynamic model in Ulph (2002). Indeed, some of the results presented here follow from Ulph (2002).

## I. INTRODUCTION

Since the re-emergence of concern for environmental protection some thirty-plus years ago, remarkable gains have been made in protecting the world's environment. The other side of that coin is that having "solved" many of the easy problems, we are left with those problems that are particularly difficult to address. Some of the toughest remaining environmental problems are those that are global problems, requiring international cooperation to solve; i.e., international treaties[1] are necessary. As we have seen on many occasions, most recently with the Kyoto Protocol on Climate Change, forging an international consensus on environmental protection is difficult, in part because all signatories must *voluntarily* accede to the treaty.

A number of arguments have been put forward regarding why it can be difficult to achieve broad international consensus on environmental treaties. An obvious reason, which has been articulated by Barrett (2003) among others, is that agreement is particularly tough when there are substantial differences from one country to another in costs and benefits from the agreement. No matter how altruistic a country may be, it is often difficult to enter into agreements that are not in the country's self-interest.[2]

Of course, if the environmental problem is real, it should be possible to fashion an agreement in which all countries are better off with the agreement than without,

---

[1] We use the word "treaty" synonymously with "agreement" here in this paper.

[2] Norway is a prominent example of an environmentally progressive country, except when domestic interests run counter to international interests, as in the international whaling convention

though political considerations often make this difficult.[3] Even if this were possible, the problem of free-riding remains: how to prevent countries from benefiting from an agreement without joining.

A subtler factor influencing the establishment of international agreements is *uncertainty*. One of the major reasons cited by US President Bush in pulling the US out of the Kyoto Protocol was uncertainty.[4] He suggested that more information was needed before he could support committing the US to the agreement. Thus not only did uncertainty play into his decision but also entering the decision was the fact that he anticipated that more would be known in the future – both uncertainty and learning were significant. Although it is unclear how dominant these factors ultimately were to Bush's decision to withdraw from Kyoto, the fact remains that uncertainty and learning may play roles in the timing of an international agreement. Interestingly, uncertainty also plays a role in the perspective of the EU, though reaching opposite conclusions: that something must be done before it is too late.[5]

Such international environmental treaties are termed in the scholarly literature, *self-enforcing international environmental agreements* (SEIEA).[6] The term

---

[3] For example, if the beneficiaries of a global climate agreement are largely in the South with the North incurring emission control costs, then a Pareto-improving agreement would involve the South paying the North – obviously a political non-starter.

[4] For instance: "I oppose the Kyoto Protocol...we must be very careful not to take actions that could harm consumers. This is especially true given the incomplete state of scientific knowledge..." (letter from George Bush to Senators Hagel, Helms, Craig and Roberts, dated March 13, 2001; available as White House Press Release on www.whitehouse.gov/news/releases/2001/03/20010314.html)

[5] For instance: "Both the IPCC and the NAS reports recognize that there are uncertainties still. But there is also agreement that the scientific evidence is solid enough to warrant concrete and urgent action. Delaying action could increase both the rate and the eventual magnitude of climate change and hence adaptation and damage costs." (Delegation of the European Commission to the US, 9/2001: http://www.eurunion.org/legislat/climatechange.htm)

[6] Hoel (1992), Carraro and Siniscalco (1993) and Barrett (1994) were some of the first authors to focus on the problem of supporting international environmental cooperation.

self-enforcing is used because there can be no appeal to a higher authority for enforcement: the terms of the agreement must be such that enforcement and incentives to adhere to the agreement are implicit in the agreement. The primary focus of research on this topic is in understanding what it takes to support a real welfare-improving SEIEA. The typical issue is what does it take to construct an agreement that a significant number of countries will voluntarily seek to join.[7]

Several authors have examined the role of uncertainty and learning in the timing of SEIEA's and have reached conflicting conclusions. Some research suggests agreements tend not to be consummated until uncertainty is resolved; other work suggests uncertainty facilitates agreement. The primary issue is timing: do uncertainty and learning accelerate or retard (or neither) the formation of an agreement? The theoretical and case study literature is mixed.[8]

That is the question we address in this paper: how do uncertainty and learning affect the formation of an SEIEA? We address this issue by introducing 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.

---

[7] One of the most comprehensive recent treatments of this topic is Barrett (2003).

[8] Cooper (1989) has examined two centuries of international treaties to control the spread of communicable diseases and similarly concludes that it is only when uncertainty is largely resolved will countries enter into international treaties. The political scientist Oran Young (1993), on the other hand, suggests that it may be easier to enter into treaties when parties are uncertain about their individual net benefits from an agreement than when that uncertainty has been resolved.

Although there are many ways of representing learning and uncertainty, this is one of the most obvious and simple.

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.

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 are introduced. The paper closes with conclusions.

## II. BACKGROUND

There is a significant economic literature, primarily post-1990, on self-enforcing international environmental agreements.[9] In this section of the paper, we review what that literature tells us about the formation of these agreements. We divide our discussion into two parts. We first consider the question of SEIEA's in an environment of certainty. It is in this context that most theoretical results have been derived. We then turn our attention to the much more specialized literature dealing with the effect of uncertainty and learning on the formation of SEIEA's.

---

[9] Wagner (2001) provides a survey.

## A. Self-Enforcing International Environmental Agreements.

The literature on international environmental agreements (IEA's) has grown over the past decade.[10] Most of the literature focuses on self-enforcing agreements; ie, agreements which are structured so that they are effective and cohesive or stable[11] 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.[12] 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. A low incentive to free-ride is the flip-side of a small gain from cooperating.

A fundamental issue in this entire literature is how big will an agreement be: what is the size of a "stable" agreement? This hinges on what holds an agreement together, what keeps countries in the agreement. Assumptions can range from complex commitment procedures, to punishments for defecting to simple self-interest without commitment.

The most common, and simplest, notion of stability draws on the cartel stability literature (eg, d'Aspremont et al, 1983; Donsimoni et al, 1986), wherein a

---

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

[11] The literature on SEIEA uses the term "stable" to refer to coalitions of countries that will tend to stay together and not break up. This is a somewhat unfortunate choice of words, since stability is generally a dynamic term referring to the tendency of an equilibrium or coalition to remain unchanged when conditions are perturbed slightly. That is not the meaning here. In the interests of clarity, we use the standard term "stability" here to describe coalitions that are cohesive, recognizing the less-than-satisfactory nature of the term.

[12] 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.

stable cartel is defined as a cartel for which there are no incentives for any individual members to leave nor any outsiders to join. This turns out to be a very strong stability assumption in the sense that many potential cartels fail the test. Chander and Tulkens (1992, 1994) adopt a stronger assumption that should any individual member of a voluntary agreement choose to leave, the entire agreement would be null and void. (Using this definition, the authors propose an innovative way of sharing the gains from the agreement such that the agreement ends up consisting of all countries.) This is of course a weak definition of stability in that the threat of total disbanding of the agreement should one country defect is what keeps the coalition together. This threat may not be entirely credible; the remaining countries might be expected to do the best they can absent the defector. Between these two definitions is the concept of farsighted stability (Ecchia and Mariotti, 1997; Eyckmans, 2003). The idea is that an agreement is stable if no country has an incentive to leave or join, but in evaluating those incentives, countries look beyond their act of joining or leaving to the credible additional actions that other countries may take. For instance, a country may see a gain from unilaterally leaving an agreement (assuming the others take no further action). But if the country leaves, then other countries may leave. The potential defector compares current payoffs with the ultimate payoff, taking into account all the subsequent moves by other players. This is a strong condition which provides stronger glue to hold an agreement together. Other authors have sought

Another issue is the extent to which participating countries are homogeneous vs. heterogeneous. Most of the results in this literature rely on homogeneity of participating countries. An exception is Barrett (2001), who shows that heterogeneity of countries can reduce the free-riding problem and thus help support larger coalitions. Heterogeneity facilitates commitment. And commitment is the big problem in self-enforcing agreements.

## B. Uncertainty.

The results reviewed in the previous section focus on forming agreements in an environment of certainty. The literature on uncertainty in the context of international agreements is sparser 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." In other words, uncertainty about the *distribution* of gains and losses can facilitate agreement.[13] 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.

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 international macroeconomic agreements. Generally, he focuses less on distributional uncertainty (as does Young) and more on uncertainty that is common to all players. And by doing so, comes up with opposite conclusions. Iida distinguishes between *strategic* uncertainty, which is uncertainty about the types of opponents, and *systematic* (also called *analytic* or *model*) uncertainty, which amounts to commonly shared uncertainty about your own payoffs (as well as the payoffs of others). Although systematic uncertainty is the focus of his paper, Iida points out that strategic uncertainty has dominated the literature on international agreements. Iida's interpretation of systematic 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, through the use of a simple example, that systematic uncertainty will tend to retard international cooperation. In this

---

[13] In the context of trade reform, Fernandez and Rodrik (1991) show that this type of uncertainty can in fact retard agreement, favoring the *status quo*.

vein, Frankel and Rockett (1988) examine international macroeconomic agreements and conduct an empirical analysis of systematic uncertainty. To do this, they examine ten different econometric models of the global economy, examining what happens to welfare if negotiating countries use the “wrong” model. They show that uncertainty can introduce significant welfare losses (relative to no agreement) when there is this type of systematic uncertainty. This fact alone can serve to retard agreement until uncertainty is resolved. Cooper (1989) explores this same issue, though by analogy with a quite different international forum: public health agreements. In analyzing a century of such agreements, he comes up with the same 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.

Several authors focus specifically on uncertainty in SEIEA’s. Na and Shin (1998) compare cooperation from both an *ex ante* (before uncertainty is resolved) and *ex post* (after uncertainty is 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 *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. *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 with non-cooperative bargaining than in full *ex ante* cooperation. Thus their result.

In the specific literature on SEIEA's, 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. Under uncertainty, both countries are identical ex ante and thus cooperation is the equilibrium. The model is constructed in such a way that with perfect knowledge of type, there is sufficient heterogeneity for noncooperation to be the Nash equilibrium. Though the paper is important, the nature of the two-country model makes it difficult to generalize results to N-country SEIEA's.

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 effect 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.[14] 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, primarily because it can lead to noncooperative equilibria with lower aggregate levels of utility. 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

---

[14] There is now a fairly large literature on how learning affects a single regulator, controlling an externality. At the simplest level, if decision makers are risk neutral, then uncertainty and learning should have no effect (Simon, 1956) unless actions are irreversible. With irreversibilities and learning occurring over time, the environmental protection decision may be biased upwards or downwards, depending on the nature of the learning process and the irreversiblities (Epstein, 1980; Kolstad, 1997; Ulph and Ulph, 1997).

In the specific context of climate change, he "timing" of regulatory action in the presence of learning has been addressed by a number of authors (e.g., Kolstad, 1994, 1996; Ulph and Ulph, 1997; 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 irreversiblilties 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 irreversilbilities imply the optimality of the delay of control. Several authors have analyzed this systematically in the context of climate change 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. Consistent with the literature discussed in the previous section, we will examine a very specific type of uncertainty – systematic uncertainty (uncertainty over a variable which is common to all countries). Omitted from our analysis is the more complex type of uncertainty, strategic or distributional uncertainty (uncertainty wherein the realization may vary from one country to another).

### A. A Self-Enforcing IEA

We will adapt the classic model of a self-enforcing international environmental agreement to include uncertainty and learning.[15] Consider i=1,...,N countries, each emitting pollution ($q_i$) which contributes to the global commons ($Q=\sum_j q_j$). Initially, we assume all countries are homogeneous, in the sense of having the same payoff function; we later relax this assumption. For simplicity, assume each country makes a discrete choice regarding how much pollution to emit, which without loss of generality may be restricted to 0 or 1: 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:

---

[15] 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 basic model presented here is a simplification of a model due to Ulph (2002). Ulph (2002) considers a stock pollutant, a generalization of the static case considered here. Some of the results presented here follow directly from the Ulph (2002) paper; other results here are more general than those found in Ulph (2002).

$$
\begin{aligned}
\Pi_i(q_i, Q_{-i}) &\equiv cq_i - bQ \\
&= cq_i - b(q_i + Q_{-i}) \\
&= q_i - \gamma(q_i + Q_{-i}) \qquad (1)
\end{aligned}
$$

where $Q_{-i}=\sum_{j\neq i} q_j$ and $\gamma = b/c$ and, without loss of generality, we let $c \equiv 1$. Thus $\gamma$ is simply the benefit-cost ratio for emissions control – the ratio of own marginal environmental damage from emissions to the marginal cost of emissions control. Clearly we wish to focus on the case of $\gamma < 1$, and we make that assumption; 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. The membership game is an announcement game in which countries decide whether or not to join the IEA.[16] In the emissions game, the membership of the IEA is given and 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 and a Nash equilibrium results. Membership of the coalition cannot change in the emissions game.

Initially anyway, we assume all countries are identical. In this case, the primary question we ask is how big will the IEA be; i.e., how many countries? Drawing on the cartel stability literature mentioned earlier, we define two conditions

---

[16] In the announcement game, each country announces "in" or "out." A Nash equilibrium is a set of announcements for which no country will do better by unilaterally changing its announcement.

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). This is simply the definition of a Nash equilibrium in the announcement/membership game. 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 $\Pi^c(n) > \Pi^f(n-1)$.

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

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.[17]

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.[18] It is useful to introduce the function I(x):

Defn: Define I(x) as the smallest integer greater than x.

---

[17] Carraro and Siniscalco (1993) show that for a model with a continuous choice of emissions levels and with the second stage emissions 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 *a priori*.

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

$$\Pi^c(n,\gamma) = \begin{cases} -\gamma(N-n) & \textit{if coalition members abate} \\ 1-\gamma N & \textit{if coalition members pollute} \end{cases} \qquad (2)$$

as shown in Table I. This implies that members of the coalition will abate if $-\gamma(N-n) > 1-\gamma N \Leftrightarrow 1-\gamma n < 0 \Leftrightarrow n > 1/\gamma$. Thus the payoff to each member of the coalition is given by

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

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 $n$ – the number of members of the coalition. 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 all countries into the "everyone pollutes" section of the figure, where payoffs are lowest. Similarly, any $n$ to the left of $1/\gamma$ involves everyone polluting, which is not an effective coaltion. The discrete nature of the problem (an integer number of countries) gives us the result. If $n$ were continuous, then stability, as defined here, 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 (relative to abatement costs), a large coalition is likely to form, though the damages and abatement costs will not justify much action. On the other hand if damage is significant, the coalition is likely to be small, with not much action either. So either way, the coalition doesn't help much. This is a well-known, though depressing, result.[19]

## B. Systematic uncertainty

---

[19] One can show that the net welfare gains of an SEIEA are larger when $\gamma$ is larger (smaller size IEA).

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 $\gamma_H > \gamma_L$:

$$\gamma = \begin{cases} \gamma_H & \text{with probability } \pi \\ \gamma_L & \text{with probability } 1-\pi \end{cases} \tag{4}$$

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 $\gamma$. Let $\Gamma \equiv \pi\,\gamma_H + (1-\pi)\,\gamma_L$, the expected value of $\gamma$.

We are interested in three cases, one where actions can be fully conditioned on the state-of-the-world (which we term no uncertainty, NU), one where actions cannot be conditioned on states-of-the-world and uncertainty is never resolved (uncertainty with no learning, NL), and one where uncertainty is resolved between the membership and emissions games (uncertainty with learning, L). 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. This follows directly from Prop. 1, with $\gamma$ in Table I replaced by either $\gamma_L$ or $\gamma_H$, depending on the state-of-the-world. Thus 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 $\Gamma$ was defined earlier as the expected value of $\gamma$. Payoffs are in Table I with $\gamma$ replaced with $\Gamma$.

Proposition 2. If $n^*_{NL}$ and $n^*_{NU}$ are defined as in Eqn. 5 and 6, then $n^*_{NL}-1 \leq n^*_{NU}$, .

Proof: Follows from the convexity of the 1/x function. However, $i(x) \equiv I(1/x)$ is not "locally" convex because it is a step function (due to its integer nature). Thus in some cases, the function i() of the convex combination of $\gamma_L$ and $\gamma_H$ may be higher than the convex combination of $i(\gamma_L)$ and $i(\gamma_H)$, due to rounding off to the integer value. But this difference will always be less than 1. ■

This proposition in essence says that the number of coalition members in the uncertainty without learning case (NL) is less than the number of coalition members in the no uncertainty case (NU), where actions can be conditioned on the state of the world. In other words, uncertainty tends to reduce the size of the stable coalition. The conclusion must be qualified by saying that due to the integer nature of the number of members of the coalition, it is possible to construct special cases where the two numbers may be essentially the same (i.e., the difference is less than 1). The intuition is that small gamma regimes generate much bigger IEA's. Thus a dispersion in the gammas tends to raise the expected value of the size of the IEA more than just the size of the IEA from the expected value of gamma.

The learning (L) case must be solved by backwards induction. The emissions game does not involve 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 $I(1/\gamma_H) \leq n < I(1/\gamma_L)$, then coalition members abate in state-of-the-world H and pollute in state-of-the-world L. Finally, if $n \geq I(1/\gamma_L)$, 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. This case is not very interesting and will not be considered, though our results apply to this case as well. The other case is where integers can be inserted between $I(1/\gamma_H)$ and $I(1/\gamma_L)$. In the following discussion of intuition, we assume this second case applies.

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 coalition profits are higher, holding n constant. One might think this would provide an incentive to increase the number of members of the coalition. Table III 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 III, 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 involves abatement. At $n= I(1/\gamma_H)-1$, 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.

<u>Proposition 3:</u> Provided $\pi > 0$, then $n^*_L = I(1/\gamma_H)$ is a stable coalition. At most there is one additional stable coalition, at $n^*_L = I(1/\gamma_L)$.

<u>Proof</u>: That internal stability holds at $n=I(1/\gamma_H)$ follows from the logic of Prop. 1. From Table III, 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. All that remains is to show that no other abating coalition is stable, with the possible exception of $n_L = I(1/\gamma_L)$. All $n < I(1/\gamma_H)$ result in no abatement and thus

can be eliminated. Internal stability fails for n such that $I(1/\gamma_H) < n < I(1/\gamma_L)$ since a coalition member will gain $\pi$ by defecting to the fringe and only lose $\pi\gamma_H < \pi$ by leaving the coalition. For all $n > I(1/\gamma_L)$, internal stability fails by the same logic: a defector gains 1 by joining the fringe and only loses $\gamma < 1$. ■

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

<u>Proposition 4.</u> If $\pi > \Gamma$ then internal stability fails at $n = I(1/\gamma_L)$, resulting in the equilibrium number of countries in the effective coalition of $n_L^* = I(1/\gamma_H) \leq n_{NL}^*$.

<u>Proof:</u> From Proposition 3, we know that $n_L^* = I(1/\gamma_H)$ is a stable coalition and the only other possible stable coalition is $n' = I(1/\gamma_L)$. Internal stability will fail at this point if $\Delta P \equiv \Pi^c(n') - \Pi^f(n' - 1) < 0$; ie, there is an incentive to leave the coalition for the fringe at n' if $\Delta P < 0$. From Table III, we see that

$$\Delta P = (1-\pi)\, n\gamma_L + \pi\gamma_H - 1 \tag{7}$$

From the definition of the $I(\bullet)$ function, we know that

$$1/\gamma_L \leq n' < 1/\gamma_L + 1 \tag{8}$$

which can be rearranged into

$$\pi(\gamma_H - 1) \leq \Delta P < \Gamma - \pi \tag{9}$$

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.

<u>Proposition 5.</u> 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 $|1/\gamma_L - 1/\gamma_L^{**}| < 1$, 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,

$$\Delta P \approx - \gamma[N-(1/\gamma_L + 1)] - [1 - \Gamma N + \pi\gamma_H/\gamma_L] = \Gamma - \pi > 0 \qquad (10)$$

This proves both parts of the proposition. $\blacksquare$

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-riding in the fringe.

The variable $\Gamma$ on the other hand is the ratio of the expected environmental benefits from abatement to the abatement costs. It is 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 ($\Gamma$) with the advantages of not having to pay to abate ($\pi$). 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 and learning tend 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 learning 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, and this will be the only stable coalition. 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, the advantages of being in the fringe grow, and grow more rapidly than the advantages of being in the coalition. Ultimately, 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. In fact, it drops below the size of the coalition with uncertainty but no learning.

For problems which will almost solve themselves, for which the IEA does not bring much to the table, we would expect $\Gamma$ to be significantly greater than 0, probably closer to 1 (though this would involve a small IEA). This suggests that Prop. 5 would probably apply and a larger IEA may be stable. For tough problems, with a much smaller Γ, the larger IEA will not generally be stable. (It has been suggested that climate change fits into this category.) In this case, uncertainty has the effect of reducing the overall size of a stable IEA. Thus learning tends to expand the size of the IEA when the IEA is small (large Γ) and shrink the size of the IEA when the IEA is large (small Γ).

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 π relative to Γ, 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.

This result is somewhat more ambiguous than the conclusions of others that systematic uncertainty unequivocally reduces the size of the coalition. We see that on the one hand uncertainty and learning unequivocally generates a smaller stable coalition than would be the case with no learning (a result which supports the prior literature). But under certain conditions, the coalition can actually expand.

How does this relate to the literature on systematic uncertainty? That literature generally views learning as reducing the size of a stable agreement, due to the fact that there can be big errors associated with acting on the basis of the "wrong" gamma. Although there may not be enough richness in our model to reach the same conclusion, if one considers the case where there is substantial uncertainty about the state of the world (π takes on a mid-range value), then our results suggest learning results in an unequivocal reduction in the size of stable IEA.

## 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 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.

First considering the case of uncertainty without learning, we find that uncertainty tends to decrease the size of the cooperating coalition in an international environmental agreement. We are comparing the size of a coalition formed with uncertainty in the private benefit-cost ratio for abatement with the expected size of coalitions which can be conditioned on the actual value of the cost-benefit ratio. This is simply a result of the nonlinear nature of coalition formation. In cases where the benefit-cost ratio of abatement is large, the coalition will be small; as the benefit-ratio drops, the size of the stable coalition rises more rapidly than the ratio drops. Thus uncertainty in the benefit-cost ratio, based on the expected value of the cost-benefit ratio, results in a smaller coalition than the probabilistic average of two coalition sizes associated with two different benefit-cost ratios.

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 occurs. We find that the possibility of several stable coalitions emerges under learning. One of the stable coalitions is unequivocally smaller than it would be without learning. Thus one result is that learning can reduce the size of the stable coalition. But there is a second possible stable coalition. When the advantages of a coalition are modest (i.e., the benefit-cost ratio is high), the most likely situation is that learning tends to result in the emergence of a larger stable coalition of participants in an IEA (though the smaller IEA remains a stable outcome as well). When one most needs a coalition (i.e., the benefit-cost ratio of abatement is low), then the most likely outcome is that learning results in a smaller coalition than with uncertainty but no learning. This discouraging result mirrors the known result in the case of certainty, that SEIEA's tend to be smallest when you need them the most.

Although we have been able to reach definitive conclusions here about the effect of uncertainty and learning on SEIEA's, our representation of learning is clearly not general. In particular, learning is by nature usually viewed as a dynamic process. Clearly a static model cannot fully address this issue. Thus there is ample opportunity for further work on this issue. Furthermore, the case of distributional or uncorrelated uncertainty has not been explored here.

## REFERENCES

Barrett, Scott, “ Self-Enforcing International Environmental Agreements,” *Oxford Economic Papers*, **46**:878-94 (1994).

Barrett, Scott, “International Cooperation for Sale,” *European Economic Review*, **45**:1835-50 (2001).

Barrett, Scott, *Environment and Statecraft* (Oxford University Press, Oxford, 2003).

Bliss, Christopher and Barry Nalebuff, “Dragon-Slaying and Ballroom Dancing: The Private Supply of a Public Good,” *J. Public Econ.*, **25**:1-12 (1984).

Brennan, Geoffrey, and James M. Buchanan, *The Reason of Rules: Constitutional Political Economy* (Cambridge University Press, Cambridge, UK, 1985).

Carraro, Carlo and Domenico Siniscalco, “Strategies for the International Protection of the Environment,” *J. Public Econ.*, **52**:309-28 (1993).

Chander, Parkash and Henri Tulkens, “Theoretical Foundations of Negotiations and Cost-Sharing in Transfrontier Pollution Problems,” *Eur. Econ. Rev.,* **36**:288-99 (1992)

Chander, Parkash and Henri Tulkens, “A Core-theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution,” *Intl Tax and Public Finance*, **2**:279-93 (1994).

Cooper, Richard N., “International Cooperation in Public Health as a Prologue to Macroeconomic Cooperation,” pp 178-254 in R. N. Cooper et al (ed.), *Can Nations Agree?* (Brookings Institution, Washington, DC, 1989).

d’Aspremont, C., A. Jacquemin, J. Jaskold-Gabszewicz and J. Weymark, “On the Stability of Collusive Price Leadership,” *Canadian J. Economics*, **16**:17-25 (1983).

Donsimoni, M.-P., N.S. Economides, and H.M. Polemarcharkis, “Stable Cartels,” *International Economic Rev.*, **27**:317-27 (1986)

Ecchia, Giulio and Marco Mariotti, “The Stability of International Environmental Coaltions with Farsighted Countries: Some Theoretical Observations,” Ch. 10 in Carol Carraro (Ed), *International Environmental Negotiations* (Edward Elgar, Cheltenham, UK, 1997).

Epstein, Larry G, “Decision Making and the Temporal Resolution of Uncertainty,” *Intl Economic Review*, **21**:269-83 (1980).

Eyckmans, Johan, “On the farsighted stability of international climate agreements,” University of Leuven working paper (Jan 2003).

Fernandez, Raquel and Dani Rodrik, “Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty,” *Amer. Econ. Rev.*, **81**:1146-55 (1991).

Finus, Michael, *Game Theory and International Environmental Cooperation* (Edward Elgar, Cheltenham, UK, 2001).

Folmer, Henk and P. von Mouche, “Interconnected Games and International Environmental Problems,” *Annals of Operations Research*, **54**:97-117 (1994).

Frankel, Jeffrey A. and K. Rockett, “International Macroeconomic Policy Coordination: When Policy Makers Do Not Agree on the True Model,” *Amer. Econ. Rev.*, **78**:318-40 (1988).

Haas, E.B., “Why Collaborate – Issue Linkage and International Regimes,” *World Politics*, **32**:357-405 (1980).

Helm, Carsten, “International Cooperation Behind the Veil of Uncertainty,” *Environmental and Resource Economics*, **12**:185-201 (1998).

Hoel, Michael, “International Environmental Conventions: The Case of Uniform Reductions of Emissions,” *Environmental and Resource Economics*, **2**:141-59 (1992).

Iida, Keisuke, “Analytic Uncertainty and International Cooperation: Theory and Application to International Economic Policy Coordination,” *International Studies Quarterly*, **37**:431-57 (1993).

Kelly, David and Charles Kolstad, “Bayesian Learning, Growth and Pollution,” *J. Economic Dynamics and Control*, **23**:491-518 (1999).

Kolstad, Charles D., “George Bush vs. Al Gore: Irreversibilities in Greenhouse Gas Accumulation and Emission Control Investment,” *Energy Policy*, **22**:771-78 (1994).

Kolstad, Charles D., “Learning and Stock Effects in Environmental Regulation: The Case of Greenhouse Gas Emissions,” *J. Env. Econ. & Mgmt*, **31**:1-18 (1996a).

Kolstad, Charles, “Fundamental Irreversibilities in Stock Externalities,” *J. Public Econ.,* **60**:221-33 (1996b).

Kolstad, Charles D., “The Difficulty in Reaching Cooperative Agreement and the Veil of Uncertainty in Transboundary Pollution,” unpublished manuscript, University of California, Santa Barbara (August 2003).

Krutilla, John V., *The Columbia River Treaty: The Economics of International River Basin Development* (Johns Hopkins University Press for Resources for the Future, Baltimore, 1967).

McGinnis, M.D., “Issue Linkage and the Evolution of International Cooperation,” *J. Conflict Resolution*, **30**:141-70 (1986).

Na, Seong-lin, and Hyun Song Shin, “International Environmental Agreements under Uncertainty,” *Oxford Economic Papers*, **50**:173-85 (1998).

Rubio, Santiago and Alistair Ulph, “A Simple Dynamic Model of International Environmental Agreements with a Stock Pollutant,” unpublished manuscript, University of Valencia (November 2001).

Rubio, Santiago and Alistair Ulph, “Leadership and Self-Enforcing International Environmental Agreements with Non-negative Emissions,” unpublished manuscript, University of Valencia (9 January 2003).

Ulph, Alistair, “International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning,” unpublished manuscript, University of Southampton (August 2002).

Ulph, Alistair, and David Maddison, “Uncertainty, Learning and International Environmental Policy Coordination,” *Environmental and Resource Economics*, **9**:451-66 (1997).

Ulph, Alistair and David Ulph, “Who Gains from Learning about Global Warrming?,” Ch 3 (pp 31-67) in Ekko C. van Ierland and Kazimierz Gorka (Eds), *Economics of Atmospheric Pollution* (Springer-Verlag, Berlin, 1996).

Ulph, Alistair and David Ulph, “Global Warming, Irreversibility and Learning,” Economic Journal, 107:636-50 (1997).

Wagner, Ulrich J., “The Design of Stable International Environmental Agreements,” *J. Economic Surveys*, 15:377-411 (2001)

Young, Oran, *International Governance: Protecting the Environment in a Stateless Society* (Cornell University Press, Ithaca, NY, 1994)

| n: | $n < I(1/\gamma)$ | $n \geq I(1/\gamma)$ |
|---|---|---|
| $\Pi^c(n)$ | $1 - \gamma N$ | $- \gamma (N - n)$ |
| $\Pi^f(n)$ | $1 - \gamma N$ | $1 - \gamma (N - n)$ |

Table I: Payoffs without uncertainty, Eqn.2.

| State-of-World: | H | | L | |
|---|---|---|---|---|
| n: | $n < I(1/\gamma_H)$ | $n \geq I(1/\gamma_H)$ | $n < I(1/\gamma_L)$ | $n \geq I(1/\gamma_L)$ |
| $\Pi^c(n)$ | $1 - \gamma_H N$ | $- \gamma_H (N - n)$ | $1 - \gamma_L N$ | $- \gamma_L (N - n)$ |
| $\Pi^f(n)$ | $1 - \gamma_H N$ | $1 - \gamma_H (N - n)$ | $1 - \gamma_L N$ | $1 - \gamma_L (N - n)$ |

Table II: Expected payoffs in Membership Game with Systematic uncertainty, NU.

| n: | $n < I(1/\gamma_H)$ | $I(1/\gamma_H) \leq n < I(1/\gamma_L)$ | $n \geq I(1/\gamma_L)$ |
|---|---|---|---|
| $\Pi^c(n)$ | $1 - \Gamma N$ | $1 - \Gamma N + \pi(\gamma_H n - 1)$ | $- \Gamma (N - n)$ |
| $\Pi^f(n)$ | $1 - \Gamma N$ | $1 - \Gamma N + \pi \gamma_H n$ | $1 - \Gamma (N - n)$ |

Table III: Expected payoffs in Membership Game with Systematic uncertainty, L.

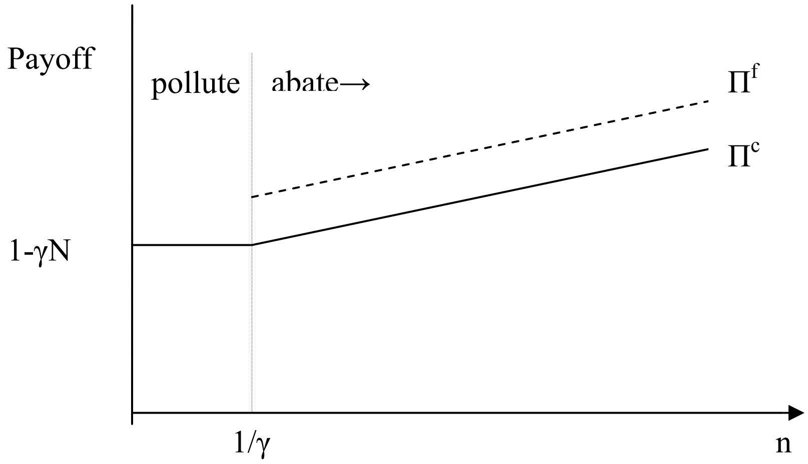

Figure 1: Payoffs in Emissions game

Figure 2: Payoff in Emissions Game, Systematic uncertainty