Climatic Change (2008) 89:125–141
DOI 10.1007/s10584-008-9399-8

Learning and international environmental agreements

Charles Kolstad · Alistair Ulph

Received: 12 January 2007 / Accepted: 13 December 2007 /

Abstract In this paper we synthesise and extend our earlier analysis (Ulph, J Risk Uncertain 29(1):53–73, 2004; Kolstad, J Environ Econ Manage 53(1):68–79, 2007) of the formation of an International Environmental Agreement (IEA) under uncertainty about the damages that might be caused by climate change and different models of learning, in which better information about these damage costs become available. Our results are generally pessimistic: the possibility of either complete or partial learning generally reduces the level of global welfare that can be achieved from forming an IEA. This suggests that information can have negative value. This may seem strange, since for a single decision-maker information cannot have negative value, because it can always be ignored. However in this case there are strategic interactions between a number of decision-makers responding to information, and it is these strategic interactions which can give information a negative value.

1 Introduction

One of the major issues in climate change policy is how to deal with ubiquitous uncertainty: though much is known, nearly everything is uncertain. There are many uncertainties in the physical science of climate, including the likelihood of catastrophic or abrupt change. The costs of mitigation and adaptation are poorly known. The response of biological systems to climate change is also uncertain. Finally, how society will respond either in terms of preventing change or in adapting to change is imperfectly understood.

An earlier version of this paper was presented at the conference “Learning and Climate Change”, IIASA, Laxenburg, Austria, 10–11 April 2006.

C. Kolstad (✉)
Department of Economics and Bren School of Environmental Science and Management,
University of California, Santa Barbara, CA, USA
e-mail: kolstad@econ.ucsb.edu

A. Ulph
Faculty of Humanities, University of Manchester, Manchester, England

Uncertainty is of course a fact of life and its existence should be no real obstacle to formulating sound public policy—policymakers can simply act on expectations. But what complicates things for climate policy is that uncertainty itself is changing—we are learning about the science and economics of the problem as time passes. Furthermore, we are taking active steps to increase our knowledge of the processes through significant R&D programs. In a decade we will be much better informed about the problem and in two decades, even better informed.

This process of learning raises an important timing question. On the one hand it can be argued that society should delay taking action to reduce greenhouse gas emissions, until more is known. After all, if we act and subsequently learn that climate change is less serious than we had thought, we will have taken steps unnecessarily. On the other hand, if we do learn that climate change is more serious than we had thought, we can always accelerate action later. However it is often argued that this 'learn-then-act' approach only makes sense if the accumulation of greenhouse gas emissions is *reversible*, so that if we make a policy "mistake," we can 'undo' the effects of the decision. But the accumulation of greenhouse gases is often viewed as *irreversible*,[1] so that by the time we learn that climate change is a serious issue we may have built up such concentrations of greenhouse gases that we are faced with drastic consequences which cannot be readily undone.

This irreversibility in the climate process leads to calls for implementing a *precautionary approach*—that, far from delaying taking steps to reduce greenhouse gas emissions while we wait for better information, we should take more steps now, to guard against getting bad news in the future and finding it is too late to do anything about it. However, there are other issues at play, including the irreversibility in accumulating emission control capital. Overall, the economics literature is ambiguous about the applicability of the precautionary principle to climate (see Kolstad 1996b; Ulph and Ulph 1997; Ingham et al. 2005).

There is one subtlety here that prompts this paper. We can be concerned with how uncertainty, learning and irreversibilities affect a single decision maker, and that is the focus of much of the relevant literature. A related, though fundamentally different, problem concerns how agents interact strategically in such an environment of uncertainty and learning. That is the subject of this paper. When uncertain, learning agents are strategically interacting in the context of negotiating a global climate agreement/treaty, how do outcomes change in comparison with the certain, non-learning case? The problem is quite different. Some have argued (e.g., Young 1994) that uncertainty and learning facilitates agreement before negotiating positions become hardened by knowing exactly which particular agents win or lose—uncertainty is liberating. Yet others (e.g., Cooper 1989) have suggested that it is only after uncertainty is largely resolved that countries will come together and agree to solve a global commons problem.

The goal of this paper is to better understand how uncertainty and learning affect the size and effectiveness of environmental agreements. One approach to this problem would be to construct a large complex model of agreement formation. The main problem with such an approach is that while one may gain realism, it will be at the expense of transparency. An alternative approach is to use a highly simplified model of strategic interaction in an attempt to better understand the basic forces at work when uncertainty and learning interact with

---

[1] The process of accumulating greenhouse gases in the atmosphere is not literally irreversible, but the usual models of climate change (e.g. underlying the DICE model) suggest that if emissions of CO_2 are stopped, atmospheric concentrations would decline, but very slowly, only reverting to pre-industrial levels after 1,000 years (Fisher et al. 2002).

self-interest and strategy. Taking such a simplified approach allows deeper understanding of the underlying strategic forces involved, though at some sacrifice in predictive power.

We take the latter simplified path and extend the standard two-stage game theoretic model of such agreements first introduced by Barrett (1994). To non-specialists, the framework may seem overly simplistic and a poor representation of the subtleties of crafting an international treaty. However, the goal of the paper is not to accurately represent the entire process of developing environmental treaties but rather to improve our understanding of the essential role of uncertainty and learning in shaping incentives to forge agreements. We seek to understand the underlying forces introduced by uncertainty and learning, rather than develop a predictive tool. Although a number of the results presented here have appeared elsewhere, this represents a new synthesis of knowledge on this important problem, aimed at a broader audience than the typical literature on international environmental agreements.

In the next section we shall provide background on relevant literature on uncertainty and learning in the context of IEA's. Then in section 3 we introduce our simple model of climate change. We introduce the model without uncertainty and thus no learning. In section 4, we show what outcomes arise for three cases, all without uncertainty: (1) when there is no cooperation among countries, (2) when there is full cooperation among countries and (3) when there is partial cooperation among countries in the form of an IEA. In section 5 we modify our basic model, introducing uncertainty and learning. We again show what outcomes arise when there is no cooperation between countries, when there is full cooperation between countries and, the heart of this paper, when there is an IEA. In the latter case we shall study three models of learning: No Learning, Complete Learning and Partial Learning. Section 6 provides conclusions.

2 Background

The precautionary principle with respect to environmental protection roughly states that when environmental risks are uncertain, not well understood and irreversible, then regulatory actions should be biased towards avoiding those risks, rather than approaching the problem as a standard case of decision-making under uncertainty.[2] The precautionary principle has a good deal of intuitive appeal and is often employed in debates on climate-change policy to support calls for increased immediate action by governments to reduce greenhouse gas emissions. However, there is a substantial body of economics literature which shows that this simplistic version of the precautionary principle is not always a correct approach to climate change policy.[3] There are several arguments for such ambiguities.

At the most basic level, there are two different effects involved in the precautionary principle: the effect of the irreversibility of the accumulating stock of greenhouse gases, which does indeed imply cutting current greenhouse gas emissions relative to the case where there is no irreversibility; and the effect of learning, which has ambiguous effects on current emissions policy depending on factors such as the degree of relative risk aversion and the degree to which decision-makers are willing to substitute consumption now for

---

[2] There are many definitions of the precautionary principle; see Foster et al. (2000).

[3] Key papers are Arrow and Fisher (1974), Henry (1974a, b), Epstein (1980), Kolstad (1996a, b), Ulph and Ulph (1997), Ha-Duong (1998), Gollier et al. (2000), Fisher et al. (2002), Webster (2002) and Lange (2003). See Ingham and Ulph (2005) for a survey.

consumption later (intertemporal substitution in consumption). Ulph and Ulph (1997) showed that with the most standard model of climate change the learning effect implies that current emissions policy should be *laxer* than with no learning. A more readily appreciated argument is that in policy terms there are many quasi-irreversible stocks involved in climate change—stocks of greenhouse gases, stocks of capital which emit greenhouse gases and stocks of capital which reduce emissions of greenhouse gases. How uncertainty and learning affects emissions will depend on how it affects all these stocks, and they have different effects on emissions. Kolstad (1996b) examined how the irreversibilities of the capital and greenhouse stocks interact and reached a conclusion similar to Ulph and Ulph (1997), that current emissions policy should be slightly laxer than the case with no learning. The arguments just cited assume that the only action that can be taken to address climate change is to reduce emissions (mitigation). But an important part of climate change policy is adaptation to a changed climate. If it is assumed that adaptation is not subject to irreversibility constraints, then allowing for adaptation (a substitute for mitigation) weakens the irreversibility effect (Ingham et al. 2005). Of course an important question is how these ambiguities work out in empirical models of climate-change policy. Ingham and Ulph (2005) survey the empirical literature and show that this literature does not provide much support for the precautionary principle.

It is important to emphasise that although the literature on uncertainty and learning with a single decision maker yields ambiguous conclusions about how the decision maker's current emissions policy should respond to the possibility of obtaining better information in the future, all the models using the standard approach to decision-making under uncertainty predict that the value of information must be non-negative. This is a standard result with a single decision-maker (see Lange and Treich (2008) in this issue), for the simple reason that a decision-maker can always ignore the information and be no worse off, so if the decision-maker responds to the information it must be at least as well off as ignoring it.

In all of the above literature it is assumed that there is a single decision-maker, so the analysis is applicable to either an individual national government, or to a putative world government. Neither is directly relevant to climate change which is a global pollution problem that must be solved through negotiations among a large number of individual national governments acting in their own self-interest. In this case we are interested in what kinds of outcomes emerge from multiple independent countries negotiating in self-interest. In examining the outcome of such negotiations, we are interested in the size of successful agreements, as well as the aggregate emission reductions achieved.

In the context of multi-country agreements to address common environmental problems, Ulph and Ulph (1996) and Ulph and Maddison (1997) consider a two-country model in which countries can either act non-cooperatively or cooperatively. They use this framework to compare outcomes with and without learning. They show that in the non-cooperative equilibrium the value of information may be negative if the uncertainty about damage costs is negatively correlated between countries.[4] In other words, being better informed can actually lead to an outcome that is collectively worse.

With only two countries, the cooperative and non-cooperative equilibria are the only two possibilities. With more than two countries one can consider intermediate cases where some countries decide to join an International Environmental Agreement (IEA) in which

---

[4] The intuition is that the country that learns that it has low damage costs will raise its emissions, the country that learns it has high damage costs cuts emissions. But in the non-cooperative equilibrium these responses trigger further strategic reactions by the two countries: the low cost country further expands its emissions, the high cost country further cuts its emissions.

signatories determine their emissions to maximize their collective self-interest, while non-signatories act non-cooperatively. These agreements have come to be called *self-enforcing* agreements because joining or not joining such an agreement has to be in the interests of all countries and cannot be enforced by a supra-national body (since none exists). Since the seminal papers by Barrett (1994) and Carraro and Siniscalco (1993), a large literature, both theoretical and empirical, has developed in this field.[5] Although concepts of the formation of cooperating blocs of countries (coalition formation) have become considerably more sophisticated, the issue of uncertainty and learning has been largely ignored, and there is an obvious additional question to be posed: how do uncertainty and the prospect of learning affect the incentives for countries to join an IEA?

To our knowledge the first paper to address this question is Na and Shin (1998). They consider a three country model in which, ex ante, the countries have the same expected net benefits from emissions of a global flow pollutant. Uncertainty is about the *distribution* across the three countries of net benefits from emissions; total global net benefits are known with certainty. They employ a model of coalition formation which is a variant of the Barrett (1994) model in which countries first decide whether to join an IEA (*the membership game*), and then decide their emissions levels (*the emissions game*), and in both games they look for a Nash equilibrium (the Nash equilibrium of the membership game is often referred to as a *stable* IEA). Na and Shin consider two models of learning. With Uncertainty and No Learning, countries make both their membership decisions and their emissions decisions under uncertainty about the true state of the world. With Uncertainty and Complete Learning, countries make their membership and emissions decisions knowing the true state of the world. Na and Shin show that with Uncertainty and No Learning, the unique stable IEA of the membership game is the *grand coalition* of all three countries. With Complete Learning the grand coalition is never stable, and there will be either a two-country coalition, or no coalition at all. So in their model, learning is bad for cooperation.

A limitation of the Na and Shin model is that if one models the emissions game with continuous strategies and equilibrium as a Nash equilibrium then the maximum number of countries who will join an IEA is three (see Finus 2001); by contrast, if a Stackelberg equilibrium is used in the emissions game, in which IEA signatories take as given the emissions *strategies* of the non-signatories, rather than the emissions *levels* of the non-signatories, then the equilibrium in the membership game can have a number of signatories anywhere between two and the grand coalition of all countries. So the fact that Na and Shin, using a Nash equilibrium in the emissions game can obtain an outcome in the membership game of the grand coalition is an artifact of their assumption that there are only three countries. To assess the impact of learning on IEA membership it would seem more sensible to employ a model with an arbitrary number of countries greater than three in which the number of signatories in equilibrium could also be greater than three. Ulph (2004) and Kolstad (2007) employ a special case of the Barrett (1994) model in which the emission strategies are discrete and thus the Stackelberg equilibrium of the emissions game is identical to the Nash equilibrium, but the equilibrium of the membership game allows the size of a stable IEA to take any value between two and the grand coalition.

A second important difference between the models of Ulph (2004) and Kolstad (2007) and that of Na and Shin is that countries are identical ex ante and ex post, and uncertainty is about the unit damage cost from climate change, which is the same for all countries. So in

---

[5] See Barrett (2002) and Finus (2001) for excellent overviews, with Bohringer and Finus (2005) applying such analysis to the Kyoto protocol.

this case uncertainty is about the global net benefits from emissions, but the distribution of these benefits across countries is known.

Ulph (2004) uses an explicitly dynamic, two-period, model with stock accumulation and irreversibility, with damages depending on the stock of emissions at the end of period 2. Countries are identical and at the start of the first period face uncertainty about whether damage costs will be high or low. If learning occurs, then at the start of period 2 countries will learn whether damage costs are high or low. Ulph (2004) considers two types of membership commitment. Under one, *fixed membership*, countries decide in period 1 whether or not to join an IEA, and are committed to this for both periods. The other type of membership he considers is *variable membership*, wherein countries change their mind regarding membership after uncertainty is resolved; i.e., they can condition both their emissions strategies and membership strategies on the state of the world.

Kolstad (2007) uses a very similar model, but, like Na and Shin, with only one period, so there is no role for stock accumulation or irreversibility. At the start of the period countries are uncertain whether damage costs will be high or low. He considers three possibilities for learning: what we shall call *Partial Learning* in which countries learn the true state of the world after the membership game but before the emissions game; what we shall call *Complete Learning* in which countries learn the true state of the world before the membership game; the base case comparison is *No Learning*, which effectively means that uncertainty is resolved after the emissions game. It is clear that partial learning is the static equivalent of fixed membership in Ulph, while full resolution corresponds to variable membership.

While the details of the two models are clearly different they come to broadly similar conclusions. Focusing on the expected number of signatories, it is largest with Complete Learning and lowest with No Learning. Partial Learning is a little messier in that it results in two possible stable IEAs, the same two as with Complete Learning: one in which both the number of signatories and welfare with Partial Learning is less than with No Learning, which is always stable, and one in which the number of signatories and welfare with Partial Learning are higher than with No Learning, which will be stable only if the probability of low damage costs is sufficiently high. The general conclusion is that the possibility of learning reduces welfare, so the value of information is generally negative.

For non-economists, the possibility that information can have negative value may appear strange. We noted earlier that with a single decision-maker information must have non-negative value. But it is a well-known result in economics that if there are many agents acting strategically, then the strategic responses of agents to new information can cause information to have negative value. Early examples can be found in the 1968 work of Braess in transport economics[6] and the work of Hirshleifer (1971) in financial economics. The papers by Ulph and Ulph (1996) and Ulph and Maddison (1997) already referred to above demonstrated the possibility of a negative value of information in a non-cooperative game model of climate change. This paper, and its antecedents Ulph (2004) and Kolstad (2007), are the first to demonstrate that the possibility of a negative value for information can also arise in the context of the formation of an IEA.

Finally we note that both Ulph (2004) and Kolstad (2007) consider the simple case of uncertainty about a variable that is common to all countries. Countries face the same expected damage costs, but there are only two possible states of the world—either all countries have high damage costs, or all have low damage costs, so countries are also

---

[6] The work of Braess was published in German; Steinberg and Zangwill (1983) present a useful survey in English.

identical ex post. The case in which countries may individually draw a high or low damage cost parameter is more complex; we defer that case for another paper (Kolstad and Ulph 2007).

3 The basic model

The model set out in this section is common to both Ulph (2004) and Kolstad (2007). The objective in defining the model is to provide the simplest framework which captures the essence of the problem. A country is represented by a single decision-maker with simple objectives. Furthermore, our decision maker may only make one decision: whether to abate the pollution he/she generates or not.

For readers less familiar with the formal mathematical approach we shall be using it is perhaps worth giving some guidance to our terminology. We shall identify key assumptions in the form of numbered assumptions, such as *Assumption 1* and similarly for definitions. Our key results will be stated as Propositions. Less important results, or results which are intermediate steps in deriving our propositions, will be called Lemmas. Finally results which are consequences of our main results will be stated as Corollaries.

Our general framework is that of N identical countries, indexed $i=1,\ldots,N$. Country i's emissions are denoted q_i, which for simplicity we assume can take one of two values: $q_i=0$ (abate) or $q_i=1$ (pollute). We denote by $Q \equiv \sum_j q_j$, the aggregate emissions of all countries and by $Q_i \equiv \sum_{j\neq 1} q_j$, the aggregate emissions of all countries other than i. Country i's net benefit function is:

$$V_i(q_i, Q_i) = q_i - \gamma_i(q_i + Q_i), \tag{1}$$

where γ_i denotes the amount of environmental damage a unit of emissions causes to country i relative to the private benefit of emitting a unit. At this stage we assume $\gamma_i = \gamma\,\forall i$, and that:

$$\text{Assumption 1}: \frac{1}{N} < \gamma < 1 \tag{2}$$

This assumption is fairly benign. If $\gamma \geq 1$, this would mean the private benefit–cost ratio is greater than one and thus that all countries would unilaterally abate—not such an interesting case. As we will see, the case of $\gamma \leq 1/N$ is similarly uninteresting since in this case the benefit–cost ratio is so low that even in a fully cooperative case, with all countries in an agreement, abating is not desirable.

We will be examining four fundamentally different versions of this basic model

- No Uncertainty
- Uncertainty with Complete Learning
- Uncertainty with Partial Learning
- Uncertainty with No Learning

Furthermore, we will be interested in benefits to individual countries, denoted by V, and aggregate payoffs, summed over all countries, denoted by W.

4 Outcomes with no uncertainty

As a base case, consider three situations: no cooperation, complete cooperation, and a cooperative agreement of a subset of countries.

4.1 Non-cooperative equilibrium

Country i takes as given emissions of other countries, Q_i, and chooses its own emissions, q_i, to maximize net benefits, V_i. From Assumption 1 we immediately derive:

Lemma 1 In the non-cooperative equilibrium with no uncertainty, all countries pollute ($q_i = 1, \forall i$) and aggregate world net benefit is $N(1-\gamma N)<0$.

The intuition is simply that each country considers the benefits it gets from emitting one unit of pollution, Eq. 1, against the damage cost it alone would suffer (γ), and since $\gamma<1$, it pays to pollute.

4.2 Co-operative equilibrium

The N countries choose $q_1, \ldots, q_N$ to maximize $W = \sum_i V_i = \sum_i q_i - N\gamma \sum_i q_i$. From Assumption 1 we immediately derive:

Lemma 2 In the cooperative equilibrium with no uncertainty, $q_i = 0, \forall i$, and aggregate world net benefit is 0.

Again the intuition is simple. Each country now considers the benefit it gets from emitting one unit of pollution, Eq. 1, against the damage it causes to all countries, $N\gamma$. Since $\gamma>1/N$ it pays each country not to pollute.

Note that in this very simple model, the non-cooperative equilibrium produces the worst possible outcome for all countries.

4.3 International environmental agreement

In this case, joining a coalition is voluntary—some countries may join, others may not. Following Barrett (1994) and others, we model this as a two-stage game. In stage 1 (membership game), each country decides whether or not to join the agreement (an IEA). The result of this is a set of signatories to the IEA and a set of fringe members, outside the IEA. We seek a Nash equilibrium in "announcements" (i.e., "in" or "out") in which no country wishes to unilaterally leave or join the coalition. In stage 2, (emission game) each non-signatory, or fringe, country, denoted by superscript f, takes as given the emissions of all other countries and chooses its emissions to maximize its individual net benefit; the signatory countries, denoted by superscript s, collectively choose their emissions to maximize the aggregate net benefit of the signatory countries taking as given the emission strategy of the non-signatories. The outcome of stage 2 is a Stackelberg equilibrium involving the IEA (acting as one) as leader and the fringe countries (acting as individuals) as followers, though since it will turn out that the dominant strategy for non-signatories is always to pollute no matter what the IEA does the Stackelberg equilibrium of the emissions game is equivalent to a Nash equilibrium.

To "solve" this problem, we work backwards from the second stage to the first stage. Thus we determine what emissions will emerge in stage 2 as a function of the size of the IEA. This yields the payoffs to different sized IEA's and allows us to move to stage 1 and determine what sized IEA will emerge.

Definition 1. Define the function $I(x)$ as the smallest integer greater than or equal to $1/x$. The next result follows immediately from Lemmas 1 and 2.

Lemma 3 (emission game equilibrium) Given N countries, of which $n \geq 2$ are signatories to an IEA: non-signatory countries always pollute. Furthermore, if $n \geq I(\gamma)$, then all signatory countries will abate, and the net benefit to a signatory and non-signatory country respectively will be:

$$V^s(n) = -(N-n)\gamma; \tag{3a}$$

$$V^f(n) = 1 - (N-n)\gamma \tag{3b}$$

If $n < I(\gamma)$, all signatory countries pollute and the net benefit to a signatory and non-signatory country respectively will be:

$$V^s(n) = V^f(n) = 1 - N\gamma. \tag{3c}$$

Non-signatory countries always pollute for the same reason as in the non-cooperative equilibrium: it is a dominant strategy. Signatory countries know this, and each signatory country abates as long as the benefit it would get from emitting one unit of pollution Eq. 1 is less than the damage that would cause to all signatory countries ($n\gamma$).

There are two ways of defining the equilibrium of the membership game. The first, as presented in Barrett (1994), borrows the concept of a *stable coalition* from the literature on oligopoly and defines a *stable IEA* as follows:

Definition 2. An IEA with n signatories is *stable* if it satisfies the two conditions:

$$\text{Internal stability}: \quad V^s(n) > V^f(n-1) \tag{4a}$$

$$\text{External stability}: \quad V^f(n) > V^s(n+1) \tag{4b}$$

i.e. no signatory country has any incentive to unilaterally leave the IEA, and no non-signatory has any incentive to unilaterally join the IEA, taking as given the membership decisions of all other countries. This definition is equivalent to saying that a stable IEA is a Nash equilibrium of the membership game.

This might seem to be a simplistic notion of stability. In fact, other authors have come up with more complex conditions for stability of an IEA (e.g., Chander and Tulkens 1994), including ways of committing countries to participate (e.g., Carraro and Siniscalco 1993). What this definition of stability gives us is a very basic, and perhaps weakest, concept of what it takes to hold an agreement together. It is a good starting point for explorations of the size of voluntary international environmental agreements.

The following is a basic result on the size of an IEA:

Lemma 4 (membership game equilibrium) The unique stable IEA of the membership game has $n^* = I(\gamma)$ signatory countries with aggregate world net benefit of

$$W(\gamma) = \left(N - n^*\right)(1 - N\gamma) < 0. \tag{5}$$

The intuition is straightforward. No IEA with $n > n^*$ signatories is internally stable. In such a case, there is an incentive for a signatory country to quit the IEA, because if it does so there will still be at least n^* counties in the IEA and thus the remaining members will continue to abate. Consequently, by quitting it gains 1 in benefit and loses γ in additional damage costs, and so the gain outweighs the cost. However when there are only n^* signatories, if one signatory country leaves the IEA the remaining countries will choose to pollute; thus the defector will gain 1 and lose $n^*\gamma$, and by definition of n^* the loss outweighs the gain. So n^* is a stable IEA. At n^*, every country in the IEA is pivotal and that is what keeps the IEA together—if any country in the IEA leaves, all of the IEA members stop abating. For all $n<n^*$, signatories will pollute and all countries get the non-cooperative payoff, which is the same for all $n<n^*$. By the definition of stability, no such IEA can be internally or externally stable. So n^* is the unique stable IEA.

Aggregate world net benefits from the stable IEA, $W(\gamma)$, lie between the aggregate net benefits of the non-cooperative and cooperative equilibria (from Lemmas 1 and 2). We define three measures of gain: the *full gain from cooperation* (FGC), defined as the difference between aggregate net benefits in the cooperative and non-cooperative equilibrium; the *absolute partial gain from cooperation* (APGC), defined as the difference in aggregate net benefits between the net benefits of the stable IEA and the net benefits in the non-cooperative equilibrium; and the *relative partial gain in cooperation* (RPGC), defined as APGC relative to FGC, i.e. a measure of how much of the maximum potential gains in cooperation are realised by the IEA:

$$\mathrm{FGC} = N(N\gamma - 1), \tag{6a}$$

$$\mathrm{APGC} = n^*(N\gamma - 1) \approx N - n^* \tag{6b}$$

$$\mathrm{RPGC} = n^*/N, \tag{6c}$$

Note that while FCC and APGC are increasing in γ, since n^* is decreasing in γ,[7] RPGC is decreasing in γ. In other words, the more an environmental problem *needs* an IEA to be solved (higher γ), and hence the greater are potential gains from cooperation, the lower is the fraction of those potential gains that an IEA will achieve—a discouraging result.

To complete this section it is worth noting the properties of $I(\gamma)$ and $W(\gamma)$. If we approximate $I(\gamma)\approx 1/\gamma$, and thus assume it is differentiable, then it is straightforward to see that $I(\gamma)$ is a decreasing and convex function, and $W(\gamma)$ is a decreasing and concave function.[8] Of course, more precisely, $I(\gamma)$ is an integer function, and not differentiable, and so is not strictly a convex function, nor is $W(\gamma)$ strictly concave. For the purposes we need for this paper we shall assume that we can treat $I(\gamma)$, $W(\gamma)$ as, respectively, convex and concave.

The concavity of $W(\gamma)$ is a key property of our model leading to the negative value of information, so it is worth emphasizing what lies behind this result. It be seen from Eq. 5 that if, as γ increases, the number of signatories and hence the level of aggregate emissions, stayed constant, then aggregate welfare would be a decreasing linear function of γ. In that case complete information would have zero value. One might expect that as γ increased the number of signatories would increase, or at least the aggregate level of emissions would fall, and this would somewhat offset the effect of the increase in γ, making $W(\gamma)$ a

---

[7] Approximation assumes that $n^*\gamma = \gamma I(\gamma) \approx 1$.

[8] $I' = -1/\gamma^2$; $I'' = 2/\gamma^3$; $W = 2N - N^2\gamma - 1/\gamma$; $W' = (1/\gamma)^2 - N^2 = n^{*2} - N^2 < 0$; $W'' = -2/\gamma^3$.

decreasing convex function of γ. But in our model, emissions by signatories and non-signatories are fixed, and as γ increases the number of signatories falls, and this exacerbates the reduction in welfare caused by the increase in γ, making $W(\gamma)$ a decreasing concave function.

5 Uncertainty and learning in IEAs

In this section we introduce uncertainty in the benefit–cost ratio (γ) into our model and then superimpose several types of learning on the model. Recall that the nature of uncertainty that we consider is that there is a single shared variable over which there is uncertainty. Specifically, there is uncertainty about γ and all countries realize the same value of γ.

5.1 Introducing uncertainty and learning

5.1.1 Uncertainty

To introduce uncertainty, we now assume that at the outset, for each country the true value of the parameter γ is unknown. With probability p, γ takes the value γ_l and with probability $1-p$, γ takes the value γ_h. Whichever the realization of γ, all countries share the same value. Furthermore,

$$\text{Assumption 2}:\ 1/N < \gamma_l < \gamma_h < 1. \tag{6}$$

This assumption is simply a variant on Assumption 1 (Eq. 2). We now define:

$$\overline{\gamma} = p\gamma_l + (1-p)\gamma_h \tag{7}$$

as the *expected* damage cost for each country. Note that Assumption 2 implies Assumption 1.

Define $n_l^* \equiv I(\gamma_l)$, $n^* = I(\overline{\gamma})$, $n_h^* \equiv I(\gamma_h)$, $W_l^* \equiv W(\gamma_l)$, $W^* = W(\overline{\gamma})$, $W_h^* = W(\gamma_h)$. These are the membership levels and aggregate welfare levels that would obtain if we knew for certain that γ was equal to its low value, expected value and high value respectively. We want to ensure that there is a significant amount of uncertainty so that γ_l, $\overline{\gamma}$, $\gamma_h$are sufficiently distinct from each other, in the sense that:

$$\text{Assumption 3}:\ n_l^* > n^* > n_h^*; W_l^* > W^* > W_h^*. \tag{8}$$

That is if we knew for certain that the true value of γ was the low value, the expected value, or the high value, then, applying Lemma 4, the resulting stable IEAs would have distinctly different membership sizes and different aggregate welfare levels: uncertainty matters.

5.1.2 Learning

To introduce the possibility of learning we assume that learning takes the form of perfect learning—i.e. the true state of the world is revealed to all countries. Again this is a very special model of learning. The crucial issue is the timing at which such information becomes available, and we define three possible cases:

No learning the true state of the world is revealed to all countries after all their decisions have been taken.

Complete learning The true state of the world is revealed to all countries before any decision is taken.

Partial learning The true state of the world is revealed to all countries after they have taken their decisions on whether or not to join an IEA (countries are thus committed to their membership decisions) but before they choose their emissions.

By Assumption 2 and using the arguments in Lemmas 1 and 2, we immediately derive the following:

Proposition 1 Uncertainty and learning (complete, partial or none) have no effect on the cooperative and non-cooperative equilibria. There are no gains from learning in these equilibria.

In the non-cooperative equilibrium, all countries always pollute both in the case of No Learning and in the case of Complete Learning, no matter what true state of the world is revealed, and so expected aggregate net benefits are $N(1-N\overline{\gamma})$. Similarly, in the co-operative equilibrium all countries always abate, for all types of learning, no matter what true state of the world is revealed, and so expected aggregate net benefits again are 0. Note that this result is in stark contrast to those in Ulph and Ulph (1996), Ulph and Maddison (1997), and Baker (2005), where learning does affect the outcome of both the non-cooperative and cooperative equilibria, and where the degree of correlation between damage costs in different countries plays an important role. The reason is that our simplifying assumptions of a dichotomous choice of emission levels by countries combined with Assumption 2 means that, as noted, the choice of emissions by countries is unaffected by uncertainty and the resolution of that uncertainty. Since behaviour is unaffected, there are no benefits to be had from learning.[9]

However uncertainty and learning do have important effects on outcomes with IEA formation. This implies that our model of uncertainty and learning has successfully isolated effects of uncertainty and learning that work solely through the structure of the IEA games.

5.2 The analysis of uncertainty, learning and IEAs

We analyse in turn the three cases of No Learning, Partial Learning and Complete Learning.

5.2.1 Uncertainty with no learning

Given that the net benefit function is linear in γ, certainty equivalence applies, which means that when countries take their decisions in the membership game and the emissions game using expected net benefits, they act as if they know for certain that γ is equal to its expected value, $\overline{\gamma}$. Therefore the outcomes are equivalent to those that obtained in the model of certainty in Section 3, with γ now equal to $\overline{\gamma}$, the expected unit damage cost. Thus from Lemma 4 we have:

Proposition 2 With No Learning the unique stable IEA has $n^{*}=I(\overline{\gamma})$ signatories and aggregate world net benefits are $W^{*}=\left(N-n^{*}\right)(1-\overline{\gamma}N)<0$.

---

[9] These results, while straightforward, were not included in either Ulph (2004) or Kolstad (2007).

5.2.2 Uncertainty with complete learning

In this case the true state of the world is revealed before countries decide whether or not to join an IEA and then what emission levels to set. So if the state of the world revealed is that all countries have low (high) damage costs, then all countries know for certain that damage costs are low (high), and we can apply the certainty analysis of Lemma 4 to argue that the unique stable IEA in that state of the world will have $n_l^*\left(n_h^*\right)$ members and aggregate welfare will be $W_l^*\left(W_h^*\right)$. Using Assumption 3, it immediately follows that:

Proposition 3 With Uncertainty and Complete Learning, the membership and aggregate world net benefits of the unique stable IEA in the low and high damage cost state are: (n_l^*, W_l^*), (n_h^*, W_h^*) respectively, where, by Assumption 3, $n_l^* > n^* > n_h^*$ and $W_l^* > W^* > W_h^*$. Expected membership is: $\overline{n} \equiv pn_l^* + (1-p)n_h^*$ and expected aggregate world net benefits are: $\overline{W} \equiv pW_l^* + (1-p)W_h^*$.

We stated at the end of section 2 that we shall treat the function $I(\gamma)$ as if it is convex, and $W(\gamma)$, as if it is concave, which is likely to be justifiable given our Assumption 3 that uncertainty is sufficiently great. So we immediately conclude that $\overline{n} > n^*, \overline{W} < W^*$. Hence:

Corollary 1 Uncertainty with complete learning leads to higher expected membership of IEA but lower expected aggregate world net benefits than the case of uncertainty with no learning.

The result on expected membership was derived in Kolstad (2007) and on expected aggregate net benefits in Ulph (2004). Thus we have shown that with complete learning information has negative value. The intuition for this follows from the intuition we gave earlier for the fact that aggregate world welfare with an IEA is a decreasing concave function of unit damages: if the number of signatories of an IEA stayed constant as unit damage costs increased, and hence the overall level of emissions stayed the same as the unit damage costs increased, then aggregate world welfare would be a decreasing *linear* function of unit damage costs, for the simple reason that we assume damage costs are a linear function of the unit damage cost parameter. But there is the further indirect effect that as unit damage costs increases, the number of signatories of an IEA decreases, and hence aggregate emissions increase. This reinforces the direct effect of an increase in unit damage costs to make welfare a decreasing and concave function of unit damage costs. Thus expected welfare when the true value of unit damage costs is revealed prior to the membership game is less than welfare when membership and emissions are based on the expected level of unit damage costs.

5.2.3 Uncertainty with partial learning

We now suppose that the true state of the world is revealed after countries decide whether to join an IEA but before they choose their emission strategies, so it is now possible for countries to condition their emissions on the state of the world.

By Assumption 2 and Lemma 3, it is straightforward to derive:

Lemma 5 (emission game—partial learning) With Uncertainty and Partial Learning, non-signatory countries always pollute no matter what the true state of the world or the number of signatories. The emission strategies of signatories, and expected net benefits of signatories and non-signatories are as follows:

a. for $n \geq n_l^*$, signatories abate in both states of the world and expected net benefits are: $V^s(n) = -(N-n)\overline{\gamma},\ \ V^f(n) = 1-(N-n)\overline{\gamma}$;
b. for $n < n_h^*$, signatories pollute in both states of the world, and expected net benefits are: $V^s(n) = V^f(n) = 1 - N\gamma$;
c. for $n_h^* \leq n < n_l^*$, signatories abate in the high damage cost state of the world and pollute in the low damage cost state, and expected net benefits are: $V^s(n) = -N\overline{\gamma} + p + n(1-p)\gamma_h$, $V^f(n) = -N\overline{\gamma} + 1 + n(1-p)\gamma_h$.

By the arguments given for Lemma 4, it is straightforward to see that no IEA with membership strictly greater than n_l^* could be stable (each signatory country has an incentive to quit), and no IEA with membership strictly less than n_h^* could be stable (members have no incentive to stay in the agreement). It is also straightforward to see that no IEA with membership lying strictly between n_h^* and n_l^* could be stable. This is because signatory countries have a reason to quit for the same reason as before. If a signatory county quits, it knows that the remaining signatories will continue to abate in the high damage cost state, so by quitting it gains the benefit of polluting in the high damage costs state (a benefit of 1), but only adds one unit of pollution in that state, costing γ_h, and by Assumption 2, the gain outweighs the loss. It is only when membership drops to n_h^* members do we get stability, because if one signatory left the IEA, then the loss would be that all signatories would now pollute in the high damage cost state, and by definition $n_h^* \gamma_h \geq 1$, so there is no gain to quitting.

This leaves the case of an IEA with membership n_l^*. Signatories abate in all states, so if one signatory left, it would get the benefit of polluting in all states, an expected benefit of 1. However in the low damage cost state all the previous signatories will now pollute and in the high damage cost state it will now pollute, so the expected cost of defecting is $pn_l^* \gamma_l + (1-p)\gamma_h$. Since $n_l^* \gamma_l \geq 1$, for large enough values of p the expected damage cost of quitting the IEA could exceed the benefit of quitting, and so an IEA with n_l^* members would also be stable. To be more precise, define $\varepsilon \equiv n_l^* - \frac{1}{\gamma_l}$, the difference between $\frac{1}{\gamma_l}$ and the smallest integer not less than $\frac{1}{\gamma_l}$, so obviously $0 \leq \varepsilon < 1$; and define

$$\overline{p} \equiv \frac{1-\gamma_h}{1-\gamma_h+\varepsilon\gamma_l}. \tag{9}$$

Then we have:

Proposition 4 Assume Uncertainty and Partial Learning and $\overline{p}$ defined in Eq. 9. If $p < \overline{p}$, then there is a unique stable IEA with membership n_h^* and aggregate world expected net benefit: $\hat{W}_h = p[N(1-\gamma_1 N)] + (1-p)\left[\left(N - n_h^*\right)(1-\gamma_h N)\right]$; if $p \geq \overline{p}$ there is a second stable IEA with n_h^* members and aggregate world expected net benefits: $\hat{W}_l = \left(N - n_l^*\right)(1-\overline{\gamma}N)$.

It is straightforward to see that $n_l^* > \overline{n} > n^* > n_h^*$, and that $W_l^* > \hat{W}_l > W^*$ A little manipulation shows that $\hat{W}_h = \overline{W} - pn_l^*(N\gamma_l - 1)$ so that by Assumption 2, $\hat{W}_h < \overline{W}$. Putting these together we get: $\hat{W}_h < \overline{W} < W^* < \hat{W}_l < W_l^*$.

Corollary 2 With Uncertainty and partial learning, for all values of p there is a stable IEA with membership and expected aggregate net benefits lower than uncertainty with either complete learning or no learning. For sufficiently high values of $p(\geq \overline{p})$), there is a second stable IEA with membership and expected aggregate net benefits greater than either complete learning or no learning.

Assumption 3 rules out the possibility of extremely high or extremely low values of p, in which case it is likely that we can rule out the second stable IEA with Partial Resolution. Again the result on membership can be found in Kolstad (2007) while the implications for welfare are new to this paper.

We can summarise the results as follows: Uncertainty with Complete Learning leads to higher expected membership but lower expected aggregate net benefits than No Learning, while Partial Learning almost certainly leads to lower membership and even lower expected aggregate net benefits. The simple intuition why partial learning is (almost certainly) worse than complete learning is that the fact that signatories will continue to abate in the high damage cost state when membership falls below n_l^* weakens the effect of defection by a signatory country when there are n_l^* signatories, and (almost certainly) rules this out as possible stable IEA, just leaving the IEA equilibrium with a low membership, lower than membership with expected damage costs. The resolution of uncertainty, whether partial or full, reduces aggregate net benefits relative to no resolution of uncertainty.

6 Conclusions

In this paper we have synthesised and extended our earlier work on the formation of international environmental agreements under uncertainty and different models of learning. In particular the analysis in this paper has elaborated the welfare implications of different models of learning, rather than just focusing on IEA membership. Our results are generally pessimistic—in almost all cases learning, whether partial or complete, reduces the level of welfare that can be achieved by forming an IEA. This suggests that information can have negative value. This may seem strange, since for a single decision-maker information cannot have negative value, because it can always be ignored. However in this case there are strategic interactions between a number of decision-makers responding to information, and it is these strategic interactions which can give information a negative value.

Of course the analysis is extremely simple. In a companion paper (Kolstad and Ulph 2007) we extend the analysis in this paper for the case of perfect correlation between damage costs for different countries, to no correlation between damage costs, but the pessimistic results remain. It would be desirable to test the robustness of these results using richer models of coalition formation (see Bloch (1997) and Finus (2001) for discussions of such extensions, and Dutta and Radner (2004) for a dynamic model of a non-cooperative equilibrium where the dynamics allow a richer set of strategies to be employed, resulting in some equilibria which make all countries better off than in the simple static non-cooperative model used in much of the literature on IEAs and climate change). It would also be important to apply this analysis to empirical models of climate change, and it is worth noting that in a first attempt to do so Dellink et al. (2006) find some limited support for our finding of a negative value of information with complete learning.

Acknowledgements We are grateful to Brian O'Neill, two anonymous referees and participants for comments which have improved this paper.

References

Arrow K, Fisher A (1974) Environmental preservation, uncertainty and irreversibility. Q J Econ 88:312–319

Baker E (2005) Uncertainty and learning in a strategic environment: global climate change. Resour Energy Econ 27:19–40

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:878–894

Barrett S (2002) Environment and statecraft. Oxford University Press, Oxford

Bloch F (1997) Non-cooperative models of coalition formation in games with spillovers. In: Carraro C, Siniscalco D (eds) New directions in the economic theory of the environment. Cambridge University Press, Cambridge, pp 311–352

Bohringer C, Finus M (2005) An economic analysis of the Kyoto Agreement. In: Helm D (ed) Climate change and policy response. Oxford University Press, Oxford

Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Public Econ 52:309–328

Chander P, Tulkens H (1994) A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. Int Tax Public Financ 2:279–293

Cooper R (1989) International cooperation in public health as a prologue to macroeconomic cooperation. In: Cooper RN et al (ed) Can nations agree. Brookings Institution, Washington, DC, pp 187–254

Dellink R, Finus M, Olieman N (2006) The stability likelihood of an international environmental agreement. Unpublished manuscript, Wageningen University, The Netherlands

Dutta P, Radner R (2004) Self-enforcing climate-change treaties. Proc Natl Acad Sci U S A 101:5174–5179

Epstein L (1980) Decision-making and the temporal resolution of uncertainty. Int Econ Rev 21:269–284

Finus M (2001) Game theory and international environmental cooperation. Edward Elgar, Cheltenham, UK

Fisher A, Hanemann M, Narain U (2002) The irreversibility effect: necessary and sufficient conditions. In: van Ierland E, Weikard H, Wesseler J (eds) Risk and uncertainty in environmental and resource economics. Environmental Economics and Natural Resources Group, Wageningen University, The Netherlands

Foster K, Vecchia P, Repacholi M (2000) Science and the precautionary principle. Science 288:979–981

Gollier C, Jullien B, Treich N (2000) Scientific progress and irreversibility: an economic interpretation of the ‘precautionary principle’. J Public Econ 75:229–253

Ha-Duong M (1998) Quasi-option value and climate change policy choices. Energy Econ 20:599–620

Henry C (1974a) Investment decisions under uncertainty: the irreversibility effect. Am Econ Rev 64:1006–1012

Henry C (1974b) Option values in economics of irreplaceable assets. Rev Econ Stud 41(S):89–104

Hirshleifer J (1971) The private and social value of information and the reward to inventive activity. Am Econ Rev 61:561–574

Ingham A, Ulph A (2005) Uncertainty, irreversibility, precaution, and the social cost of carbon. In: Helm D. Climate change and policy response. Oxford University Press, Oxford

Ingham A., Ma J, Ulph A (2005) Climate change, mitigation, and adaptation with uncertainty and learning. Paper presented at the Symposium NCCR CLIMATE 2005 Interfaces Between Climate and Economic Dynamics, Interlaken, 3–4 March 2005

Kolstad C (1996a) Fundamental irreversibilities in stock externalities. J Public Econ 60:221–233

Kolstad C (1996b) Learning and stock effects in environmental regulations: the case of greenhouse gas emissions. J Environ Econ Manage 31:1–18

Kolstad C (2007) Systematic uncertainty in self-enforcing international environmental agreements. J Environ Econ Manage 53(1):68–79

Kolstad C, Ulph A (2007) International environmental agreements under uncertainty when learning reveals differences between countries. Unpublished manuscript, University of Manchester, Manchester, England

Lange A (2003) Climate change and the irreversibility effect—combining expected utility and MaxiMin. Environ Resour Econ 25:417–434

Lange A, Treich N (2008) Uncertainty, learning and ambiguity in economic models on climate policy: some classical results and new directions. Climatic Change, DOI 10.1007/s10584-008-9401-5

Na S-L, Shin HS (1998) International environmental agreements under uncertainty. Oxf Econ Pap 50:173–185

Rubio S, Ulph A (2006) Self-enforcing international environmental agreements revisited. Oxf Econ Pap 58:233–263

Steinberg R, Zangwill W (1983) The prevalence of Braess’ paradox. Transp Sci 17:301–318

Ulph A (2004) Stable international environmental agreements with a stock pollutant, uncertainty and learning. J Risk Uncertain 29(1):53–73

Ulph A, Maddison D (1997) Uncertainty, learning and international environmental policy coordination. Environ Resour Econ 9:451–466
Ulph A, Ulph D (1996) Who gains from learning about global warming. In: Ierland EV, Gorka K (eds) The economics of atmospheric pollution. Springer, Heidelberg, pp 31–67
Ulph A, Ulph D (1997) Global warming, irreversibility and learning. Econ J 107:636–650
Webster M (2002) The curious role of learning: should we wait for more data. Energy J 23:97–119
Young O (1994) International governance: protecting the environment in a stateless society. Cornell University Press, Ithaca, NY