The Role of Uncertainty and Learning for the Success of International Climate Agreements

Michael Finus*

and

Pedro Pintassilgo**

Abstract

Transnational externalities (e.g. transboundary pollution, trade, contagious diseases and terrorism) warrant coordination and cooperation between governments, but this proves often difficult. One reason for meager success is the public good character of many of these economic problems, encouraging free-riding. Another reason one might suspect is uncertainty, surrounding most environmental problems, and in particular climate change. This provides often an excuse for remaining inactive. Paradoxically, some recent papers have concluded just the opposite: the "veil of uncertainty" can be conducive to the success of international environmental cooperation. In this paper, we explain why and under which conditions this can be true. However, we argue that those conditions are rather the exception than the rule. Most important, we suggest a mechanism for those conditions where learning has a negative effect on the success of cooperation which removes this effect or even turns it into a positive effect. Our results apply beyond the specifics of climate change to similar problems where cooperation generates positive externalities.

Keywords: voluntary provision of public goods, transnational cooperation, self-enforcing agreements, uncertainty, learning

JEL-Classification: C72, D62, D81, H41, Q20.

* Corresponding Author: Department of Economics, University of Bath, Bath, BA2 7AY, UK; Tel: #44-1225386228, Fax: #44-1225383423, E-mail: m.finus@bath.ac.uk.

** Faculty of Economics and Research Centre for Spatial and Organizational Dynamics, University of Algarve, Faro, Portugal.

1. Introduction

Transnational externalities (e.g. transboundary pollution, trade, contagious diseases and terrorism; see e.g. Sandler 2004 and Yi 1996) warrant coordination and cooperation between governments, but this proves often difficult. One of the greatest challenges to international co-operation and public good provision the world is currently facing is climate change, as emphasized by the two prominent studies, the Stern and the IPCC Reports (Stern 2006 and IPCC 2007). International response to global warming can be traced back to 1988 when the IPCC was founded – an international body that gathers and synthesizes current world-wide scientific evidence on climate change. However, it was not until 1997 that 38 countries agreed to specific emission ceilings under the Kyoto Protocol, which was not ratified before 2002, after several concessions had been granted to various participants and after the US had declared that it would withdraw from the treaty alto-gether. Currently, a "Post-Kyoto" agreement for the period after 2012 is being negotiated, with the aim to tighten emission ceilings, encourage the participation of the US and the "emerging" polluters China and India.

One important problem for effective cooperation is free-riding. For a common property resource this is well-known since Hardin (1968) and has been reiterated in the specific context of self-enforcing international environmental agreements (SEIEAs) by Barrett (1994), Carraro and Siniscalco (1993) and Hoel (1992). Later papers, using richer models, either with an empirical (e.g. Bosello et al. 2003, Finus and Tjøtta 2003 and Weikard et al. 2006) or theoretical (e.g. Asheim et al. 2006, Barrett 2001, 2006, Rubio and Ulph 2007 and Ulph 2004) focus, have suggested some possibilities to make SEIEAs more effective, but have confirmed the general negative conclusion more or less.[1]

Another important problem is the large uncertainty surrounding environmental damages caused by greenhouse gases. Predictions about mitigation costs are also difficult (IPCC 2007 and Stern 2006). For instance, the former US President George Bush used uncertainty as one argument for his decision to withdraw from the Kyoto Protocol. In a letter to Senators, dated March 13, 2001, as quoted by Kolstad (2007), he wrote: "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".

[1] Surveys are provided in Barrett (2003) and Finus (2003).

Recently, a literature has emerged, which combines free-riding with uncertainty and learning, using simple SEIEA-models with a static payoff function[2] (Kolstad 2007, Kolstad and Ulph 2008, 2011 and Na and Shin 1998). Their main conclusion is that in the strategic context of the formation of climate agreements, learning leads to worse outcomes (measured in terms of the aggregate payoff and public good provision level) than no learning. This "negative" result about the role of learning for the success of SEIEAs, though interesting, is intriguing as it runs counter to increased research efforts on climate change world-wide. This motivates the three research questions posed in this paper. What are the driving forces to generate this result? How general is this conclusion? Is there a way to fix the problem? The following paragraphs provide a brief introduction to how we address these questions.

What are the driving forces to generate this result?

In order to understand the driving forces it is important to note that the "standard" SEIEA-model assumes a two-stage coalition formation game. In the first stage, countries choose their membership and in the second stage their contributions to the public good. The game is solved backwards. In the second stage, non-members derive their contributions from maximizing their own payoffs but coalition members from maximizing the aggregate payoff over all their members. Hence, in this model, the social optimum is reproduced if all countries join the agreement (grand coalition) and the Nash equilibrium if all countries remain singletons (singletons coalition structure). Intermediate cases of cooperation are also captured, with some, though not all, countries joining the agreement. In the first stage, stable coalitions are determined based on the concept of internal and external stability.

In this paper, we suggest a three step procedure in order to understand how uncertainty and learning affects the success of SEIEAs. In the first step, we look at stable coalitions (their sizes and membership) with and without learning in the first stage. In the second step, we consider a generic coalition and compare total provision levels and payoffs with and without learning in the second stage. In the third step, we combine steps 1 and 2 to derive overall conclusions. In many cases, this allows already for immediate conclusions.

We say that the *second stage effect from learning* is positive (negative) if the total payoff with learning is larger (smaller) than without learning for any possible coalition that has formed

[2] That is, it captures the public bad nature of greenhouse gases but not their dynamics as stock pollutants.

in the first stage. For the understanding of this second stage effect two benchmarks are useful. The first benchmark is the grand coalition. When it forms, there is no strategic interaction, between the coalition and non-members, and the second stage effect from learning is positive, and in most cases strictly positive. This is in line with the general wisdom that "learning is good". The second benchmark is the singletons coalition structure corresponding to the Nash equilibrium. Now there is strategic interaction between players and the value of information can become negative as shown for instance in Gollier and Treich (2003) for three economic examples. For instance, suppose as in Ulph (1998) that all players have ex-ante symmetric expectations but turn out to be asymmetric ex-post. Assume that asymmetry means only different marginal benefits from the total public good provision but symmetric marginal costs from individual public good contributions. Then equilibrium contributions with learning are asymmetric but are symmetric without learning. Both equilibria are inefficient (each player sets marginal cost only equal to own but not the sum of marginal benefits) but the equilibrium under no learning is at least cost-effective (i.e. marginal contribution costs equalize). Thus, with strategic interaction, the second stage effect from learning can be negative. In our coalition model, which captures not only the grand coalition and the singletons coalition structure but also intermediate cases of cooperation, such a negative second stage effect from learning occurs if there is pure uncertainty about the distribution of the benefits from public good provision.

By the same token, we say that the *first stage effect from learning* is positive (negative) if the size of stable coalitions is larger (smaller) with than without learning. Again, the example from above is useful to illustrate this point. With learning those players with high marginal benefits gain more than proportionally from cooperation and those with low marginal benefits less than proportionally. If differences are sufficiently pronounced, no or only small coalitions will be stable.[3] Without learning, and ex-ante symmetric expectations, the distribution is symmetric and larger coalitions are stable. The intuition of this phenomenon is along the lines of Young (1994), borrowing the concept of the veil of uncertainty from Brennan and Buchanan (1985), who argues that agreements are easier if potential participants do not know the distributional consequences.

[3] This idea is also illustrated in a simple two-player model in Helm (1998) and in Kolstad (2005).

How general is this conclusion?

A natural approach to answer this question is to set up a model that takes the best ingredients of existing models and generalizes them in several ways.

First, we consider the *three learning scenarios* full, partial and no learning as proposed in Kolstad (2007) and Kolstad and Ulph (2008, 2011) where learning means to learn the values of the uncertain parameters of the payoff functions. The additional scenario of partial learning compared to Na and Shin (1998) turns out to be useful when breaking down the effects of learning between both stages.

Second, we consider a *strictly concave payoff function* which we take from Na and Shin (1998). Compared to the papers by Kolstad and Ulph, which assume a linear payoff function, this avoids boundary solutions in the second stage and allows to derive clear-cut second stage effects from uncertainty and learning.

Third, we consider *three types of uncertainty*. Taking a broader view, we argue that the models by Kolstad (2007) and Kolstad and Ulph (2008) essentially capture *pure uncertainty about the level of the benefits* from the public good provision and the model by Na and Shin (1998) *pure uncertainty about the distribution of these benefits*. We find these benchmarks useful for analytical clarity but consider also the more realistic case with *simultaneous uncertainty about the level and the distribution of these benefits*. Moreover, we neither restrict attention to only three players as in Na and Shin (1998), nor do we assume only two values of the random benefit variables as in Kolstad (2007) and Kolstad and Ulph (2008, 2011), but allow for any possible values of the random variables.

In this more general setting, we derive sufficient conditions when learning has a positive impact on the success of SEIEAs, when it has a negative impact and also point out the circumstances under which conclusions are not clear-cut. It will become apparent that those cases where learning has a negative impact on the success of SEIEAs are rather the exception than the rule.

Is there a way to fix the problem?

If the first stage effect from learning is negative because of an unequal distribution of the gains from cooperation among coalition members, then we show that an appropriate transfer mechanism can fix this problem. In fact, even larger coalitions than for symmetry can be stabilized because the relative benefit from participating in a coalition compared to free-riding increases.

In the following, we outline the general setting of our coalition model in section 2. Section 3 derives the model solutions for stages 1 and 2 and section 4 compares results. Section 5 summarizes the main conclusions and proposes some issues for future research.

2. Model Outline

2.1 Coalition Formation Game

2.1.1 General Remarks

International environmental agreements are typical agreements where countries have to decide whether they are in or out, participation is voluntary and membership is open, i.e. a country can neither be forced into nor excluded from participation. Therefore, we model coalition formation as a two-stage open membership single coalition game. In the first stage, players (i.e. countries in our context) decide whether to join an agreement (i.e. a climate treaty in our context) or to remain outsiders as singletons. In the second stage, players choose their policy levels (i.e. contributions to the public good which we call abatement in our context). The game is solved backward, assuming that strategies in each stage must form a Nash equilibrium.

This game has also been called cartel formation game with non-members called fringe players. It originates from the literature in industrial organization (d' Aspremont et al. 1983) and has been applied widely in this literature (e.g. Deneckere and Davidson 1985, Donsimoni et al. 1986 and Poyago-Theotokay 1995; see Bloch 2003 and Yi 1997 for surveys) but also in the literature on self-enforcing international environmental agreements (e.g. Barrett 1994, Carraro and Siniscalco 1993 and Rubio and Ulph 2007; see Barrett 2003 and Finus 2003 for surveys).

In the first stage, players' membership decisions lead to a coalition structure, $K = \{S, 1_{n-m}\}$, which is a partition of players, with n being the total number of players, S the coalition of size m, $m \leq n$, 1_{n-m} the $n-m$ singletons, and N the set of players, $S \subseteq N$. Because coalition formation is only interesting if there are at least three players, we assume subsequently $n \geq 3$. Due to the simple structure of this coalition formation game, i.e. there can be at most one non-trivial coalition (i.e. all players that do not belong to S are singletons with a non-trivial coalition being a coalition of at least two players), coalition structure K is entirely determined by coalition S. We will sometimes call the members of S signatories and the non-members of S non-signatories.

In the second stage, given that some coalition S has formed, players choose their abatement levels q_i. The decision is based on the following payoff function:[4]

$$(1) \qquad \Pi_i = B_i\left(\sum_{k=1}^{n} q_k\right) - C_i\left(q_i\right) \ , \ i \in N$$

where $B_i(\bullet)$ is country i's concave benefit function from global abatement (in the form of reduced damages, e.g. measured against some business-as-usual-scenario) and $C_i(\bullet)$ its convex abatement cost function from individual abatement. The global public good nature of abatement is captured by the benefit function which depends on the sum of all abatement contributions. For a start, we assume that all functions and their parameters are common knowledge and introduce uncertainty later.

Working backward, we assume that the equilibrium strategy vector in the second stage constitutes a (coalitional) Nash equilibrium between coalition S with its m members and the $n-m$ singletons. Specifically, the coalition acts de facto as a single or meta player (Haeringer 2004), maximizing the aggregate payoff of its members

$$(2) \qquad \underset{q^S}{max}. \sum_{i \in S} \Pi_i\left(S\right) \Rightarrow \sum_{i \in S} B_i^{'}\left(\sum_{k=1}^{n} q_k\right) = C_i^{'}\left(q_i\right) \ \ \forall\, i \in S$$

whereas as all singletons maximize their own payoffs

$$(3) \qquad \underset{q_i}{max}. \Pi_i\left(S\right) \Rightarrow B_i^{'}\left(\sum_{k=1}^{n} q_k\right) = C_i^{'}\left(q_i\right) \ \ \forall\, i \notin S$$

where q^S in (2) is the abatement vector of those players that belong to coalition S, $B_k^{'}$ and $C_k^{'}$ are the derivatives of B_k and C_k with respect to q_k, respectively. The simultaneous solution of the first order conditions (F.O.C.s) in (2) and (3) delivers equilibrium abatement levels $q_{i \in S}^{*}(S)$ of signatories and $q_{i \notin S}^{*}(S)$ of non-signatories.[5] The F.O.C.s in (2) are

---

4 An alternative specification of payoff functions, comprising damage cost functions from global emissions and benefit functions from individual emissions, produces equivalent results. This equivalence holds as long as non-negativity constraints are observed, as discussed for instance in Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006).

5 The simultaneous choice in the second stage has been called Nash-Cournot assumption in the literature. An alternative assumption could be a sequential choice which has been called Stackelberg assumption. In the context of coalition formation the latter assumption is not innocuous for several reasons. For the specific type of payoff functions which we introduce in section 2.3, both assumptions are equivalent. For a discussion see Finus (2003).

the Samuelson conditions for the optimal provision of a public good, though they hold only for coalition members; the F.O.C.s in (3) are those in a non-cooperative equilibrium.[6] Substituting the equilibrium abatement levels for a given coalition S into payoff functions (1) delivers the payoffs of signatories, $\Pi^*_{i\in S}(S)$, and non-signatories, $\Pi^*_{i\notin S}(S)$, in the second stage of the coalition formation game.

In the first stage, stable coalitions are determined by invoking the concept of internal and external stability, which is de facto a Nash equilibrium in membership strategies.

(4) internal stability: $\Pi_i^*(S) \geq \Pi_i^*(S \setminus \{i\}) \;\; \forall\, i \in S$

(5) external stability: $\Pi_i^*(S) > \Pi_i^*(S \cup \{i\}) \;\; \forall\, i \notin S$.

That is, no signatory should have an incentive to leave coalition S to become a non-signatory and no non-signatory should have an incentive to join coalition S. In order to avoid knife-edge cases, we assume that if players are indifferent between joining coalition S and remaining outside, they will join the agreement. Coalitions which are internally and externally stable are called stable.

Note that the singletons coalition structure is stable by definition and hence existence of a stable coalition is guaranteed: if all players announce not to join the agreement, then a single deviation by one player will make no difference. If there is more than one stable coalition, we apply the Pareto-dominance criterion and delete all those coalitions which are Pareto-dominated by other stable coalitions. In our model it is easy to show that the singletons coalition structure is Pareto-dominated by any other coalition.

2.1.2 Symmetry, Asymmetry and Transfers

In the case of ex-ante symmetric players, i.e., all players have the same payoff function, things simplify: payoffs among the group of signatories are identical and the same applies to the group of non-signatories. Hence, we can talk about a coalition of size m being stable, with the understanding that all combinations of m players form a stable coalition. Moreover, internal stability can be written as $\Pi_i^{*M}(m) \geq \Pi_i^{*NM}(m-1)$ and external stability as $\Pi_i^{*NM}(m) > \Pi_i^{*M}(m+1)$ with the superscript "M" standing for "Member" and the superscript "NM" standing for "Non-Member". This simplification emphasizes that if a

[6] Note that if $S = N$ (i.e. all players form the grand coalition) the equilibrium abatement vector corresponds to the social optimum and if either $S = \{i\}$ or $S = \varnothing$ (i.e. all players act as singletons) this corresponds to the Nash equilibrium.

coalition of size m is externally stable, then we can conclude that a coalition of size $m+1$ will not be internally stable. Similarly, if a coalition of size m is internally stable, then a coalition of size $m-1$ will not be externally stable.

If players are not ex-ante symmetric, then this simplification is not appropriate. In fact, given the assumption of joint welfare maximization of coalition members (implying an efficient allocation of abatement duties among coalition members), if players have different payoff functions, payoffs among coalition members may be quite different (implying an asymmetric distribution of the total payoff of the coalition among its members). This may upset the stability of coalitions. A way to circumvent this problem is transfers. In principle, many transfer schemes are perceivable which typically will lead to different sets of stable coalitions. In order to avoid this sensitivity, we consider transfers in its most general form, based on the concept of an almost ideal transfer scheme (AITS) proposed by Eyckmans and Finus (2009) with similar notions in Fuentes-Albero and Rubio (2009), McGinty (2007) and Weikard (2009).

Suppose we ask the question whether there exists a transfer scheme that can stabilize a coalition S. Then a necessary condition for internal stability is that the sum of payoffs in the coalition (summing over the left-hand side of inequality (4)) exceeds the sum of free-rider payoffs when leaving the coalition (summing over the right-hand side of inequality (4)). We call this potential internal stability which is reflected in inequality (6) below. By the same token, a necessary condition for external stability is that the sum of non-members' payoffs (summing over the left-hand side of inequality (5)) exceeds the sum of payoffs when joining the coalition (summing over the right-hand side of inequality (5)). We refer to this condition as potential external stability in inequality (7).

(6) potential internal stability: $\sum_{i\in S} \Pi_i^*(S) \geq \sum_{i\in S} \Pi_i^*(S\setminus\{i\})$

(7) potential external stability: $\sum_{i\notin S} \Pi_i^*(S) > \sum_{i\notin S} \Pi_i^*(S\cup\{i\}) \;\; \forall\, i\notin S$

The interesting aspect is that with the AITS these two necessary conditions become also sufficient conditions. The AITS works as follows: every coalition member $i\in S$ receives his free-rider payoff when leaving the coalition, $\Pi_i^*(S\setminus\{i\})$, plus a positive share λ_i, $\sum_{i\in S}\lambda_i = 1$, of the surplus obtained by the coalition as a whole over the sum of free-rider payoffs, $\omega(S) = \sum_{i\in S}\Pi_i^*(S) - \sum_{i\in S}\Pi_i^*(S\setminus\{i\})$. Hence, the payoff of a coalition member with transfers is given by $\Pi_i^{*T}(S) = \Pi_i^*(S\setminus\{i\}) + \lambda_i\omega(S)$ where the superscript "T" stands for transfers. Replacing $\Pi_i^*(S)$ by $\Pi_i^{*T}(S)$ in the condition for internal stability in

(4) shows that if the surplus of coalition S is (weakly) positive, $\omega(S) \geq 0$, i.e., coalition S is potentially internally stable, then coalition S will be internally stable with the AITS. By the same token, replacing $\Pi_i^*(S \cup \{i\})$ by $\Pi_i^{*T}(S \cup \{i\})$ in the condition for external stability in (5) shows that if the surplus of all enlarged coalitions $S \cup \{i\}$ is negative, $\omega(S \cup \{i\}) < 0$ for all $i \notin S$, i.e. coalition S is potentially externally stable, then coalition S will be externally stable with the AITS. Hence, a similar link between internal and external stability observed above for symmetric players can be established with the AITS. That is, if coalition S is externally stable, then all coalitions $S \cup \{i\}$, for all $i \notin S$, are not internally stable; if coalition S is internally stable, then all coalitions $S \setminus \{i\}$, for all $i \in S$, are not externally stable.

There are three more interesting features of the AITS worthwhile mentioning of which we make use later. First, the design of the AITS suggests that it aims at stabilizing a coalition internally. This seems sensible as in the context of public good provision one is more concerned about players leaving a coalition than players joining a coalition. Every coalition that can be potentially internally stabilized will be internally stable with the AITS. Without transfers or with other transfer scheme (e.g. Nash-Bargaining solution or Shapley Value; see Eyckmans and Finus 2009 for examples), this may not necessarily be the case. Hence, we can conclude that if a coalition S is internally stable without transfers, it will also be internally stable with the AITS. Conversely, if a coalition is not internally stable with the AITS, it will also not be stable without transfers (or any other transfer scheme).

Second and immediately related to this, if a coalition S is stable (internally and externally stable) without transfers, then with the AITS coalition S will either also be stable or a larger coalition with additional players R, $S \cup R$, will be stable. The reason is that if S was not externally stable with the AITS, then a coalition $S \cup \{i\}$ would be internally stable. If $S \cup \{i\}$ was also externally stable, then it would be stable. If it was not externally stable, then the argument of adding players is repeated, noting that eventually one coalition must be externally stable, simply because the grand coalition is externally stable by definition.

Third, as we will argue in section 3, in our public good game the total payoff (summing payoffs over all players) when coalition $S \cup \{i\}$ forms is strictly higher than when coalition S forms. That is, the formation of larger coalitions translates into higher global payoffs. This explains the word "ideal" in the name of the transfer scheme. Among those coalitions which can be potentially internally stabilized, the AITS stabilizes the coalition with the highest global payoff. Because this may not be the grand coalition if free-rider incentives are too strong, the word "almost" is part of the name of the transfer scheme.

From the three features we can conclude that with AITS, stable coalitions will be at least as large and successful (in terms of aggregate payoffs) than without transfers.

2.2 Three Learning Scenarios

Suppose that some parameter values of the payoff functions are uncertain. Following Kolstad and Ulph (2008, 2011), we assume risk-neutral agents and distinguish *three learning scenarios*: 1) full learning, 2) partial learning and 3) no learning. *F*ull *L*earning (abbreviated to FL) can be considered as a benchmark case in which players learn about the true parameter values before taking the membership decision in the first stage. Hence, uncertainty is fully resolved at the beginning of the game. For *P*artial *L*earning (abbreviated to PL) it is assumed that players decide about membership under uncertainty but know that they will learn about the true parameter values before deciding upon abatement levels in the second stage. Hence, the membership decision is based on expected payoffs, under the assumption that players will take the correct decision in the second stage. Finally, under *N*o *L*earning (abbreviated to NL) also the abatement decision has to be taken under uncertainty. That is, players derive their abatement strategies by maximizing expected payoffs. The membership decisions are also based on expected payoffs, though these payoffs differ from those under partial learning, given that less information is available.

It is worthwhile pointing out that our assumption implies that learning takes the form of perfect learning (Kolstad and Ulph 2008, 2011). That is, if players learn about parameter values, no uncertainty remains. Hence, partial learning is de facto delayed learning, though we stick to the terminology introduced by Kolstad and Ulph. Furthermore, partial learning requires assuming that players stick to their membership decision taken in the first stage even though additional information becomes available in the second stage.[7] Full learning is certainly an optimistic and no learning a pessimistic benchmark about the role of learning in the context of climate change. Partial learning approximates (because beliefs are not updated in a Bayesian sense) the fact that information becomes available over time.

[7] This assumption has been called fixed membership in the literature (Rubio and Ulph 2007 and Ulph 2004). The record of international environmental agreements shows that countries usually do not leave agreements once they have ratified them, though new members may join at a later stage (Finus 2003). This is also suggested by Iida (1993) in other areas of international policy coordination.

2.3 Three Types of Uncertainty

2.3.1 Introduction

The assumption about the uncertain parameter values in the payoff functions of players are summarized in *three types of uncertainty*. Due to the complexity of coalition formation, the consideration of a particular payoff function is required. In order to avoid binary equilibrium choices "abate" or "not abate", as for instance in Kolstad (2007) and Kolstad and Ulph (2008, 2011), which renders the model not well-behaved in the first and second stage (see section 4.2), we consider a strictly concave payoff function which is still simple enough to derive analytical results:[8]

$$(8) \qquad \Pi_i = b_i \sum_{k=1}^{n} q_k - c_i \frac{q_i^2}{2} \ , \ i \in N \, , \ b_i > 0 \, , \ c_i > 0$$

where b_i is a benefit parameter, $b_i \sum_{k=1}^{n} q_k$ is the benefit from global abatement, c_i is a cost parameter, and $c_i \frac{q_i^2}{2}$ is the abatement cost from individual abatement.

Generally, the benefit as well as the cost parameters could be uncertain. However, following Kolstad (2007), Kolstad and Ulph (2008, 2011) and Na and Shin (1998), in the climate context, uncertainty about the benefits from reduced damages appears to be more important than uncertainty about abatement costs. Hence, we simplify the model, by dividing payoffs by the cost parameter c_i, define the benefit-cost ratio by $\gamma_i = b_i / c_i$, and hence payoff function (8) reads:

$$(9) \qquad \Pi_i = \gamma_i \sum_{k=1}^{n} q_k - \frac{q_i^2}{2} \ , \ i \in N \, , \ \gamma_i > 0 \ .$$

Henceforth, we call γ_i the benefit parameter. If this parameter is uncertain, then it is represented by the random variable Γ_i, with associated distribution f_{Γ_i}. The assumptions regarding our three types of uncertainty are displayed in Table 1.

All three types of uncertainty capture an important aspect surrounding climate change (IPCC 2007). There is much uncertainty about the absolute level of the benefits from abatement in the form of reduced damages (type 1). It is still not clear how serious the damages of climate change will be overall. But there is also much debate about the regional distribution of damages, i.e. which countries will benefit most from greenhouse gas emission reductions (type 2)? Hence, uncertainty of type 3, considering simultaneously uncer-

[8] A similar payoff function has been used for instance by Barrett (2006) and Na and Shin (1998) but also by many others.

tainty about the level and the distribution of the benefits from abatement, reflects the most comprehensive type of uncertainty. Nonetheless, types 1 and 2 turn out to be useful benchmarks in the analysis. As the random variable Γ_i is the benefit-cost ratio and the only variable in our simple model, it exclusively determines the gains from cooperation. Hence, we de facto model uncertainty about the level and/or distribution of the gains from cooperation – a problem which certainly applies to many economic problems with externalities.

Table 1: Three Types of Uncertainty about the Benefit Parameters

Type of Uncertainty	Definition of Parameters	Interpretation of Parameters	Ex-ante Expectations of Parameters	Ex-post Realizations of Parameters
1) pure uncertainty about the level of benefits	one random player draws a benefit parameter for all players from some distribution	common	symmetric	symmetric
2) pure uncertainty about the distribution of benefits	each player draws a benefit parameter from a set $\{\gamma_1, ..., \gamma_n\}$ without replacement	individual	symmetric	asymmetric
3) simultaneous uncertainty about the level and the distribution of benefits	each player draws a benefit parameter from a set $\{\gamma_1, ..., \gamma_n\}$ with replacement	common and individual	symmetric	asymmetric

Note that for all three types, *uncertainty is symmetric* as all players know as much or little about their own as about their fellow players' payoff functions. Moreover, *expectations are symmetric* as all players share the same beliefs about the distribution of the uncertain parameters. This requires that some coordination has taken place ex-ante. Even under learning, disagreement about optimal policy levels is not an issue (e.g. because of asymmetric benefit parameters) as coalition members maximize their joint welfare. We comment on this simplification in section 5. Nevertheless, as we will see in section 3, disagreement will figure indirectly into our model when it comes to decide on the participation in the agreement. Then under full learning, asymmetry may cause little participation if not balanced by transfers.

2.3.2 Assumptions

Uncertainty of Type 1: Pure Uncertainty about the Level of Benefits

Uncertainty of type 1 is inspired by Kolstad (2007) and Kolstad and Ulph (2008, 2011), which the authors call systematic uncertainty as it relates to a *common parameter*. All players have the same expectations ex-ante, and once uncertainty is resolved, all players have the same benefit parameter ex-post. Regarding the distribution of the benefit parameter, Kolstad (2007) and Kolstad and Ulph (2008) assume a simple Bernoulli distribution, in which the benefit parameter can take on only two values: low and high.

For uncertainty of type 1, all benefits are equal, i.e., $\Gamma_i = \Gamma_j$ for all $i, j \in N$. This may be viewed as a draw from an urn where a random player draws a ticket with a value γ_i which applies to all players. The main idea captured by these assumptions is that there is *pure uncertainty about the level of the benefits* from global abatement. Distributional issues play no role as realizations are completely symmetric. This may also be viewed as uncertainty about the global benefits from abatement. For the later analysis it is helpful to point out that because γ_i represents the marginal benefits from abatement, uncertainty about the level of benefits translates into a positive variance of the sum of marginal benefits but a zero variance across the realized marginal benefits. That is, the sum of marginal benefits is random but, for each realization, the benefit parameters are equal across players.

Compared to the studies mentioned above, we can be more general in two respects. First, we do not have to assume any particular distribution of the uncertain benefit parameter. Second, our payoff function does not imply binary equilibrium abatement strategies and hence optimal abatement strategies are a function of the benefit parameter. This leads to more interesting effects of uncertainty and learning on second stage outcomes and also helps to derive clear-cut conclusions.

Uncertainty of Type 2: Pure Uncertainty about the Distribution of Benefits

Uncertainty of type 2 draws on the idea of Na and Shin (1998). They consider three players with ex-ante symmetric expectations about the random variable Γ_i. The ex-post realizations are three different values $\gamma_1 < \gamma_2 < \gamma_3$. This can be viewed as players sequentially drawing one ticket from an urn with three tickets, each ticket representing a different value γ_i, $i = 1, 2, 3$. Tickets are not returned to the urn.

The main idea captured by this experiment is that uncertainty relates to *individual parameters*. Moreover, though expectations about the benefit parameters are symmetric, their realizations are asymmetric. Each player is allocated a different value from the set $\{\gamma_1, \gamma_2, \gamma_3\}$.

Thus the sum of all γ_i's over all players is known, i.e. the sum of marginal benefits is constant. As a result, the variance of the sum of marginal benefits is zero but the variance of the realized marginal benefits across players is positive.

In contrast to uncertainty of type 1, type 2 can be interpreted as *pure uncertainty about the distribution of the benefits* from global abatement. That is, the vector of random variables $\Gamma = (\Gamma_1,...,\Gamma_n)$ can be interpreted as the shares of the global benefits from abatement, as for instance modeled in Dellink et al. (2008), simply by dividing each random variable by the sum of marginal benefits, $\sum_{i=1}^{n} \Gamma_i$.

We would like to generalize the idea of Na and Shin in four respects. First, we do not restrict attention to only three players but assume any arbitrary number of players n. Second, not all realizations γ_i have to be different, though at least two, otherwise no uncertainty would remain, i.e. we assume $\gamma_1 \leq \gamma_2 \leq ... \leq \gamma_n$ with at least one inequality being strict. Third, we also consider the scenario of partial learning, whereas Na and Shin compare only full and no learning. Fourth, we consider the possibility to mitigate asymmetries through transfers.

Uncertainty of Type 3: Simultaneous Uncertainty about the Level and Distribution of Benefits

Uncertainty of type 3 combines the features of uncertainty of types 1 and 2: there is *simultaneous uncertainty about the level and the distribution of the benefits* from global abatement. That is, there is uncertainty about a *common parameter and individual parameters*. This idea is captured by assuming that all random variables, Γ_i, $i = 1,...,n$, are independently distributed, though identically (because we assume symmetric expectations). This may be viewed as players sequentially drawing tickets from an urn with tickets being returned after each draw. As for uncertainty of type 2, we assume that at least two tickets are different. The assumptions of uncertainty of type 3 imply that the sum of marginal benefits is random and its variance is larger (smaller) than under uncertainty of type 2 (type 1). That is, the level effect is maximal for uncertainty of type 1, minimal for uncertainty of type 2 and intermediate for type 3. Moreover, the variance across the realized parameter values for uncertainty of type 3 is, on average, lower (larger) than for type 2 (type 1). That is, the distribution effect is maximal for uncertainty of type 2, minimal for uncertainty of type 1 and intermediate for type 3.[9] Different from Kolstad and Ulph (2011), we do not restrict the uncertain

[9] We view assumptions about uncertainty of type 3 as the most "natural" way to combine type 1 and 2, though it is clear that an infinite number of possible combinations could be constructed which are either closer to type 1 or 2.

parameter values to two values (high and low), equilibrium abatement strategies are not binary and we study the role of transfers.

3. Model Solution

In this section, we solve the model for the three learning scenarios and the three uncertainty cases. As a reference point, we start by considering certainty.

3.1 Certainty

According to the procedure of backwards induction, we first solve the coalition formation game for second stage outcomes. Using payoff function (9), and following the instructions of how signatories choose their equilibrium abatement levels as stated in (2) and non-signatories as stated in (3), we obtain the following equilibrium abatement levels:

$$(10) \qquad q_i^*(S) = \sum_{\ell \in S} \gamma_\ell \quad \forall\, i \in S\,,\; q_j^*(S) = \gamma_j \quad \forall\, j \notin S\,,\; \sum_{k \in N} q_k^*(S) = m \sum_{i \in S} \gamma_i + \sum_{j \notin S} \gamma_j \quad .$$

Casual inspection of (10) reveals that individual and total abatement increase in the benefit parameters. It is also evident that when a coalition S of size m has formed, and a non-signatory $j \notin S$ joins the coalition, total abatement will increase. We call this last property global efficiency from cooperation (GEF), implying that a sequence of accessions to a coalition S increases global abatement, with the highest global abatement being obtained in the grand coalition.

If abatement levels in (10) are substituted into payoff functions (9), we obtain payoffs without transfers of signatories $\Pi_{i \in S}^*(S)$, non-signatories $\Pi_{j \notin S}^*(S)$ and total payoffs, $\sum_{k \in N} \Pi_k^*(S) = \sum_{i \in S} \Pi_i^*(S) + \sum_{j \notin S} \Pi_j^*(S)$. The same observations made above for abatement can be observed for payoffs. That is, payoffs increase in the benefit parameter, and total payoffs increase with an enlargement of the coalition (GEF), with the highest total payoff being obtained by the grand coalition, corresponding to the social optimum.[10]

For the understanding of the incentives to form coalitions, two more properties are interesting. If a non-signatory j joins coalition S such that $S \cup \{j\}$ forms, then total payoffs of this group increases, $\sum_{i \in S} \Pi_i^*(S \cup \{j\}) > \sum_{i \in S} \Pi_i^*(S) + \Pi_{j \notin S}^*(S)$. This property, called *superadditivity* (SAD), underlines that there is generally an incentive for cooperation, at least if the gains from cooperation are shared "fairly" among coalition members. However, also all remaining non-signatories $k \notin S \cup \{j\}$ benefit from this enlargement of coalition

[10] Property GEF, as well as SAD and PE mentioned below, are easy to prove using payoffs provided in Appendix 1.

S, $\Pi^*_{k\notin S\cup\{j\}}\left(S\cup\{j\}\right) > \Pi^*_{k\notin S\cup\{j\}}\left(S\right)$, a property called *positive externality* (PE). Hence, even if we consider transfers and find that the grand coalition is not stable , then this can be interpreted as the positive externality effect being stronger than the superadditivity effect. In the absence of transfers, an additional reason for small coalitions may be that the gains from cooperation are shared unequally as Lemma 1 illustrates.

Lemma 1: Equilibrium Coalition Size under Certainty and No Transfers

Consider a coalition S *which is composed of* m *members with parameters* $\gamma_{i_1} \le \gamma_{i_2} \le ... \le \gamma_{i_m}$, *and* $n-m$ *singletons with parameters* $\gamma_{j_1} \le \gamma_{j_2} \le ... \le \gamma_{j_{n-m}}$. *In the absence of transfers, no stable coalition comprises more than three players, and a coalition is stable if and only if:*

$a)\ \gamma_{i_1} = \gamma_{i_2} = \gamma_{i_3} \quad if \quad m=3 \ \wedge \ n=3$

$b)\ \gamma_{i_1} = \gamma_{i_2} = \gamma_{i_3} \ \wedge 2\gamma_{j_{n-m}} < 3\gamma_{i_1} \quad if \quad m=3 \ \wedge \ n\ge 4$

$c)\ \frac{\gamma_{i_2}}{\gamma_{i_1}} \le \sqrt{2} \ \wedge \ \gamma_{j_{n-m}} < \frac{\gamma_{i_1}+\gamma_{i_2}}{2} \quad if \quad m=2 \ \wedge \ n\ge 3$

$d)\ \forall\gamma_i,\ i\in N \quad if \quad m=1$

Proof: Follows using payoffs in Appendix 1 and applying the conditions of internal and external stability in (4) and (5), respectively. **Q.E.D.**

The main conclusions from Lemma 1 are straightforward. First, no coalition larger than three players is stable. Second, a three player coalition is only stable if all coalition members have the same benefit parameter (conditions a and b). If the total number of players exceeds three, then the symmetric benefit parameters of the coalition members must be sufficiently large compared to those benefit parameters of outsiders, such that this coalition is also externally stable (condition b). Third, if a coalition among different players is formed, the maximum stable coalition size is two and players cannot be too different (condition c). Moreover, the mean of the two benefit parameters of the coalition members must be sufficiently large compared to the benefit parameters of outsiders, otherwise external stability would be violated. Finally, the singleton coalition structure is stable by definition (condition d).

The upshot of Lemma 1 is that in the absence of transfers, heterogeneity is an obstacle to cooperation. This is quite different with transfers.

Lemma 2: Equilibrium Coalition Size under Certainty and Transfers

Consider a coalition S *which is composed of* m *members and let the vector of benefit parameters be denoted by* γ^S*, the standard deviation of* γ^S *by* σ_S*, the arithmetic mean of* γ^S *by* μ_S *and define the coefficient of variation by* $CV_S = \frac{\sigma_S}{\mu_S}$*. Assume that coalition members apply the almost ideal transfer scheme.*

a) There is a stable coalition of at least three players; no stable coalition comprises less than three players.

b) A coalition of size $m \geq 4$ *is internally stable if and only if* $CV_S \geq A(m)$ *where*

$A(m) = \sqrt{\frac{m^2 - 4m + 3}{2m - 3}}$ *with* $A(m)$ *increasing in* m*.*

c) A coalition of size $m \geq 4$ *is externally stable if and only if for all coalitions* $S \cup \{j\}$*,* $j \notin S$*,* $CV_{S \cup \{j\}} < A(m+1)$*.*

Proof: See Appendix 2.

Comparing Lemmas 1 and 2 shows that the largest coalition without transfers comprises three players, which is the smallest coalition size in the presence of transfers. Without transfers, heterogeneity is an obstacle for forming large coalitions, with transfers it is an asset and if the coefficient of variation, CV_S, is large enough, even the grand coalition can be stable. Without transfers, payoffs among coalition members are asymmetrically distributed, making free-riding attractive. With transfers, an asymmetric distribution of payoffs is no longer an issue. Moreover, heterogeneity generates additional surplus from cooperation. More precisely, a low mean μ_S and a high standard deviation σ_S are conducive to cooperation as the coefficient of variation increases. The intuition is the following.

First, the larger the average benefit parameter, the larger will be equilibrium abatement levels of signatories and the more ambitious are coalitional abatement efforts compared to no cooperation. Ambition translates into a high free-rider incentive.

Second, the higher the standard deviation of the coalition members' benefit parameters, the larger are the gains from cooperation for coalition members, increasing the attractiveness of cooperation (or decreasing the attractiveness of free-riding). Recall, if countries behave non-cooperatively, they set their own marginal benefits equal to their own marginal costs from abatement, whereas if they form a coalition they set the sum of marginal benefits of coalition members equal to their marginal cost. Hence, the move from no to coalitional cooperation is associated with two effects. First, the externality is internalized among coali-

tion members. This internalization effect has two components, the benefit to the coalition is the superadditivity component, but because this benefit is *non-exclusive* also non-signatories benefit from higher abatement efforts of the coalition, which is the positive externality component. Second, abatement burdens are allocated cost-effectively among coalition members. This cost-effectiveness effect is an *exclusive* benefit accruing to the coalition and hence contributes only to the superadditivity component.

Now suppose all coalition members are symmetric, i.e. have the same benefit parameter. Because all players have the same abatement cost function in our model, no cooperation already implies a cost-effective allocation of abatement burdens. Hence, coalition members benefit from the internalization effect but not from the cost-effectiveness effect. In other words, the superadditivity component is not very strong compared to the positive externality component, resulting in only small coalitions being stable.

In contrast, if players have different marginal benefit parameters γ_i, no cooperation is no longer cost-effective and cooperation is also associated with a cost-effectiveness effect. In fact, the larger the difference in benefit parameters, the larger will be the cost-effectiveness effect. In other words, the larger the asymmetry, the stronger is the superadditivity compared to the positive externality component, allowing for larger stable coalitions.

3.2 Uncertainty

3.2.1 Full Learning

The scenario of full learning (FL) is (almost) a direct application of what has been derived for certainty. For the purpose of later comparison with the scenarios of partial and no learning, we will evaluate outcomes from an ex-ante perspective. This has no impact on the properties superadditivity, positive externality and global efficiency from cooperation.

In terms of equilibrium coalitions, with reference to our three types of uncertainty (see subsection 2.3.2), the following predictions follow directly from Lemmas 1 and 2.

Proposition 1: Equilibrium Coalitions under Full Learning

Under the full learning scenario, the expected equilibrium coalition size $E\left[m^{*FL}\right]$ *is given by:*

Uncertainty of Type 1 (Pure Uncertainty about the Level of Benefits)

$E\left[m^{*FL}\right]=3$ *where all possible 3-player coalitions are stable, with and without transfers.*

Uncertainty of Type 2 (Pure Uncertainty about the Distribution of Benefits)

Without transfers: $1 \le E\left[m^{*FL}\right] \le 3$. *In particular:*

a1) $E\left[m^{*FL}\right] = 1$, *only the singletons coalition structure is stable if neither of conditions a, b and c in Lemma 1 hold;*

a2) $E\left[m^{*FL}\right] > 1$, *with at least one non-trivial coalition being stable, though no stable coalition comprises more than three players if at least one of the conditions a, b and c in Lemma 1 hold.*

With transfers (AITS): $E\left[m^{*FL}\right] \ge 3$, *all stable coalitions comprise at least three players. In particular, recalling that the coefficient of variation of the benefit vector of coalition* S, γ^S, *is denoted by* CV_S:

b1) $E\left[m^{*FL}\right] = 3$, *all three-player coalitions are stable if* $CV_S < \sqrt{3/5}$, $\forall S : \#S = 4$;

b2) $E\left[m^{*FL}\right] > 3$, *stable coalitions comprise three or more players if* $\exists S$, $\#S = 4 : CV_S \ge \sqrt{3/5}$, $n \ge 4$;

b3) $E\left[m^{*FL}\right] \ge 4$, *stable coalitions comprise four or more players if* $CV_S \ge \sqrt{3/5}$ $\forall S : \#S = 4$, $n \ge 4$.

Uncertainty of Type 3 (Simultaneous Uncertainty about the Level and the Distribution of Benefits)

Without transfers: $1 < E\left[m^{*FL}\right] \le 3$.

With transfers (AITS): $E\left[m^{*FL}\right] \ge 3$, *all stable coalitions comprise at least three players. In particular, recalling that the coefficient of variation of the benefit vector* γ^S *of coalition* S *is denoted by* CV_S:

c1) $E\left[m^{*FL}\right] = 3$, *all three-player coalitions are stable if for all possible realizations of the vector of benefit parameters,* $\gamma = (\gamma_1, ..., \gamma_n)$, $CV_S < \sqrt{3/5}$, $\forall S : \#S = 4$.

c2) $E\left[m^{*FL}\right] > 3$: *all stable coalitions comprise three or more players and at least one comprises four or more players if there are realizations of the vector of benefit parameters,* $\gamma = (\gamma_1, ..., \gamma_n)$, *such that* $\exists S : CV_S \ge \sqrt{3/5}$, $n \ge 4$.

If there is pure uncertainty about the level of benefits (uncertainty of type 1), then all players will have the same benefit parameter γ_i. For symmetric players, transfers make no difference and all coalitions of three players are stable. If there is pure uncertainty about the distribution of benefits or simultaneous uncertainty about the level and distribution of benefits (uncertainty of type 2 and 3), then we assumed that at least two γ_i-values are

different, but we did not rule out the possibility that some players may have the same γ_i-value. Hence, without transfers, coalitions comprising two and/or three players can be stable but also the singletons coalition structure. A subtle detail of uncertainty of type 3 is that even if the distribution of the random benefit parameters is spread over a wide range of different values, there is a positive probability that all players receive exactly the same value. That is, there is the possibility that all players draw the same ticket from the urn, as tickets are returned. This explains the statement $E\left[m^{*FL}\right]>1$ for uncertainty of type 3. With transfers, we know from Lemma 2 that all stable coalitions will comprise at least three players. The specific conditions listed in Proposition 1 will prove useful for the comparison of results in section 4.

3.2.2 Partial Learning

Partial Learning (PL) assumes that players know the parameter values once they enter the second stage of coalition formation. Hence, equilibrium abatement levels in the second stage are exactly those stated under certainty in (10) and equilibrium payoffs are those provided in Appendix 1. Thus, PL and FL are identical in the second stage and, as a consequence, properties GEF, PE and SAD also hold for PL, which can be easily verified using payoffs in Appendix 3.

However, PL and FL are different regarding the first stage, in which countries choose their membership. Under PL, countries decide on their membership not based on actual payoffs but based on expected payoffs which are provided in Appendix 3. Since all players share the same beliefs, expected payoffs of all signatories are identical and the same applies to all non-signatories. Consequently, membership is based on symmetric ex-ante payoffs. This allows for three simplifications. First, we can discard transfers, i.e. they would not make any difference. Second, the actual identity of coalition members does not matter for the outcome. In other words, if a coalition of size m is stable, all coalitions with m members are stable. Third, there is a direct link between internal and external stability as explained in section 2.1 for symmetric players.

Proposition 2: Equilibrium Coalitions under Partial Learning

Let the vector of random benefit parameters, including all players, be denoted by $\Gamma=\left(\Gamma_1,\Gamma_2,...,\Gamma_n\right)$, *with* $\Gamma_1,\Gamma_2,...,\Gamma_n$ *being identical distributed because of symmetric beliefs of all players. Denote the standard deviation of* Γ_i *by* σ, *the expected value by* μ *and define the coefficient of variation by* $CV=\frac{\sigma}{\mu}$.

Under the partial learning scenario, with and without transfers, the expected equilibrium coalition size $E\left[m^{*PL}\right]$ *is given by*:

Uncertainty of Type 1 (Pure Uncertainty about the Level of Benefits)

$E\left[m^{*PL}\right]=3$ *where all possible 3-player coalitions are stable.*

Uncertainty of Type 2 (Pure Uncertainty about the Distribution of Benefits)

a1) If $CV<B(m,n)$ *where* $B(m,n)=\sqrt{\frac{(m-3)(n-1)}{n+m-3}}$, *with* $B(m,n)$ *increasing in* m *and* n, *then* $E\left[m^{PL}\right]=3$ *where all possible 3-player coalitions are stable.*

a2) If $CV\geq B(m,n)$, $n\geq 4$, *then* $E\left[m^{*PL}\right]\geq 4$ *and all possible* m *-player coalitions are stable for which* $B(m,n)\leq CV<B(m+1,n)$ *if* $m<n$ *and* $CV\geq B(m,n)$ *if* $m=n$ *hold.*

Uncertainty of Type 3 (Simultaneous Uncertainty about the Level and Distribution of Benefits)

b1) If $CV<C(m)$ *where* $C(m)=\sqrt{m-3}$ *with* $C(m)$ *increasing in* m, *then* $E\left[m^{PL}\right]=3$ *where all possible 3-player coalitions are stable.*

b2) If $CV\geq C(m)$, $n\geq 4$, *then* $E\left[m^{PL}\right]\geq 4$ *and all possible* m *-player coalitions are stable for which* $C(m)\leq CV<C(m+1)$ *if* $m<n$ *and* $CV\geq C(m)$ *if* $m=n$ *hold.*

Proof: See Appendix 4.

The qualitative conclusions under partial learning have much resemblance to what we found for certainty and full learning if transfers among coalition members are applied. Asymmetric distributions of the random benefit parameters with a large standard deviation as well as a low expected value (implying together a large coefficient of variation) are conducive to the stability of large coalitions. The mechanism is similar to the one laid out for certainty and transfers (section 3.1). The only difference is that not the standard deviation and the mean of the random benefit parameters of coalition members but of all players matter, as under PL, all players are ex-ante symmetric when deciding upon membership.

Note that the threshold of the coefficient of variation for uncertainty of type 2, $B(m,n)$, depends not only on the coalition size m but also on the total number of players as the draw of "benefit parameter tickets" from the urn is not independent (draw without replacement). Moreover, the threshold of the coefficient of variation for uncertainty of type 3, $C(m)$ is higher than for uncertainty of type 2, $B(m,n)$, i.e. $C(m)>B(m,n)$. The reason is that for any given distribution of the random benefit parameters, the variance

across realized benefit parameter values is larger for uncertainty of type 2 than type 3, as explained in section 2.3.

3.2.3 No Learning

No learning (NL) assumes that players do not know the realizations of the random variables Γ_i in the second stage of coalition formation. Hence, they derive their equilibrium abatement levels from maximizing expected payoffs, i.e. payoffs in (2) and (3) have to be replaced by expected payoffs. For our specific payoff functions in (9), replacing γ_i by the random variable Γ_i, this implies that payoffs are linear in the random variables and hence certainty equivalence holds. That is, the maximization of expected payoffs is equivalent to the maximization of payoffs under certainty for $\gamma_i = E[\Gamma_i]$. Moreover, we can make use of the fact that for all three types of uncertainty players are assumed to have identical beliefs, i.e. $E[\Gamma_i] = E[\Gamma_j]$ for all $i, j \in N$. This delivers equilibrium abatement levels:

$$(11) \qquad q_i^{**}(S) = mE[\Gamma_k] \;\; \forall i \in S\,,\; q_j^{**}(S) = E[\Gamma_k] \;\; \forall \; j \notin S\,, \qquad \sum_{k \in N} q_k^{**} = \left(m^2 - m + n\right) E[\Gamma_k]$$

where we use two asterisks in order to stress the difference to the abatement levels under certainty (see (10), which are also those under full and partial learning). It is evident that individual and total abatement levels increase in the expected value of the benefit parameter. Moreover, total abatement increases with the size of the coalition. Computing expected payoffs (see Appendix 5) confirms this also for payoffs. Hence, property GEF holds under NL and as it is easy to confirm, using the payoffs in Appendix 5, properties SAD and PE as well.

When it comes to determine stable coalitions, two pieces of information are useful. First, due to symmetric beliefs of all players, not only the abatement levels (as is evident from (11)) but also the payoffs (as is evident from Appendix 5) of all signatories are identical and the same is true for all non-signatories. Hence, transfers have no influence on stable coalitions. Second, because abatement levels and payoffs only depend on the (symmetric) expected value of the benefit parameters (and not on its variance), the type of uncertainty (types 1, 2 and 3) does not matter for the size of equilibrium coalitions.

Proposition 3: Equilibrium Coalitions under No Learning

Under the no learning scenario, with and without transfers, for all three types of uncertainty (i.e. pure uncertainty about the level of benefits, pure uncertainty about the distribution of benefits, and simultaneous uncertainty about the level and distribution of benefits), the expected equilibrium coalition size is $E\left[m^{*NL}\right]=3$ *where all possible 3-player coalitions are stable.*

Proof: Inserting payoffs in Appendix 5 into the condition for internal and external stability as defined in (4) and (5), respectively, delivers the result. **Q.E.D.**

Compared to full and partial learning, stable coalitions are robust to the variance and expected value of the benefit parameters.

4. Comparison of Results

4.1 Preliminaries

In this section, we compare results for the three learning scenarios for each of the three types of uncertainty. For a sensible comparison across the three learning scenarios, it is important to evaluate outcomes from an ex-ante perspective. That is, we compute and compare expected coalitions sizes and expected total payoffs across the three learning scenarios.

Analytically, we proceed in three steps. First, we compare coalitions relating to the first stage of coalition formation. We say that the first stage effect from learning is positive (negative) if the expected coalition size with learning is larger (smaller) than without learning. For this comparison we draw heavily on the results established in section 3. Second, we compare expected aggregate payoffs in the second stage of coalition formation, considering a generic coalition S of size m. We say that the second stage effect from learning is positive (negative) if the expected aggregate payoff with learning is larger (smaller) than without learning. Third, we pull the information from steps 1 and 2 together in order to derive conclusions regarding the overall outcome. As we consider three scenarios of learning, the effect of learning is viewed along the sequence of no, partial and full learning (NL, PL and FL).

In many cases, conclusions are derived along the following simple reasoning. If the effect of learning is positive (negative) in both stages, the effect of learning will be positive (negative) overall. The essential link between both stages is related to the property global efficiency from cooperation (GEF). If for any generic coalition S of size m learning scenario A performs better than scenario B in stage 2, then if scenario A leads to equal or larger

coalitions than scenario B in the first stage, we can conclude that A performs better than B overall. Moreover, it is helpful to recall that FL and PL are identical in the second stage and hence any difference regarding the overall performance of these two learning scenarios must stem from the first stage.

In some cases this simple reasoning does not work, even though the break-down in steps 1 and 2 is always useful when it comes to understand the driving forces. One reason is that sometimes effects in the two stages work in the opposite direction and hence overall conclusions are either not clear-cut or need further analysis. Another reason is that in some cases under full learning and no transfers the link between stages 1 and 2 is not direct. That is, not only the expected coalition size in the first stage matters, but also the identity of players in stable coalitions. Hence, the concept of a generic coalition of size m, which means all combinations of m players forming a coalition of size m, is not useful. The expected total abatement and payoff of a generic coalition of size m may be lower than if a particular coalition with m or less players forms, provided this particular coalition comprises players with high benefit values γ_i.

Finally, regarding the overall evaluation, we exclusively focus on aggregate payoffs, neglecting aggregate abatement. Nevertheless, it is useful to be aware of the following relations regarding total expected abatement in the second stage. Taking expectations over total abatement level under PL and FL in (10) for a generic coalition S of size m and comparing it with the expected total abatement level under NL in (11) reveals that both are equal. Hence, in the second stage, expected total abatement levels of the three learning scenarios are equal for each type of uncertainty. The reason is that equilibrium abatement levels are linear in the benefit parameters and, for a given number of players and coalition size, only depend on expected values. This implies that the second stage effect from learning regarding total payoffs can be exclusively attributed to different allocations of abatement levels across players under different learning scenarios. This simplification turns out to be useful for the interpretation of different second stage outcomes.

4.2 Pure Uncertainty about the Level of Benefits

Pure uncertainty about the level of benefits (uncertainty of type 1) allows for the derivation of straightforward results. From section 3 we know that all three scenarios of learning lead to coalitions of size three as players are ex-ante and ex-post symmetric. However, as Lemma 3 shows, there is no equivalence in terms of second stage outcomes.

Lemma 3: First and Second Stage Outcomes (Pure Uncertainty about the Level of Benefits)

First Stage of Coalition Formation: The expected stable coalition size under full learning (FL), partial learning (PL) and no learning (NL) are ranked as follows:

$$FL=PL=NL.$$

Second Stage of Coalition Formation: Consider a generic coalition S of size m, $1 \le m \le n$. For every generic coalition S, the total expected payoffs under the three learning scenarios are ranked as follows:

$$FL=PL>NL$$

where the difference between the two learning scenarios (FL and PL) and no learning (NL) increases in the variance of the random benefit parameter Γ_i.

Proof: First stage outcomes follow directly from Propositions 1, 2 and 3. Second stage outcomes follow from FL=PL by definition and by comparing expected payoffs under PL and NL, as provided in Appendices 3 and 5, respectively, noting that $\Gamma_i = \Gamma_j \quad \forall i, j \in N$ for uncertainty of type 1. **Q.E.D.**

The intuition regarding second stage outcomes is the following. Because of symmetric realizations of all benefit parameters, all signatories choose the same abatement level under FL and PL and the same holds for non-signatories. If the realized global benefit parameter is high (low), then all players will chose high (low) abatement levels. This means that with learning, though abatement levels across signatories and non-signatories are different, collectively the abatement levels across all players will be optimally adjusted according to the realization of the global benefit parameter.

Under NL, abatement levels within the group of signatories and non-signatories are also symmetric because abatement levels are chosen based on the expected value of the benefit parameters, which are identical due to symmetric beliefs. However, due to less information compared to FL and PL, abatement levels are based on the expected value of the global benefit parameter, and hence not adjusted according to the realization of the global benefit parameter. This means that if the realized global benefit parameter is high (compared to the expected value), abatement levels across all players fall short of optimal abatement levels (undershooting) and if the realized global benefit parameter is low (compared to the expected value), abatement levels are too high (overshooting). The collective under- and overshooting under NL compared to FL and PL leads to a loss of net benefits (i.e. payoffs) which is more pronounced the larger the variance of the global benefit parameter. The reason is that the total payoff in the second stage is a strictly convex function in the benefit

parameter γ.[11] Therefore, the total payoff from adopting an abatement level based on the expected value of the global benefit parameter is lower than the expected value of total payoffs.

Proposition 4: Overall Outcome (Pure Uncertainty about the Level of Benefits)

For total expected payoffs, the outcome of the two-stage coalition formation game is ranked as follows:

$$FL=PL>NL.$$

Proof: Follows directly from Lemma 3.

Thus, under pure uncertainty about the level of benefits, the overall outcome is that learning leads to higher expected payoffs. This contrasts with results by Kolstad (2007), and Kolstad and Ulph (2008). They find that though full learning leads to larger stable coalitions than no learning, expected total payoffs are smaller. For partial learning they find multiple equilibria for some parameter values, and conclude that the most likely equilibrium leads to lower membership and a lower expected aggregate payoff than full and no learning. Thus, in terms of payoffs, they suggest: NL>FL>PL. The question that arises is why their results are so different from ours? The following paragraphs provide a brief explanation, which is supported by a short formal exposition and examples in Appendix 7.

Kolstad and Ulph assume a linear payoff function and hence equilibrium abatement levels are corner solutions, i.e. binary strategies "abate" or "not abate". The model is calibrated such that, on the one hand, abatement pays in the social optimum and in a coalition above a threshold number of players. On the other hand, abatement does not pay for non-signatories and particularly in the Nash equilibrium. The stable coalition is a knife-edge equilibrium: once a coalition member leaves, the coalition breaks apart as for the remaining coalition members it does no longer pay to abate. This threshold depends on the benefit parameter γ; the larger γ, the higher the benefits from cooperation but less coalition members are needed to form a profitable coalition. Hence, the size of stable coalitions decreases in the benefit parameter γ.

Though this model looks simple at the outset, it appears to us that this simplicity comes at a price. Regarding the first stage effect from learning, the claim that expected membership is higher under FL than NL rests on the assumption that the equilibrium size of coalitions

[11] This is easily proved by using total payoffs as stated in Appendix 1, assuming symmetry of all benefit parameters and deriving the first and second derivatives with respect to the common benefit parameter.

can be approximated as a continuous variable. However, as Karp (2011) has pointed out, accounting for the fact that the number of signatories can only be an integer value, Kolstad and Ulph's conclusion may not be valid. More important, the second stage effect from learning changes in their model depending on the coalition size and can be positive, negative and nil. Therefore, one would expect that also overall effects may not be clear-cut. This expectation is not confirmed using the approximation in Kolstad and Ulph (2008) because total expected payoffs of the overall game are strictly concave in γ. However, again, Karp (2011) has shown that the approximation may be misleading and that the ranking FL>NL is possible. Clearly, in our model, effects in the first and second stage are well-behaved and always clear-cut, and payoffs are strictly convex in γ and therefore "learning is good".

4.3 Pure Uncertainty about the Distribution of Benefits

From section 3 we know that pure uncertainty about the distribution of benefits (uncertainty of type 2) can lead to smaller coalitions under FL than under PL and NL if transfers are not used to balance a possible asymmetric distribution of the gains from cooperation. With transfers, this shortcoming can be fixed. In fact, if the asymmetry of the benefit parameters across players is large enough, coalitions under FL will be larger than under NL. The same is true for PL, though no transfers are needed. All of this is compactly summarized in Lemma 4 regarding the first stage of coalition formation.

Lemma 4: First and Second Stage Outcomes (Pure Uncertainty about the Distribution of Benefits)

First Stage of Coalition Formation: The expected stable coalition size under full learning (FL), partial learning (PL) and no learning (NL) are ranked as follows:

$$\textit{No transfers:} \qquad PL \geq NL > FL;$$

$$\textit{Transfers:} \qquad PL, FL \geq NL.$$

Second Stage of Coalition Formation: Consider a generic coalition S of size m, $1 \leq m \leq n$. For every generic coalition S, the total expected payoffs under the three learning scenarios are ranked as follows:

$$NL \geq FL = PL \text{ with strict inequality if } m < n$$

where the difference between the two learning scenarios and no learning increases in the variance of the random benefit parameters Γ_i.

Proof: First stage outcomes follow directly from Propositions 1, 2 and 3. Second stage outcomes follow from FL=PL by definition and by comparing expected payoffs under PL

and NL, as provided in Appendix 3 and 5, respectively, noting that $\sum_{k=1}^{n} \Gamma_k$ is a constant under uncertainty of type 2 and using the terms in (III) in Appendix 4. **Q.E.D.**

Interestingly, regarding second stage outcomes, results are completely reversed compared to pure uncertainty about the level of benefits (compare Lemmas 3 and 4). The intuition is along the lines of the example mentioned in the introduction. Recall from section 4.1 that expected total abatement in the second stage is the same under all three scenarios of learning and hence total benefits, which are linear in total abatement, will be identical. Thus, differences must be related to abatement costs.

Total abatement costs will also be the same for all scenarios of learning if and only if the grand coalition forms because all players will choose abatement levels such that marginal abatement costs are equal to the sum over all marginal benefits. Since not only under PL and FL, but also under NL, the sum of marginal benefits is known (i.e. there is no uncertainty about the level of benefits), the second stage effect from learning is nil in the grand coalition, which is also the social optimum.

However, for any coalition smaller than the grand coalition, expected total costs under NL are strictly lower than under PL and FL and hence the second stage effect from learning is strictly negative. In the singletons coalition structure this is evident: under FL and PL, players will choose different abatement levels due to different realizations of the benefit parameter whereas under NL all players will chose the same abatement level due to identical expected values of the benefit parameter. Only a symmetric allocation of abatement duties is cost-effective as all players have the same abatement cost functions.

Finally, for larger coalitions than the singleton coalition structure, a similar argument applies. Signatories choose higher abatement levels than non-signatories, the difference between both groups is smaller on average under NL than under FL and PL. In other words, the allocation of abatement between signatories and non-signatories is, on average, "more symmetric" under NL than under PL and FL. The advantage of NL over PL and FL increases with the variance of the random benefit parameters.

When deriving conclusions regarding overall outcomes, it is clear that if the negative second stage effect from learning is not upset through a positive first stage effect from learning, then NL leads to better outcomes than FL and PL. PL also performs better than FL if the first stage effect from learning is negative, which can happen without transfers. All of this is summarized under point i) in Proposition 5 below. As the identity of members of stable coalition matters under FL and no transfers, the sufficient condition for NL>FL

and PL>FL is rather restrictive (i.e. only the singleton coalition structure is stable under FL and no transfers). For instance, point ii) suggests a class of distributions for which also NL>PL>FL (if $n \geq 4$) holds, despite under FL a two-player coalition among the two players with the largest γ_i-values is stable. For this class of distributions, all three-player coalitions are stable under PL and NL.

In order to reverse the result that "learning is bad" (see point iii) in Proposition 5), it is clear that the negative second stage effect from learning has to be overcompensated by a positive first stage effect. Not surprisingly, because of opposing effects of asymmetry on both stages, the sufficient conditions for FL>NL appear to be rather strong. Nevertheless, there is clear evidence that asymmetry improves upon overall outcomes under both learning scenarios, though under FL transfers are required. Finally, point iv) confirms that even without transfers, "learning can be good". The example we use in Appendix 6 to show this is quite informative. It assumes 100 countries with a distribution of the benefit parameter such that 98 countries have a low benefit parameter $\gamma_i = 1$ and two countries a very large benefit parameter $\gamma_j = 100$. This large asymmetry, even without transfers, leads to a two-player coalition among the two countries with the highest benefit parameter under FL, which performs better than a generic coalition of size 3 under NL and even of size 4 under PL.

Proposition 5: Overall Outcome (Pure Uncertainty about the Distribution of Benefits)

Let the outcome of the two-stage coalition formation game be evaluated in terms of expected total payoffs.

i) Sufficient conditions for "learning being bad" are

- $NL > FL$ *if n=3, or if only the singletons coalition structure is stable under FL and no transfers (condition a1 in Prop. 1), or if no coalition larger than 3 players is stable under FL with transfers (condition b1 in Prop. 1).*
- $NL \geq PL$ *if no coalition larger than 3 players is stable under PL (condition a1 in Prop. 2), with strict inequality if* $n \geq 4$.
- *PL>FL if only the singletons coalition structure is stable under FL and no transfers (condition a1 in Prop. 1) or if no coalition larger than 3 player is stable under FL with transfers (condition b1 in Prop. 1) and stable coalitions are strictly larger than 3 players under PL (condition a2 in Prop. 2).*

ii) For any discrete uniform distribution of the benefit parameter over equidistant values, i.e $\{\gamma_1, \gamma_2, ..., \gamma_n\}$, *with* $\gamma_{i+1} = \gamma_i + b,\ b > 0,\ i = 1,...,n-1,$ *the following ranking holds under no transfers:*

$NL \geq PL > FL$ *with strict inequality if* $n \geq 4$.

iii) Sufficient conditions for "learning being good" are

- $FL > NL$ *if stable coalitions are strictly larger than 3 players under FL and transfers (condition b3, Prop. 1).*
- $PL > NL$ *if stable coalitions are strictly larger than 3 players under PL (condition a2 in Prop. 2).*

iv) There are distributions of the random benefit variable for which even without transfers "learning is good" with ranking FL>PL>NL for which a necessary condition is that at least one non-trivial coalition is stable under FL (condition a2 in Prop. 1).

Proof: See Appendix 6.

From Proposition 5 it appears that Na and Shin's result that "learning is bad" is a rather special result. As mentioned in point i), it only holds for any distribution if there are only three players. Then, the stable coalition under NL is the grand coalition, corresponding to the social optimum, whereas FL will lead only to a two- or one-player coalition and the second stage disadvantage of FL cannot be overturned. However, considering also the learning scenario PL in Na and Shin's setting, with $n = 3$, would already lead to less clear-cut results as then NL=PL. Moreover, as Proposition 5 suggests, there are many situations when "learning can be good", even if there is pure uncertainty about the distribution of the benefits from cooperation. In particular, Na and Shin's negative result about the role of learning for the success of SEIEAs is driven by asymmetry, but as we show, in a more general setting, asymmetry can have the opposite effect.

4.3 Simultaneous Uncertainty about the Level and Distribution of Benefits

Simultaneous uncertainty about the level and the distribution of benefits constitutes a mix of the two previous types of uncertainty (see section 2.3.2, in particular footnote 9). The uncertainty part about the distribution shows up in the same first stage outcome as for uncertainty of type 2 (compare Lemmas 4 and 5) and the uncertainty part about the level shows up in the same second stage outcome as for uncertainty of type 1 (compare Lemmas 3 and 5). In particular, the last result stresses that the negative second stage effect from learning disappears once there is not pure uncertainty about the distribution of the benefits but also some uncertainty about the level of the benefits, which is certainly true for climate change but can also be expected for many other economic problems associated with positive externalities.

Lemma 5: First and Second Stage Outcomes (Simultaneous Uncertainty about the Distribution of Benefits)

First Stage of Coalition Formation: The expected stable coalition size under full learning (FL), partial learning (PL) and no learning (NL) are ranked as follows:

$$\textit{No transfers:} \quad PL \geq NL > FL;$$

$$\textit{Transfers:} \quad PL, FL \geq NL$$

Second Stage of Coalition Formation: Consider a generic coalition S *of size* m, $1 \leq m \leq n$. *For every generic coalition* S, *the total expected payoffs under the three learning scenarios are ranked as follows:*

$$FL = PL > NL$$

where the difference between the two learning scenarios and no learning increases in the variance of the random benefit parameters Γ_i.

Proof: First stage outcomes follow directly from Propositions 1, 2 and 3. Second stage outcomes follow from FL=PL by definition and by comparing expected payoffs under PL and NL, as provided in Appendix 3 and 5, respectively, noting that benefit parameters are independently distributed. **Q.E.D.**

Combining the effects from both stages as stated in Lemma 5, allows for straightforward conclusions.

Proposition 6: Overall Outcome (Simultaneous Uncertainty about the Level and Distribution of Benefits)

i) $PL > NL$.

ii) FL>NL with transfers.

iii) There are distributions of the random benefit variable for which without transfers learning is bad: NL>FL.

Proof: i) and ii) follow directly from Lemma 5. iii) assume for instance that the random benefit parameters are equidistantly distributed, e.g. $\gamma_i = \{1,2,3\}$, $\forall i \in N$, $n = 3$, then there are 27 possible combinations of realizations of the random variables for which stable coalitions can be determined using Lemma 1 under FL and no transfers and aggregate payoffs can be computed using Appendix 1. Computing the average aggregate payoff over all possible realizations gives an expected aggregate payoff of 39 whereas under NL, using Appendix 5, and noting that $m^{*NL} = 3$, the expected aggregate payoff is 54. **Q.E.D.**

Though without transfers FL may lead to worse outcomes than NL, this never happens with transfers and PL is anyway always better than NL. Overall, if there is some uncertainty about the level of the benefits from cooperation, then even in a strategic context "learning is good", though a compensation scheme may be needed under FL. This clearly qualifies the negative conclusions about the role of learning for the success of SEIEAs derived in previous papers.

5. Summary and Conclusions

This paper addressed the role of uncertainty and learning for the formation of self-enforcing international environmental agreements (SEIEAs). The central question was whether the veil of uncertainty impacts positively or negatively on the success of such agreements. The previous literature suggested a positive impact with the conclusion that "learning is bad". From our model the answer appears to be less straightforward. In fact, we showed that the negative impact of learning on the success of SEIEAs requires a couple of special assumptions and can also be mitigated in some cases by a transfer mechanism that is designed to minimize free-rider incentives.

The essential ingredients of our SEIEA-model is a (pure) public good provision game in which there is an incentive to cooperate but also an incentive to free-ride, as the benefits of public good provision are non-exclusive. Hence, full participation may not be a stable outcome. We assumed two-stages: in the first stage countries choose their membership in an agreement (becoming a signatory or remaining a non-signatory) and in the second stage choose their contributions to the public good. Uncertainty is related to the parameters of the benefit functions of individual players, with benefits depending on the sum of contributions. We considered uncertainty about the level and the distribution of benefits. Each type of uncertainty is considered first in isolation and then combined, giving rise to three types of uncertainty altogether. Uncertainty about the level of benefits is related to a common parameter and uncertainty about the distribution of the benefits is related to individual parameters. We also considered three scenarios of learning, with the two benchmark scenarios, full and no learning, and the intermediate scenario partial learning.

It became apparent that uncertainty and learning has an impact on both stages. There are two main reasons why "learning can be bad". First, the veil of uncertainty can be helpful when it comes to decide about participation in an agreement if the gains from cooperation could be unequally distributed. The higher the uncertainty about the distribution of the benefits from cooperation compared to the level of the benefits and the more dispersed

individual benefit shares are among players, the smaller will be stable coalitions under full learning. In contrast, under no and partial learning an asymmetric distribution of the gains from cooperation is not an issue. Due to identical beliefs about the distribution of the random variables, all players are symmetric ex-ante when choosing membership.

Second, when signatories and non-signatories choose their optimal contributions in the second stage, they interact strategically as contributions are strategic substitutes. More information would be beneficial if there was a social planner, which corresponds to the grand coalition in our model or if individual players acted in isolation. However, in our setting, players act strategically as long as not all players join an agreement and hence more information could lead to worse outcomes. We showed that in our model if there is pure uncertainty about the distribution of the benefits from cooperation, the aggregate payoff under partial and full learning falls short of that under no learning. The allocation of contributions among players is "more symmetric" under no learning (as it is based on symmetric expected values of the random variable) than under partial and full learning (as it is based on realized values of the random variable) which turns out to be a cost advantage for strictly convex and symmetric abatement cost functions.

Though these negative effects from learning on the success of cooperation are interesting, we argued in this paper that there are good reasons why we should expect the effect of learning to be much less negative. Regarding the negative impact of asymmetry on membership, we argued that a well-designed transfer scheme can hedge against an asymmetric distribution of the gains from cooperation under full learning. In fact, we showed that with transfers, membership under full learning never falls short of that under no learning and exceeds it for a sufficient degree of asymmetry. That is, the larger the degree of asymmetry, the larger the relative advantage of membership in an agreement over free-riding. Hence, the larger will be stable agreements under full learning, also compared to no learning. Under partial learning, a similar asymmetry-effect works, though no transfers are needed. A sufficient degree of asymmetry leads to larger agreements under partial than under no learning.

Regarding the possible negative impact of learning on total payoffs in the second stage, we concluded that it only holds in our model if there is pure uncertainty about the distribution of benefits. Once there is also uncertainty about the level of the benefits, which certainly applies to climate change, the effect from learning in the second stage becomes positive.

Taking a wider perspective, it is clear that our results apply beyond the specifics of climate change. The setting is that of a public good game but any economic problem which is characterized by positive externalities from cooperation would lead to similar qualitative conclusions. Knowing ex-ante the total size of the cooperative pie from coordinated action is advantageous as it allows to better target at the total level of efforts. In contrast, receiving confirmation that the individual slices of the pie of various participants might be quite unequally distributed may cause problems and therefore should be hedged against with an appropriate compensation mechanism. Heterogeneity among cooperating agents can be an asset if comparative advantages are exploited, but only if the gains from cooperation are fairly shared.

Despite we generalized some aspects of previous models, our model also shares some of their limitations. We pick only three which we believe are the most policy-relevant. First, we assumed that all players share the same beliefs, which requires some agreement or coordination about the current scientific evidence. Regarding many international policy issues, as in climate change, not all parties use and believe in the same scientific evidence. Therefore, a natural extension of our model could be to capture this phenomenon through a bargaining process in the second stage of coalition formation. It would be interesting to analyze whether an evolvement to more similar views, e.g. through intensified international research collaboration, political dialogue, and institutional coordination, through international bodies like for instance IPCC, would be conducive to the success of cooperation. Second, we assumed risk-neutrality but one could also look at the effects of risk-aversion on treaty outcomes, like in Boucher and Bramoullé (2010). They show that uncertainty about the level of benefits leads to lower abatement efforts than certainty, but higher participation, which may lead overall to higher aggregate payoffs. Whether this would also hold in our setting is difficult to predict as their model assumes a linear payoff function as Kolstad (2007) and Kolstad and Ulph (2008, 2011) and, as we have argued above, already in the context of risk-neutrality conclusions can be fundamentally different. Third, it would be interesting to model partial learning in a Bayesian sense. This would require to model coalition formation as a truly dynamic game in which players can revise their membership over time (Rubio and Ulph 2007). In order to obtain new interesting economic insight compared to our analysis and that of Rubio and Ulph (2007), it would probably also be necessary to allow for the possibility that players can reduce uncertainty through investment in R&D (learning-by-research) and implementation (learning-by-doing). For instance, the more players invest in mitigating climate change, the more they will learn

about their abatement costs over time. However, most likely, already even for simple extensions it will be difficult to obtain analytical solutions as first results by Breton and Sbragia (2011) suggest.

Acknowledgement

The research has been conducted while Pedro Pintassilgo was a visiting scholar at the Department of Economics, University of Stirling. He would like to acknowledge the hospitality of the department as well as financial support by the Portuguese Foundation for Science and Technology (FCT), grant no. BSAB/735/ 2007. Both authors have benefited from comments by Andrew Oswald, University of Warwick, and Ian Lange, University of Stirling, three anonymous referees and Roberton C. Williams III, the Co-Editor.

References

Asheim, G.B., C.B. Froyn, J. Hovi and F.C. Menz (2006), Regional versus Global Cooperation for Climate Control. "Journal of Environmental Economics and Management", vol. 51(1), pp. 93-109.

d'Aspremont, C., A. Jacquemin, J.J. Gabszewicz and J.A. Weymark (1983), On the Stability of Collusive Price Leadership. "Canadian Journal of Economics", vol. 16(1), pp. 17-25.

Barrett, S. (1994), Self-enforcing International Environmental Agreements. "Oxford Economic Papers", vol. 46, pp. 878-894.

Barrett, S. (2001), International Cooperation for Sale. "European Economic Review", vol. 45(10), pp. 1835-1850.

Barrett, S. (2003), Environment and Statecraft: The Strategy of Environmental Treaty-making. Oxford University Press, New York.

Barrett, S. (2006), Climate Treaties and "Breakthrough" Technologies. "American Economic Review", vol. 96(2), pp. 22-25.

Bloch, F. (2003), Non-cooperative Models of Coalition Formation in Games with Spillovers. In: Carraro, C. (ed.), The Endogenous Formation of Economic Coalitions. Edward Elgar, Cheltenham, UK et al., ch. 2, pp. 35-79.

Boucher, V. and Y. Bramoullé (2010), Providing Global Public Goods under Uncertainty. "Journal of Public Economics", vol. 94, pp. 591-603.

Bosello, F., B. Buchner and C. Carraro (2003), Equity, Development, and Climate Change Control. "Journal of the European Economic Association", vol. 1(2-3), pp. 601-611.

Brennan, G. and J.M. Buchanan (1985), The Reason for Rules: Constitutional Political Economy. Cambridge University Press, Cambridge, UK.

Breton, M. and L. Sbragia (2011), Learning under Partial Cooperation and Uncertainty. Papers presented at the Environment and Sustainability Forum, Exeter, UK.

Carraro, C. and D. Siniscalco (1993), Strategies for the International Protection of the Environment. "Journal of Public Economics", vol. 52(3), pp. 309-328.

Dellink, R., M. Finus and N. Olieman (2008), The Stability Likelihood of an International Climate Agreement. "Environmental and Resource Economics", vol. 39(4), pp. 357-377.

Deneckere, R. and C. Davidson (1985), Incentives to Form Coalitions with Bertrand Competition. "The RAND Journal of Economics", vol. 16(4), pp. 473-486.

Diamantoudi, E. and E.S. Sartzetakis (2006), Stable International Environmental Agreements: an Analytical Approach. "Journal of Public Economic Theory", vol. 8(2), pp. 247-263.

Donsimoni, M.-P., N.S. Economides and H.M. Polemarchakis (1986), Stable Cartels. "International Economic Review", vol. 27(2), pp. 317-327.

Eyckmans, J. and M. Finus (2009), An Almost Ideal Sharing Scheme for Coalition Games with Externalities. Stirling Discussion Paper Series, 2009-10, University of Stirling.

Finus, M. (2003), Stability and Design of International Environmental Agreements: The Case of Transboundary Pollution. In: Folmer, H. and T. Tietenberg (eds.), International Yearbook of Environmental and Resource Economics, 2003/4, Edward Elgar, Cheltenham, UK, ch. 3, pp. 82-158.

Finus, M. and S. Tjøtta (2003), The Oslo Protocol on Sulfur Reduction: The Great Leap Forward? "Journal of Public Economics", vol. 87(9-10), pp. 2031-2048.

Fuentes-Albero, C. and Rubio, S.J. (2009), Can the International Environmental Cooperation Be Bought? "European Journal of Operation Research", vol. 2002, pp. 255-64.

Gollier, C. and N. Treich (2003), Decision-making under Scientific Uncertainty: The Economics of the Precautionary Principle. "Journal of Risk and Uncertainty", vol. 27(1), pp. 77-103.

Haeringer, G. (2004), Equilibrium Binding Agreements: A Comment. "Journal of Economic Theory", vol. 117(1), pp. 140-143.

Hardin, G. (1968), The Tragedy of the Commons. "Science", vol. 162(3859), pp. 1243-1248.

Helm, C. (1998), International Cooperation behind the Veil of Uncertainty. "Environmental and Resource Economics", vol. 12(2), pp. 185-201.

Hoel, M. (1992), International Environment Conventions: The Case of Uniform Reductions of Emissions. "Environmental and Resource Economics", vol. 2(2), pp. 141-159.

Iida, K. (1993), Analytical Uncertainty and International Cooperation: Theory and Application to International Economic Policy Coordination. International Studies Quarterly, vol. 37, pp. 431-457.

IPCC (2007), Climate Change 2007, Synthesis Report.

Karp, L. (2011), The Effect of learning on Membership and Welfare in an International Environmental Agreement. *Climatic Change*, DOI 10.1007/s10584-011-0134-5.

Kolstad, C.D. (2005), Piercing the Veil of Uncertainty in Transboundary Pollution Agreements. "Environmental and Resource Economics", vol. 31(1), pp. 21-34.

Kolstad, C.D. (2007), Systematic Uncertainty in Self-enforcing International Environmental Agreements. "Journal of Environmental Economics and Managament", vol. 53(1), pp. 68-79.

Kolstad, C.D. and A. Ulph (2008), Learning and International Environmental Agreements. "Climatic Change", vol. 89(1-2), pp. 125-141.

Kolstad, C.D. and A. Ulph (2011), Uncertainty, Learning and Heterogeneity in International Environmental Agreements. "Environmental and Resource Economics", vol. 50, pp. 389-403.

McGinty, M. (2007), International Environmental Agreements among Asymmetric Nations. "Oxford Economic Papers", vol. 59, pp. 45-62.

Na, S.-L. and H.S. Shin (1998), International Environmental Agreements under Uncertainty. "Oxford Economic Papers", vol. 50(2), pp. 173-185.

Poyago-Theotoky, J. (1995), Equilibrium and Optimal Size of A Research Joint Venture in an Oligopoly with Spillovers. "The Journal of Industrial Economics", vol. 43(2), pp 209-226.

Rubio S. J. and A. Ulph (2006), Self-enforcing International Environmental Agreements Revisited. "Oxford Economic Papers", vol. 58(2), pp. 233-263.

Rubio S. J. and A. Ulph (2007), An Infinite-horizon Model of Dynamic Membership of International Environmental Agreements. "Journal of Environmental Economics and Management", vol. 54(3), pp. 296-310.

Sandler T. (2004), Global Collective Action. Cambridge University Press, Cambridge.

Sandler, T., F. P. Sterbenz and J. Posnett (1987), Free Riding and Uncertainty. "European Economic Review", vol. 31. pp. 1605-1617.

Stern, N. (2006), Stern Review: The Economics of Climate Change. Report prepared for the HM Treasury in the UK. Published 2007: Cambridge University Press, Cambridge, UK.

Ulph, A. (1998), Learning about Global Warming? In: Hanley, N. and H. Folmer (eds.), Game Theory and the Environment. Edward Elgar, Cheltenham, UK et al., ch. 13, pp. 255-286.

Ulph, A. (2004), Stable International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning. "Journal of Risk and Uncertainty", vol. 29(1), pp. 53-73.

Weikard, H.-P. (2009), Cartel Stability under Optimal Sharing Rule. "The Manchester School", vol. 77, pp. 575-93.

Weikard, H.-P., M. Finus and J.C. Altamirano-Cabrera (2006), The Impact of Surplus Sharing on the Stability of International Climate Agreements. "Oxford Economic Papers", vol. 58(2), pp. 209-232.

Yi, S.-S. (1996), Endogenous Formation of Customs Unions under Imprefect Competition: Open Regionalism is Good. "Journal of International Economics", vol. 41, pp. 153-177.

Yi, S.-S. (1997), Stable Coalition Structures with Externalities. "Games and Economic Behavior", vol. 20(2), pp. 201-237.

Young, O. (1994), International Governance: Protecting the Environment in a Stateless Society. Cornell University Press, Ithaca, New York.

Appendix[12]

Appendix 1: Payoffs under Certainty in the Second Stage

Using equilibrium abatement levels in (10) and inserting them into payoff functions in (9), gives the following payoffs:

$$\Pi^{*}_{i\in S}(S)=\gamma_i\left(m\sum_{\ell\in S}\gamma_\ell+\sum_{j\notin S}\gamma_j\right)-\frac{1}{2}\left(\sum_{\ell\in S}\gamma_\ell\right)^2$$

(I) $$\Pi^{*}_{j\notin S}(S)=\gamma_j\left(m\sum_{i\in S}\gamma_i+\sum_{k\notin S}\gamma_k\right)-\frac{1}{2}\left(\gamma_j\right)^2$$

$$\Pi^{*}(S)=\sum_{i\in S}\Pi^{*}_i(S)+\sum_{j\notin S}\Pi^{*}_j(S)=$$

$$\frac{1}{2}m\left(\sum_{i\in S}\gamma_i\right)^2+\left(\sum_{i\in S}\gamma_i\right)\left(\sum_{j\notin S}\gamma_j\right)(1+m)+\left(\sum_{j\notin S}\gamma_j\right)^2-\frac{1}{2}\sum_{j\notin S}\left(\gamma_j\right)^2$$

These payoffs are used to compute stable coalitions under certainty and full learning.

Appendix 2: Proof of Lemma 2

A coalition of size $m=1$ is internally stable by definition. For all other coalitions S of size $m\geq 2$, we use payoffs provided in Appendix 1 and insert them in the condition for potential internal stability (6), recalling that all potentially internally stable coalitions are internally stable for the almost ideal transfer scheme, which gives after some manipulation the condition $(2m-3)\sigma_S^2+\left(-m^2+4m-3\right)\mu_S^2\geq 0$. This condition always holds for $m=2$ and $m=3$. For $m\geq 4$, we solve this condition and obtain $CV_S\geq A(m)$. Noting the relation between (potential) internal and (potential) external stability, external stability of a coalition S of size m follows from the failure of potential internal stability for all coalitions $S\cup\{i\}$ of size $m+1$ where one player j outside coalition S is added to S. By the same token, a coalition of size $m=1$ and $m=2$ cannot be externally stable because all coalitions of size $m=2$ and $m=3$ are internally stable. Hence, the minimum size of stable coalitions with transfers is $m=3$. **(Q.E.D.)**

[12] For some proofs we only provide the intuition due to space limitations. Details are available upon request.

Appendix 3: Expected Payoffs under Partial Learning in the Second Stage

Replacing all γ_i's by their random counterparts Γ_i's in the payoffs in Appendix 1, taking expectations over the payoffs and noting that the random variables $\Gamma_1, \Gamma_2, ..., \Gamma_n$ are identically distributed due to symmetric beliefs, we obtain after some manipulation:

$$E\left[\Pi^*_{i\in S}(S)\right] = \frac{m}{2}E\left[\left(\Gamma_i\right)^2\right] + \left(n + \frac{m^2}{2} - \frac{3}{2}m\right)E\left[\Gamma_i\Gamma_k\right], \quad k \in N \setminus \{i\}$$

(II) $$E\left[\Pi^*_{j\notin S}(S)\right] = \frac{1}{2}E\left[\left(\Gamma_j\right)^2\right] + \left(m^2 + n - m - 1\right)E\left[\Gamma_j\Gamma_k\right], \quad k \in N \setminus \{j\}$$

$$E\left[\sum_{\ell\in N}\Pi^*_\ell(S)\right] = \left(\frac{n + m^2 - m}{2}\right)E\left[\left(\Gamma_i\right)^2\right] + \left(-\frac{m^3}{2} + \left(n - \frac{1}{2}\right)m^2 + (1-n)m + n^2 - n\right)E\left[\Gamma_i\Gamma_k\right], i \neq k \in N$$

These payoffs are used to compute stable coalitions under partial learning. They also represent second stage expected payoffs under partial and full learning for a generic coalition S of size m.

Appendix 4: Proof of Proposition 2

A coalition of size $m=1$ is internally stable by definition. For all other cases, $m \geq 2$, we use payoffs in Appendix 3 and apply the definition of internal stability in (4). After some manipulation, this condition reads $E\left[\left(\Gamma_i\right)^2\right] + (2-m)E\left[\Gamma_i\Gamma_k\right] \geq 0$ with player i and k being any two different players. This condition can be further manipulated, depending on the type of uncertainty. For uncertainty of type 1, $\Gamma_i = \Gamma_k \quad \forall i,k \in N$ and hence the condition for internal stability reads $(3-m)E\left[\left(\Gamma_i\right)^2\right] \geq 0$ and due to the link between internal and external stability, the condition for external stability reads $(2-m)E\left[\left(\Gamma_i\right)^2\right] < 0$ (by setting $m = m+1$ in the condition for internal stability and assuming that it fails). Hence, $m^* = 3$. For uncertainty type 2, we make use of the fact that $\sum_{k=1}^{n}\Gamma_k$ is a constant and hence it can be shown that

(III) $$E\left[\Gamma_i\Gamma_k\right] = \frac{n\left(E\left[\Gamma_i\right]\right)^2 - E\left[\left(\Gamma_i\right)^2\right]}{n-1}.$$

Denoting $\mu = E\left[\Gamma_i\right]$ and $\sigma^2 = Var\left[\Gamma_i\right]$ and noting that $E\left[\left(\Gamma_i\right)^2\right] = \sigma^2 + \mu^2$, the condition for internal stability reads $(n+m-3)\sigma^2 + (3-m)(n-1)\mu^2 \geq 0$. This condition always holds for $m=2$ and $m=3$. Because of the link between internal and external stability, $m=2$ is not externally stable. For $m \geq 4$, the internal stability condition can be

rewritten as $CV \geq B(m,n)$ and hence for external stability we have $CV < B(m+1,n)$. For uncertainty of type 3, $\Gamma_1, \Gamma_2, ..., \Gamma_n$ are independently distributed and hence the condition for internal stability reads $\sigma^2 + (3-m)\mu^2 \geq 0,$ from which the conclusions in Proposition 2 follow immediately. **(Q.E.D.)**

Appendix 5: Expected Payoffs under No Learning in the Second Stage

Substituting γ_i by $E[\Gamma_i]$ in payoff function (9), inserting abatement levels from (11) into these payoff functions for signatories and non-signatories, and taking expectations over payoffs, delivers:

$$E\left[\Pi^*_{i\in S}(S)\right] = \left(\frac{m^2}{2} - m + n\right)\left(E[\Gamma_k]\right)^2$$

(IV) $$E\left[\Pi^*_{j\notin S}(S)\right] = \left(m^2 - m + n - \frac{1}{2}\right)\left(E[\Gamma_k]\right)^2$$

$$E\left[\sum_{k\in N}\Pi^*_k(S)\right] = \left(-\frac{1}{2}m^3 + m^2 n + m\left(-n + \frac{1}{2}\right) + n^2 - \frac{1}{2}n\right)\left(E[\Gamma_k]\right)^2$$

These payoffs are used to compute stable coalitions under no learning. They also represent expected payoffs under no learning in the second stage for a generic coalition S of size m.

Appendix 6: Proof of Proposition 5

Statements i) follow directly from combining first and second stage effects from learning and recalling property GEF. Statements ii) requires comprehensive calculations which are available upon request. Statement iii): If condition a2 in Proposition 2 holds, then $E[m^{*PL}] \geq 4 > 3 = E[m^{*NL}]$. Using the lower bound $m^{*PL} = 4$ (because of property GEF), inserting this into expected aggregate payoffs under PL, as provided in Appendix 3, and inserting $m^{*NL} = 3$ into expected payoffs under NL, as provided in Appendix 5, delivers the result PL>NL. FL>NL follows from $E[m^{*FL}] \geq 4,$ according to condition b3 in Proposition 1 and noting that FL and PL are equal in the second stage. Statement iv): Consider $n = 100$, $\gamma_1 = \gamma_2 = ... = \gamma_{98} = 1,$ and $\gamma_{99} = \gamma_{100} = 10$. Computing the expected payoff under NL using Appendix 5 and setting $m^{*NL} = 3$ because of Proposition 3, gives 14,673.1. Under FL and no transfers, only the two players with $\gamma_{99} = \gamma_{100} = 10$ form a stable coalition according to Lemma 1, and hence inserting this information into aggregate payoffs, as provided in Appendix 1, delivers a payoff of 15,835. Under PL, $m^{*PL} = 4$ according to Proposition 2, condition a2, and inserting this information into aggregate pay-

offs provided in Appendix 3, gives 15,395 and hence overall the ranking FL>PL>NL follows. **(Q.E.D)**

Appendix 7: First and Second Stage Effects in Kolstad and Ulph's Model

Consider the payoff function $\Pi_i = \gamma \sum_{k \in N} q_k - q_i$, assume $1/n \le \gamma < 1$, and a normalized binary abatement strategy such that $q_i \in \{0,1\}$. Then, in the social optimum $q_i^{*S} = 1$ and in the Nash equilibrium $q_i^{*N} = 0$. A coalition will abate and make a difference to no cooperation, provided $\gamma \ge 1/m$ or $m \ge 1/\gamma$. The internally and externally stable coalition of size m^* is the largest integer equal or larger than $1/\gamma$, i.e. $I\left(1/\gamma\right)$. Under uncertainty, let the probability of a low value γ_ℓ be p and of a high value γ_h be $1-p$ and denote the expected value by $\bar{\gamma} = p\gamma_\ell + (1-p)\gamma_h$. Then, $m^{*NL} = I\left(1/\bar{\gamma}\right)$ and $m^{*FL} = pI\left(1/\gamma_\ell\right) + (1-p)I\left(1/\gamma_h\right)$. Moreover, non-signatories abatement levels are always zero whereas signatories abatement levels are given by:

$$q_{i\in S}^{*NL} = \begin{cases} 1 & if \quad m \ge m^{*NL} = \bar{m}^* = I\left(1/\bar{\gamma}\right) \\ 0 & if \quad m < m^{*NL} = \bar{m}^* = I\left(1/\bar{\gamma}\right) \end{cases}$$

$$q_{i\in S}^{*FL} = \begin{cases} 1 & if \quad m \ge m_s^* = I\left(1/\gamma_s\right), \quad s \in \{\ell, h\} \\ 0 & if \quad m < m_s^* = I\left(1/\gamma_s\right), \quad s \in \{\ell, h\} \end{cases}.$$

Consider possible coalition sizes, m_1, m_2, m_3 and m_4 such that $m_1 < m_h < m_2 < \bar{m} < m_3 < m_\ell < m_4$, and denote total abatement by Q and the total payoff by Π, then for m_1: $Q^{*NL} = 0 = Q^{*FL}$ and $\Pi^{*NL} = \Pi^{*FL} = 0$; m_2: $Q^{*NL} = 0 < (1-p)m_2 = Q^{*FL}$ and $\Pi^{*NL} = 0 < m_2(1-p)(n\gamma_h - 1) = \Pi^{*FL}$; m_3: $Q^{*NL} = m_3 > (1-p)m_3 = Q^{*FL}$ and $\Pi^{*NL} = nm_3\bar{\gamma} - m_3 > m_3(1-p)(n\gamma_h - 1) = \Pi^{*FL}$ and m_4: $Q^{*NL} = m_4 = Q^{*FL}$ and $\Pi^{*NL} = nm_4\bar{\gamma} - m_4 = (1-p)(m_4 n\gamma_h - m_4) + p(m_4 n\gamma_l - m_4) = \Pi^{*FL}$ which shows that the second stage effect from learning depends on the size of the coalition. Consider $\gamma_\ell = 0.3$, $\gamma_h = 0.5$, then if $p = 0.95$, $\bar{\gamma} = 0.31$ and $\bar{m}^* = m^{*NL} = 4$ whereas the expected size under FL is $m^{*FL} = 3.9$. If $p = 0.5$ expected membership under both learning scenarios is 3 whereas if $p = 0.75$, expected membership under FL is 3.5 whereas under NL only 3. This shows that first stage effects are not always clear-cut. For a more general analysis, see Karp (2011).

Research Highlights

- sufficient conditions that that "learning can be bad" for public good provision
- sufficient conditions that "learning is good" for public good provision
- suggest transfer mechanism to fix problem when learning is bad
- explain driving forces of the impact of uncertainty and learning
- show that asymmetry can be conducive to cooperative agreements.