UNIVERSITY OF EXETER

Economics Department Discussion Papers Series

ISSN 1473 – 3307

International Environmental Agreements under Uncertainty: Does the Veil of Uncertainty Help?

Michael Finus and Pedro Pintassilgo

Paper number 10/03

URL: http://business-school.exeter.ac.uk/economics/papers/

International Environmental Agreements under Uncertainty: Does the Veil of Uncertainty Help?

Michael Finus*

and

Pedro Pintassilgo**

Abstract

Na and Shin (1998) showed that the veil of uncertainty can be conducive to the success of self-enforcing international environmental agreements. Later papers confirmed this negative conclusion about the role of learning. In the light of intensified research efforts worldwide to reduce uncertainty about the environmental impact of emissions and the cost of reducing them, this conclusion is intriguing. The purpose of this paper is threefold. First, we analyze whether the result carries over to a more general setting without restriction on the number of players and which considers not only no and full learning but also partial learning. Second, we test whether the conclusion also holds if there is uncertainty about abatement costs instead of uncertainty about the benefits from global abatement. Third, we propose a transfer scheme that mitigates the possible negative effect of learning and which may even transform it into a positive effect.

Keywords: transnational cooperation, self-enforcing international environmental agreements, uncertainty, learning

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

* Department of Economics, Business School, University of Exeter, UK; m.finus@exeter.ac.uk.

** Faculty of Economics, University of Algarve, Faro, Portugal; ppintas@ualg.pt. 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.

1. Introduction

Many economic problems are characterized by externalities across agents at the national and international level. Examples include imperfect competition, research and development with imperfect appropriation of research output, international trade, contagious diseases, international terrorism and transboundary pollution (Arce and Sandler 2004). It is well-known that in the presence of externalities cooperation between agents can improve upon uncoordinated action. As shown for instance in Bloch (2003) and Yi (1997) self-enforcing cooperation proves easy (difficult) in the context of negative (positive) externalities from cooperation. If the enlargement of coalitions has a negative impact on outsiders, there are strong incentives to participate in cooperation and typically the grand coalition forms. The opposite is true for a positive impact, which applies to many examples of pure and impure public good provision. Then free-riding is encouraged and the formation of large and effective self-enforcing agreements proves difficult. Prominent and well-studied examples in the literature are international environmental agreements (IEAs).[1] The difficulties of establishing effective global cooperation are underlined by the current efforts of negotiating a Post-Kyoto Protocol.

Na and Shin (1998) point out that two issues have received little attention: a) the role of asymmetry and b) the role of uncertainty. They construct a public good model of coalition formation with three players. Players are symmetric with respect to abatement costs from individual abatement but receive different benefits from global abatement. Differences are due to different realizations of an individual benefit parameter. They show that if decisions are taken after uncertainty about the parameters is resolved, the grand coalition comprising all three players is never stable: the outcome of the game is either a two-player coalition or complete non-cooperation, depending on the degree of heterogeneity of the benefit

1 The literature dates back to Barrett (1994), Carraro and Siniscalco (1993) and Hoel (1992) and is surveyed for instance in Barrett (2003) and Finus (2003, 2008).

parameters. They contrast this case of full learning, which they call ex-post negotiations, with the case of no learning, which they call ex-ante negotiations. In the latter case, under the assumption of ex-ante symmetric expectations about the realization of the uncertain benefit parameters, they show that the grand coalition is stable. The coalitional equilibria under the two models of learning are compared from an ex-ante perspective. It is shown that the expected total payoff over all players under learning is lower than under no learning. This leads to the conclusion that the veil of uncertainty (a term coined by Brennan and Buchanan 1985) is conducive to the success of self-enforcing cooperation. Later papers with slightly different models have confirmed this conclusion (Kolstad 2007 and Kolstad and Ulph 2008 and 2009).

Considering the model of Na and Shin (1998) one wonders which are the crucial assumptions that drive results, how general they are and whether there is a way to avoid this negative outcome. One extension which we rule out from the outset is to consider a different payoff function. Though Na and Shin's payoff function is simple (linear benefits from global abatement and quadratic costs from individual abatement), it captures the main driving forces. The evaluation of the success of coalition formation requires the consideration of a particular function anyway and more complicated functions would make analytical solutions difficult to obtain in the context of uncertainty. However, we also do not find it attractive to switch to the simpler payoff function with linear benefit and cost functions considered for instance in Kolstad (2007) and Kolstad and Ulph (2008, 2009) as this leads to corner solutions in terms of equilibrium abatement choices, irrespective of the type of uncertainty.

The most obvious extension is to allow for any number of players instead of only three players. For more than three players, even under no learning the grand coalition will not necessarily form. Consequently, strategic interaction between coalition members and non-members will be present when choosing abatement levels.

The second extension is inspired by the work of Kolstad (2007), and Kolstad and Ulph (2008, 2009) who consider not only the polar cases of no and full learning (as in Na and Shin 1998), but also the intermediate case which they call partial learning. Partial learning means that players have to take their membership decision in the first stage under uncertainty, but will learn the parameter values of the payoff function before they take abatement decisions in the second stage. This implies de facto delayed learning as no uncertainty remains once players learn.[2] Hence, in these models learning takes the form of perfect learning. That is, if players learn about parameter values, no uncertainty remains.

The third and fourth extensions are motivated by the suspicion that asymmetry may be a crucial factor that leads to small coalitions under full learning. As we show in more detail below, Na and Shin's assumption implies that there is pure uncertainty about the distribution of the benefits from global abatement, but the aggregate level of benefits is known. Hence, the third extension looks at the other extreme case, namely pure uncertainty about the level of benefits with symmetric realizations of the benefit parameters. This is what Kolstad (2007) calls systematic uncertainty. We consider a general version of this assumption in the context of Na and Shin's model. In contrast, the fourth extension sticks to pure uncertainty about the distribution of benefits but considers transfers to mitigate the asymmetric distribution of the gains from cooperation. Both extensions (third and fourth) qualify the negative conclusion about the role of learning. Interestingly, it turns out that with transfers asymmetry becomes even an asset.

---

2 We stick to the terminology "partial learning" in order to relate our work to the literature on IEAs and uncertainty.

Finally, the fifth extension considers a mirror image of the Na and Shin (1998) model: pure uncertainty about the distribution of the costs from individual abatement, instead of uncertainty about the benefits from global abatement. Again, it turns out that the negative role of learning has to be qualified. For all extensions, we identify three effects (information, strategic and distributional effect) which help to explain the role of learning.

In the following, we outline our model in section 2. Section 3 analyzes the outcome under the various extensions. Section 4 summarizes the main conclusions and proposes some issues for future research.

2. Model

2.1 Coalition Formation

Consider as in Na and Shin (1998) and in many other models on IEAs that countries decide in the first stage whether to join an agreement (in which case they are called signatories) or to remain an outsider as a singleton (in which case they are called non-signatories). 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 (without restriction to $n = 3$ as in Na and Shin), m the size of coalition S, $m \leq n$, and N the set of players, $S \subseteq N$. In this simple coalition formation game, coalition structure K is entirely determined by coalition S.

In the second stage, given that some coalition S has formed in the first stage, players choose their abatement levels y_i. For a start, assume no uncertainty and that as in Na and Shin (1998) the decision is based on the following payoff function:

(1) $$\Pi_i = b_i\left(\sum_{k=1}^{n} y_k\right) - c_i y_i^2\ ,\ i \in N$$

where b_i is the parameter of the benefit function from global abatement (in the form of reduced damages, e.g. measured against some business-as-usual-scenario) and c_i the parameter of the cost function from individual abatement. Both parameters are assumed to be strictly positive.

For signatories it is assumed that they derive their equilibrium abatement levels by maximizing the aggregate payoff to their coalition

$$(2) \qquad \max_{y^S}. \sum_{k \in S} \Pi_k(S) \Rightarrow \sum_{k \in S} b_k = c_i y_i \Rightarrow \frac{\sum_{k \in S} b_k}{c_i} = y_i^*(S), \ \forall\, i \in S$$

whereas as all singletons maximize their own payoff

$$(3) \qquad \max_{y_j}. \Pi_j(S) \Rightarrow b_j = c_j y_j \Rightarrow \frac{b_j}{c_j} = y_j^*, \ \forall\, j \notin S$$

where y^S is the abatement vector of signatories, $y_i^*(S)$ is individual equilibrium abatement of a signatory, which depends on S and y_j^* is individual equilibrium abatement of a non-signatory, which is independent of S. Comparing (2) and (3), it is evident that a country ℓ will abate more if it belongs to the coalition than if it remains outside. Equilibrium payoffs of the second stage are derived by substituting equilibrium abatement levels derived from (2) and (3) into payoff functions (1), which we denote by $\Pi^*_{i \in S}(S)$ and $\Pi^*_{j \notin S}(S)$, respectively, noting that also non-signatories payoffs depend on S as benefits depend on total abatement.

Since the entire game is solved by backward induction, we now move back to the first stage, determining stable coalitions by applying the following definition:

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

(5) external stability: $\Pi_j^*(S) > \Pi_j^*(S \cup \{j\}) \ \ \forall\, j \notin S$.

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 and the set of stable coalitions is denoted by S^*. In case there is more than one stable coalition, we apply the Pareto-dominance selection criterion. We denote the set of Pareto-undominated stable coalitions by $\Psi^* \subseteq S^*$. If non-trivial coalitions are stable, they Pareto-dominate the singleton coalition structure. Note that the coalition structure comprising only singletons is stable by definition and hence existence of an equilibrium is guaranteed.[3]

Note that this coalition formation model possesses three interesting properties which are helpful for the subsequent analysis and summarized in Lemma 1.

Lemma 1: Properties of the Coalition Game

Consider payoff functions (1) with equilibrium abatement levels derived from (2) and (3). Let S *and* $\hat{S} = S \cup \{j\}$ *be two coalitions formed in the first stage where* $\hat{S}$ *is derived by one non-signatory* j *joining coalition* S.

a) Positive Externality (PE): The payoff of a country k, *which is neither a signatory of* S *nor of* $\hat{S}$, *will be strictly higher under* $\hat{S}$ *than* S, *i.e.* $\Pi^*_{k \notin S}(S) < \Pi^*_{k \notin \hat{S}}(\hat{S})$.

b) Superadditivity (SAD): The aggregate payoff of the signatories of coalition S and of a non-signatory j *is strictly lower than the aggregate payoff of the signatories of* $\hat{S}$, *including the previous non-signatory* j, *i.e.* $\sum_{i \in S} \Pi^*_i(S) + \Pi^*_{j \notin S}(S) < \sum_{i \in S} \Pi^*_i(\hat{S}) + \Pi^*_{j \in \hat{S}}(\hat{S}) = \sum_{i \in \hat{S}} \Pi^*_i(\hat{S})$.

[3] The reason is that the singleton coalition structure can be generated by $S = \varnothing$, i.e. all players announce not to join the agreement. Then if one player changes her announcement, such that $\tilde{S} = \{i\}$, the coalition structure remains the same.

c) Global Efficiency from Cooperation (GEC): The aggregate payoff and the aggregate abatement level of all countries is strictly lower under S *than* $\hat{S}$, *i.e.* $\sum_{i \in S} \Pi_i^*(S) + \sum_{k \notin S} \Pi_k^*(S) < \sum_{i \in \hat{S}} \Pi_i^*(\hat{S}) + \sum_{k \notin \hat{S}} \Pi_k^*(\hat{S})$ *and* $\sum_{i \in S} y_i^*(S) + \sum_{k \notin S} y_k^*(S) < \sum_{i \in \hat{S}} y_i^*(\hat{S}) + \sum_{k \notin \hat{S}} y_k^*(\hat{S})$.

Proof: From the F.O.C. in (2) it is evident that if country j joins coalition S, such that $\hat{S}$ forms, it will choose a higher abatement level, as well as all signatories of S, but all remaining non-signatories' abatement levels remain constant. Hence, total abatement will be higher if $\hat{S}$ than if S forms (GEC with respect to abatement). Moreover, countries which are neither signatories of S nor of $\hat{S}$ will have higher benefits but the same costs and hence higher payoffs (PE). SAD follows from $\max_{y^S} . \sum_{i \in S} \Pi_i\left(S\right) + \max_{y_j} . \Pi_j\left(S\right) < \max_{y^{\hat{S}}} . \sum_{i \in \hat{S}} \Pi_i\left(\hat{S}\right)$ because of the presence of externalities and because abatement strategies of all non-merging singletons remain constant. Finally, GEC with respect to aggregate payoffs follows from combining properties PE and SAD. (**Q.E.D.**)

Global Efficiency from Cooperation (GEC) makes it interesting from a normative point of view to analyze the prospects of cooperation. The highest (lowest) global abatement level and the (lowest) highest total payoff is obtained in the grand coalition (if all countries play as a singleton) which corresponds to the social optimum (Nash equilibrium). Note that the property GEC also holds under uncertainty, which we consider later by taking expectations over the uncertain parameters. This will be useful when evaluating the success under the three models of learning below. The Positive Externality (PE) explains why the formation of large stable coalition is difficult, despite Superadditivity (SAD) holds. Starting from no cooperation and forming gradually large coalitions, the SAD-effect is gradually outweighted by the PE-effect. Again, these effects are also present under uncertainty.

2.2 Learning Scenarios

We now assume that some parameter values of the payoff functions are uncertain. Following Na and Shin (1998) and many other papers, we assume risk-neutral agents as players are governments and not individuals. Additional to Na and Shin (1998) and as in Kolstad and Ulph (2008, 2009), we also consider the scenario of partial learning, which gives rise to the following three *learning scenarios*: 1) full learning, 2) partial learning and 3) no learning. *F*ull *L*earning (abbreviated 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 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 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 taken based on expected payoffs, though these payoffs differ from those under partial learning, given that less information is available.

Thus, viewed together, uncertainty is symmetric: all players know as much or little than their fellow players. FL and PL are identical regarding the second stage. Hence, differences and similarities between these two learning scenarios in terms of overall outcomes must be related to the first stage. Both scenarios differ from NL as abatement decisions under NL are based on expected payoffs. All three scenarios differ with respect to the first stage. The determination of stable coalitions is based on known payoffs under FL, expected payoffs given that abatement decision will be based on realized parameter values under PL and expected payoffs based on expected parameter values under NL.

In the following, we analyze four cases. Case 1 is the Na and Shin (1998) case, assuming (pure) uncertainty about the distribution of the benefits from global abatement. Case 2 is the Kolstad (2007) case, assuming (pure) uncertainty about the level of the benefits from global abatement. Case 3 is the Na and Shin (1998) case with transfers. Case 4 is a mirror image of the Na and Shin (1998) case, assuming pure uncertainty about the distribution of the costs from individual abatement, considering both, no transfers and transfers.

3. Results

3.1 Case 1: Uncertainty about the Distribution of Benefits without Transfers

Na and Shin's case requires assuming cost symmetry and hence we set $c_i = c_j = c$, $\forall\, i \in N,$ in payoff function (1) and define $\theta_i = {}^{b_i}\!/_{c}$, which we call from now onwards the benefit parameter. If this parameter is uncertain, it is represented by the random variable Θ_i with associated distribution f_{Θ_i}. Like in Na and Shin (1998), θ_i is viewed as an individual parameter. Moreover, expectations about Θ_i are symmetric, though realizations are asymmetric as the random variables are correlated. Na and Shin (1998) consider only three players with benefit parameters uniformly distributed over a set of three positive values. We also adopt a uniform distribution, but as we consider an indefinite number of players, we have to define a specific set of benefit parameter values over the set of players. Hence, we assume the following probability distribution:

$$(6) \qquad f_{\Theta_i}\left(\theta_i\right) = \begin{cases} \dfrac{1}{n} & for \quad \theta_i = k,\ k \in N \\ 0 & otherwise \end{cases}$$

which implies the following expected value, $E\left[\Theta_i\right]$, and variance, $Var\left[\Theta_i\right]$:

$$(7) \qquad E\left[\Theta_i\right] = \frac{n+1}{2} \text{ and } Var\left[\Theta_i\right] = \frac{n^2-1}{12} .$$

In this setting, correlation means that all players have a different benefit parameter: $\theta_i \neq \theta_k, \ \forall\, i \neq k \in N$. Thus, vector $\Theta = (\Theta_1, \ldots, \Theta_n)$ is composed of all the elements of N, i.e. $\bigcup_{i=1}^{n} \Theta_i = N$.

For the interpretation of this and the following cases, it is helpful to define the level of benefits from global abatement as $L = \bigcup_{i=1}^{n} \Theta_i$. Hence the marginal benefit of player i can be written as $\Theta_i = \lambda_i L$ where λ_i represents the individual share in global benefits. In case 1, the level L is constant. Thus, uncertainty is purely about the shares, or *the distribution of benefits* from global abatement.

For FL and PL, equilibrium abatement levels of the second stage follow directly from (2) and (3), assuming $c_i = c_j = c, \ \forall\, i \in N,$ and letting $\theta_i = b_i / c$. For NL, payoffs in (2) and (3) have to replaced by expected payoffs. However, as payoffs are linear in the random variables Θ_k, certainty equivalence holds. That is, the maximization of expected payoffs is equivalent to the maximization of payoffs under certainty for $\theta_k = E[\Theta_k]$.

Before proceeding to the first stage, it is already informative to compare second stage outcomes. For this, we take an ex-ante perspective and compute equilibrium expected abatement and payoff levels also in the case of FL and PL. Note that there is an immediate link between individual and total expected levels: individual levels are a fraction n of total levels. This is due to the ex-ante symmetry of all countries – they do not know whether they will be signatories or non-signatories. Thus, in the following analysis and proofs, we concentrate on total expected levels.

Lemma 2: Expected Abatement and Payoffs in the Second Stage in Case 1

Let $K=\{S,1_{n-m}\}$ *be some coalition structure with coalition* S *of size* m*. Under all three learning scenarios, the following relations hold ex-ante in case 1:*

Individual and Total Expected Abatement Levels: $FL=PL=NL$.

Individual and Total Expected Payoff Levels: $FL=PL\leq NL$ *with strict inequality if* $S\neq N$.

Proof: See Appendix 1. **(Q.E.D.)**

It is clear that FL and PL are identical as there are no differences in the second stage. With respect to abatement, there is also no difference to NL. This suggests that despite there is over- and undershooting in terms of optimal abatement levels under NL, compared to the realizations of the random benefit parameter Θ_i, on average this cancels out. This is different for payoffs.

Consider first the grand coalition, $S=N$, corresponding to the social optimum. Then there is no strategic interaction between players. We call the payoff difference between learning and no learning in the grand coalition the information effect from learning. In other words, the information effect measures the value of information in the absence of any strategic interaction and stability considerations. A priori we know that this information effect cannot be negative – it can only be zero at worst. For $S=N$ the first order conditions require setting the sum of marginal benefits over all players equal to individual marginal abatement costs. By assumption, in case 1, also under NL the sum of marginal benefits is known. Hence, the information effect from learning is zero.

Consider now any other coalitions structure different from the grand coalition $S\neq N$ where there is interaction between players. We call the payoff difference between learning and no learning in these coalition structures the strategic effect from learning. This effect is negative according to Lemma 2. In order to explore the intuition behind this result, let us take the extreme case where no non-trivial coalition forms, which corresponds to the Nash equilibrium.[4] Under FL and PL, all countries will choose a different abatement level, as all have a different parameter $\theta_i = {}^{b_i}\!/_{c}$ due to different b_i's. This is not cost-effective as all have the same cost parameter c. In contrast, under NL, due to symmetric expectations about Θ_i, all countries choose the same abatement level which is also inefficient as under FL and PL (i.e. marginal abatement costs are not set equal to the sum of marginal benefits), but at least cost-effective (i.e. all marginal abatement costs are equal). Thus, the negative strategic effect from learning is a cost-effectiveness effect here. Put differently, getting it on average right across all players in terms of costs is more important than getting it individually right.

We now move to the first stage and determine stable coalitions. In order to make coalition formation interesting, we assume henceforth $n \geq 3$. We find:

Lemma 3: Equilibrium Coalitions in the First Stage in Case 1

In case 1, under the three scenarios of learning, the expected equilibrium coalition size $E[m^*]$ *is given by:*

$$E\left[m^{*PL}\right] = E\left[m^{*NL}\right] = 3 \quad \text{and} \quad E\left[m^{*FL}\right] = \begin{cases} 1 & if \quad n = 3 \\ 2 & if \quad n \geq 4 \end{cases}$$

where under FL for $n \geq 4$ *the only stable coalition is formed by the two players with the highest* θ_i.

Proof: See Appendix 2. **(Q.E.D.)**

For NL the intuition is straightforward. As pointed out above, due to certainty equivalence, equilibrium abatement levels correspond to those under certainty if the parameters θ_k are equal to the expected value of Θ_k. Due to ex-ante symmetric expectations, we have

4 The idea is illustrated for two players in an emission game in Ulph (1998). Other examples with negative value of information in non-cooperative equilibria are discussed for instance in Gollier and Treich (2003).

symmetric payoff functions. For payoff function (1) it is well-known from the literature (see e.g. Finus 2003) that the stable coalition comprises three signatories if $n \geq 3$.[5] Also under PL the ex-ante symmetry leads to the same stable coalitions, though certainty equivalence does not hold. This is different for FL where due to asymmetric realizations of Θ_i signatories receive asymmetric payoffs, implying that only smaller coalitions are stable. If $n \geq 4$, only the two countries with the highest benefit parameter find it attractive to form a coalition.

The driving force of this result is what we call the distributional effect from learning: the payoff difference of various degrees of learning due to the stability of different coalitions. The intuition 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.[6] Since stable coalitions depend on second stage outcomes, it is generally difficult to disentangle the distributional from the strategic effect. However, this poses no problem when FL and PL are compared as second stage outcomes are identical: it is the payoff difference between FL and PL due to different coalition sizes, resulting from different distributions of the gains from co-operation among coalition members.

Hence, compared to Na and Shin (1998), we also confirm for PL the stable coalition size of 3, but this is not the grand coalition as long as there are more than three countries. Like in Na and Shin (1998) the coalition size under FL falls short of the coalition size under NL, and, as just confirmed, also under PL.

---

5 Note that similar small coalitions are obtained for other strictly concave payoff functions as long as one does not assume Stackelberg leadership of signatories (see Finus 2003).

6 The idea is also illustrated in a simple two-player model in Helm (1998) and in Kolstad (2005).

We now combine the first and second stage outcomes to evaluate the overall success of IEAs.

Proposition 1: Outcome in Case 1 (Uncertainty about the Distribution of Benefits without Transfers)

In case 1, under the full, partial and no learning scenario, expected equilibrium total abatement levels and expected total payoffs are ranked as follows:

1) *Total Abatement:* $NL = PL > FL$

2) *Total Payoff:* $\begin{cases} NL = PL > FL & if \quad n = 3 \\ NL > PL > FL & if \quad n \geq 4 \end{cases}$.

Proof: Relations between NL and PL follow directly from Lemmas 2 and 3. For the relation between PL and FL, we note that second stage outcomes are the same. For $n = 3$, $E\left[m^{*PL}\right] = 3 > 1 = E\left[m^{*FL}\right]$ and hence $PL > FL$ follows from property GEC (Lemma 1). For $n > 3$, though $E\left[m^{*PL}\right] = 3 > 2 = E\left[m^{*FL}\right]$, the identity of players matters for FL, as a two-player coalition among the players with the highest θ_i produces a higher abatement and payoff level than an average two-player coalition. Hence, we compute expected abatement and payoffs under PL assuming $E\left[m^{*PL}\right] = 3$ and compare them with those levels under FL where a coalition among the two countries with the highest benefit parameters is formed, $n-1$ and n, which delivers the result. The relations between NL and FL follow directly from the relations between NL and PL as well as PL and FL. **(Q.E.D.)**

Hence, the negative role of learning as concluded by Na and Shin (1998) is confirmed for more than three players including the intermediate case of partial learning. The strategic effect from learning is negative, leading to worse outcomes under FL and PL. Additionally, the distributional effect from learning makes FL even worse than PL.

3.2 Case 2: Uncertainty about the Level of Benefits

It became evident that the asymmetric realization of the random benefit variable Θ_i is a driving force for the negative outcome under learning, regardless whether we consider PL or FL. Hence, we consider the other extreme assumption with pure uncertainty about the level of the benefits from global abatement. In order to capture this, we again assume cost symmetry, $c_i = c_j = c \;\; \forall\, i \in N$, and again define $\theta_i = b_i / c$. However, now uncertainty does not relate to an individual but a common parameter, which Kolstad (2007) and Kolstad and Ulph (2008, 2009) call systematic uncertainty. Again, all players have the same expectations ex-ante, but now, once uncertainty is resolved, all countries have also the same benefit parameter ex-post, $\Theta_i = \Theta_k \;\; \forall\, i, k \in N$. Thus, uncertainty is perfectly and positively correlated. More important, however, it is to note that in case 2 uncertainty is de facto about the *level of the benefits* from global abatement. For the subsequent analysis, no assumption about the functional form of the probability distribution f_{Θ_i} is required (as for instance in Kolstad 2007). We proceed as in case 1, looking first at second stage outcomes, then at first stage outcomes and finally combine both.

Lemma 4: Expected Abatement and Payoffs in the Second Stage in Case 2

Let $K = \{S, 1_{n-m}\}$ *be some coalition structure with coalition* S *of size* m*. Under all three learning scenarios, the following relations hold ex-ante in case 2:*

Individual and Total Expected Abatement Levels: $FL = PL = NL$.

Individual and Total Expected Payoff Levels: $FL = PL > NL$.

Proof: See Appendix 1. **(Q.E.D.)**

Again, over- and undershooting under NL compared to the realizations of Θ_i, which is now systematic, cancels out on average. However, the relations in terms of payoffs are reversed to case 1. The symmetric realizations of Θ_i implies that also under FL and PL abatement is cost-effectively allocated in all possible coalitions structures. Though overshooting under NL is associated with additional benefits, this is costly due to strictly convex cost functions. The additional costs are higher than the cost savings when there is undershooting. Thus, more information is beneficial if there is systematic uncertainty. This is confirmed when evaluating overall outcomes in Proposition 2, as Lemma 5 shows that first stage outcomes are the same under all three scenarios of learning.

Lemma 5: Equilibrium Coalitions in the First Stage in Case 2

In case 2, under the three scenarios of learning, the expected equilibrium coalition size $E[m^*]$ *is given by:*

$$E\left[m^{*FL}\right] = E\left[m^{*PL}\right] = E\left[m^{*NL}\right] = 3 \ .$$

Proof: See Appendix 2. **(Q.E.D.)**

Proposition 2: Outcome in Case 2 (Uncertainty about the Level of Benefits)

In case 2, under the full, partial, and no learning scenario, expected equilibrium total abatement levels and expected total payoffs are ranked as follows:

1) *Total Abatement:* $FL = PL = NL$

2) *Total Payoff:* $FL = PL > NL$.

Proof: Follows immediately from Lemma 4 and 5. **(Q.E.D.)**

Taken together, our intuition that Na and Shin's negative result hinges on the asymmetric realization of the random benefit variable Θ_i is confirmed. The strategic effect from learning is now positive and hence no longer upsets second stage outcomes under FL and PL (see Lemma 4). Additionally, the distributional effect from learning is zero, posing no longer a disadvantage to FL in terms of the size of stable coalitions (see Lemma 5).

3.3 Case 3: Uncertainty about the Distribution of Benefits with Transfers

Another possibility to avoid the negative outcome of case 1 is transfers. Transfers can address the negative distributional effect from learning. They are relevant for FL, but have no effect on PL and NL as membership decisions are based on expected payoffs which are symmetric among signatories and also among all non-signatories. Therefore, transfers will affect the ranking between FL and PL, FL and NL, but not the ranking between PL and NL.

More precisely, the analysis proceeds as follows. First, note that the maximization procedure in the second stage (see (2) and (3)) implies transferable utility and hence equilibrium abatement levels are not affected by transfers. (Hence, transfers have no impact on the information and strategic effect.) This is different for first stage outcomes. Generally speaking, many transfer schemes have been considered in the literature on IEAs (e.g. Barrett 2001, Botteon and Carraro 1997, Bosello et al. 2003, Eyckmans and Finus 2006 and Weikard et al. 2009). However, in order to avoid the sensitivity of outcomes regarding particular assumptions, we apply the concept of an almost ideal sharing scheme (AISS) proposed by Eyckmans and Finus (2009) of which similar notions are found in Fuentes-Albero and Rubio (2009), McGinty 2007 and Weikard (2009). They argue that if and only if:

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

holds, then there exists a transfer system which makes S internally stable. A transfer system for which every potentially internally stable coalition is internally stable belongs to the AISS, which gives each coalition member its free-rider payoff, $\Pi_i^*(S\setminus\{i\})$, plus some positive share of the surplus $\sigma(S)=\sum_{i\in S}\Pi_i^*(S)-\sum_{i\in S}\Pi_i^*(S\setminus\{i\})$. For every transfer system belonging to the AISS, coalition S is externally stable if and only if all larger coalitions $S \cup \{j\}$, considering all possible fringe player $j \notin S$ joining coalition S, are not potentially internally stable. Taken together, this means that every transfer scheme belonging to AISS will lead to the same set of internally and externally stable coalitions and hence set of stable coalitions (robustness). Most important, among those coalitions that can be potentially internally stabilized, i.e. $\sigma(S) \geq 0$ (which may not be possible for large coalitions because $\sigma(S) < 0$), AISS stabilizes (in the sense of internal and external stability) those with the highest aggregate welfare over all players (optimality). This property hinges on only one structural property, namely the (weakly) positive externality (PE) from coalition formation. As known from Lemma 1, this property holds even in its strong version in our coalition formation game. Applying AISS to our setting, we derive the following result.

Lemma 6: Equilibrium Coalitions in the First Stage in Case 3

In case 3, under the three scenarios of learning, the expected equilibrium coalition size $E[m^*]$ *is given by:*

$$E\left[m^{*PL}\right] = E\left[m^{*NL}\right] = 3 \quad \text{and} \quad E\left[m^{*FL}\right] = \begin{cases} 3 & \text{if} \quad n \leq 8 \\ f(n) > 3 & \text{if} \quad n \geq 9 \end{cases}$$

where under FL *all possible 3-player coalitions are stable if* $n \leq 8$*, no stable coalition comprises less than three players if* $n \geq 9$ *and* $f(n)$ *increases in* n.

Proof: See Appendix 2. **(Q.E.D.)**

As pointed out above, the outcomes of PL and NL are not affected by transfers. The coalition size under FL does no longer fall short of those under PL and NL. To the contrary, if n is large enough (i.e. $n \geq 9$), the coalition size will be larger than under PL and NL. Due to the assumption about the distribution of the variables Θ_i, the degree of asymmetry among players (measured as the variance of the elements of the vector Θ)

increases with the number of players n. This asymmetry is conducive to the size of stable coalitions if accompanied by an appropriate transfer scheme. The intuition is the following.

Cooperation among some players compared to the non-cooperative status quo typically serves two purposes. First, internalizing an externality among coalition members by choosing higher abatement levels than under no cooperation. This is a benefit every coalition member enjoys and, in fact, also non-signatories, as exemplified by the property PE (see Lemma 1). Second, equalizing marginal abatement costs across coalition members and hence reaping the gains from cost-effectiveness. This is an exclusive benefit only the coalition enjoys as a group that does not spread to non-signatories. This is captured by the property superadditivity (see Lemma 1). This exclusive benefit is higher for heterogeneous than for symmetric players. It is particularly pronounced here as without cooperation payoffs would be very low: the Nash equilibrium would be particularly inefficient as marginal abatement costs differ between players under FL (as they do under PL). Taken together, compared to case 1, under FL, transfers reduce the negative distributional effect from learning or may even transform it into a positive effect. Overall, this leads to the following result.

Proposition 3: Outcome in Case 3 (Uncertainty about the Distribution of Benefits with Transfers)

In case 3, under the full, partial, and no learning scenario, expected equilibrium total abatement levels and expected total payoffs are ranked as follows:

1) Total Abatement:

$$\begin{cases} FL = PL = NL & \text{if} \quad n \le 8 \\ FL > PL = NL & \text{if} \quad n \ge 9 \end{cases}$$

2) Total Payoff:

$$\begin{cases} FL = PL = NL & \text{if} \quad n = 3 \\ NL > FL = PL & \text{if} \quad 4 \le n \le 8 \\ NL > FL > PL & \text{if} \quad n = 9 \\ FL > NL > PL & \text{if} \quad n \ge 10 \end{cases}$$

Proof: The ranking between PL and NL follows from Proposition 2 as transfers have no effect. The ranking between PL and FL follows from the same second stage outcomes (Lemma 2), the same or different first stage outcomes (depending on n) as given in Lemma 6 and applying property GEC (Lemma 1). The ranking between FL and NL is established in three steps. Step 1: For total abatement, we combine second stage outcomes in Lemma 2, first stage outcomes in Lemma 6 and apply property GEC. Step 2: For $n = 3$ and $4 \leq n \leq 8$, the same reasoning as in step 1 applies for the relation between total payoffs under FL and NL. Step 3: For $n \geq 9$, NL produces better second stage outcomes than FL (Lemma 2), but FL produces larger stable coalitions (Lemma 6). Hence, under FL, for each $n \geq 9,$ we consider all possible θ-vectors and compute the expected total payoff over all Pareto-undominated stable coalitions.[7] Then, we compare this to the expected total payoff under NL with $E\left[m^{*NL}\right] = 3$. **(Q.E.D.)**

Taken together, the ranking between FL and NL, as established by Na and Shin (1998) and which is our case 1, changes with transfers if the degree of asymmetry between players is large enough. Then the veil of uncertainty is no longer conducive to cooperation but has a negative impact. Thus, transfers may be seen as a successful safety valve or hedging strategy in the presence of uncertainty about the distribution of benefits.

3.4 Case 4: Uncertainty about the Distribution of Costs

In this final section, we consider a mirror image of Na and Shin's case (case 1), namely uncertainty about the distribution of abatement costs. For completeness, we consider a subcase without and a subcase with transfers. The mirror image means benefit symmetry, i.e. $b_i = b_j = b \;\; \forall\, i \in N$, but cost asymmetry in payoff function (1). For simplicity, we again define $\theta_i = b/c_i$ and assume the same distribution as in (6). Now, correlated uncertainty

7 As we could not obtain a closed form solution for the expected total payoff, the values in step 3 were obtained through an algorithm programmed with the software package Matlab.

stems from the cost parameter c_i, with $\theta_i \neq \theta_k$, $\forall\, i \neq k \in N$, $\bigcup_{i=1}^{n} \Theta_i = N$. For second stage outcomes, we find:

Lemma 7: Expected Abatement and Payoffs in the Second Stage in Case 4

Let $K = \{S, 1_{n-m}\}$ be some coalition structure with coalition S of size m. Under all three learning scenarios, the following relations hold ex-ante in case 4:

Individual and Total Abatement Levels: $FL = PL > NL$.

Individual and Total Payoff Levels: $FL = PL > NL$.

Proof: See Appendix 3. **(Q.E.D.)**

Comparing case 4 (Lemma 7) with case 1 (Lemma 2) clearly shows that relations in terms of expected total payoffs are reversed. Both the information and the strategic effect, when comparing FL or PL with NL, are now positive. In the presence of asymmetric abatement cost functions, additional information allows non-signatories to better target abatement levels and signatories to allocate abatement duties cost-effectively. The cost savings compared to NL show up in higher payoffs and allows choosing higher abatement levels on average under FL and PL. For first stage outcomes, we establish the following.

Lemma 8: Equilibrium Coalitions in the First Stage in Case 4

In case 4, under the three scenarios of learning, the expected equilibrium coalition size $E[m^*]$ is given by:

$E\left[m^{*PL}\right] = E\left[m^{*NL}\right] = 3$ *and*

$E\left[m^{*FL}\right] = 2$ *in the case of no transfers where the only stable coalition is formed by the two players with lowest θ_i and*

$E\left[m^{*FL}\right] = 3$ *in the case of transfers, where all coalitions of three players are stable.*

Proof: See Appendix 3. **(Q.E.D.)**

Again, the result for coalition sizes under the scenarios NL and PL are not surprising due to symmetric expected payoffs on which membership decisions are based. Without transfers, FL suffers from the negative distributional effect from learning, leading to smaller coalitions than under PL and NL. In contrast to case 1, it is now the two players with the lowest parameter θ_i, i.e. the steepest abatement cost function, which form a stable coalition under FL. For all other players, the free-rider incentive is stronger than the cost-saving potential from cooperation. With transfers, the distributional effect from learning is zero as the size of stable coalitions is the same as under PL and NL. Different from case 3, under FL no larger coalition than three will be stable. The intuition is that now the cost-saving potential from cooperation with learning is lower than in case 3 as the benchmark case without cooperation, i.e. the Nash equilibrium, is no longer cost-ineffective. Thus, through a higher cost-saving potential, uncertainty about the distribution of benefits (case 3) provides a higher incentive for cooperation than uncertainty about the distribution of costs (case 4).

Pulling Lemma 7 and 8 together, an overall comparison follows almost immediately.

Proposition 4: Outcome in Case 4 (Uncertainty about the Distribution of Costs)

In case 4, under the full, partial, and no learning scenario, expected equilibrium total abatement levels and expected total payoffs are ranked as follows:

<u>*No Transfers*</u>

1) *Total Abatement:* $\begin{cases} PL > NL > FL & if \quad 3 \le n \le 8 \\ PL > FL > NL & if \quad n \ge 9 \end{cases}$

2) *Total Payoff:* $\begin{cases} PL > NL > FL & if \quad 3 \le n \le 7 \\ PL > FL > NL & if \quad n \ge 8 \end{cases}$

<u>*Transfers*</u>

1) *Total Abatement:* $FL = PL > NL$

2) *Total Payoff:* $FL = PL > NL$.

Proof: For transfers all relations follow directly from Lemma 7 and 8. For no transfers, relations between PL and NL also follow directly from Lemma 7 and 8. For FL, we compute expected total abatement and payoff levels using the information that the players with the lowest θ_i values, i.e. players 1 and 2, form a two player coalition. For PL and NL we compute expected payoffs using the information that $E\left[m^{*PL}\right]=E\left[m^{*NL}\right]=3$. A comparison then delivers the ranking between FL and PL as well as FL and NL. **(Q.E.D.)**

Hence, with transfers, learning, in the form of FL or PL, is always better than no learning, NL, though there is no difference between PL and FL. Without transfers, PL is always better than FL and NL. FL is better than NL if n is sufficiently large. The intuition is that for sufficiently large n the first stage advantage of NL over FL becomes less important (the effect of a slightly larger coalition diminishes in the presence of many non-signatories), giving more weight to the second stage advantage of FL over NL.

4. Summary and Conclusions

The starting point of our analysis was the negative conclusion of Na and Shin (1998) about the value of information in a public good game of coalition formation. We addressed two general questions. How general is this result? What are the driving forces? The first question was analyzed by pointing out that in Na and Shin (1998) uncertainty is symmetric and about the parameters of the payoff functions. In particular, uncertainty is about the distribution of the benefits from the provision of the public good "global abatement" with ex-ante symmetric expectations but ex-post asymmetric realizations of the benefit parameter. We confirmed Na and Shin's negative result in a more general setting which allowed for any number of players and considered not only the learning scenarios no and full learning, but also the intermediate scenario partial learning. In a two-stage coalition formation game, no learning means that players never learn the true parameter values, partial learning that they find out about them after the first (membership decision) but before the second stage (abatement decision), and full learning that they know them before the first stage. We denoted this setting case 1. This was contrasted with cases 2, 3 and 4. Case 2 did not focus on uncertainty about the distribution but the level of the benefits from public good provision. Case 3 stuck to uncertainty about the distribution of the benefits but introduced a transfer scheme to balance asymmetric gains from cooperation among coalition members. Case 4 was the mirror image of case 1 and considered uncertainty about the distribution of the costs of public good provision. For completeness, this was done not only without but also with transfers.

It became apparent that a departure from Na and Shin's setting leads to more positive results about the role of learning. In case 2 the value of information is always positive, when comparing full or partial learning with no learning. In case 3, the negative value of information of case 1 could be partly mitigated. Transfers left no and partial learning unaffected but improved upon full learning. The larger the asymmetries among players, in the form of different realizations of the benefit parameter, the larger are the gains from cooperation which can now be fully reaped through a transfer scheme. In other words, diversity becomes an asset if accompanied by an appropriate transfer scheme. Also in case 4, the mirror image of case 1, which assumed uncertainty about the distribution of the costs, the negative effect of learning was less pronounced and was clearly positive with transfers.

In all four cases, three main effects (though to a different extent and with different signs) were at work, called the information, strategic and distribution effect. The information effect basically means that, in the absence of strategic interaction, i.e. when the grand coalition forms, more information cannot decrease the expected aggregate payoff. The strategic effect refers to the impact of learning on expected payoffs for all coalition structures where there is strategic interaction between players, i.e. the grand coalition does not form. We showed that in Na and Shin's setting this strategic effect from learning is negative such that full and partial learning rank worse than no learning in terms of expected payoffs. Finally, once stability of coalitions is considered, the distribution of the gains from cooperation among coalition members becomes an issue. In Na and Shin's setting this distributional effect from learning is negative under full learning. Asymmetry of the benefit parameter translates into asymmetry of the gains from cooperation, upsetting large stable coalition. With transfers, however, this asymmetries can be balanced. The distributional effect may become even positive – leading to larger stable coalitions than under no and partial learning – as the aggregate gains from cooperation to coalition members increase with diversity.

Taken together, our results suggest that the veil of uncertainty may be good if the uncertainty about the distribution is larger than about the level of the gains from cooperation. This is particularly true if the distribution of the gains from cooperation is more asymmetric ex-post than ex-ante expected. For such cases, a transfer scheme is helpful which hedges against possible asymmetries. Since asymmetries are the result of diversity, and diversity allows for larger comparative advantages from cooperation, transfers may even turn an apparent disadvantage into an advantage. Moreover, it appears that Na and Shin's setting of pure uncertainty about the distribution of the benefits from the provision of a public good is an interesting but also a special assumption. For many economic problems, uncertainty about the level of the benefits as well as uncertainty about the costs of public good provision will be important too. In such cases the positive effect of the veil of uncertainty is less evident from our model which is certainly a relief to all those that believe that learning should be beneficial.

References

Arce M., D. G. and T. Sandler (2001), Transnational Public Goods: Strategies and Institutions. "European Journal of Political Economy", vol. 17, pp. 493-516.

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.

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.

Botteon, M. and C. Carraro (1997), Burden-sharing and Coalition Stability in Environmental Negotiations with Asymmetric Countries. In: Carraro, C. (ed.), International Environmental Negotiations: Strategic Policy Issues. Edward Elgar, Cheltenham, UK et al., ch. 3, pp. 26-55.

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

Eyckmans, J. and M. Finus (2006), Coalition Formation in a Global Warming Game: How the Design of Protocols Affects the Success of Environmental Treaty-making. "Natural Resource Modeling", vol. 19(3), pp. 323-358.

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. (2008), Game Theoretic Research on the Design of International Environmental Agreements: Insights, Critical Remarks and Future Challenges. "International Review of Environmental and Resource Economics", vol. 2(1), pp. 29-67.

Fuentes-Albero, C. and S.J. Rubio (2009), Can the International Environmental Cooperation Be Bought? Forthcoming European Journal of Operation Research.

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.

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.

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 (2009), Uncertainty, Learning and Heterogeneity in International Environmental Agreements. Mimeo.

McGinty, M. (2007), International Environmental Agreements among Asymmetric Nations. "Oxford Economic Papers", vol. 59(1), 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.

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.

Weikard, H.-P. (2009). Cartel Stability under Optimal Sharing Rule. The Manchester School, vol. 77(5), 575-593.

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. (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[8]

Appendix 1: Proof of Lemmas 2 and 4

For FL and PL, abatement and payoffs in cases 1, 2 and 3, are given by (using (2) and (3) in the text).

(I) $$Y^*(S) = m\sum_{i\in S}\theta_i + \sum_{j\notin S}\theta_j$$

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

(II) $$\Pi^*_{j\notin S}(S) = \theta_j\left(m\sum_{i\in S}\theta_i + \sum_{k\notin S}\theta_k\right) - \frac{1}{2}(\theta_j)^2 = \frac{1}{2}(\theta_j)^2 + m\theta_j\sum_{i\in S}\gamma_i + \theta_j\sum_{k\neq j\notin S}\theta_k$$

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

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

Taking expectations gives:

Cases 1 and 3:

(III) $$E\left[Y^{FL=PL}(S,\Theta)\right] = \left(m^2 - m + n\right)E\left[\Theta_i\right] = \frac{\left(m^2 - m + n\right)(n+1)}{2}$$

$$E\left[\Pi^{FL=PL}_{i\in S}(S,\Theta)\right] = \frac{(n+1)\left(6n^2 + \left(3m^2 - 5m + 4\right)n + 2m^2 - 4m\right)}{24}$$

(IV) $$E\left[\Pi^{FL=PL}_{j\notin S}(S,\Theta)\right] = \frac{(n+1)\left[3n^2 + \left(3m^2 - 3m + 1\right)n + 2m^2 - 2m - 1\right]}{12}$$

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

$$E\left[\Pi^{FL=PL}\left(S,\Theta\right)\right]=\frac{\left(n+1\right)\left(6n^3+\left(6m^2-6m+2\right)n^2+\left(-3m^3+5m^2-2m-2\right)n-2m^3+2m\right)}{24}$$

Case 2:

(V) $$E\left[Y^{FL=PL}\left(S,\Theta\right)\right]=\left(m^2-m+n\right)E\left[\Theta_i\right]$$

$$E\left[\Pi_{i\in S}^{FL=PL}\left(S,\Theta\right)\right]=\left(\frac{m^2}{2}-m+n\right)E\left[\Theta_i^2\right]$$

(VI) $$E\left[\Pi_{j\notin S}^{FL=PL}\left(S,\Theta\right)\right]=\left(m^2+n-m-\frac{1}{2}\right)E\left[\Theta_j^{\;2}\right]$$

$$E\left[\Pi^{FL=PL}\left(S,\Theta\right)\right]=\left(m^2\left(n-\frac{m}{2}\right)+\left(n-\frac{1}{2}\right)\left(n-m\right)\right)E\left[\Theta_k^2\right]$$

In this case $E\left[\Theta_k^2\right]$ remains unspecified, as no assumption about the distribution of the random variables is necessary for the analysis.

For NL, certainty equivalence holds. Thus, in cases 1, 2 and 3, payoffs and abatement are the same as those under certainty with $\theta_k=E\left[\Theta_k\right]\ \forall\, k\in N$:

(VII) $$E\left[Y^{NL}\left(S,\Theta\right)\right]=\left(m^2-m+n\right)E\left[\Theta_k\right]$$

$$E\left[\Pi_{i\in S}^{NL}\left(S,\Theta\right)\right]=\left(\frac{m^2}{2}-m+n\right)\left(E\left[\Theta_i\right]\right)^2$$

(VIII) $$E\left[\Pi_{j\notin S}^{NL}\left(S,\Theta\right)\right]=\left(m^2-m+n-\frac{1}{2}\right)\left(E\left[\Theta_j\right]\right)^2$$

$$E\left[\Pi^{NL}\left(S,\Theta\right)\right]=\left(m^2\left(n-\frac{m}{2}\right)+\left(n-\frac{1}{2}\right)\left(n-m\right)\right)\left(E\left[\Theta_k\right]\right)^2$$

where in cases 1 and 3 $E\left[\Theta_k\right]=\left(n+1\right)/2$.

The equality of expected total abatement in the three learning scenarios follows directly from (III) and (VII) for case 1 (Lemma 2), and (V) and (VII) for case 2 (Lemma 4). Regarding expected total payoffs, using (IV) and (VIII), we find for case 1 (Lemma 2):

$E\left[\Pi^{FL=PL}\left(S,\Theta\right)\right]-E\left[\Pi^{NL}\left(S,\Theta\right)\right]=-\frac{\left(n+1\right)\left(n+m^{2}-1\right)\left(n-m\right)}{24}$ which is strictly negative for $n>m$, implying $S\neq N$, and zero if $n=m$, implying $S=N$, for all $m,n\in N$ and $m\leq n$. For case 2 (Lemma 4), using (VI) and (VIII), we find:

$E\left[\Pi^{FL=PL}\left(S,\Theta\right)\right]-E\left[\Pi^{NL}\left(S,\Theta\right)\right]=Var\left[\Theta_{k}\right]$ where $Var\left[\Theta_{k}\right]>0$ by assumption.

Appendix 2. Proof of Lemmas 3, 5 and 6

Lemma 3: For PL and NL we use expected payoffs in (IV) and (VIII), respectively, and apply definition (4) of internal and definition (5) of external stability which delivers the result. For FL, in case 1 we note that there are θ-vectors with asymmetric entries. Consequently, $E[m^{*FL}]<3$, as it can be shown, using payoffs in (II) and the definition of internal stability in (4) that for all non-symmetric θ-vectors no coalition of three or more players is internally stable. The particular result that only the single coalition is stable if $n=3$ and comprises the two players with the highest θ_i if $n\geq 4$ is also immediately derived by using (II) and (4).

Lemma 5: For PL and NL the result follows from applying definition (4) of internal and definition (5) of external stability to payoffs (VI) and (VIII), respectively. For FL the equilibrium coalition size immediately follows from symmetry.

Lemma 6: Transfers do not affect the outcomes under PL and NL. For FL we first prove that all coalitions of three or less players are potentially internally stable using payoffs in (II) and the definition of potentially internal stability in (8). Given the relation between potential internal stability and external stability, it follows that all coalitions strictly smaller than 3 must be externally unstable and hence cannot be stable. Thus, $E[m^{*FL}] \geq 3$ follows. Now it suffices to show that potential internal stability is violated for all coalitions larger than three players if $n \leq 8$, considering all possible θ-vectors in case 3 up to these thresholds (and hence $E[m^{*FL}] = 3$ follows). Above these thresholds, we show that there is at least one θ-vector of the form $(1,2,3,n)$ for which potential internal stability holds. Since either one these 4-player coalitions or larger coalitions are externally stable, we can conclude that $E[m^{*FL}] > 3$.

Appendix 3. Proof of Lemmas 7 and 8

Lemma 7: In case 3, assuming $b_i = b_j = b \;\; \forall\, i \in N$, using $\theta_i = b / c_i$, the payoff function can be written as

(IX) $$\Pi_i = \left(\sum_{k=1}^{n} y_k \right) - \frac{1}{\theta_i} y_i^2 \;,\; i \in N$$

with abatement and payoffs under FL and PL as follows:

(X) $$\begin{cases} y_i = m\theta_i, & \forall\, i \in S \\ y_j = \theta_j, & \forall\, j \notin S \end{cases} \text{ and } Y^*(S) = m \sum_{i \in S} \theta_i + \sum_{j \notin S} \theta_j$$

$$\Pi^*_{i \in S}(S) = \left(m \sum_{\ell \in S} \theta_\ell + \sum_{j \notin S} \theta_j \right) - \frac{m^2 \theta_i}{2}$$

(XI) $$\Pi^*_{j \notin S}(S) = \left(m \sum_{\ell \in S} \theta_\ell + \sum_{k \notin S} \theta_k \right) - \frac{\theta_j}{2}$$

$$\Pi^*(S) = m\left(n - \frac{m}{2} \right) \sum_{\ell \in S} \theta_\ell + \left(n - \frac{1}{2} \right) \sum_{k \notin S} \theta_k$$

The corresponding expected values are:

(XII) $$E\left[Y^{FL=PL}\left(S,\Theta\right)\right]=\left(m^2-m+n\right)E\left[\Theta_i\right]=\frac{\left(m^2-m+n\right)\left(n+1\right)}{2}$$

$$E\left[\Pi_{i\in S}^{FL=PL}\left(S,\Theta\right)\right]=\frac{\left(m^2-2m+2n\right)\left(n+1\right)}{4}$$

(XIII) $$E\left[\Pi_{j\notin S}^{FL=PL}\left(S,\Theta\right)\right]=\frac{\left(2m^2-2m+2n-1\right)\left(n+1\right)}{4}$$

$$E\left[\Pi^{FL=PL}\left(S,\Theta\right)\right]=\frac{\left(n+1\right)\left[2n^2+\left(2m^2-2m-1\right)n-m^3+m\right]}{4}$$

Under NL, given a coalition structure $K=\left\{S,1_{(n-m)}\right\}$ has formed in the first stage, equilibrium abatement strategies are given by:

(XIV) $$\begin{cases} E\left[y_i\left(S,\Theta\right)\right]=m\left(E\left[\dfrac{1}{\Theta_i}\right]\right)^{-1}, & \forall i\in S \\ E\left[y_j\left(S,\Theta\right)\right]=\left(E\left[\dfrac{1}{\Theta_j}\right]\right)^{-1}, & \forall j\notin S \end{cases}$$

$$E\left[Y^{NL}\left(S,\Theta\right)\right]=\left(m^2-m+n\right)\left(E\left[\frac{1}{\Theta_k}\right]\right)^{-1}$$

which give rise to the following expected payoffs in the second stage

$$E\left[\Pi_{i\in S}^{NL}\left(S,q^{**},\Theta\right)\right]=\left(\frac{m^2}{2}-m+n\right)\left(E\left[\frac{1}{\Theta_i}\right]\right)^{-1}$$

(XV) $$E\left[\Pi_{j\notin S}^{NL}\left(S,q^{**},\Theta\right)\right]=\left(m^2-m+n-\frac{1}{2}\right)\left(E\left[\frac{1}{\Theta_j}\right]\right)^{-1}$$

$$E\left[\Pi^{NL}\left(S,q^{**},\Theta\right)\right]=\left(m\left(-\frac{m^2}{2}+nm-n+\frac{1}{2}\right)+n\left(n-\frac{1}{2}\right)\right)\left(E\left[\frac{1}{\Theta_k}\right]\right)^{-1}$$

Using (XII) and (XIV) we get:

$$E\left[Y^{FL=PL}\left(S,\Theta\right)\right]-E\left[Y^{NL}\left(S,\Theta\right)\right]=\left(\frac{m^2}{2}-m+n\right)\left(E\left[\Theta_k\right]-\left(E\left[\frac{1}{\Theta_k}\right]\right)^{-1}\right)$$

$$=\left(\frac{m^2}{2}-m+n\right)\left(n\left(\frac{1}{2}-\frac{1}{\sum_{i=1}^{n}\frac{1}{i}}\right)+\frac{1}{2}\right)>0,\quad \forall\, n\geq 3\wedge m\leq n$$

Using (XIII) and (XV) we get:

$$E\left[\Pi^{FL=PL}\left(S,\Theta\right)\right]-E\left[\Pi^{NL}\left(S,\Theta\right)\right]$$

$$=\frac{2n^2+\left(2m^2-2m-1\right)n-m^3+m}{2}\left(n\left(\frac{1}{2}-\frac{1}{\sum_{i=1}^{n}\frac{1}{i}}\right)+\frac{1}{2}\right)>0,\ \forall n\geq 3\wedge m\leq n$$

Lemma 8: For PL and NL the result follows from applying definitions (4) and (5) of stability to payoffs (XIII) and (XV), respectively. For FL with no transfers the equilibrium coalition size follows from applying the definitions (4) and (5) of stability to payoffs (XI); with transfers, by using payoffs (XI) and the definition of potential internal stability (8) we show that only coalitions with size $m\in\{1,2,3\}$ are internally stable under the AISS. From the relation between internal and external stability for this sharing scheme, we conclude that the only externally stable coalitions are those of size $m\geq 3$. Hence, the only Pareto undominated stable coalitions are those of size $m=3$.