International environmental agreements under uncertainty: Does the 'veil of uncertainty' help?

By Michael Finus* and Pedro Pintassilgo†

* University of Exeter Business School, UK; m.finus@exeter.ac.uk
† Faculty of Economics and Research Centre for Spatial and Organizational Dynamics, University of Algarve, Faro, Portugal

Na and Shin showed that the 'veil of uncertainty' can be conducive to the success of self-enforcing international environmental agreements. Later papers confirmed this conclusion about the negative impact 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 analyse 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.

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

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, 2001). 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 two issues that play an important role in the negotiations over the provision of global public goods: (i) asymmetry and (ii) uncertainty. For instance, in climate change, there are large differences between industrialized, developing and transition countries in terms of abatement costs and perceived benefits from greenhouse gas emission reduction. These differences typically translate into asymmetric gains from cooperation, adding to the difficulties associated with free-rider incentives. An obvious way to mitigate these difficulties is transfer payments as for instance analysed in Botteon and

---

1 The literature dates back to Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994), and is surveyed for instance in Barrett (2003) and Finus (2001, 2003).

Carraro (1997), Barrett (2001), Bosello *et al.* (2003), Eyckmans and Finus (2006), and Weikard *et al.* (2006).

Also the uncertainties, not only about expect abatement costs, but in particular about the damages from emissions are still quite large, despite ongoing research in this area, as it is evident from the regular updates of IPCC reports (IPCC 2007). 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'. The example suggests that uncertainty may actually hamper cooperation. This raises the question whether this is general true or whether the veil of uncertainty (a term coined by Brennan and Buchanan, 1985) may also be conducive to the success of self-enforcing cooperation. Interestingly, the theoretical literature on uncertainty, learning and the formation of IEAs (e.g. Na and Shin, 1998; Kolstad, 2007; and Kolstad and Ulph, 2008, 2011) stresses exactly this second possibility: in the strategic context of treaty-formation, resolving uncertainty through learning may lead to worse outcomes than if uncertainty prevails or is only partially resolved. This result, though interesting, is intriguing for at least two reasons. First, it does not provide an explanation for situations in which uncertainty actually hampers cooperation. Second, it runs counter to intensified research efforts worldwide to reduce uncertainty about the environmental impact of emissions and the cost of reducing them. Therefore, this paper addresses three research questions: How general are the results obtained in previous papers? What are the crucial driving forces? Is there a way to avoid the negative impact of learning when it occurs?

Our analysis builds on the seminal contribution by Na and Shin (1998) who analyse asymmetry and uncertainty within one framework. 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 (in the form of reduced damages). 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 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. For no learning and 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 learning and no 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 can be useful in a strategic setting.

In order to explore our three research questions mentioned above, we extend Na and Shin's model in several directions. As will be apparent from the subsequent discussion, the number of possible modifications is large, which requires to focus on the most interesting extensions. Our choice is guided by the attempt to pick stylized scenarios at the border of the possibility frontier, which allow working out the underlying driving forces clearly. Possible qualitative conclusions for intermediate cases are indicated in later sections.

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, 2011) as this leads to corner solutions in terms of equilibrium abatement choices, irrespective of the type of uncertainty.

Another extension could be to consider not only risk-neutrality but also allow for the possibility of risk-aversion. However, the effect of different degrees of risk-aversion on treaty formation outcomes is complex and hence a topic in its own right (e.g. Bramoullé and Treich, 2009; Boucher and Bramoullé, 2010; and Finus *et al.*, 2010).

The most obvious and first extension appears to be lifting the restriction on the number of players and to allow for any number of 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, 2011) 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 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. Hence, what we call partial learning is essentially delayed learning. It should be noted that once players learn about parameter values, all of them learn everything and no uncertainty remains (i.e. perfect learning).[2]

The third and fourth extensions are motivated by the suspicion that asymmetry may be a crucial factor in Na and Shin's model that leads to small coalitions under full learning. As

---

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

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. 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 under learning. Both extensions (third and fourth) qualify the negative conclusion about the role of learning. Interestingly, we show that in the presence of transfers, asymmetry can even be beneficial for the success of cooperation.

Finally, the fifth and sixth extensions are motivated by the suspicion that the results from the uncertainty about benefits may not carry over to uncertainty about costs. Hence, the fifth extension considers a mirror image of the Na and Shin's assumption: instead of uncertainty about the distribution of benefits from global abatement, we consider uncertainty about the distribution of the costs from individual abatement. The sixth extension considers the mirror image Kolstad's assumption: instead of uncertainty about the level of benefits we consider uncertainty about the level of costs. Again, it turns out for both extensions that the negative role of learning has to be qualified.

For all extensions, we will 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 analyses 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: Setting

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, $m \le n$, 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:

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

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

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

whereas all singletons maximize their own payoff

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

where y^S is the abatement vector of signatories, $y_i^*(S)$ is the individual equilibrium abatement of a signatory, which depends on S and y_j^* is the individual equilibrium abatement of a non-signatory, which is independent of S.[3] 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 to the first stage, determining stable coalitions by applying the following definition:

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

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

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

3 The implicit assumption in (2) about the 'cooperative behaviour' of coalition members is common in the literature, though certainly optimistic regarding the 'cooperative sense' of players who decide to be a member of a coalition.

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.[4]

The payoffs of signatories $\Pi^*_{i\in S}(S)$ and non-signatories $\Pi^*_{j\notin S}(S)$, which are derived from solving (2) and (3) simultaneously, and substituting equilibrium abatement levels into payoff functions (1), may be modified through transfers in a TU-fashion (Transferable Utility). This is relevant if parameters b_i and/or c_i are different across players. Even if the total payoff of the coalition $\sum_{i\in S} \Pi^*_i(S)$ is large but unequally distributed among its members, only small coalitions may be stable. From the literature (e.g. Botteon and Carraro, 1997; Barrett 2001; Bosello *et al.*, 2003; Eyckmans and Finus, 2006; and Weikard *et al.*, 2006) it is apparent that the rules of splitting the total coalitional pie will crucially affect the size and the membership of stable coalitions and hence the success of agreements (measured as the total payoff overall players). Recently, Eyckmans and Finus (2009) proposed an almost ideal sharing scheme (AISS; for similar notions see McGinty, 2007; Fuentes-Albero and Rubio, 2009; and Weikard, 2009) which avoids some of the disadvantages of previously considered transfer rules. They show that if AISS is applied: (i) for any transfer rule belonging to AISS, the set of stable coalitions will be same (robustness); (ii) among those coalitions which can be potentially internally stabilized (which may not include some large coalitions), the coalition that corresponds to the highest aggregate payoff over all players

4 The reason is that the singleton coalition structure can be supported as an equilibrium 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.

will be stabilized (optimality). Eyckmans and Finus propose that each coalition member should receive payoff $\Pi^T_{i\in S}$ after transfers which is defined as follows:[5]

$$\Pi^T_{i\in S} = \Pi^*_i\left(S\setminus\{i\}\right) + \lambda_{i\in S}\sigma(S) \text{ with } \sigma(S) = \sum\nolimits_{i\in S}\Pi^*_i\left(S\right) - \sum\nolimits_{i\in S}\Pi^*_i\left(S\setminus\{i\}\right), \quad \sum_{i\in S}\lambda_i = 1. \tag{6}$$

That is, every signatory i receives his/her free-rider payoff when leaving coalition S, $\Pi^*_i\left(S\setminus\{i\}\right)$, plus a share λ_i of the surplus $\sigma(S)$ which is the difference between the total coalitional payoff and the sum of free-rider payoffs. Any transfer rule is said to belong to AISS if it has the structure of (6).

In order to appreciate the properties of AISS, we may note the following. First, $\sum\nolimits_{i\in S}\Pi^T_i\left(S\right) = \sum\nolimits_{i\in S}\Pi^*_i\left(S\right)$. That is, transfers are paid only among coalition members. Second, the surplus $\sigma(S)$ may be positive, nil or negative. If $\sigma(S)$ is negative, coalition S will not be internally stable as $\Pi^T_i\left(S\right) < \Pi^*_i\left(S\setminus\{i\}\right)$ for all $i\in S$ (see (4)). However, no other transfer scheme could do any better, as (4) would fail for at least one signatory. If $\sigma(S)$ is positive or nil, which Eyckmans and Finus call potential internal stability, coalition S will be internally stable for all members of S through AISS (which may not be the case for other transfer schemes like the Shapley value or the Nash bargaining solution). Third, internal stability of S does not depend on the choice of weights λ_i. Hence, any transfer rule of the form in (6) will lead to the same set of internally stable coalitions. Fourth, with AISS, there is a direct link between internal and external stability. For every transfer rule belonging to 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. Hence, like internal stability, also external stability S does not

5 Hence, the transfer system takes the form $t_i = \Pi^*_i\left(S\setminus\{i\}\right) - \Pi^*_i\left(S\right) + \lambda_{i\in S}\sigma(S)$ with $\Pi^T_{i\in S} = \Pi^*_i\left(S\right) + t_i$ and $\sum\nolimits_{i\in S} t_i = 0$.

depend on the choice of weights. Consequently, the set of stable coalitions is the same for all transfer rules belonging to AISS to which we referred above as robustness. Fifth, as shown in Carraro *et al.* (2006) an expansion from a stable coalition through transfers between current signatories and non-signatories is not possible as AISS has already exhausted all possibilities.[6] Finally, it is for a similar reason that if S generates the highest global payoff among all potentially internal coalitions, it must also be externally stable.[7] We referred to this above as optimality.

2.2 Coalition Formation: Properties

The coalition formation model possesses three interesting properties which will be helpful for the subsequent analysis and are summarized in Lemma 1. As these properties are related to the second stage, and because we are operating in a TU-framework, these properties hold irrespective of transfers.

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.

(i) 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. $\Pi^*_{k \notin S}(S) < \Pi^*_{k \notin \hat{S}}(\hat{S})$.

---

6 If S is stable (and hence externally stable), $S \cup \{j\}$ cannot be potentially internally stable, i.e. $\sigma(S \cup \{j\}) < 0$. In other words, the enlarged coalition $S \cup \{j\}$ does not generate sufficient resources to satisfy all free-rider claims.

7 If it were not externally stable, there would be a fringe player $j \notin S$, such that $\sigma(S \cup \{j\}) \geq 0$. Thus, $S \cup \{j\}$ would be internally stable and due to the property of positive externalities (Lemma 1), $S \cup \{j\}$ would Pareto-dominate S.

(ii) 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.

$$\sum_{i\in S}\Pi_i^*(S)+\Pi_{j\notin S}^*(S)<\sum_{i\in S}\Pi_i^*(\hat{S})+\Pi_{j\in\hat{S}}^*(\hat{S})=\sum_{i\in\hat{S}}\Pi_i^*(\hat{S}).$$

(iii) 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) and (3) it is evident that if non-signatory 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 as this is evident from (3). Hence, we can conclude: total abatement will be higher if $\hat{S}$ than if S forms (GEC with respect to abatement); 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 (see (3)). Finally, GEC with respect to aggregate payoffs follows from combining properties PE and SAD. □

Global Efficiency from Cooperation (GEC) makes it interesting from a normative point of view to analyse the prospects of cooperation. The highest (lowest) global abatement level and the highest (lowest) total payoff is obtained in the grand coalition (if all countries play as singletons) which corresponds to the social optimum (Nash equilibrium). Note that the property GEC also holds under uncertainty, which we consider later by using expected payoffs over the uncertain parameters. This will be useful when evaluating the success under the three models of learning below. Superadditivity (SAD) makes it interesting for players to form a coalition, but due to the Positive Externality (PE) the formation of large stable coalition is difficult. 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.

2.3 Three 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. Additional to Na and Shin (1998), and as in Kolstad and Ulph (2008, 2011), we also consider the scenario of partial learning, which gives rise to the following three learning scenarios: (i) full learning, (ii) partial learning and (iii) no learning.

Full Learning (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 Partial Learning (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 No Learning (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, players are *ex ante* symmetric regarding uncertainty: all players know as much or little as their fellow players. FL and PL are identical regarding the second stage, but not regarding the first 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.

2.4 Five Cases of Uncertainty

In the following, we analyse five stylized cases of uncertainty of which their assumptions are displayed in Table 1.

Table 1 Five Cases of Uncertainty

Cases	*Uncertainty*	*Assumption*	*Ex ante Expectations*	*Ex post Realization*
Case 1	distribution of benefits	known symmetric cost no transfers	symmetric	asymmetric
Case 2	level of benefits	known symmetric cost	symmetric	symmetric
Case 3	distribution of benefits	known symmetric cost with transfers	symmetric	asymmetric
Case 4	distribution of costs	known symmetric benefits with and without transfers	symmetric	asymmetric
Case 5	level of costs	known symmetric benefits	symmetric	symmetric

We start by generalizing the setting of Na and Shin (1998), which is our case 1, and then sequentially explore alternative assumptions (cases 2 to 5). This will help us to better understand the driving forces of Na and Shin's result and its generality. Case 1 assumes (pure) uncertainty about the distribution of the benefits from global abatement. Case 2 is the Kolstad (2007) case and assumes (pure) uncertainty about the level of the benefits from global abatement. Case 3 is the Na and Shin case with transfers. Case 4 is the mirror image of the Na and Shin case, assuming pure uncertainty about the distribution of the costs from individual abatement. For completeness and to save space, no transfers and transfers are considered together under case 4. Finally, case 5 is the mirror image of the Kolstad case, assuming pure uncertainty about the level of costs.

Despite the five cases cover quite a lot of assumptions, it is clear that this is only a small selection of possibilities. After we derive the results for the five stylized cases in sections 3.1 to 3.5, we will come back to these assumptions and indicate the impact of alternative assumptions in section 3.6.

3. Results

It is worth noting that the proofs of the results in section 3 typically proceed in three steps, reflecting the fact that the game is solved by backward induction. They look at outcomes in the second stage (step 1) and the first stage (step 2) and subsequently combine both stages (step 3). In many (though not all) cases, comparisons between different learning scenarios are straightforward: if one learning scenario performs equally or better in one stage and better in another stage than another scenario, it will perform better overall.

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. 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. Since the only restriction to observe is that these parameters are non-negative, we assume parameters to be uniformly distributed over a set of equidistant values: $\{1,2,...,n\}$. Hence, we assume the following probability distribution:[8]

$$f_{\Theta_i}(\theta_i) = \begin{cases} \frac{1}{n} & \text{for} \quad \theta_i = k,\ k \in \{1,2,...,n\} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

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

$$E[\Theta_i] = \frac{n+1}{2} \text{ and } Var[\Theta_i] = \frac{n^2-1}{12}. \tag{8}$$

In this setting, as in Na and Shin (1998), the variables are correlated such that all players have a different benefit parameter: $\theta_i \neq \theta_k,\ \forall\, i \neq k \in \{1,2,...,n\}$. Thus, vector $\Theta = (\Theta_1,...,\Theta_n)$ is composed of all values of the benefit parameter: $\bigcup_{i=1}^{n} \Theta_i = \{1,2,...,n\}$.

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 of global benefits. In case

8 In all cases explored in this paper, the equilibrium coalition structures as well as the rankings of total abatement levels and total payoff across the three learning scenarios do not change if all elements in the set of parameter values are multiplied by a positive constant: $\{1\delta, 2\delta, ..., n\delta\}$, $\delta > 0$. This is due to the fact that for the three learning scenarios, second stage payoffs are homogeneous functions of the same degree and the same holds for total abatement (see the Appendix). Hence, results are robust over an infinite number of sets with equidistant parameters.

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 be 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 $\theta_k = E\left[\Theta_k\right]$.

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.

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:

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

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

Proof See Appendix 1.

It is clear that FL and PL are identical regarding the second stage. For payoff function (1), 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, depending on the realization of the random benefit parameter Θ_i, on average this cancels out. This is different for payoffs and as we explain below.

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 coalition 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.[9] 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.

---

9 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).

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

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.

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 symmetric payoff functions. For payoff function (1) it is well-known from the literature that the stable coalition comprises three signatories if $n \geq 3$.[10] Also under PL *ex ante* symmetric payoffs 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

---

10 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) for an overview.

potential participants do not know the distributional consequences.[11] 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: the distributional effect is the payoff difference between FL and PL due to different coalition sizes, resulting from different distributions of the gains from cooperation among coalition members. Here the distributional effect from learning is negative.

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.

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:

(i) Total Abatement: $NL = PL > FL$

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

Proof[12] *NL versus PL*: follows directly from Lemmas 2 and 3. *PL versus FL*: we note the same second stage outcomes. 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

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

12 For this and other proofs in the text as well as in the Appendix detailed computations are available upon request from the authors.

$E\left[m^{*PL}\right]=3>2=E\left[m^{*FL}\right]$, 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, which delivers the result. The relation between NL and FL follows directly from the relations between NL and PL as well as PL and FL. □

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 than NL. Additionally, the negative 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 idea, we again assume cost symmetry, $c_i=c_j=c \quad \forall\, i\in N$, and 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, 2011) call systematic uncertainty. Again, all players have the same expectations *ex ante*, but now, once uncertainty is resolved, all countries also have the same benefit parameter *ex post*, $\theta_i=\theta_k \quad \forall\, i,k\in N$. Hence, 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 combining 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:

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

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

Proof See Appendix 1.

Again, over- and undershooting of abatement under NL depending on the realizations of Θ_i, which is now systematic, cancels out on average. However, the relations in terms of payoffs are reversed compared 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 about benefits. 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 $E\left[m^*\right]$ is given by $E\left[m^{*FL}\right] = E\left[m^{*PL}\right] = E\left[m^{*NL}\right] = 3$.

Proof See Appendix 2.

The driving force of this result is that players are symmetric both *ex ante* and *ex post* in the three learning scenarios. Hence, in all scenarios, symmetry makes all coalitions of size 3 stable.

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:

(i) Total Abatement: $FL = PL = NL$

(ii) Total Payoff: $FL = PL > NL$.

Proof Follows immediately from Lemma 4 and 5. □

Taken together, our intuition that Na and Shin's result about the negative impact of learning hinges on the asymmetric realization of the random benefit variable Θ_i is confirmed. As opposed to case 1, in case 2, the strategic effect from learning is no longer negative but positive and hence does not upset 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 (though not the strategic effect in a TU-framework). 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. Because transfers are based on realized parameter values under FL, there is no room for strategic misrepresentation of preferences. We apply the transfer scheme AISS (see section 2.1) to the setting of case 1 (see section 3.1) which leads to the following results.

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

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.

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 our 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 (see equation (8)). 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 which only the coalition as a group enjoys and which 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. As no cooperation is particularly inefficient under FL, due to the differences of marginal abatement costs among players, the additional benefit from cooperation is larger under FL than NL, leading to larger stable coalitions. 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:

(i) Total Abatement: $\begin{cases} FL = PL = NL & \text{if} \quad n \leq 8 \\ FL > PL = NL & \text{if} \quad n \geq 9 \end{cases}$

(ii) Total Payoff: $\begin{cases} FL = PL = NL & \text{if} \quad n = 3 \\ NL > FL = PL & \text{if} \quad 4 \leq n \leq 8 \\ NL > FL > PL & \text{if} \quad n = 9 \\ FL > NL > PL & \text{if} \quad n \geq 10 \end{cases}$

Proof The ranking between PL and NL follows from Proposition 1 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 (higher expected total payoffs) than FL (Lemma 2), but FL produces better first stage outcomes (larger stable coalitions) than NL (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.[13] Then, we compare this to the expected total payoff under NL with $E\left[m^{*NL}\right]=3$. □

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 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 (7). Hence, all players have a different cost parameter c_i, with $\theta_i \neq \theta_k$, $\forall\, i \neq k \in N$, $\bigcup_{i=1}^{n} \Theta_i = \{1,2,...,n\}$. 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:

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

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

---

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

Proof See Appendix 3.

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 from learning, 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 than under NL. 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 $E\left[m^{*}\right]$ 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 the 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.

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. With transfers, the distributional effect from learning is zero as the size of stable coalitions under FL 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:

No Transfers

(i) Total Abatement: $\begin{cases} PL > NL > FL & \text{if} \quad 3 \leq n \leq 8 \\ PL > FL > NL & \text{if} \quad n \geq 9 \end{cases}$

(ii) Total Payoff: $\begin{cases} PL > NL > FL & \text{if} \quad 3 \leq n \leq 7 \\ PL > FL > NL & \text{if} \quad n \geq 8 \end{cases}$

Transfers

(i) Total Abatement: $FL = PL > NL$

(ii) 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 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. □

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.

3.5 Case 5: Uncertainty about the Level of Costs

Finally, we consider the mirror image of Kolstad's case (case 2), namely uncertainty about the level of abatement costs. The mirror image means benefit symmetry, i.e. $b_i = b_j = b$ $\forall\, i \in N$, defining again $\theta_i = b/c_i$ and assuming uncertainty about the level of costs, with *ex ante* symmetric expectations and the same cost parameters *ex post*, $\Theta_i = \Theta_k \quad \forall\, i, k \in N$. Uncertainty about a common cost parameter is equivalent to uncertainty about a common $\theta = b/c$. Thus, there is equivalence between cases 2 and 5 and consequently the outcomes under cost uncertainty are exactly those in Lemmas 4 and 5 and Proposition 2.

Uncertainty about the level of costs translates into symmetry under our assumptions. Hence, the distributional effect from learning is zero and first stage outcomes are the same for all scenarios of learning. Information about the level of costs provides an advantage which translates into better second stage outcomes in terms of payoffs for FL and PL compared to NL. Hence, pulling both stages together implies that learning (FL and PL) perform equally but better than no learning (NL) in terms of aggregate payoffs.

3.6 Putting Results in Perspective

As mentioned in section 2.4 and displayed in Table 1, the five stylized cases which we consider are useful to work out the driving forces of uncertainty and learning but depend on some simplifying assumptions. Now, we discuss the impact of departing from these assumptions.

1) We considered either uncertainty about the level or the distribution of a parameter. On the cost side, the strategic effect from learning was always positive and the possible negative distributional effect from learning could be fixed through transfers. Hence, the qualitative conclusions would not change if we considered intermediate cases with simultaneous uncertainty about the level and the distribution of costs. On the benefit side, intermediate cases would reveal that the larger the uncertainty is about the level of the global benefits from abatement compared to their distribution, the higher the likelihood that the strategic effect from learning will be less negative or even positive and the smaller, in absolute terms, will be the negative (positive) distributional effect from learning without transfers (with transfers). For instance, if we assumed the setting of case 1, with an asymmetric distribution of the benefit parameter $f_{\Theta_i}(\theta_i)$ (see equation (7), section 3.1), but with independently distributed parameters θ_i, then uncertainty would be about the level and the distribution simultaneously. For this assumption, it turns out that the strategic effect from learning is positive and the distributional effect from learning is generally less pronounced (less negative without transfers and less positive with transfers) with the overall conclusion that learning leads to better outcomes with and without transfers.[14]

2) We considered uncertainty either on the benefit or cost side. However, our results can be used to indicate the impact of simultaneous uncertainty about benefits and costs from abatement. The key conclusion would be that the veil of uncertainty can be conducive to the success of cooperative agreements, only if the uncertainty is predominately about the benefits from abatement and in particular about their distribution. However, even then can the negative impacts from learning be remedied through transfers.

---

14 Results are available from the authors upon request.

3) In all cases we assumed the non-random parameter to be symmetric. It can be expected that the positive strategic effect from learning with cost uncertainty would not disappear with asymmetric benefits as the driving force is a cost-effectiveness effect. With uncertainty about the distribution of costs, costly misallocation of abatement duties among coalition members can be avoided through learning. When uncertainty is about the level of costs, costly over- and undershooting of global abatement can be avoided through learning. The same applies for uncertainty about the level of benefits if we assume asymmetric costs. Only the implications of our case 1 with uncertainty about the distribution of the benefits would qualitatively change if there was cost asymmetry. In our case 1, NL compared to FL and PL benefits from symmetric expectations but asymmetric realizations of the random benefit variables due to cost symmetry. NL is cost-effective compared to PL and FL and hence the strategic effect from learning is negative. This cost advantage of NL over FL and PL would be weakened or would even disappear if we allowed for cost asymmetry.

   Regardless whether we assume that the non-random cost or benefit parameter is asymmetric, it would be natural to consider transfers, not only for FL but also for PL and NL and hence distributional issues should not work to the disadvantage of learning. The upshot of all this is that asymmetry of the non-random parameter makes it even more likely that learning is beneficial.

4) Finally, all five cases assumed symmetric *ex ante* expectations about the uncertain parameter. This simplifies computations but should not affect the qualitative conclusions. In our model, the driving force for the distributional effect from learning is the difference between the variance of expected parameter values *ex ante* and of the *ex post* realizations. In cases 1 and 4 *ex post* realizations are asymmetric which, without transfers, leads to smaller coalitions under FL than under PL and NL. If we started

with asymmetric expectations *ex ante* but with a higher (lower) degree of asymmetry *ex post*, the distributional effect would also be negative (would now be positive). Naturally, however, we could improve upon first stage outcomes not only under FL but also under PL and NL by considering transfers.

Like in our cases, also with asymmetric *ex ante* expectations we could model uncertainty about the level or distribution of a parameter. One version directly related to our assumptions could be to assume cost symmetry and benefit parameters given by $b_i = i + \theta_i$, with i a certain benefit component different for all players, e.g. i could be an index running from 1 to n. The random component θ_i could be uniformly distributed in some interval $[-a,+a]$. For θ_i s positively correlated like in case 2, such that all players have the same θ_i, $\theta_i = \theta_k \ \ \forall\, i,k \in N$, we would model pure uncertainty about the level of benefits. For θ_i s negatively correlated like in case 1, such that $\theta_i \neq \theta_k$ $\forall\, i,k \in N$, we would have a pure distributional effect. Finally, for θ_i s independently distributed we would have uncertainty about the level and distribution of the benefit parameter, simultaneously. Assuming $a \leq 1$, the authors could confirm the qualitative conclusions outlined above.

4. Summary and Conclusions

The starting point of our analysis was the conclusion of Na and Shin (1998) about the negative value of information in a public good coalition formation game, which was confirmed in later studies (Kolstad, 2007; Kolstad and Ulph, 2008, 2011). We addressed three general questions. How general is this result? What are the driving forces? If the negative outcome occurs, can it be mitigated?

Even though our model made reference to international environmental agreements, it is evident that the applications are much wider. The crucial ingredient is an economic problem with externalities where cooperation is associated with global welfare gains. Because these gains are non-exclusive and there is no authority or third party that can enforce cooperative agreements, strong free-rider incentives make it difficult to establish cooperation. Moreover, at the time of negotiations, the benefits and costs from cooperation may not be exactly known, may be revealed only later and be asymmetrically distributed. Hence, applications comprise the provision of many global public goods, like nuclear non-proliferation treaties, the coordination of peace and humanitarian missions as well as the fight against international terrorism through the UN, avoiding the spread of contagious diseases across the globe through the WHO, and negotiations of multilateral trade agreements within the institutions of the WTO.

By generalizing the model of Na and Shin (1998), we confirmed that the common wisdom of more information being always better than less information may not be true if the enforcement of an agreement is not simply assumed but the strategic behavior of agents is accounted for. On the one hand, though more information is useful for individual agents and/or for members of an agreement, at the aggregate this may no longer be true. The strategic effect from learning may be negative as captured by our second stage of coalition formation. On the other hand, if agents learn that the gains from cooperation are *ex post* more asymmetrically distributed than *ex ante* expected, learning leads to a smaller participation in an agreement and hence the veil of uncertainty can be helpful. In other words, the distributional effect from learning may be negative as captured by our first stage of coalition formation. However, our model also captures the common expectation that more information is beneficial: clearly if the grand coalition forms and hence there is no strategic interaction, the information effect from learning is always positive.

We put Na and Shin's conclusion in a wider perspective by systematically considering four alternative assumptions. On the one hand, if the assumption of pure uncertainty about the distribution of benefits is given up and instead either pure uncertainty about the level of benefits (case 2) or uncertainty about costs, be it about the level or distribution of costs (cases 4 and 5), the strategic effect from learning is no longer negative. On the other hand, if either distributional issues are not important (cases 2 and 5) or are addressed through transfers (case 3 and case 4), the distributional effect from learning is no longer negative and may become even positive. Hence, taken together, the positive effect of the veil of uncertainty as stressed by the previous literature is less evident from our model which is certainly a relief to all those that believe that learning should be beneficial.

Our analysis suggests that policy recommendations should not only address efficiency and cost-effectiveness but also strategic and distributional issues if an agreement is based on voluntary participation and contributions. Moreover, we demonstrate that diversity may not be an obstacle for cooperation but an asset, but only if accompanied by an appropriate transfer scheme. Clearly, in the climate context asymmetries across developed and developing countries are large. Hence, the establishment of the global environmental facility (GEF) with payment obligations of industrialized countries and developing countries being beneficiaries seems a rational strategy. Moreover, our results stress that not only the allocative but also the distributional implications of emission permit trading for treaty formation has to move more into the focus of the policy debate. After all, the transfer scheme we considered can be replicated through the initial allocation of emission rights.

Finally, for future research an interesting extension of our model could be to consider the effect of risk-aversion on treaty outcomes like in Boucher and Bramoullé (2010). They show that, in an abatement game, uncertainty leads to lower abatement efforts than certainty, but higher participation which may lead to higher aggregate welfare. Whether this would also hold for our setting is difficult to predict as their model setting is quite different from ours: they consider a linear payoff function with two discrete optimal abatement choices (abate or not abate), uncertainty is about the level of a parameter with only two possible values and partial learning is not considered. There are also other possible avenues to analyse how the veil of uncertainty impacts on the formation and success of international environmental agreements, such as dynamic games of coalition formation.

Acknowledgments

Both authors would like to thank Christos Kotsogiannis for helpful comments on an earlier draft. The research was 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.

Funding

Portuguese Foundation for Science and Technology (BSAB/735/2007 to P.P.).

References

Arce M., D.G. and Sandler, T. (2001) Transnational public goods: strategies and institutions, *European Journal of Political Economy*, **17**, 493-516.

Barrett, S. (1994) Self-enforcing international environmental agreements, *Oxford Economic Papers*, **46**, 878-94.

Barrett, S. (2001) International cooperation for sale, *European Economic Review*, **45**, 1835-50.

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

Bloch, F. (2003) Non-cooperative models of coalition formation in games with spillovers, in C. Carraro (ed.) *The Endogenous Formation of Economic Coalitions*, Edward Elgar, Cheltenham, 35-79.

Bosello, F., Buchner, B., and Carraro, C. (2003) Equity, development, and climate change control, *Journal of the European Economic Association*, **1**, 601-11.

Botteon, M. and Carraro, C. (1997) Burden-sharing and coalition stability in environmental negotiations with asymmetric countries, in C. Carraro (ed.) *International Environmental Negotiations: Strategic Policy Issues*, Edward Elgar, Cheltenham, 26-55.

Boucher, V. and Bramoullé, Y. (2010) Providing global public goods under uncertainty, *Journal of Public Economics*, **94**, 591-603.

Bramoullé, Y. and Treich, N. (2009) Can uncertainty alleviate the commons problem? *Journal of the European Economic Association*, **7**, 1042-67.

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

Carraro, C., Eyckmans, J., and Finus, M. (2006) Optimal transfers and participation decisions in international environmental agreements, *Review of International Organizations*, **1**, 379-96.

Carraro, C. and Siniscalco, D. (1993) Strategies for the international protection of the environment, *Journal of Public Economics*, **52**, 309-28.

Eyckmans, J. and Finus, M. (2006) Coalition formation in a global warming game: how the design of protocols affects the success of environmental treaty-making, *Natural Resource Modeling*, **19**, 323-58.

Eyckmans, J. and Finus, M. (2009) An almost ideal sharing scheme for coalition games with externalities, Stirling Discussion Paper Series, 2009-10, University of Stirling.

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

Finus, M. (2003) Stability and design of international environmental agreements: the case of transboundary pollution, in H. Folmer and T. Tietenberg (eds) *International Yearbook of Environmental and Resource Economics 2003/4*, Edward Elgar, Cheltenham, 82-158.

Finus, M., Pintassilgo, P., and Ulph, A. (2010) International environmental agreements with uncertainty, learning and risk aversion, SCI Discussion Papers, No. 6.2010. University of Manchester, UK.

Fuentes-Albero, C. and Rubio, S.J. (2009) Can the international environmental cooperation be bought? *European Journal of Operation Research*, **2002**, 255-64.

Gollier, C. and Treich, N. (2003) Decision-making under scientific uncertainty: the economics of the precautionary principle, *Journal of Risk and Uncertainty*, **27**, 77-103.

Helm, C. (1998) International cooperation behind the veil of uncertainty, *Environmental and Resource Economics*, **12**, 185-201.

Hoel, M. (1992) International environment conventions: the case of uniform reductions of emissions, *Environmental and Resource Economics*, **2**, 141-59.

IPCC (2007) *Climate Change 2007: Mitigation of Climate Change. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge.

Kolstad, C.D. (2005) Piercing the veil of uncertainty in transboundary pollution agreements, *Environmental and Resource Economics*, **31**, 21-34.

Kolstad, C.D. (2007) Systematic uncertainty in self-enforcing international environmental agreements, *Journal of Environmental Economics and Management*, **53**, 68-79.

Kolstad, C.D. and Ulph, A. (2008) Learning and international environmental agreements, *Climatic Change*, **89**, 125-41.

Kolstad, C.D. and Ulph, A. (2011) Uncertainty, learning and heterogeneity in international environmental agreements, *Environmental and Resource Economics,* forthcoming.

McGinty, M. (2007) International environmental agreements among asymmetric nations, *Oxford Economic Papers*, **59**, 45-62.

Na, S.-L. and Shin, H.S. (1998) International environmental agreements under uncertainty, *Oxford Economic Papers*, **50**, 173-85.

Ulph, A. (1998) Learning about global warming? In N. Hanley and H. Folmer (eds) *Game Theory and the Environment*, Edward Elgar, Cheltenham, 255-86.

Weikard, H.-P. (2009) Cartel stability under optimal sharing rule, *The Manchester School*, **77**, 575-93.

Weikard, H.-P., Finus, M., and Altamirano-Cabrera, J.C. (2006) The Impact of surplus sharing on the stability of international climate agreements, *Oxford Economic Papers*, **58**, 209-32.

Yi, S.-S. (1997) Stable coalition structures with externalities, *Games and Economic Behavior*, **20**, 201-37.

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

Appendix

For some proofs we only provide a sketch and intuition due to space limitations. Details are available from the authors on request.

Appendix 1: Proof of Lemmas 2 and 4

We consider second stage outcomes in cases 1, 2 and 3, noting that cases 1 and 3 are identical in the second stage. For FL and PL, abatement and payoffs are given by (using (2) and (3) in the text):

$$Y^*(S) = m\sum_{i\in S}\theta_i + \sum_{j\notin S}\theta_j \tag{A1}$$

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

$$\Pi^*_{j\notin S}(S) = \theta_j\left(m\sum_{i\in S}\theta_i + \sum_{k\notin S}\theta_k\right) - \frac{1}{2}\left(\theta_j\right)^2 = \frac{1}{2}\left(\theta_j\right)^2 + m\theta_j\sum_{i\in S}\theta_i + \theta_j\sum_{k\neq j\notin S}\theta_k \tag{A2}$$

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

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

Taking expectations gives:

Cases 1 and 3:

$$E\left[Y^{FL=PL}(S,\Theta)\right] = \left(m^2 - m + n\right)E\left[\Theta_i\right] = \frac{\left(m^2-m+n\right)(n+1)}{2} \tag{A3}$$

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

$$E\left[\Pi_{j\notin S}^{FL=PL}(S,\Theta)\right]=\frac{(n+1)\left[3n^2+\left(3m^2-3m+1\right)n+2m^2-2m-1\right]}{12} \qquad \text{(A4)}$$

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

Case 2:

$$E\left[Y^{FL=PL}(S,\Theta)\right]=\left(m^2-m+n\right)E\left[\Theta_i\right] \qquad \text{(A5)}$$

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

$$E\left[\Pi_{j\notin S}^{FL=PL}(S,\Theta)\right]=\left(m^2+n-m-\frac{1}{2}\right)E\left[\Theta_j^{\,2}\right] \qquad \text{(A6)}$$

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

Note that $E\left[\Theta_k^2\right]$ can remain unspecified, as no assumption about the distribution of the random variables is necessary for the analysis.

Finally, we compute the second stage outcomes for NL. Under this learning scenario, certainty equivalence holds. Thus, in cases 1, 2 and 3, payoffs and abatement are the same as those under certainty with $\theta_k=E\left[\Theta_k\right]\ \forall\, k\in N$:

$$E\left[Y^{NL}(S,\Theta)\right]=\left(m^2-m+n\right)E\left[\Theta_k\right] \qquad \text{(A7)}$$

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

$$E\left[\Pi_{j\notin S}^{NL}\left(S,\Theta\right)\right]=\left(m^2-m+n-\frac{1}{2}\right)\left(E\left[\Theta_j\right]\right)^2 \tag{A8}$$

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

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

Using the expected outcomes from above, the comparisons undertaken in Lemma 2 and 4 are proved. The equality of expected total abatement in the three learning scenarios follows directly from (A3) and (A7) for case 1 (Lemma 2), and (A5) and (A7) for case 2 (Lemma 4). Regarding expected total payoffs, using (A4) and (A8), 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 (A6) and (A8), 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

We consider first stage outcomes in cases 1, 2 and 3, based on the second stage outcomes determined in Appendix 1.

Lemma 3 (Case 1): For PL and NL we use expected payoffs in (A4) and (A8), respectively, and apply definition (4) of internal and definition (5) of external stability which delivers the result. For FL, we note that there are θ-vectors with asymmetric entries. Consequently, $E[m^{*FL}]<3$, as it can be shown, using payoffs in (A2) 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 for $n=3$ the only stable coalition structure is the one formed by singletons, and that for $n \geq 4$ stable coalitions comprise the two players with the highest θ_i, is also derived using (A2) and (4).

Lemma 5 (Case 2): For PL and NL the result follows from applying definition (4) of internal and definition (5) of external stability to payoffs (A6) and (A8), respectively. For FL the equilibrium coalition size follows from symmetry, payoffs (A2) and definitions (4) and (5).

Lemma 6 (Case 3): 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 (A2) and the definition of potentially internal stability in section 2.1 (i.e. $\sigma(S) \geq 0$). 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 of these 4-player coalitions or even larger coalitions are externally stable, we can conclude that $E[m^{*FL}] > 3$.

Appendix 3: Proof of Lemmas 7, 8

Lemma 7: In order to determine second stage outcomes in case 4, assuming $b_i = b_j = b$ $\forall\, i \in N$, using $\theta_i = b / c_i$, the payoff function can be written as

$$\Pi_i = \left(\sum_{k=1}^{n} y_k \right) - \frac{1}{\theta_i} \frac{y_i^2}{2} \ , \ i \in N \tag{A9}$$

with second stage equilibrium abatement and payoffs under FL and PL as follows:

$$\begin{cases} y_i = m\theta_i, & \forall\, i \in S \\ y_j = \theta_j, & \forall\, j \notin S \end{cases} \text{ and } Y^*(S) = m\sum_{i\in S}\theta_i + \sum_{j\notin S}\theta_j \tag{A10}$$

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

$$\Pi^*_{j\notin S}(S) = \left(m\sum_{\ell\in S}\theta_\ell + \sum_{k\notin S}\theta_k \right) - \frac{\theta_j}{2} \tag{A11}$$

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

The corresponding expected values are:

$$E\left[Y^{FL=PL}(S,\Theta)\right] = \left(m^2 - m + n\right)E\left[\Theta_i\right] = \frac{\left(m^2 - m + n\right)\left(n+1\right)}{2} \tag{A12}$$

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

$$E\left[\Pi^{FL=PL}_{j\notin S}(S,\Theta)\right] = \frac{\left(2m^2 - 2m + 2n - 1\right)\left(n+1\right)}{4} \tag{A13}$$

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

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

$$E\left[y_i\left(S,\Theta\right)\right]=m\left(E\left[\frac{1}{\Theta_i}\right]\right)^{-1},\quad \forall i\in S$$

$$E\left[y_j\left(S,\Theta\right)\right]=\left(E\left[\frac{1}{\Theta_j}\right]\right)^{-1},\quad \forall j\notin S \tag{A14}$$

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

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

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

$$E\left[\Pi_{j\notin S}^{NL}\left(S,q^{**},\Theta\right)\right]=\left(m^2-m+n-\frac{1}{2}\right)\left(E\left[\frac{1}{\Theta_j}\right]\right)^{-1} \tag{A15}$$

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

A comparison of second stage outcomes of total abatement, using (A12) and (A14), gives:

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

$$=\left(\frac{m^2}{2}-m+n\right)\left(n\left(\frac{1}{2}-\frac{1}{\sum_{i=1}^{n}\frac{1}{i}}\right)+\frac{1}{2}\right)>0,\quad \forall\, n\geq 3\wedge m\leq n. \tag{A16}$$

A comparison of second stage outcomes of total payoffs, using (A13) and (A15), gives:

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

$$=\frac{2n^2+\left(2m^2-2m-1\right)n-m^3+m}{2}\left(n\left(\frac{1}{2}-\frac{1}{\sum_{i=1}^{n}\frac{1}{i}}\right)+\frac{1}{2}\right)>0,\ \forall n\geq 3\wedge m\leq n. \tag{A17}$$

Lemma 8: In order to determine first stage outcomes in case 4, we note that for PL and NL the result follows from applying definitions (4) and (5) of stability to payoffs (A13) and (A15), respectively. For FL with no transfers, the equilibrium coalition size follows from applying the definitions (4) and (5) of stability to payoffs (A11). For FL with transfers, we use payoffs (A11) and the definition of potential internal stability, requiring the surplus $\sigma(S)$ as defined in (6) to be non-negative and 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 $m = 3$.