Designing International Environmental Agreements under Participation Uncertainty

Mao, Liang

College of Economics, Shenzhen University

15 May 2017

Designing International Environmental Agreements under Participation Uncertainty

Liang Mao[*]

Abstract

We analyze the design of international environmental agreement (IEA) by a three-stage coalition formation game. In stage one, a designer chooses an IEA rule which, depending on the coalition of signatories formed in stage two, specifies the action that each signatory should take in stage three. A certain degree of participation uncertainty exists in that each country intending to sign the IEA for its best interest has a probability to end up a non-signatory. An IEA rule is said to be optimal if it maximizes the expected payoff of each signatory. We provide an algorithm to determine an optimal rule, and show its advantage over some rules used in the literature.

Keywords: International environmental agreement, coalition formation, participation uncertainty, *ex ante* and *ex post* stable
JEL codes: Q54, H41, C72

[*] College of Economics, Shenzhen University, Shenzhen, Guangdong, 518060, China. Email: maoliang@szu.edu.cn.

1 Introduction

The increasing emission of transboundary pollution has threatened the human society from many aspects. For a country, reducing the emission of a global pollutant can be regarded as providing a public good benefiting the whole world. However, voluntary abatement of the pollutant is typically not sufficient, because countries have incentives to free ride on the abatement effort of other countries. One possible method to overcome this problem is to form a coalition wherein the members sign a self-enforcing international environmental agreement (IEA) and follow certain abatement rules.

The formation of such a coalition is sometimes modeled as a two-stage open membership game played by some self-interested countries.[1] In stage one (participation stage), each country decides whether to join the coalition and sign the IEA. In stage two (abatement stage), the signatories follow given rules of the IEA, while each non-signatory decides its own action.

Unfortunately, it is reported that IEAs do not always work very well. The welfare improvements created by IEAs are typically not as significant as one might expect, both in theory and in practice. For instance, Barrett (1994) suggests that "... an IEA may achieve a high degree of cooperation, but only when the difference between global net benefits under the noncooperative and full cooperative outcomes is small". Kellenberg and Levinson (2014) report that "Most empirical work ... finding that IEAs result in no improvements beyond what would have occurred in their absence". There could be many reasons for the failure of IEAs, but in this paper, we focus on the following two of them.[2]

One reason is the existence of participation uncertainty: a country initially intending to sign the IEA for its own interest has a chance to make a mistake and end up a non-signatory. Here, making a mistake means behaving irrationally: the country does not sign the IEA even though signing

[1] See Finus (2001) and Carraro (2003) for some reviews.

[2] Many other reasons have been discussed in the literature, such as failing to realize or be uncertain about the fact that under some conditions climate damages could shift to a significantly larger value (Barrett, 2013; Nkuiya et al., 2015), using inappropriate forms or parameters for payoff functions (Karp and Simon, 2013), and omitting the R&D cooperation in adaption to climate change (Masoudi and Zaccour, 2017).

is a better choice for it. This uncertainty, which makes it more difficult to form a large coalition and achieve an efficient outcome, may be due to some political reasons. For example, ratification of the IEA may be prevented by some interest groups (Köke and Lange, 2017).

In this paper, the participation uncertainty is assumed to be one-way. That is, we assume that only countries intend to sign the IEA may make mistakes, while the probability that a country not intending to sign the IEA becomes a signatory, which rarely happens in reality, is zero. In fact, there may not exist a stable coalition if uncertainty in both directions can happen.[3]

In the IEA literature, there are several other types of risk or uncertainty, such as uncertainty about the payoff functions (Kolstad, 2007; Dellink et al., 2008; Nkuiya et al., 2015; Meya et al., 2017), uncertainty about the threshold that triggers climate catastrophe (Barrett, 2013), and uncertainty about the realization of other players' mixed strategies (Hong and Karp, 2014). To keep analysis simple, in this paper we omit all these kinds of uncertainty, and focus on participation uncertainty only. In reality, participation uncertainty has drawn much attention since the recent withdraw of the United States from the Paris Agreement. In theory, this type of uncertainty turns out to be a key factor in determining what a "good" IEA looks like.[4]

Several recent studies have discussed the political backgrounds behind participation uncertainty. For example, Köke and Lange (2017) analyze the ratification process of an IEA by introducing public choice approach into the model. They assume that the preference of the agent who is responsible for the ratification of IEA is subject to uncertainty, leading to a possibility of ratification failure. Cazals and Sauquet (2015) suggest that the ratification probability of a country may depend on whether it is a developed country or a developing one, and whether leaders have incentive to postpone the ratification. By contrast, in this paper we ignore the detailed political reason or process that creates the uncertainty, and simply introduce a given parameter

[3] The introduction of one-way uncertainty has already imposed enough restrictions on the stability of coalitions such that there exists a unique stable size of coalition (See Proposition 1 in section 3). Additional restrictions from two-way uncertainty will generally be too strong for a stable coalition to exist.

[4] See the uncertainty effect in section 5 of this paper.

to be the exogenous probability with which each cooperating country will not become a signatory.

A second reason of why IEAs are not so successful is, the IEA rules used in some early works are exogenously given and may not be justified as appropriate ones. Notably, a large body of studies[5] assume that in the abatement stage, given any coalition of signatories that has already been formed, all signatories should coordinate their actions and maximize the joint payoffs of this coalition. We call this rule the maximizing total payoff (MTP) rule. Meanwhile, some other studies[6] apply the coalition unanimity (CU) rule, which requires participation of all countries for an IEA to be effective. In section 5, we will show that generally both the MTP rule and the CU rule are not "optimal" in the presence of participation uncertainty.[7] Briefly speaking, the MTP rule is more robust against the aforementioned uncertainty, but may provide insufficient incentives for cooperation. By contrast, although the CU rule provides large participation incentives, it suffers a lot from uncertainty. Some other rules used in the literature are combinations or extensions of the MTP rule and/or the CU rule, and will be discussed in section 5.

In order to overcome the problem raised by exogenous IEA rules, several studies have analyzed the endogenous determination of the IEA rules from some specific aspects, allowing one to choose a rule among a certain set of feasible rules. Altamirano-Cabrera et al. (2008) compare the MTP rule with three other rules that involving abatement quotas. Finus and Maus (2008) assume that signatories consider only a fraction α of damages and endogenously determine the optimal value of α. Carraro et al. (2009) discuss the MTP rule with an additional restriction of minimal participation; here, the threshold of forming the coalition is endogenously determined. Köke and Lange (2017) consider an endogenous rule that simultaneously determine the threshold of cooperation and the signatories' abatement level. However, these studies analyze only certain special cases of endogenous rules and hence cannot be considered fully general.

[5] Usually called noncooperative approach. See, among others, Carraro and Siniscalco (1993), and Barrett (1994).

[6] Sometimes called cooperative approach. See Chander and Tulkens (1997) for example.

[7] The formal definition of optimal rule will be given in section 4.

Following the above literature, the main purpose of this study is to extend the traditional IEA formation model to allow for participation uncertainty and more general IEA rules where the required abatement level for each signatory is fully contingent on the number of signatories. We hope this extension will help us design a better IEA rule than those used in the literature.

To this end, we employ a three-stage coalition formation game under the setting of symmetric countries. In stage one (designing stage), a designer[8] launches a coalition and announces an IEA rule, which is a function specifying the abatement level of each signatory for every possible cardinality of the coalition of signatories formed in stage two. The designer's goal is to maximize the expected payoff of each signatory. In stage two (participation stage), each country can independently choose whether to cooperate or to free ride. There exists a one-way participation uncertainty in this stage so that each cooperator has probability $\varepsilon \geq 0$ to end up a non-signatory and probability $1-\varepsilon$ to be a signatory, while each free rider will definitely become a non-signatory. Stage three is the usual abatement stage that determines each country's abatement level and payoff.

Compared to traditional IEA formation models, our model has the following features. First, the class of rules we study are very general in the sense that the assigned actions for signatories are fully contingent on the number of signatories. In particular, this type of rules are flexible enough to incorporate many commonly used rules as special cases, some of which will be discussed in section 5. Of course, more general rules are possible if we study more general model settings such as asymmetric players, but they are beyond the scope of this paper.

Second, the rule is endogenously chosen by the designer. Since the designer is the initiator of the coalition, it is quite natural to assume that he/she does not care about the interests of non-signatories so as to discourage free-riding behavior. Therefore, it seems reasonable to assume the designer's goal is to maximize the expected payoff of signatories. Another justification of this assumption is that it guarantees that the coalition of signatories formed

[8] The designer can be a set of countries or an independent organization, for example, the United Nations.

in stage two is renegotiation-proof. That is, once this coalition is formed under the rule based on above goal, the signatories will not unanimously agree on replacing this rule by a new rule such as the MTP rule. Otherwise, the coalition that is expected to form under the new rule will change accordingly, leading to a smaller (at least not larger) expected payoff of each signatory.

Third, the setting of endogenously determined rules and the presence of participation uncertainty would largely increase the analytical complexity of the model. So we have to keep other aspects of our model as simple as possible to make it tractable and to illustrate the main idea in a more transparent manner. Specifically, we assume that all countries are *ex ante* symmetry, that the game is one-shot instead of dynamic, that the pollutant is a flow rather than a stock, and that the benefit function is linear while the cost function is quadratic. Although these simplified assumptions have been used in the IEA literature somewhere or other[9], some of them are surely not realistic and cannot be directly applied to design an IEA rule in real world. For instance, the linear benefit function results in a dominant abatement level for each non-signatory, which greatly simplify our analysis; however, it also makes this model difficult to capture the issue linkage prevails among countries. Nevertheless, starting from this basic model, some more complicated and more realistic situations can be left for future studies.

The internal and external stability concept introduced by d'Aspremont et al. (1983) has been widely applied in coalition formation literature for anticipating which coalition will be formed in the participation stage. However, there are two important reasons why this stability concept is not appropriate for our model. First, internally and externally stable coalitions are not necessarily unique. Thus, in the first stage of the game, the designer is typically not able to determine the precise payoff outcome associated with each certain rule, and consequently has difficulty in comparing different rules. Second, the existence of participation uncertainty makes it more difficult to determine the coalition formed in participation stage. Specifically, even if a coalition is internally and externally stable, it can still be vulnerable to

[9] Especially when the model involves uncertainty. For example, see Kolstad (2007) and Köke and Lange (2017).

deviation after some degree of uncertainty is realized[10]. Luckily, we find a stronger version of stability concept, namely *ex ante* and *ex post* stability, that overcomes both issues. Simply put, a coalition is *ex ante* and *ex post* stable if no country will change its decision both before and after any degree of uncertainty is realized. In addition, *ex ante* and *ex post* stable coalition is essentially unique under any IEA rule.

By using this stability concept, we can solve the three-stage IEA formation game by backward induction. We prove that given any IEA rule, the size of stable coalitions can be uniquely determined (Proposition 1). Therefore, a unique value of the designer's objective can be associated with each IEA rule. Furthermore, we provide an algorithm to derive a rule which maximizes the above objective (Theorem 1), and name it optimal rule[11]. Thus, we can determine the size of the coalition of countries that choose to cooperate in stage two, and the rule that the designer announces in stage one.

To illustrate the difference between the optimal rule and some rules used in the literature, a numerical example is discussed. This example shows that the MTP rule, the CU rule, and some other rules are all not optimal for the designer in general. Moreover, we provide some conditions under which these rules are optimal or almost optimal (Proposition 3, Proposition 4), and explain the intuition behind these outcomes. In short, the optimal rule could be better than other rules mainly because it is more flexible and thus can be more appropriately designed when facing the tradeoff between two conflicting purposes: punishing free-riding behavior and reducing uncertainty cost.

The remainder of this paper is organized as follows. Section 2 presents the setup of the model and the three-stage coalition formation game. We introduce *ex ante* and *ex post* stability to solve this game in section 3, and derive an optimal IEA rule in section 4. Several special IEA rules are discussed and compared with the optimal rule in section 5. Finally, section 6 concludes the paper.

[10] See section 3 for more detailed explanation on this issue.

[11] Of course, it is only optimal within the class of rules we discussed.

2 The model

Let $N = \{1, 2, \ldots, n\}$ be a set of homogeneous countries, where $n \geq 2$. There is a perfectly divisible good with negative externalities, for example, a pollutant. Let x_i denote country i's abatement level of the good and let $x = (x_1, \ldots, x_n)$ be an abatement combination.

Given x, country i's payoff is

$$u_i(x) = \lambda \sum_{j \in N} x_j - \frac{1}{2} x_i^2, \tag{1}$$

where $\lambda > 0$ is the common marginal benefit from total abatement $\sum_{j \in N} x_j$ due to negative externalities of the good, and $x_i^2/2$ is country i's individual abatement cost. Assume that payoffs are transferable, and the social welfare is the total payoffs of all countries:

$$U(x) = \sum_{i \in N} u_i(x) = n\lambda \sum_{i \in N} x_i - \sum_{i \in N} \frac{x_i^2}{2}.$$

An abatement combination $x^* = (x_1^*, \ldots, x_n^*)$ is said to be socially optimal if it maximizes social welfare. The first-order conditions $\partial U(x)/\partial x_i = 0$ yield

$$x_i^* = n\lambda, \quad \forall i \in N. \tag{2}$$

On the other hand, if each country i chooses x_i to maximize its own payoff u_i given the other countries' abatement levels, the first-order conditions $\partial u_i(x)/\partial x_i = 0$ lead to

$$x_i^0 = \lambda, \quad \forall i \in N. \tag{3}$$

Note that x_i^0 is a dominant abatement level of i, regardless of other countries' actions. From this, it follows that $x^0 = (x_1^0, \ldots, x_n^0)$ is the unique Nash equilibrium of this non-cooperative abatement game.

Since $x_i^* > x_i^0$, the world suffers from too much emission of the good. This is a commonly known social dilemma due to free rider problem. One possible method to overcome this problem is to form a coalition that regulates the countries' actions by a self-enforcing IEA. The formation of the coalition and the determination of the IEA rule follows a three-stage game.

- Stage one. A designer announces an IEA rule, which is a function $e(\cdot)$ assigning a real value $e(\overline{m}) \geq 0$ to each integer $\overline{m} \in [1, n]$, where $\overline{m}$ is the cardinality of the coalition of signatories $\overline{M}$ that will be formed in stage two. We shall see that $e(\overline{m})$ is indeed the required abatement level for each signatory.[12] A rule $e(\cdot)$ can also be denoted by a vector $e = \big(e(1), \ldots, e(n)\big) \in \mathbb{R}^n_+$. Let R denote the set of all rules.

- Stage two. All countries in N simultaneously decide whether to be a cooperator and intend to sign the IEA, or to be a free rider and determine not to sign. Let M denote the set of cooperators, and let $m = |M|$ denote its cardinality. However, there is a one-way uncertainty with regard to each cooperator's final participation decision. Specifically, a cooperator $i \in M$ will make a mistake and end up a non-signatory with probability ε, and become a signatory with probability $1 - \varepsilon$, where $0 \leq \varepsilon < 1$ is exogenously given. However, each free rider $j \notin M$ never makes a mistake and would certainly become a non-signatory. Figure 1 demonstrates the relationship of these types of countries. Let $\overline{M}$ denote the set of signatories, and $\overline{m} = |\overline{M}|$ be its cardinality. $\overline{M}$ is obviously a subset of M, and hence $\overline{m} \leq m$. Given m and $\varepsilon \in (0, 1)$, $\overline{m}$ follows a binomial distribution so that the probability that $\overline{m} = k$ is

  $$b(k; m, 1-\varepsilon) = \frac{m!}{k!(m-k)!}\varepsilon^{m-k}(1-\varepsilon)^k, \quad \forall k = 0, 1, \ldots, m. \tag{4}$$

  Additionally, if $\varepsilon = 0$, then $b(m; m, 1) = 1$, $b(k; m, 1) = 0$, $\forall k < m$.

- Stage three. Given rule e and coalition $\overline{M}$, each signatory $i \in \overline{M}$ carries out its abatement $x_i = e(\overline{m})$ according to rule e, while each non-signatory $j \notin \overline{M}$ chooses its dominant abatement level $x_j = \lambda$. All countries receive their respective payoffs.

Note that a rule $e(\cdot)$ is defined on the number of signatories $\overline{m}$, rather than on the number of cooperators m. This is because whether a country

[12] Due to symmetry, $e(\overline{m})$ is the same for all signatories, and only depends on $\overline{m}$.

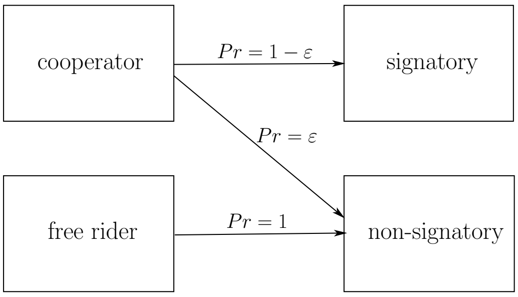

Figure 1: The relations of players

is a signatory or non-signatory is observable, but whether it has chosen to cooperate or to free ride is its own subjective decision and cannot be directly observed by others — this can only be deduced by comparing this country's expected payoffs involving different choices. If a non-signatory is better off cooperating than free-riding, then it should have chosen to be a cooperator, but somehow makes a mistake due to some political reasons, as addressed in the introduction. In sum, the unobservable set M only depends on countries' subjective choices[13], while the observable set $\overline{M}$ depends on countries' choices as well as the realization of uncertainty.

Now, let $G(n, \lambda, \varepsilon)$ denote the above three-stage game. Assume that each country is risk neutral and chooses its action (cooperate or free ride) to maximize its expected payoff. In particular, a country will choose to cooperate if it is indifferent between cooperating and free riding. Meanwhile, because of the reasons addressed in the introduction, we assume that the designer will choose a rule to maximize the (identical) expected payoff of each signatory.

3 Stable coalition

We can solve game $G(n, \lambda, \varepsilon)$ by backward induction. Consider stage three first. Given e and $\overline{m}$, let $X(\overline{m}, e) = \overline{m}e(\overline{m}) + (n - \overline{m})\lambda$ denote the total

[13] Nevertheless, we will show in Proposition 1 that essentially a stable M can be uniquely determined for each rule e.

abatement of all countries. Now, a non-signatory's payoff is

$$\overline{u}^I(\overline{m}, e) = \lambda X(\overline{m}, e) - \frac{\lambda^2}{2}, \quad \text{if} \ \ \overline{m} < n, \tag{5}$$

and a signatory's payoff is

$$\overline{u}^C(\overline{m}, e) = \lambda X(\overline{m}, e) - \frac{e(\overline{m})^2}{2}, \quad \text{if} \ \ \overline{m} \geq 1. \tag{6}$$

Additionally, when no country signs the IEA, define

$$\overline{u}^C(0, e) = u_i(x^0) = \Big(n - \frac{1}{2}\Big)\lambda^2, \tag{7}$$

so that $\overline{u}^C(\overline{m}, e)$ is well-defined for all $\overline{m} \in [0, n]$. This will be helpful in defining (15).

Now we study stage two of $G(n, \lambda, \varepsilon)$. Given e and m, consider the expected payoff of a cooperator $i \in M$. With probability $b(k; m-1, 1-\varepsilon)$ there will be k signatories among all $m-1$ cooperators other than i. In this case, country i itself has probability ε to be a non-signatory and get payoff $\overline{u}^I(k, e)$, and has probability $1-\varepsilon$ to be a signatory and get payoff $\overline{u}^C(k+1, e)$. Therefore, each cooperator's expected payoff is

$$u^C(m, e) = \sum_{k=0}^{m-1} b(k; m-1, 1-\varepsilon) \left[\varepsilon \overline{u}^I(k, e) + (1-\varepsilon)\overline{u}^C(k+1, e)\right]. \tag{8}$$

Similarly, each free rider's expected payoff is

$$u^I(m, e) = \sum_{k=0}^{m} b(k; m, 1-\varepsilon)\overline{u}^I(k, e). \tag{9}$$

We use the concept of stable coalition to predict which countries choose to cooperate in stage two. We will see that it is a stronger version of the internally and externally stability concept introduced by d'Aspremont et al. (1983). Roughly speaking, a coalition M is stable if the countries in M are the only ones choosing to cooperate before any uncertainty is realized (*ex ante* stable), and the cooperators would not change their decisions even after some degree of uncertainty is realized (*ex post* stable).

Formally, coalition M is said to be *ex ante* stable relative to e if no country $i \in M$ is willing to unilaterally leave M (internally stable) and no country $j \notin M$ is willing to unilaterally join M (externally stable) before uncertainty is realized. Hence, a coalition $M \notin \{\emptyset, N\}$ is *ex ante* stable relative to e if

$$u^C(m,e) \geq u^I(m-1,e),\ u^C(m+1,e) < u^I(m,e). \tag{10}$$

In addition, $M = \emptyset$ is *ex ante* stable relative to e if $u^C(1,e) < u^I(0,e)$, while $M = N$ is *ex ante* stable relative to e if $u^C(n,e) \geq u^I(n-1,e)$.

When there is no participation uncertainty, *ex ante* stability is identical to internal and external stability and is sufficient to identify the cooperator set M. However, the existence of participation uncertainty makes things more complex. For instance, suppose M is an *ex ante* stable coalition relative to e. After realizing that some other cooperators have become non-signatories[14], a cooperator may regret for its former decision and choose to leave M.[15] In other words, being *ex ante* stable only suggests that the coalition is stable before any uncertainty is realized, but this does not guarantee that the coalition will remain stable after some degree of uncertainty is realized. This leads to the *ex post* stability concept defined below.

A coalition $M \neq \emptyset$ is said to be *ex post* stable relative to e if, no matter how many (at least one) cooperators have become non-signatories, no other country will change its decision from cooperating to free riding under e.[16] That is,

$$u^C(s,e) \geq u^I(s-1,e),\ \forall s \in [1, m-1], \tag{11}$$

where s is the number of cooperators that have not made mistakes yet. In addition, $M = \emptyset$ is trivially defined to be *ex post* stable relative to any rule.

Finally, M is said to be stable relative to e if it is both *ex ante* and *ex post* stable relative to e. Ultimately, a stable coalition will not provide any

[14] That is, $\overline{M}$ is a proper subset of M.

[15] For example, after the United States quits the Paris agreement, China may also consider whether or not to quit.

[16] Although each cooperator still has a change to make a mistake.

incentive for any country to change its decision regarding participation under any circumstance. In sum, we have:

(a) $M = \emptyset$ is stable relative to e, if and only if

$$u^C(1, e) < u^I(0, e); \tag{12}$$

(b) $M \notin \{\emptyset, N\}$ is stable relative to e, if and only if

$$u^C(m + 1, e) < u^I(m, e),\ u^C(s, e) \geq u^I(s - 1, e),\ \forall s \in [1, m]; \tag{13}$$

(c) $M = N$ is stable relative to e, if and only if

$$u^C(s, e) \geq u^I(s - 1, e),\ \forall s \in [1, n]. \tag{14}$$

Because of the symmetry of countries, whether a coalition M is stable relative to rule e depends only on its cardinality m. If a coalition M is stable relative to e, then we say that m is a stable size relative to e. Similarly, if M is *ex ante* (*ex post*) stable relative to e, then m is said to be an *ex ante* (*ex post*) stable size relative to e.

The following proposition establishes the existence and uniqueness of stable size relative to any given rule. Given the symmetry of countries, the uniqueness of stable size implies that all stable coalitions are essentially identical. Let $m(e)$ denote this unique stable size relative to e. Thus, if the designer announces rule e in stage one, then in stage two there will be $m(e)$ cooperators. In the proof, we will provide an algorithm to derive $m(e)$, and show that $m(e)$ is exactly the smallest *ex ante* stable size for each e.

Proposition 1. *There is a unique stable size $m(e)$ relative to each rule e.*

Proof. First, we prove that there exists at most one stable size relative to any e. Assume that both m_1 and m_2 are *ex ante* stable sizes relative to e, where $m_1 < m_2$. We only need to show that m_2 cannot be an *ex post* stable size, and thus cannot be a stable size, relative to e. Assume for a contradiction that m_2 is an *ex post* stable size relative to e. Then according to (11), we have $u^C(m_1 + 1, e) \geq u^I(m_1, e)$ since $m_1 < m_2$. However, this contradicts the assumption that m_1 is an *ex ante* stable size relative to e, due to (10). Therefore, we have proved that there exists at most one stable size $m(e)$, and if it exists, $m(e)$ must be the smallest *ex ante* stable sizes relative to e.

Now we show that the stable size $m(e)$ always exists for each e, using a method introduced by d'Aspremont et al. (1983). If $u^C(1, e) < u^I(0, e)$, then we find a stable size $m(e) = 0$ according to (12); otherwise, we have $u^C(1, e) \geq u^I(0, e)$. Furthermore, if $u^C(2, e) < u^I(1, e)$, then stable size is $m(e) = 1$ according to (13); otherwise, we have $u^C(1, e) \geq u^I(0, e)$, $u^C(2, e) \geq u^I(1, e)$. Proceeding in this manner, we shall either find a stable size $m(e) < n$, or eventually have $u^C(k, e) \geq u^I(k-1, e)$, $k = 1, 2, \ldots, n$, which imply that the stable size is $m(e) = n$ according to (14). □

Given $e \in R$, let $Eu^C(e)$ denote the expected payoff of a signatory from the viewpoint of the designer in stage one. Since there are $m(e)$ cooperators, the probability that there exist exactly k signatories is $b\big(k; m(e), 1-\varepsilon\big)$ for any $k \leq m(e)$. Thus, we have

$$Eu^C(e) = \sum_{k=0}^{m(e)} b\big(k; m(e), 1-\varepsilon\big)\overline{u}^C(k, e). \tag{15}$$

Recall that from (7), when there is no signatory ($\overline{m} = 0$), the designer will take a non-signatory's payoff as a substitute for a signatory's payoff. We make this trivial assumption only to ensure that $Eu^C(e)$ is always well-defined, even though it is possible that $\overline{M} = \emptyset$.

4 Optimal rule

Now, consider stage one of $G(n, \lambda, \varepsilon)$. The objective of the designer in this stage is to maximize $Eu^C(e)$ by choosing rule e. If a rule e exists such that $Eu^C(e) \geq Eu^C(e')$ for all $e' \in R$, then we say that e is optimal. We let R^* denote the set of optimal rules, and let Eu^* be the maximal value of $Eu^C(e)$ if R^* is not empty. That is, $Eu^* = Eu^C(e^*)$ where $e^* \in R^*$.

For $s \in [0, n]$, let $R(s) = \{e \in R \,|\, m(e) = s\}$ denote the set of rules relative to whom the stable sizes are s. A rule $e \in R(s)$ is said to be locally optimal at s, if e is optimal within $R(s)$, that is, $Eu^C(e) \geq Eu^C(e')$ for all $e' \in R(s)$.

The following theorem establishes the existence of optimal rule.

Theorem 1. *For any game $G(n, \lambda, \varepsilon)$, there exists an optimal rule e^*.*

Proof. First, we try to find a locally optimal rule e^*_s at each $s \in [0, n]$. It is easy to verify that $e^*_0 = (n\lambda, \dots, n\lambda)$ is locally optimal at $s = 0$. To find a locally optimal rule at $s \in [1, n]$, define

$$\overline{R}(s) = \big\{ \big(e(1), \dots, e(s)\big) \in \mathbb{R}^s_+ \mid u^C(k, e) \geq u^I(k-1, e), \ \forall k \in [1, s] \big\}. \tag{16}$$

We prove the following lemma in the appendix.

Lemma 1. *For each $s \in [1, n]$, $\overline{R}(s)$ is a non-empty bounded closed set.*

Note that $\sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e)$ only depends on, and is continuous in, $\big(e(1), \dots, e(s)\big)$. From Lemma 1, there exists a vector $\big(e^*_s(1), \dots, e^*_s(s)\big)$ that solves the following constrained optimization problem:

$$\max_{(e(1),\dots,e(s)) \in \overline{R}(s)} \sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e). \tag{17}$$

We can derive this vector by the Kuhn-Tucker theorem.

For $s = n$, it is obvious that $(e^*_n(1), \dots, e^*_n(n)) := e^*_n \in R(n)$. For $s \in [1, n-1]$, we can find a vector $\big(e^*_s(s+1), \dots, e^*_s(n)\big)$, which can be combined with $\big(e^*_s(1), \dots, e^*_s(s)\big)$ to get $e^*_s := \big(e^*_s(1), \dots, e^*_s(n)\big)$. From (6) and (8), as long as $e^*_s(s+1)$ is sufficiently large, $u^C(s+1, e^*_s)$ will be small enough so that $u^C(s+1, e^*_s) < u^I(s, e^*_s)$. Since $\big(e^*_s(1), \dots, e^*_s(s)\big) \in \overline{R}(s)$, $u^C(s+1, e^*_s) < u^I(s, e^*_s)$, we have $e^*_s \in R(s)$.

Now, we shall show that e^*_s is locally optimal at s. Suppose, on the contrary, that there exists $e' \in R(s)$ such that $Eu^C(e') > Eu^C(e^*_s)$. Then, $\sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e') > \sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e^*_s)$. But this contradicts $\big(e^*_s(1), \dots, e^*_s(s)\big) \in \arg\max_{(e(1),\dots,e(s)) \in \overline{R}(s)} \sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e)$.

In sum, we have proved that for any $s \in [0, n]$, there exists a rule e^*_s that is locally optimal at s. Thus, $Eu^C(e^*_s) = \max_{e \in R(s)} Eu^C(e)$.

Finally, we can easily find a rule e^* in the set $\{e^*_0, e^*_1, \ldots, e^*_n\}$ such that

$$Eu^C(e^*) = \max\{Eu^C(e^*_0), Eu^C(e^*_1), \ldots, Eu^C(e^*_n)\}. \tag{18}$$

It is easy to show that e^* is an optimal rule. Suppose for a contradiction that there exists e' such that $Eu^C(e') > Eu^C(e^*)$, and $e' \in R(s)$. Then, from (18), it follows that $Eu^C(e^*) \geq Eu^C(e^*_s) \geq Eu^C(e')$, which contradicts the assumption $Eu^C(e') > Eu^C(e^*)$. This ends the proof of the theorem. □

The proof of Theorem 1 suggests a two-step algorithm to construct an optimal rule. First, find a locally optimal rule e^*_s at each $s \in [0, n]$ mainly by solving the constrained optimization problem (17). Then, among all these locally optimal rules pick an optimal rule e^* by (18).

Table 1: Locally optimal rules e^*_s for $G(5, 2, 0.1)$

s	e^*_s	$Eu^C(e^*_s)$
0	$(10, 10, 10, 10, 10)$	18
1	$(2, 10, 10, 10, 10)$	18
2	$(2, 4, 10, 10, 10)$	19.62
3	$(2, 4, 6, 10, 10)$	24.32
4	$(2, 2.95, 5.23, 8, 10)$	32.13
5	$(2, 2, 3.61, 6.44, 10)$	42.78

To illustrate how to derive an optimal rule, consider example $G(5, 2, 0.1)$. In Table 1, we list for each $s \in [0, 5]$ a locally optimal rule e^*_s and the corresponding value of $Eu^C(e^*_s)$. Since $Eu^C(e^*_5) > Eu^C(e^*_s)$ for all $s < 5$, we have an optimal rule $e^* = e^*_5 =$ (2, 2, 3.61, 6.44, 10), and $Eu^* = 42.78$.

A careful observation of Table 1 also motivates the following property, which partially describes how (locally) optimal rules look like. It also tell us that if $m(e^*) = n$ and no country makes any mistake, then each signatory's abatement level $e^*(n) = n\lambda$ is socially optimal.

Proposition 2. *If e^*_s is locally optimal at $s \geq 1$, then $e^*_s(1) = \lambda$, $e^*_s(s) = s\lambda$.*

Proof. See the appendix. □

5 Comparison of rules

Now, we discuss some special rules commonly used in the literature and compare them to the optimal rule e^*.

(a) A rule e^a is called the MTP rule if it always aims to maximize the total payoffs of all signatories, no matter which coalition $\overline{M}$ is formed. Because of the symmetry of countries, we have $\overline{m} \cdot \overline{u}^C(\overline{m}, e^a) \geq \overline{m} \cdot \overline{u}^C(\overline{m}, e')$, or $\overline{u}^C(\overline{m}, e^a) \geq \overline{u}^C(\overline{m}, e')$, for all $\overline{m} \in [1, n]$ and all $e' \in R$. That is, e^a maximizes $\overline{u}^C(\overline{m}, e)$, and thus $e^a(\overline{m}) = \lambda \overline{m}$, for all $\overline{m} \in [1, n]$.

(b) A rule e^b is called a minimal participation rule (Köke and Lange, 2017) if there exists $\widetilde{m} \in [1, n]$ and $q > \lambda$, such that $e^b(\overline{m}) = \lambda$ when $1 \leq \overline{m} < \widetilde{m}$, and $e^b(\overline{m}) = q$ when $\overline{m} \geq \widetilde{m}$. In other words, this rule requires an abatement level q for signatories when at least $\widetilde{m}$ countries sign the IEA, while there is no requirement for signatories when less than $\widetilde{m}$ countries sign. In particular, if $\widetilde{m} = n$, e^b is called a coalition unanimity (CU) rule (Chander and Tulkens, 1997).

(c) A rule e^c is called an MTP rule with minimal participation (Carraro et al., 2009) if there exists $\widetilde{m} \in [1, n]$, such that $e^c(\overline{m}) = \lambda$ for all $1 \leq \overline{m} < \widetilde{m}$, and $e^c(\overline{m}) = \lambda \overline{m}$ for all $\overline{m} \geq \widetilde{m}$. Obviously, e^c is a combination of e^a and e^b. In particular, if $\widetilde{m} = 1$, e^c is the MTP rule; if $\widetilde{m} = n$, e^c is a CU rule.

If there is a minimal participation rule e^{b*} such that $Eu^C(e^{b*}) \geq Eu^C(e^{b'})$ for all minimal participation rule $e^{b'}$, then e^{b*} is called an optimal minimal participation rule. We can derive an optimal minimal participation rule e^{b*} for any given game $G(n, \lambda, \varepsilon)$ by the following algorithm. First, for any integer $s \in [1, n]$, choose $q = q^*_s$ to maximize $Eu^C(e^b)$ under constraint $\widetilde{m} = s$, and get a corresponding rule e^{b*}_s. Then, $Eu^C(e^{b*})$ is the maximal value in $\{Eu^C(e^{b*}_s) \,|\, 1 \leq s \leq n\}$, and e^{b*} can thus be found in $\{e^{b*}_s \,|\, 1 \leq s \leq n\}$. Similarly, we can define and derive optimal MTP rule with minimal participation e^{c*} by choosing appropriate $\widetilde{m}$. For example, consider game

$G(5, 2, 0.4)$. According to Table 2, we have $e^{b*} = (2,\ 2,\ 5.37,\ 5.37,\ 5.37)$, $Eu^C(e^{b*}) = 26.12$; $e^{c*} = (2, 2, 2, 8, 10)$, $Eu^C(e^{c*}) = 25.15$.

Table 2: Calculating e^{b*} and e^{c*} for $G(5, 2, 0.4)$

$\widetilde{m} = s$	q_s^*	$m(e_s^{b*})$	$Eu^C(e_s^{b*})$	$m(e_s^{c*})$	$Eu^C(e_s^{c*})$
1	2	5	18	3	20.59
2	3.23	4	20.89	3	20.59
3	5.37	5	**26.12**	4	23.10
4	8.46	5	25.03	5	**25.15**
5	10	5	20.49	5	20.49

Now we compare rules e^a, e^{b*} and e^{c*} with optimal rule e^*. We first consider an example $G(5, 2, \varepsilon)$, and list the corresponding $m(e)$ and $Eu^C(e)$ for $\varepsilon \in \{0, 0.1, \ldots, 0.9\}$ in Table 3. For each ε in this table, the maximal value of $Eu^C(e)$, where $e \in \{e^a, e^{b*}, e^{c*}\} := R^0$, is highlighted in bold to show which of the three rules is best for the designer. Moreover, let

$$\eta = \max_{e \in R^0} \frac{Eu^C(e^*) - Eu^C(e)}{Eu^C(e)} \times 100$$

denote the percentage of payoff improvement of e^* relative to the best rule in R^0. We list the values η in Table 3 to show at least how much better the optimal rule can be than the other rules in R^0.

Table 3: Simulation for $G(5, 2, \varepsilon)$

	e^a		e^{b*}		e^{c*}		e^*		
ε	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	η
0	3	26	5	**50**	5	**50**	5	50	0
0.1	3	24.32	5	**36.90**	5	**36.90**	5	42.78	15.93
0.2	3	22.86	5	35.20	5	**35.86**	5	36.88	2.84
0.3	3	21.63	5	29.63	5	**29.86**	5	32.02	7.23
0.4	3	20.59	5	**26.12**	5	25.15	5	28.03	7.31
0.5	4	21.88	5	23.88	5	**24.31**	5	24.79	1.97
0.6	4	20.38	5	21.38	5	**21.55**	5	22.22	3.11
0.7	5	**20.26**	5	19.79	5	**20.26**	5	20.26	0
0.8	5	**18.94**	5	18.82	5	**18.94**	5	18.94	0
0.9	5	**18.22**	5	18.20	5	**18.22**	5	18.22	0

In this example, we can see from Table 3 that e^{c*} weakly dominates e^{a}, but neither e^{b*} or e^{c*} is dominated by each other. More interestingly, when ε is zero, the CU rule $e^{b*} = e^{c*}$ is optimal; when ε is large, the MTP rule $e^{a} = e^{c*}$ is optimal; when ε is positive but not too large, none of these three rules in R^{0} is optimal. These outcomes can be explained as follows.

For the designer, a rule e has two important aspects that may affect the value of the objective function $Eu^{C}(e)$. On the one hand, the designer wishes more countries to sign the IEA and that the signatories engage in a high abatement level. Thus, the rule should provide for a strong incentive for cooperation, or a strong punishment for free riding, by creating a large payoff gap between signing and not signing. On the other hand, the designer also wishes to reduce the harm that participation uncertainty brings on the expected payoffs of signatories. This can be accomplished only by designing a rule by which even when some countries do not sign the IEA due to mistakes, those signatories can still maintain a relatively high level of abatement, leading to a small payoff gap between signing and not signing.

We call these two aspects of the rules as incentive effect and uncertainty effect, respectively. A rule has a strong/weak incentive effect if it provides strong/weak incentives for countries to cooperate; a rule has a strong/weak uncertainty effect if a certain ε has a small/large impact on $Eu^{C}(e)$.

These two effects are typically contradictory. A rule with a strong uncertainty effect usually has a weak incentive effect. This is because any factor of the rule protecting the signatories from harm caused by uncertainty will require those signatories to maintain a high level of abatement regardless of the other countries' mistakes. However, this requirement would also reduce the incentive for cooperation. The designer faces a tradeoff between these two conflicting effects. To design an appropriate rule under any specific situation, the designer should evaluate the relative importance of the two effects.

From Table 3, when $\varepsilon = 0$ the optimal minimal participation rule e^{b*}, which is currently a CU rule, is optimal. This is because this rule has a strong incentive effect and weak uncertainty effect — even a very small uncertainty will cause a huge loss on $Eu^{C}(e^{b*})$. However, the uncertainty effect is irrelevant as $\varepsilon = 0$. Moreover, the next proposition shows that there is a

CU rule that is "almost" optimal when uncertainty is sufficiently small.

Proposition 3. *Suppose $e^b(n) = \lambda n$, $e^b(m) = \lambda$ for all $m < n$. For any $\mu > 0$, there exists $\gamma > 0$ such that when $\varepsilon < \gamma$, $Eu^C(e^b) > Eu^C(e') - \mu$ for all $e' \in R$.*

Proof. It is obvious that $m(e^b) = n$. Given any $\mu > 0$, when ε is sufficiently small, $Eu^C(e^b) = \sum_{k=0}^{n} b\big(k; n, 1-\varepsilon\big)\overline{u}^C(k, e^b)$ can be arbitrarily close to $\overline{u}^C(n, e^b)$, and thus $Eu^C(e^b) > \overline{u}^C(n, e^b) - \mu$. From (6), it is easy to verify that $\overline{u}^C(n, e^b) \geq \overline{u}^C(m, e')$ for all $m \leq n$ and all e'. Hence, $Eu^C(e^b) > \overline{u}^C(n, e^b) - \mu \geq \sum_{k=0}^{m(e')} b\big(k; m(e'), 1-\varepsilon\big)\overline{u}^C(k, e') - \mu = Eu^C(e') - \mu$. □

In contrast, from Table 3, the MTP rule e^a is an optimal rule only when ε is large enough. This turns out to be a general outcome according to the next proposition.

Proposition 4. *When ε is sufficiently large, $m(e^a) = n$, and e^a is optimal.*

Proof. See the appendix. □

Intuitively, Proposition 4 holds because the MTP rule has a strong uncertainty effect and a weak incentive effect, but the former is more important when ε is large. The reason why e^a has a strong uncertainty effect is, it requires signatories to maintain a relatively high level of abatement so that the joint payoff of $\overline{M}$ is maximized, even when $\overline{m}$ is small due to participation uncertainty.

The fact that the MTP rule may not be optimal under small uncertainty seems to be counterintuitive at first glance. Once a coalition $\overline{M}$ is formed, it is quite natural to require all signatories to act as one player and maximize their total payoffs. This explains why the MTP rule is so popular in the coalition formation literature. However, a shortcoming of this rule is that it has a weak incentive effect and hence may not effectively overcome the free rider problem. Indeed, a reasonable rule should require the maximization of corresponding payoff for only a stable coalition, rather than for all coalitions. It is these latter redundant requirements that lead to a weak incentive effect and undermine the MTP rule.

There has been a debate regarding which one of the two rules, the MTP rule (the noncooperative approach) or the CU rule (the cooperative approach), is more appropriate for an IEA.[17] According to Proposition 3 and 4, the answer depends on the value of ε. For the designer, the MTP rule is better than the CU rule when participation uncertainty is large enough — that is, when uncertainty effect is more important than incentive effect.

From Table 3, the optimal MTP rule with minimal participation e^{c*} always weakly dominates e^a. This is obvious because e^a is a special type of e^c. Moreover, according to Proposition 4 and Proposition 3, when ε is large enough, $e^{c*}(= e^a)$ is optimal, while when ε is sufficiently small, $e^{c*}(= e^{b*})$ is almost optimal. However, when ε is neither very large or very small, e^{c*} may be significantly away from optimal, and so do e^a and e^{b*}.

A common problem with e^a, e^{b*} and e^{c*} is, sometimes they are not flexible enough to be an optimal rule. In particular, in our model setup, these rules are somehow either constant or linear in $\overline{m}$. For example, e^a is linear in $\overline{m}$; e^{b*} is a step function that is constant both when $\overline{m} < \widetilde{m}$ and when $\overline{m} \geq \widetilde{m}$; e^{c*} is constant when $\overline{m} < \widetilde{m}$, and is linear in $\overline{m}$ when $\overline{m} \geq \widetilde{m}$. By contrast, the example in the previous section shows that e^* need not be constant or linear in $\overline{m}$. Therefore, one advantage of e^* is that it has more flexibility to fully exploit the potential of attaining a larger value of $Eu^C(e)$ when facing the tradeoff between incentive effect and uncertainty effect.

Other than uncertainty parameter ε, we can also analyze the influence of other parameters (n and λ) on the difference in signatories's welfare between optimal rule and other rules. When doing so, we are more interested in cases when ε is fixed to a value that is positive but not very large so that e^* is typically strictly better than the other three rules in R^0. In Table 4, we consider example $G(n, 2, 0.4)$ with $3 \leq n \leq 10$. In Table 5, we study another example $G(5, \lambda, 0.4)$ with $0.5 \leq \lambda \leq 4$. Note that in both examples, there is no sign that η would converge to zero as n and λ becomes larger. Without a rigorous analysis, these examples suggest that the advantage of e^* over other rules will probably remain to exist for arbitrary number of countries and any value of marginal benefit from abatement.

[17] For example, see Tulkens (1998).

Table 4: Simulation for $G(n, 2, 0.4)$

	e^a		e^{b*}		e^{c*}		e^*		
n	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	η
3	3	**12.59**	3	12.26	3	**12.59**	3	12.59	0
4	3	16.59	4	18.91	4	**19.10**	4	19.77	3.51
5	3	20.59	5	**26.12**	5	25.15	5	28.03	7.31
6	3	24.59	6	**35.20**	6	30.30	6	37.24	5.79
7	3	28.59	7	42.59	7	**42.91**	7	47.29	10.21
8	3	32.59	8	**54.11**	8	48.55	8	58.13	7.43
9	3	36.59	9	**63.90**	9	52.83	9	69.70	9.08
10	3	40.59	10	**74.73**	9	56.83	10	81.97	9.69

Table 5: Simulation for $G(5, \lambda, 0.4)$

	e^a		e^{b*}		e^{c*}		e^*		
λ	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	η
0.5	3	1.29	5	**1.63**	5	1.57	5	1.75	7.36
1.0	3	5.15	5	**6.53**	5	6.29	5	7.01	7.35
1.5	3	11.58	5	**14.69**	5	14.15	5	15.77	7.35
2.0	3	20.59	5	**26.12**	5	25.15	5	28.03	7.31
2.5	3	32.18	5	**40.81**	5	39.30	5	43.80	7.33
3.0	3	46.33	5	**58.77**	5	56.60	5	63.07	7.32
3.5	3	63.01	5	**79.99**	5	77.03	5	85.84	7.31
4.0	3	82.37	5	**104.47**	5	100.62	5	112.12	7.32

6 Conclusion

In this study, we introduce a three-stage coalition formation game to endogenously determine IEA rules that are dependent on the number of signatories under participation uncertainty. We provide an algorithm to derive a rule that is optimal in the sense of maximizing the expected payoff of signatories. Part of the past failures of IEAs can be attributed to the use of non-optimal rules, especially when the participation uncertainty is positive but not very large. In particular, the MTP rule has a weak incentive effect and is not optimal unless participation uncertainty is very large; the CU rule has a very weak uncertainty effect and is not optimal as long as there exists participation uncertainty. The main policy implication of our findings is that a good IEA rule should be flexible enough to reach a balance between providing sufficient incentive for cooperation and reducing the losses caused by uncertainty.

Some further works and extensions may be worth studying in future research. First, by now we know little about the properties of optimal rules. For instance, an open question is under what conditions are optimal rules *ex ante* efficient in the sense that they result in full participation and induce socially optimal abatement level before uncertainty is realized; that is, $m(e^*) = n$ and $e^*(n) = n\lambda$.[18] This question is important, because its answer shows when we can be more optimistic about what IEAs can accomplish.

Second, our analysis has taken advantage of the simple model setups such as homogenous countries and simple payoff function. We can study models with more general settings, for example, with heterogenous countries. Along with the extension of heterogenous countries, we may further study more complex IEA rules. For example, a rule may contain an abatement function $e(\overline{M})$ assigning a country-specific abatement level for signatories, and a transfer function $t(\overline{M})$ characterizing the money transferred among signatories.

Third, some other objectives of the designer can be studied. For instance, sometimes it makes more sense to assume that the designer will maximize expected social welfare rather than the signatories' welfare.

Finally, note that in addition to the IEA issue, the MTP rule is widely applied in some other areas involving the voluntary provision of goods with externalities, such as cartel formation in oligopoly markets (d'Aspremont et al., 1983; Donsimoni et al., 1986), cooperation in R&D (Katz, 1986; Poyago-Theotoky, 1995), and sharing natural resource (Miller and Nkuiya, 2016). In a typical application, players first decide whether to join a coalition, and then all coalition members act according to the MTP rule. However, in most of these works, participation uncertainty is implicitly assumed to be zero, which implies that the MTP rule may not always be an appropriate rule.[19] Therefore, it is reasonable and necessary to re-examine the outcome of these works by endogenizing the choice of the coalition rules.

[18] Because of Proposition 2, we only need to know when $m(e^*) = n$ holds.

[19] See Mao (2018) for a criticism of the MTP rule without participation uncertainty.

Appendix

Proof of Lemma 1.

(a) From (16), $\overline{R}(s)$ is obviously a closed set in $\mathbb{R}^s_+$ for each $s \in [1, n]$.

(b) Now, we prove that $\overline{R}(s)$ is a bounded set in $\mathbb{R}^s_+$ by induction on s. We can easily see that $\overline{R}(1) = \{\lambda\}$ is bounded in $\mathbb{R}^1_+$. Assume inductively that $\overline{R}(k)$ is bounded in $\mathbb{R}^k_+$, $1 \leq k \leq n-1$. That is, there exist $T_1, T_2, \ldots, T_k > 0$, such that for each $\big(e(1), \ldots, e(k)\big) \in \overline{R}(k)$: $e(q) < T_q$, $1 \leq q \leq k$.

Now, consider $\overline{R}(k+1)$. According to (16), for each $\big(e(1), \ldots, e(k+1)\big) \in \overline{R}(k+1)$, we have $e(q) < T_q$, $1 \leq q \leq k$. Additionally, $e(k+1)$ satisfies $u^C(k+1, e) \geq u^I(k, e)$; that is,

$$-\frac{1}{2}e(k+1)^2 + a(k+1)e(k+1) + A(k) \geqslant 0,$$

where $A(k)$ depends on $\big(e(1), \ldots, e(k)\big)$. Thus, $e(k+1)$ is also bounded, implying that $\overline{R}(k+1)$ is bounded in $\mathbb{R}^{k+1}_+$. Consequently, $\overline{R}(s)$ is bounded in $\mathbb{R}^s_+$ for each $s \in [1, n]$.

(c) It remains to prove that $\overline{R}(s)$ is not empty. Given $s \in [1, n]$, we can construct $\big(\hat{e}(1), \ldots, \hat{e}(s)\big)$ as follows: (n1) $\hat{e}(s) = \lambda s$; (n2) $\hat{e}(k) = \lambda$, $\forall k \in [1, s-1]$.

For any $m < n$ and any rule e, we have

$$\begin{aligned} & u^C(m+1, e) - u^I(m, e) \\ =& (1-\varepsilon) \sum_{k=0}^{m} b(k; m, 1-\varepsilon) \left[\overline{u}^C(k+1, e) - \overline{u}^I(k, e)\right]. \end{aligned} \tag{19}$$

Note that from (n2), $\overline{u}^C(k+1, \hat{e}) - \overline{u}^I(k, \hat{e}) = 0$, $k \in [1, s-2]$; from (n1) and (n2), $\overline{u}^C(s, \hat{e}) - \overline{u}^I(s-1, \hat{e}) = \frac{1}{2}\lambda^2(s-1)^2 \geq 0$. Hence, from (19), $u^C(m+1, \hat{e}) \geq u^I(m, \hat{e})$, $m \in [0, s-1]$. Therefore, $\big(\hat{e}(1), \ldots, \hat{e}(s)\big) \in \overline{R}(s)$, which implies $\overline{R}(s) \neq \emptyset$. □

Proof of Proposition 2.

$e^*_s(1) = \lambda$ is obviously true since otherwise we have $u^C(1, e^*_s) < u^I(0, e^*_s)$, which contradicts $e^*_s \in E(s)$. It remains to prove $e^*_s(s) = s\lambda$. Given $s \in [1, n]$, suppose for a contradiction that $e_s^*(s) \neq s\lambda$. Let e_s' be a rule such that $e_s'(k) = e_s^*(k)$ for all $k < s$, $e_s'(s) = s\lambda$, and when $s < n$, $e_s'(s+1)$ be large enough so that $u^C(s+1, e_s') < u^I(s, e_s')$.[20]

Since $\overline{u}^C(k, e_s') = \overline{u}^C(k, e_s^*)$ for all $k < s$, $\overline{u}^C(s, e_s^*) = \lambda[se_s^*(s) + (1-s)\lambda] - \frac{1}{2}e_s^*(s)^2 < \frac{1}{2}\lambda^2(s^2 - 2s + 2) = \overline{u}^C(s, e_s')$, we have $u^C(s, e_s') > u^C(s, e_s^*) \geq u^I(s-1, e_s^*) = u^I(s-1, e_s')$, and $u^C(k, e_s') = u^C(k, e_s^*) \geq u^I(k-1, e_s^*) = u^I(k-1, e_s')$ for all $k < s$. Additionally, $u^C(s+1, e_s') < u^I(s, e_s')$. These imply $e_s' \in E(s)$.

Moreover, $Eu^C(e_s') = \sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e_s') > \sum_{k=0}^{s} b\big(k; s, 1-\varepsilon\big)\overline{u}^C(k, e_s^*) = Eu^C(e_s^*)$, which contradicts e_s^* is locally optimal at s. This implies $e_s^*(s) = s\lambda$. □

Proof of Proposition 4.

From the definition of e^a, we have

$$\overline{u}^C(\overline{m}, e^a) - \overline{u}^I(\overline{m}-1, e^a) = -\frac{1}{2}\lambda^2(\overline{m}-1)(\overline{m}-3) = \begin{cases} = 0, & \text{if } \overline{m} = 1, 3 \\ > 0, & \text{if } \overline{m} = 2 \\ < 0, & \text{if } 3 < \overline{m} \leq n \end{cases} \tag{20}$$

When ε is very large, we have

$$b\big(0; m, 1-\varepsilon\big) \gg b\big(1; m, 1-\varepsilon\big) \gg \cdots \gg b\big(m; m, 1-\varepsilon\big), \ \forall m \leq n, \tag{21}$$

where $\gg$ means "far greater than". From (19), (20) and (21), it follows that $u^C(m+1, e^a) - u^I(m, e^a) \geq 0$ for all $m \in [0, n-1]$. Hence, $m(e^a) = n$.

Suppose rule e^0 is optimal and $m(e^0) = n$, then $e^0(1)$ maximizes $\overline{u}^C(1, e)$. Otherwise, we can find e' such that $\overline{u}^C(1, e') > \overline{u}^C(1, e^0)$. Hence, from (15) and (21), $Eu^C(e') > Eu^C(e^0)$, which contradicts e^0 is optimal. Assume inductively that $e^0(k)$ maximizes $\overline{u}^C(k, e)$ for each $k \in [1, m]$, where $m < n$. If e^0 is optimal for a sufficiently large ε, $e^0(k+1)$ also maximizes $\overline{u}^C(k+1, e)$. Otherwise, let e' be such that $e'(s) = e^0(s)$, $s \leq k$, and $\overline{u}^C(k+1, e') > \overline{u}^C(k+1, e^0)$. From (15) and (21), this implies that $Eu^C(e') > Eu^C(e^0)$, and contradicts the assumption that e^0 is optimal.

If $m(e^0) = s < n$, we can show that $Eu^C(e^0) < Eu^C(e^a)$ for a sufficiently

[20] Recall that in the proof of Theorem 1, we have shown that such $e_s'(s+1)$ can be found.

large ε, and thus e^0 cannot be optimal. We only have to consider the case when e^0 is locally optimal at s. Repeating the reasoning process in the previous paragraph, we know that $e^0(k)$ maximizes $\overline{u}^C(k,e)$ for any $k\in[1,s]$. Otherwise, we can find $e'\in R(s)$ such that $Eu^C(e^0)<Eu^C(e')$, which contradicts the assumption that e^0 is locally optimal at s. This implies that $e^0=e^a$, for all $k\leq s$. Also note that $\overline{u}^C(k,e^a)$ is increasing in k, hence $\sum_{k=0}^{s}\big[b\big(k;s,1-\varepsilon\big)-b\big(k;n,1-\varepsilon\big)\big]\overline{u}^C(k,e^0)\approx\big[b\big(0;s,1-\varepsilon\big)-b\big(0;n,1-\varepsilon\big)\big]\overline{u}^C(0,e^0)<\big[b\big(0;s,1-\varepsilon\big)-b\big(0;n,1-\varepsilon\big)\big]\overline{u}^C(s+1,e^a)\approx\sum_{k=0}^{s}\big[b\big(k;s,1-\varepsilon\big)-b\big(k;n,1-\varepsilon\big)\big]\overline{u}^C(s+1,e^a)=\sum_{k=s+1}^{n}b\big(k;n,1-\varepsilon\big)\overline{u}^C(s+1,e^a)$, where "$\approx$" hold because ε is very large. Thus, $Eu^C(e^a)=\sum_{k=0}^{n}b\big(k;n,1-\varepsilon\big)\overline{u}^C(k,e^a)=\sum_{k=0}^{s}b\big(k;s,1-\varepsilon\big)\overline{u}^C(k,e^0)-\sum_{k=0}^{s}\big[b\big(k;s,1-\varepsilon\big)-b\big(k;n,1-\varepsilon\big)\big]\overline{u}^C(k,e^0)+\sum_{k=s+1}^{n}b\big(k;n,1-\varepsilon\big)\overline{u}^C(k,e^a)>Eu^C(e^0)+\sum_{k=s+1}^{n}b\big(k;n,1-\varepsilon\big)\big[\overline{u}^C(k,e^a)-\overline{u}^C(s+1,e^a)\big]\geq Eu^C(e^0)$.

Thus, we have proved that if e^0 is optimal when ε is large enough, then $e^0(k)$ maximizes $\overline{u}^C(k,e)$ for any $k\in[1,n]$, which implies that $e^0=e^a$. That is, e^a is optimal when ε is sufficiently large. □

References

Altamirano-Cabrera, J.C., Finus, M., Dellink, R., 2008. Do abatement quotas lead to more successful climate coalitions? The Manchester School 76, 104–129.

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxford Economic Papers 46, 878–894.

Barrett, S., 2013. Climate treaties and approaching catastrophes. Journal of Environmental Economics and Management 66, 235–250.

Carraro, C. (Ed.), 2003. The Endogenous Formation of Economic Coalitions. Edward Elgar.

Carraro, C., Marchiori, C., Oreffice, S., 2009. Endogenous minimum participation in international environmental treaties. Environmental and Resource Economics 42, 411–425.

Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. Journal of Public Economics 52, 309–328.

Cazals, A., Sauquet, A., 2015. How do elections affect international cooperation? Evidence from environmental treaty participation. Public Choice 162, 263–285.

Chander, P., Tulkens, H., 1997. The core of an economy with multilateral environmental externalities. International Journal of Game Theory 26, 379–401.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J., 1983. On the stability of collusive price leadership. Canadian Journal of Economics 16, 17–25.

Dellink, R., Finus, M., Olieman, N., 2008. The stability likelihood of an international climate agreement. Environmental and Resource Economics 39, 357–377.

Donsimoni, M.P., Economides, N., Polemarchakis, H.M., 1986. Stable cartels. International Economic Review 27, 317–327.

Finus, M., 2001. Game theory and international environmental cooperation. Edward Elgar.

Finus, M., Maus, S., 2008. Modesty may pay! Journal of Public Economic Theory 10, 801–826.

Hong, F., Karp, L., 2014. International environmental agreements with endogenous or exogenous risk. Journal of the Association of Environmental and Resource Economists 1, 365–394.

Karp, L., Simon, L., 2013. Participation games and international environmental agreements: A non-parametric model. Journal of Environmental Economics and Management 65, 326–344.

Katz, M.L., 1986. An analysis of cooperative research and development. The RAND Journal of Economics 17, 527–543.

Kellenberg, D., Levinson, A., 2014. Waste of effort? International environmental agreements. Journal of the Association of Environmental and Resource Economists 1, 135–169.

Köke, S., Lange, A., 2017. Negotiating environmental agreements under ratification constraints. Journal of Environmental Economics and Management 83, 90–106.

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

Mao, L., 2018. A note on stable cartels. Working paper.

Masoudi, N., Zaccour, G., 2017. Adapting to climate change: Is cooperation good for the environment? Economics Letters 153, 1–5.

Meya, J.N., Kornek, U., Lessmann, K., 2017. How empirical uncertainties influence the stability of climate coalitions. International Environmental Agreements: Politics, Law and Economics (online first), 1–24.

Miller, S., Nkuiya, B., 2016. Coalition formation in fisheries with potential regime shift. Journal of Environmental Economics and Management 79, 189–207.

Nkuiya, B., Marrouch, W., Bahel, E., 2015. International environmental agreements under endogenous uncertainty. Journal of Public Economic Theory 17, 752–772.

Poyago-Theotoky, J., 1995. Equilibrium and optimal size of a research joint venture in an oligopoly with spillovers. Journal of Industrial Economics 43, 209–226.

Tulkens, H., 1998. Cooperation vs. free riding in international environmental affairs: two approaches, in: Hanley, N., Folmer, H. (Eds.), Game theory and the environment. Elgar, pp. 330–344.