Dyn Games Appl (2019) 9:122–141
https://doi.org/10.1007/s13235-018-0255-9

Catastrophe and Cooperation

Pim Heijnen[1] · **Lammertjan Dam**[1]

Published online: 20 March 2018

Abstract We study international environmental agreements in the presence of a potential climate catastrophe between sovereign countries that are heterogeneous in their exposure to climate change. We do so by analyzing a stochastic game with an absorbing state. The equilibrium structure of this game is very different from the infinitely repeated games that are usually studied in the literature on environmental agreements. In particular, there is no "folk theorem" that guarantees that the social optimum can be sustained in a Nash equilibrium as long as players are sufficiently patient. However, often, it is feasible to implement an abatement scheme with the same level of aggregate abatement as in the social optimum, but the distribution of abatement among countries is sub-optimal. Moreover, the discount rate has a non-monotonic effect on the optimal environmental agreement.

Keywords International environmental agreement · Voluntary participation · Abatement · Catastrophes · Non-cooperative game

JEL Classification C63 · C73 · H41 · Q54

P. Heijnen: We would like to thank Florian Wagener, Priscilla Man, Allard van der Made and seminar participants at the University of Amsterdam, the University of Queensland, the Tinbergen Institute, the University of Groningen, Umeå University, the ISDG workshop (Barcelona, July 2013), the 13th workshop on optimal control, dynamic games and nonlinear dynamics (Vienna, May 2015), SING11-GTM2015 (St. Petersburg, July 2015), the 22nd annual conference of the EAERE (Zürich, June 2016), and the Vintage Workshop (Vienna, December 2016) for their helpful comments and remarks. We especially thank two anonymous referees for their helpful suggestion that have improved the presentation.

✉ Pim Heijnen
p.heijnen@rug.nl

Lammertjan Dam
l.dam@rug.nl

[1] Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

1 Introduction

The two main economic questions regarding climate change are (i) which policy measures should be taken to combat the negative effects of climate change and (ii) how do we design international environmental agreements to implement these policy measures? In this paper, our focus is on the latter question. We develop a parsimonious dynamic model of international environmental agreements. We argue that the three key issues that shape the form of international environmental agreements are that climate change may be catastrophic, that countries are sovereign, and that countries differ in their exposure to climate change. In this setting, we characterize stable environmental agreements and show that they can be close to the social planner outcome.

By catastrophic climate change we mean an abrupt and permanent change in the climate with which large economic costs are associated.[1] For instance, the rise in global temperatures could trigger the melting of the Siberian permafrost. The subsequent release of methane would lead to a further increase in temperature, leading to the release of more methane and even further increase in temperature. This example is just one possible scenario, but catastrophic shifts in ecological systems are a well-documented phenomenon [28]. Because catastrophes involve a great deal of uncertainty, in both when they will happen and what precisely will happen, the economic costs are large compared to the cost due to any gradual change. For tractability, we focus on the case where the only cost of climate change is the cost of a catastrophic shift. Moreover, the catastrophe is a random event and the probability that the catastrophe occurs decreases if resources are allocated to abatement. The recent literature on regime shifts [9,11,21,27] has devoted much attention to this aspect of climate change, focusing on the optimal choice of one decision maker. However, it is the joint (or aggregate) level of abatement that determines by how much this probability decreases, and our contribution is to extend the analysis to multiple decision makers (countries).

Climate change is global in scale, so limiting the negative effects requires international cooperation. The Kyoto protocol shows that the world is aware of the necessity for cooperation. Unfortunately, as the failure of the USA to ratify the Kyoto protocol illustrates, it also shows that any international environmental agreement needs to entice countries to participate: given that all other countries join, it should be optimal for a country to join as well. This imposes constraints on the form an international environmental agreement can take, as was first recognized by the pioneering work of Carraro and Siniscalco [8], Hoel [17] and Barrett [1,2].[2]

Participation constraints will differ between countries, since some countries will be more severely affected by a climate catastrophe. For instance, a rise in sea levels is a serious issue for a low-lying country like the Netherlands, whereas the direct cost for a country without coastal areas, like Switzerland, will be zero. This renders it more likely that the Netherlands will participate in international environmental agreements.

[1] Note that we are not using the terms "catastrophe" and "catastrophic shift" in the mathematical sense, i.e., the destabilization or vanishing of a steady state of a dynamic system when a system parameter crosses a critical value. In particular, a catastrophe in the mathematical sense of the word is a purely deterministic event. In our setting, the time at which the catastrophe occurs is a random event, perhaps due to uncertainty about the critical value.

[2] Note that, of these pioneering studies, our work is most closely related to Hoel [17], who also starts from the premise that countries are heterogeneous.

These three features of the economic mechanism—catastrophic change, need for cooperation, and heterogeneity in exposure—are modeled in the following way.[3] At each point in time, there are two possible states of the world: pre-catastrophe and post-catastrophe. Pre-catastrophe, each period all countries have the same level of net production and a catastrophe happens with some probability. When the catastrophe occurs, it permanently destroys a fraction of within-period net production and this fraction differs between countries. Countries can mitigate the threat of climate change by allocating resources to abatement: the higher aggregate abatement, the lower the probability of a catastrophe.[4]

First, we compare the social optimum to the stationary Nash equilibrium. As usual, the Nash equilibrium is inefficient. The first source of inefficiency is that there is insufficient abatement in the Nash equilibrium. The second source of inefficiency is more subtle. In our framework, welfare decreases if prior to the catastrophe some countries abate more than others. That is, given an aggregate level of abatement, welfare is highest when all countries abate the same amount. However, in the Nash equilibrium the level of abatement will differ between countries, since the incentive to abate is stronger if a country is hurt more by the catastrophe. This difference in abatement levels leads to an additional decrease in welfare. The intuition for the decrease in welfare is as follows. Before the catastrophe, all countries have identical preferences and the same level of net production. Hence, the marginal cost of abatement (i.e., the marginal utility of a decrease in consumption) is higher in countries that abate more. This implies that if two countries differ in their abatement level, then joint welfare can be increased by shifting abatement from a high abatement country to a low abatement country.

Second, we examine stable international environmental agreements, i.e., an international environmental agreement in which every country joins and cooperation is sustained by trigger strategies. Since the outside option for some countries is more attractive than for others, the distribution of abatement among countries tends to be unbalanced. This imbalance implies that in general the social optimum cannot be implemented by a stable international environmental agreement. However, in most circumstances, it is feasible to implement an abatement scheme with the same level of aggregate abatement as the social optimum. The difficulty is to persuade all countries to join this abatement scheme. Countries with little exposure to the negative effects of climate change will only join an international environmental agreement if their abatement requirements are low. The burden then falls disproportionately on countries that are severely impacted by the catastrophe. As discussed in the previous paragraph, welfare decreases if abatement is unequally distributed among countries. Therefore, in the optimal stable international environmental agreement aggregate abatement will be somewhat less compared to the social optimum—but substantially higher than in the Nash equilibrium.

Third, unlike most models in the literature on international environmental agreements, we have a stochastic game with an absorbing state. In this setting the usual folk theorems do not apply and we show that a higher degree of patience among the players may actually lead to lower levels of abatement in a stable international environmental agreement. Note that one critique of the Stern report [29] has been that it overemphasizes the cost of climate change by choosing a very low discount rate (Nordhaus [26] is the most vocal critic). We

[3] Dutta and Radner [13] claim to address the same three features in their model of international environmental agreements. However, their model is deterministic and abatement enters both the objective function and the
Footenote 3 continued
state equation in a linear fashion. While this allows them to fully characterize the set of Nash equilibria even when countries are heterogeneous, none of the features of a catastrophic shift appear in their approach.

[4] While countries differ in their exposure to climate change, in the pre-catastrophe state they have the same marginal cost of abatement.

provide one reason why this critique can be challenged: if abatement is mainly an instrument to prevent catastrophes, then its benefits are not in the long-run, but rather they occur before the catastrophe takes place. This encourages a somewhat impatient decision maker to invest in abatement, but a very patient decision maker will disregard it.

The reason that in our setting patience is not necessarily a virtue is that the main benefit of abatement is the postponement of an immediate catastrophe. However, the catastrophe cannot be postponed indefinitely. A very patient decision maker will take the latter into account, thereby decreasing the incentive to abate. This contrasts with the usual logic that abatement accumulates over time and therefore a more patient decision maker will put more emphasis on the benefits of abatement.

Our paper bridges two strands of the literature.[5] There is an extensive literature on the stability of international environmental cooperation [7,15,32]. In this literature, there are usually immediate benefits of abatement, since abatement marginally improves the state of the environment. We focus on the possibility of a sudden shift in the state of the environment, since one of the benefits of abatement is that it might postpone (or even avoid) a catastrophe. We are not the first to investigate catastrophic shifts: see for instance [16,23,25,30,33]. However, most of these papers focus on a single decision maker or, occasionally, multiple decision makers. But when this literature considers the case of multiple decision makers, they do not focus on the question whether the cooperative outcome can be sustained in a Nash equilibrium.[6] This paper is an attempt to jointly investigate these issues in a tractable framework.

As a final remark, abatement is a natural choice of instrument when studying prevention of a catastrophic shift. However, after a catastrophic shift has occurred, adaptation becomes the natural response. This is an important issue that is receiving wider attention in the literature, see, e.g., Benchekroun et al. [6] who show that an increase in the effectiveness of adaptation can diminish the incentive to free-ride.

The outline of the paper is as follows. Section 2 introduces the model. Theoretical results are presented in Sect. 3. A numerical example is presented in Sect. 4, and we discuss the role of the discount factor, country heterogeneity and efficient implementation in that section. Section 5 concludes. All proofs are in the appendix.

2 Model

2.1 The Environment

In order to get tractable results, the representation of the environmental catastrophe will be rather parsimonious, i.e., we only distinguish between a pre-catastrophe state of the world

[5] Due to the inherent dynamic nature of the problem, we focus on the part of the literature, where the dynamics are explicit. Another approach is Barrett [3], who models the climate catastrophe as a (static) threshold public good game, where passing the (potentially unknown) threshold is interpreted as a climate catastrophe. While this captures the idea that improvements (or deteriorations) are non-marginal, it disregards the fact that it is more costly to reverse climate change. Moreover, there is research [19,20,24], where, in a static game, uncertainty about the benefit-cost ratio of abatement is resolved either before or after the signing of an international environmental agreement and it is explored how this influences the scope for cooperation.

[6] A good, recent example is van der Ploeg and de Zeeuw [31]: their modeling framework can be seen as a more general version of our model. However, they only compare the cooperative and the non-cooperative outcome without addressing the question whether cooperation is stable. A rare paper that investigates whether cooperation can arise as an equilibrium phenomenon is Battaglini and Harstad [4], but they do not allow for catastrophic shifts.

and a post-catastrophe state of the world. While there are differences in environmental quality within each of these states, these are minor compared to the huge deterioration of environmental quality as a result of the catastrophe.

The timing of the catastrophe is stochastic. Formally, the state of the environment at time $t = 0, 1, 2, \ldots$ is denoted by Ω_t. The environment is either in a high-production state ($\Omega_t = H$) or the environment is in a low-production state ($\Omega_t = L$). The high-production state is pre-catastrophe, and the low-production state is post-catastrophe. The game starts in the pre-catastrophe world: $\Omega_0 = H$. In each period there is a probability p of staying in the high-production state, and the low-production state is irreversible. Here, p is endogenous and depends on the aggregate level of abatement. Given aggregate levels of abatement, we thus obtain a Markov chain in which the low-production state, L, is an absorbing state.

2.2 The Economy

There are n countries, indexed by $i = 1, \ldots n$. Each country internally follows the Golden Rule, that is savings are such that resources allocated to capital are optimal and each country maximizes within-period net production. The catastrophe reduces net production because it reduces the marginal productivity of capital. Net production in country i is y (if $\Omega_t = H$) and $\alpha_i y$ (if $\Omega_t = L$), where $\alpha_i \in (0, 1)$.[7] The effect of a catastrophe is a decrease in net production and α_i is a measure of how much country i is hit by the catastrophe. Countries are labeled such that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$, i.e., we rank countries from least to most sensitive to the catastrophe. Country i at time t invests $m_{it} \geq 0$ in abatement. The remainder of net production is consumed and gives country i a per-period utility of $u(y - m_{it})$ (if $\Omega_t = H$) or $u(\alpha_i y - m_{it})$ (if $\Omega_t = L$), where the utility function $u(\cdot)$ is increasing, strictly concave and satisfies the Inada conditions. Moreover, the countries are prudent: $u''' \geq 0$.[8] Countries maximize the normalized discounted sum of per-period utility, which is referred to as the welfare of country i.

Let

$$Y_i(\Omega_t) = \begin{cases} y & \text{if } \Omega_t = H \\ \alpha_i y & \text{if } \Omega_t = L \end{cases}$$

denote net production of country i at time t. Then the welfare of country i at time t is

$$V_i(\Omega_t, \{m_{is}\}_{s \geq t}) = (1 - \delta)\mathbb{E} \sum_{s=t}^{\infty} \delta^{t-s} u(Y_i(\Omega_s) - m_{is}),$$

where $\delta \in (0, 1)$ is the discount factor and with the expectation taken over the Markov chain induced by the countries' choices of abatement. In principle, country i may choose a different

[7] Net production is production minus investment in capital. For example, suppose K is the capital stock and the production function is $\xi(K) = \sqrt{\alpha K}$, where $\alpha = 1$ pre-catastrophe and $\alpha = \alpha_i$ post-catastrophe. Moreover, let ψ denote the depreciation rate. Then net production is $\xi(K) - \psi K$. Under the golden rule, net production is maximized: $\max_K \xi(K) - \psi K = \frac{\alpha}{4\psi}$. Define $y = \frac{1}{4\psi}$ and we see that net production is y before the catastrophe and $\alpha_i y$ after.

[8] Given that in our model countries abate to minimize the probability of a catastrophe, we have this assumption in common with the literature on optimal loss prevention [14]. However, that literature deals with static loss prevention with a single decision maker, whereas we study dynamic loss prevention with multiple decision makers. Note that in our setting this assumption ensures that countries have an incentive to invest in catastrophe prevention. This is different from Karp and Simon [18], who show that if marginal abatement costs are convex (and with linear marginal benefits), a stable international environmental agreement will have at most three members. In their model, there model there is no uncertainty and, therefore, the link between a positive third derivative of the abatement cost function and prudency is not the most obvious one.

level of abatement each period. However, our focus will be on stationary behavior, where abatement only depends on the state. Consequently, the time subscript is frequently dropped.

2.3 Abatement and Welfare

Let $M = \sum_i m_i$ denote aggregate abatement. We assume that the transition probability depends on aggregate abatement: $p(M)$, where $p(\cdot)$ is increasing, concave and $p(M) < 1$ for all $M \geq 0$.

Due to the irreversibility of the low-production state, there will be no abatement post-catastrophe. Furthermore, in all cases we examine abatement is time-invariant. Therefore, m_i will denote the level of abatement of country i pre-catastrophe. An abatement scheme is a vector $(m_1, m_2, \ldots, m_n, M)$, where $M = \sum_i m_i$. Note that welfare of country i depends only on m_i and M and it can be calculated using a recursive formulation:

$$V_i(m_i, M) = (1-\delta)u(y - m_i) + \delta p(M)V_i(m_i, M) + \delta(1 - p(M))u(\alpha_i y),$$

where welfare is a weighted average of utility now, $u(y - m_i)$, and welfare in the next period, which is $V_i(m_i, M)$ with probability $p(M)$ and $u(\alpha_i y)$ with probability $1 - p(M)$. Note that $u(\alpha_i y)$ is the post-catastrophe per-period utility and the more patient the countries are the higher the weight on next-period welfare. Then we obtain:

$$V_i(m_i, M) = \frac{(1-\delta)u(y - m_i) + \delta(1 - p(M))u(\alpha_i y)}{1 - \delta p(M)}. \tag{1}$$

It can easily be shown that V_i is decreasing in m_i and increasing in M.

Note that we are dealing with a stochastic game. Folk theorems that apply to (infinitely) repeated games do not necessarily carry over to stochastic games. For instance, in the setting of repeated games, we know that if for each country welfare in the socially optimal outcome exceeds welfare in the Nash equilibrium of the stage game and the discount factor is sufficiently close to 1, then the socially optimal outcome can be enforced. Dutta [12] shows for stochastic games this is only true if the Markov chain over the state space (as induced by the players' strategies) is irreducible (for any choice of the players' strategies).[9]

In the game presented here, the Markov chain is reducible (due to the low-production state being absorbing). Intuitively, Dutta's [12] results do not apply in our setting, since it becomes difficult to punish very patient players. Note that punishment is only possible in the high-production state (in the low-production state utility is always equal to $u(\alpha_i y)$). Since very patient players put little weight on the present (high-production) state, it may not be possible to set punishments at an appropriately high level. Hence, in general, the socially optimal outcome cannot be sustained as part of a Nash equilibrium.

3 Theoretical Results

In this section, we derive the equilibrium conditions for three different scenarios and characterize their properties. The benchmark is the social planner solution (SP), where abatement levels are chosen such that joint welfare is maximized. Then we examine a stationary Nash equilibrium (NE), where all countries choose abatement independently. Finally, we examine

[9] More precisely, Dutta [12] shows that if this Markov chain is irreducible for any choice of the player's strategies, then the set of equilibrium payoffs approaches the entire individually rational set of payoffs as the discount factor approaches one. See Levine [22] for an example, where the Markov chain is reducible and not all individually rational payoffs are equilibrium payoffs as the discount factor approaches one.

the joint welfare-maximizing Nash equilibrium that can be sustained using trigger strategies. We will refer to the final scenario as a stable international environmental agreement (SA).[10]

3.1 Social Planner

The social planner maximizes

$$\sum_i V_i(m_i, M)$$

subject to

$$\sum_i m_i = M \text{ and } m_i \geq 0 \text{ for all } i$$

Before we present the solution, note that if the social planner wants to implement a desired level of aggregate abatement, then the efficient way to achieve this is by setting the same level of abatement in each country. This is the equimarginal principle: the social planner allocates abatement to the country with the lowest marginal cost of abatement. In the optimum, the marginal costs of abatement need to be the same. But since the countries are identical before the catastrophe, this means that abatement is the same in all countries. Let $m_1 = m_2 = \cdots = m_n \equiv \mu$. Define $M^{SP} = n\mu$ as the level of aggregate abatement in the social planner solution. We make the following assumption:

Assumption 1 There is an interior social planner solution: $\mu > 0$.

This is obviously the interesting case to examine: the divergence between the socially optimal outcome and the stationary Nash equilibrium arises because usually in the latter case there is underabatement. This situation only occurs if the social planner abates a strictly positive amount.

Note that if W is the maximum aggregate welfare (i.e., welfare in the social planner solution), then by the principle of optimality:

$$W = \max_{\mu} n(1-\delta)u(y-\mu) + \delta p(n\mu)W + \delta(1-p(n\mu))\sum_i u(\alpha_i y)$$

Hence, the optimal level of abatement in each country is determined by

$$-n(1-\delta)u'(y-\mu) + n\delta p'(n\mu)[W - \sum_i u(\alpha_i y)] = 0. \tag{2}$$

Substituting $W = \sum_i V_i(\mu, n\mu)$ yields, cf. (1), an implicit expression for abatement per country in the social planner solution:

$$\left[nu(y-\mu) - \sum_i u(\alpha_i y)\right] f(n\mu) = u'(y-\mu), \tag{3}$$

where $f(M) \equiv \delta p'(M)/(1-\delta p(M))$.[11] The socially optimal level of abatement is determined as follows. On the right-hand side (RHS) of (3), we have the marginal cost of

[10] By trigger strategy we mean that all countries agree on an abatement scheme. If a country deviates, then a punishment regime is entered. The punishment regime could be the stationary Nash equilibrium, but it could also be a more severe punishment. By stable we mean that no country has an incentive to deviate. For details, see Sect. 3.3.

[11] Note that due to concavity of u and p, the first-order condition in (2) is a necessary and sufficient for a maximizer. In the appendix, we show that (3) has a unique solution.

abatement. On the left-hand side (LHS) of (3), there are two terms: the term between brackets is the benefit of avoiding the catastrophe, which is multiplied by a "hazard rate" that measures the marginal probability of avoiding the catastrophe. In the optimum, a country abates until the point where the benefit of reducing the probability of a catastrophe is equal to marginal cost of avoiding the catastrophe.

3.2 Stationary Nash Equilibrium

In the Nash equilibrium, in each period every country independently sets its abatement level. A common equilibrium concept is a stationary equilibrium, where the strategy does not depend on history or time. In our setting, this means that we have to determine the level of abatement for each country when the environment is in the high-production state. Recall that an abatement scheme $(m_1, m_2, \ldots, m_n, M)$ will yield welfare $V_i(m_i, M)$ to country i. Then we apply the one-stage deviation principle to find the equilibrium level: for each country, it should not be welfare-improving to deviate from m_i at any single stage of the game.[12] Formally:

Definition 1 An abatement scheme $(m_1^{NE}, m_2^{NE}, \ldots, m_n^{NE}, M^{NE})$ is a stationary Nash equilibrium when, for all i,

$$m_i^{NE} \in \arg\max_{m\geq 0}(1-\delta)u(y-m)+\delta p(M_{-i}^{NE}+m)V_i^{NE}+\delta(1-p(M_{-i}^{NE}+m))u(\alpha_i y),$$

where $V_i^{NE} = V_i(m_i^{NE}, M^{NE})$ and $M_{-i}^{NE} = \sum_{j\neq i} m_j^{NE}$.

Using the definition, we see that m_i^{NE} is determined by

$$-(1-\delta)u'(y-m_i^{NE})+\delta p'(M^{NE})\left[V_i^{NE}-u(\alpha_i y)\right]\leq 0, \tag{4}$$

where the inequality holds with equality if $m_i^{NE} > 0$.[13] Substituting

$$V_i^{NE} = \frac{(1-\delta)u(y-m_i^{NE})+\delta(1-p(M^{NE}))u(\alpha_i y)}{1-\delta p(M^{NE})},$$

we get

$$-(1-\delta)u'(y-m_i^{NE})+\delta p'(M^{NE})\left[\frac{(1-\delta)u(y-m_i^{NE})+\delta(1-p(M^{NE}))u(\alpha_i y)}{1-\delta p(M^{NE})}-u(\alpha_i y)\right]\leq 0$$

which simplifies to

$$\left[u(y-m_i^{NE})-u(\alpha_i y)\right]f(M^{NE})\leq u'(y-m_i^{NE}) \text{ for all } i. \tag{5}$$

Compared to (3) we see that in the Nash equilibrium country i only takes into account its own benefit of abatement (i.e., $\left[u(y-m_i^{NE})-u(\alpha_i y)\right]$).

Then we have the following result:

Proposition 1 *If f is decreasing, then there is a unique stationary Nash equilibrium.*

[12] Observe that although each country selects a single abatement level, this level is determined by dynamic considerations.

[13] Note that due to concavity of u and p, the first-order condition in (4) is a necessary and sufficient for a maximizer.

Note that, when f is decreasing, abatement is a strategic substitute, i.e., a country abates less if other countries abate more. This seems realistic, and we maintain this assumption throughout the paper:

Assumption 2 f is decreasing.

The stationary Nash equilibrium has the following properties:

Proposition 2 *Abatement is weakly increasing in a country's exposure to climate change:* $0 \leq m_1^{NE} \leq \cdots \leq m_n^{NE}$. *In particular,*

1. *Suppose* $j < k$. *Then* $m_k^{NE} = 0$ *implies* $m_j^{NE} = 0$.
2. $m_k^{NE} = m_j^{NE} > 0$ *if and only if* $\alpha_k = \alpha_j$.

Countries that are more severely affected by the catastrophe will abate more. Moreover, it is possible that the least affected countries do not abate at all.

Proposition 3 *In the stationary Nash equilibrium the aggregate level of abatement is less than in the social planner solution:* $M^{NE} < M^{SP}$.

This shows that the Nash equilibrium is inefficient in two ways. There is not enough abatement and the abatement is not distributed efficiently among countries.

3.3 Stable International Environmental Agreements

In an international environmental agreement, the countries jointly agree on an abatement scheme. The agreement is supported by trigger strategies, i.e., if any country deviates from the agreement, then from that period onward we enter a punishment regime. We assume that if country i deviates, then it will be punished in such a manner that its welfare after deviation $\tilde{V}_i$ is at most the welfare it would receive in the stationary Nash equilibrium, i.e., $\tilde{V}_i \leq V_i^{NE}$. In the next section, where we present a numerical example, different punishment regimes are discussed.

The incentive constraints have two peculiar features. First, incentive constraints are not independent: overabatement by one country changes the incentives for the other countries. In particular, it makes it more attractive for other countries to deviate. Therefore, if one country voluntarily abates more, other countries may deviate from the optimal scheme. It is tempting to argue that if a country wants to abate more, then welfare can be increased by letting this country abate more and reducing the levels for the other countries. In general this is not true, since a greater spread in the abatement levels will decrease joint welfare. Hence, any deviation from the abatement scheme, including upward deviations, needs to be punished. Second, it is not necessarily true that more patient players have less strict incentive constraints (for reasons outlined at the end of Sect. 2.3). Therefore, we expect that the discount rate to have a non-monotonic effect and it will be easier to get countries to cooperate if they are a bit impatient.

An abatement scheme $(m_1, \ldots, m_n, M)$ leads to an incentive constraint for each country. If country i does not deviate, then it receives welfare $V_i(m_i, M)$. The most attractive deviation gives welfare:

$$\max_{m \geq 0} (1-\delta)u(y-m) + \delta p(M_{-i}+m)\tilde{V}_i + \delta(1-p(M_{-i}+m))u(\alpha_i y).$$

Then the incentive constraint for country i is

$$IC_i: \; V_i(m_i, M) \geq \max_{m \geq 0} (1-\delta)u(y-m) + \delta p(M_{-i}+m)\tilde{V}_i + \delta(1-p(M_{-i}+m))u(\alpha_i y).$$

Observe that the RHS of the incentive constraint is a function of M_{-i}, i.e., abatement by all countries except i. Since $M_{-i} = M - m_i$, we see that both the LHS and the RHS of IC_i are functions of m_i and M. We can show the following.

Lemma 1 *Conditional on the aggregate level of abatement $M > M^{NE}$, there exist a bound on abatement $z_i(M)$ such that country i will join an environmental agreement when its contribution m_i does not exceed $z_i(M)$, i.e., $IC_i \implies 0 \leq m_i \leq z_i(M)$.*

We say that an international environmental agreement is stable if the incentive constraint for all countries is satisfied.[14] An optimal international environmental agreement is a stable international environmental agreement that maximizes joint welfare:

$$\max_{m_i, M} \sum_i V_i(m_i, M)$$

such that

$$\sum_i m_i = M$$
$$0 \leq m_i \leq z_i(M) \quad \text{for all } i$$

To solve this maximization problem, we employ a two-step procedure. First, we investigate if a certain level of aggregate abatement can be sustained by a stable international environmental agreement. Then, we address the question what the optimal aggregate level of abatement is.

Take the aggregate level of abatement M as given. Since it is trivial to enforce an abatement scheme in which $M = M^{NE}$, and since welfare can be increased by abating more, we focus on abatement schemes where $M > M^{NE}$. Conditional on the aggregate level of abatement, we can characterize how the burden will be shared among the countries. To find the optimal allocation, we make use of the following lemma:

Lemma 2 *Let $\tilde{m}$ be a feasible vector of abatement levels. Suppose that for some j and k, $0 \leq \tilde{m}_j < \tilde{m}_k \leq z_k$. Let $\widehat{m} = \tilde{m} + \varepsilon v$, where v is a vector such that $v_j = 1$, $v_k = -1$ and all remaining entries are zero. Then there exists $\varepsilon > 0$ such that $\widehat{m}$ will strictly improve welfare and $\widehat{m}$ is feasible.*

The lemma implies the following

1. All countries for which the upper bound is not binding ($m_i < z_i$) have the same level of abatement.
2. If for country i the upper boundary is binding ($m_i = z_i$), then this level of abatement is smaller than the level of abatement for the countries for which the upper bound is not binding.

Roughly speaking, in an optimal international environmental agreement the burden will be shared equally. However, the requirement that the agreement is stable may lead to deviations

[14] Observe that in the literature following Barrett [1], Hoel [17] and Carraro and Siniscalco [8] an environmental agreement is stable when no country participating in the agreement has an incentive to leave (internal stability) and no country, currently not participating in the agreement, wishes to join (external stability), cf. d'Aspremont et al.'s [10] definition of a stable cartel. Our incentive constraint is the condition under which no country wants to leave the agreement (internal stability). Since we are only concerned with agreements where all countries join, external stability plays no role here. Note that there are different ways to model stability, e.g., Benchekroun and Chaudhuri [5] use the concept of *farsightedness*, where countries anticipate that leaving the agreement may lead other countries to leave as well.

from this principle. In particular, countries with a binding incentive constraint are allowed to abate less to ensure that they will not deviate from the environmental agreement. Formally stated:

Proposition 4 *Suppose* $M^{NE} < M < \sum_i z_i(M)$. *The solution to the maximization problem*

$$\max_{m_i} V_i(m_i, M)$$

such that

$$\sum_i m_i = M$$

and

$$0 \leq m_i \leq z_i(M) \quad \text{for all } i$$

is unique and can be determined as follows. Construct the function

$$\mathcal{F}(\gamma) = \gamma \sum_{i \notin U(\gamma)} 1 + \sum_{i \in U(\gamma)} z_i - M,$$

where

$$U(\gamma) = \{i \mid z_i \leq \gamma\}.$$

There exists a unique $\gamma^* > 0$ *such that* $\mathcal{F}(\gamma^*) = 0$. *The solution is given by*

$$m_i = z_i \quad \text{for all } i \in U(\gamma^*),$$
$$m_i = \gamma^* \quad \text{otherwise.}$$

Now we turn to the question which aggregate level of abatement is optimal under the restriction that the corresponding international environmental agreement is stable.

Note that it is possible for the social planner solution and the optimal stable agreement to coincide. Observe that if

$$\mu \leq \min_i z_i(M^{SP}), \tag{6}$$

then the social optimum is a stable agreement and must therefore be the optimal agreement. When (6) does not hold, stable agreements can still reach the same level of aggregate abatement as the social optimum. This is feasible if

$$M^{SP} \leq \sum_i z_i(M^{SP}). \tag{7}$$

However, this may not be the optimal agreement, since in general it will require some countries to abate less than other countries. Ceteris paribus, a greater divergence of abatement among countries leads to a loss in welfare (in the sense of Lemma 2). By lowering the aggregate level of abatement, abatement per country can be more homogeneous. This leads to a tradeoff between the optimal amount of aggregate abatement and the efficient implementation of such a scheme. We expect that at the socially optimal level of abatement, the latter effect dominates the first, as the numerical results in the next section confirm.[15]

[15] We have assumed that the utility function is strictly concave. Most of our results hold when the utility function is linear with the notable exception of Lemma 2. With linear utility, the social welfare function only depends on aggregate abatement, i.e., the distribution of abatement is not of importance. Social welfare has a unique maximum at $M = M^{SP}$ and in the optimal stable agreement, aggregate abatement is as close to M^{SP} as the incentive constraints allow. Then (7) is the condition under which the social optimum and the optimal stable agreement coincide.

Table 1 Incentive constraints when aggregate abatement is at the social planner level

Country	z	
	Nash	Maxmin
1	0.0072	0.0236
2	0.0129	0.0365
3	0.0189	0.0476
4	0.0297	0.0579
5	0.0639	0.0686

The column "Nash" is Nash punishment, and the column "Maxmin" is the maxmin punishment

4 Numerical Example

In this section, we discuss how the three different scenarios behave with the aid of a numerical example.[16] The parameter values and functional forms are as follows. For the transition probability, we use:

$$p(m) = \frac{\tau m + \varphi}{\tau m + 1},$$

where $\tau = 100$ and $\varphi = 0.1$. Note that without abatement the probability of staying in the high-production state is φ. The utility function is $u(c) = \sqrt{c}$. Moreover, $y = 1$ and $\delta = 0.8$. We set $n = 5$ and $\alpha_i = 0.95 - 0.0375(i - 1)$. Country 1 loses 5% of net production due to the catastrophe and country 5 loses 20%. We consider two punishment regimes. In the Nash punishment scenario after deviation countries will play the stationary Nash equilibrium. In the maxmin punishment, all countries (except the deviator) will stop abatement completely. These two punishment regimes represent the two extremes: maxmin punishment is the harshest punishment that the countries can inflict upon a deviator, while Nash punishment is the most lenient one (without actually rewarding deviators).

The aggregate level of abatement in the social planner solution is 0.1116, and hence, the abatement level per country is 0.0223. See Table 1 for the incentive constraints at this level of aggregate abatement. We see that under Nash punishment the social planner solution is not feasible, since country 1, 2 and 3's maximum abatement level is below 0.0223. However, since $\sum_i z_i = 0.1326$, it is feasible to have the same level of aggregate abatement in the SA. Under maxmin punishment, the social planner solution is feasible. This shows that if punishment is severe enough then the social planner solution can be enforced by a stable environmental agreement. To see what shape the optimal agreement takes, we now focus our attention on the Nash punishment scenario.

Table 2 shows the abatement level for each country in each different scenario, as well as aggregate abatement and the probability of staying in the high-production state. Though it is feasible to have the same level of aggregate abatement as in the SP, it is optimal to abate a bit less in the SA. In this case, the incentive constraint for the first four countries is binding and country 5 provides the remainder of the abatement. In the NE, abatement is considerable lower with the first three countries not abating at all. In both the SP and the SA, the probability of staying in the high-production state is approx. 92.6%. In the NE, this figure is a bit lower at 85.1%. While this may seem a relatively small difference, it implies that on average it takes

[16] MATLAB code for all computations are available on request.

Table 2 Abatement in three different scenarios

Country	SP	SA	NE
1	0.022	0.007	0.000
2	0.022	0.013	0.000
3	0.022	0.019	0.000
4	0.022	0.030	0.006
5	0.022	0.042	0.044
Aggregate	0.112	0.111	0.051
p	0.926	0.926	0.851

In the stable environmental agreement Nash punishments are used

Table 3 Welfare in three different scenarios

Country	SP	SA	NE
1	0.9856	0.9914	0.9906
2	0.9811	0.9848	0.9833
3	0.9766	0.9779	0.9759
4	0.9720	0.9690	0.9664
5	0.9672	0.9592	0.9476

In the stable environmental agreement Nash punishments are used

13.5 periods to transition to the low-production state in both the SP and the SA, but it only takes 6.7 periods in the NE.[17]

Table 3 shows the welfare for each country. Strikingly, in the SP scenario, countries 1 and 2 receive lower welfare than in the NE (which of course is compensated by the huge welfare gain of country 5). This is the reason why (even for small discount rates) the social planner solution cannot be enforced by a trigger strategy. Hence, country heterogeneity is an obstruction to reaching the first-best outcome.

In the previous section, we argued that there may be a discount rate that is most conducive to cooperation. When the discount factor is 0.8 and maxmin punishments are used, the social planner solution is a stable environmental agreement. The hypothesis is then that this ceases to be true when the discount factor is sufficiently close to one. In the numerical example, this happens at $\delta = 0.99996$. Hence, it is not true that if the social planner solution is a stable environmental agreement for a discount rate $\bar{\delta}$, then it is also stable for all discount rates $\delta > \bar{\delta}$. In that sense the effect of the discount rate on the stability of the social planner solution is non-monotonic.

Finally, we investigate the role of heterogeneity between countries and the tradeoff between efficiency, i.e., an agreement where the difference in abatement between countries is kept small, and an agreement that implements the same level of aggregate abatement as in the social planner outcome. We keep mainly the same parameter values as before and use Nash punishments. The only difference is that now there are six countries, $n = 6$, and three

[17] Note that the aggregate level of abatement is severely restricted by country 1, 2 and 3, whose willingness to contribute is much lower than country 4 and 5. Potentially, a partial coalition of country 4 and 5 could perform better than the "grand coalition" since it faces less strict incentive constraints. However, in the example, a partial coalition where country 4 and 5 cooperate performs worse than the grand coalition. Calculation show that if country 4 and 5 cooperate, then in the welfare-maximizing outcome (subject to the incentive constraint) the levels of abatement of country 4 and 5 are resp. 0.0216 and 0.0552 (and the associated welfare levels are 0.9675 and 0.9493). The welfare of the participating countries is lower than when all countries cooperate. Moreover, because aggregate abatement is also substantially lower, the welfare of the non-participating countries also decreases.

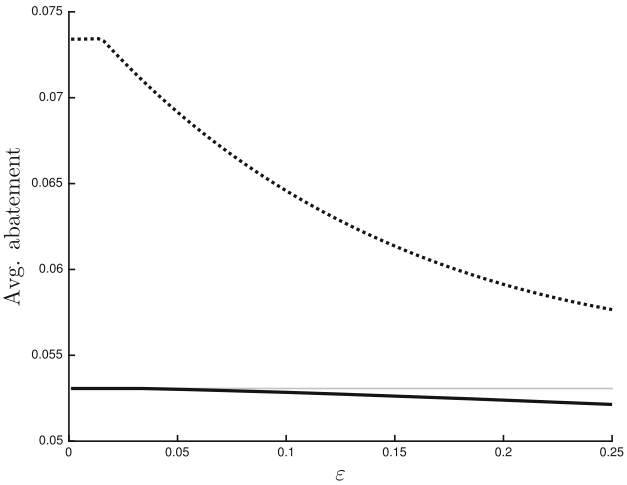

Fig. 1 The gray line is the level of abatement in the social planner solution as function of ε, the heterogeneity parameter. The solid black line indicates the average level of abatement in the optimal stable environmental agreement. The dashed black line is the maximum average level of abatement that could be sustained by a stable environmental agreement

Table 4 'Loss high' is the loss in welfare for the high exposure countries and 'gain low' is the gain in welfare for the low exposure countries when aggregate abatement is increased from M^{SA} to M^{SP} (but such that the incentive constraints hold)

ε	Loss high ($\times 10^{-5}$)	Gain low ($\times 10^{-5}$)
0.15	1.8	1.6
0.20	1.8	1.2
0.25	0.8	0

countries have low exposure to the catastrophe and three countries have high exposure. To be precise, for $\varepsilon \in [0, 0.25]$ we have $\alpha_i = 0.75 + \varepsilon$ for $i = 1, 2, 3$ and $\alpha_i = 0.75 - t(\varepsilon)$ for $i = 4, 5, 6$. The function t is such that the level of abatement in the social planner solution is the same for every value of ε, namely $\mu = 0.0531$. An increase in the value of ε is an increase in the level of heterogeneity between countries.

Figure 1 shows the average level of abatement in the SA and the maximum average level of abatement that could be sustained by a stable environmental agreement vis-à-vis the average level of abatement in the social planner solution. Observe that for low values of ε, it is possible to implement the social planner; for higher values of ε (roughly beyond $\varepsilon = 0.05$) this is no longer true and in the SA the average abatement levels will be lower than μ. Since it is feasible to design a stable agreement with the same level of average abatement as in the social planner solution—the maximum average level of abatement that could be sustained is larger than μ for all values of ε—we conclude that there is a tradeoff between efficiency and more abatement. To increase the average level of abatement to μ, the amount that the high exposure countries abate needs to be increased (the incentive constraint is already binding for the low exposure countries). However, this is inefficient as Table 4 illustrates. For different values of ε, we see the welfare loss of the high exposure countries exceeds the welfare gain of low exposure countries gain in welfare.

5 Concluding Remarks

In this paper, we develop a parsimonious model of international environmental agreements, incorporating three key issues: climate change is catastrophic, countries are sovereign (and hence there are participation constraints in designing international environmental agreement), and countries differ in their exposure to climate change. Due to the irreversibility of the catastrophe, this leads to a stochastic game with an absorbing state and our intuition on discounting does not work. Since the catastrophe is irreversible, the payoff of very patient players will be mainly determined by the payoff in the low-production state. This limits the extent to which a player can be punished when it deviates from an abatement scheme. Hence, international environmental agreements could actually be easier to implement if decision makers are a bit myopic. If catastrophes are reversible, then "folk theorems", such as the one presented in Dutta [12], again apply and the main obstacle to implementing the social planner solution is the heterogeneity of countries: in this case side payments may be essential to foster international cooperation.

Appendix: Proofs

Proof that the social planner solution is unique We have to show that (3) has a unique solution. Consider

$$\left[nu(y-m)-\sum_i u(\alpha_i y)\right] f(nm) = u'(y-m)$$

as a function of m. Observe that the LHS and the RHS of the equation are continuous and differentiable functions in m. Note that the RHS is increasing in m. We show that evaluated at any solution the LHS is decreasing in m. Since by Assumption 1 a solution exists, this implies uniqueness. The LHS is decreasing in m if

$$\left[nu(y-m)-\sum_i u(\alpha_i y)\right] f'(nm) < u'(y-m) f(nm). \tag{8}$$

From (3) we see that for any solution:

$$\left[nu(y-m)-\sum_i u(\alpha_i y)\right] = \frac{u'(y-\mu)}{f(n\mu)} \tag{9}$$

Evaluating (8) at $m=\mu$, substituting (9) and simplifying, we get:

$$f'(n\mu) < (f(n\mu))^2$$

Using the definition of f, this simplifies to $\delta p''(1-\delta p') < 0$ which is true.

Proof of Proposition 1 Suppose that the aggregate level of abatement is M. If this is the aggregate abatement of a stationary Nash equilibrium, then either m_i is the solution to

$$[u(y - m_i) - u(\alpha_i y)]\, f(M) = u'(y - m_i).$$

or $m_i = 0$ when this solution does not exist. This defines a continuous function $m_i = \zeta_i(M)$. If f is decreasing, then it is straightforward to verify that there exists $\bar{M}_i$ such that $\zeta_i(M) = 0$ for all $M \geq \bar{M}_i$, and ζ_i is decreasing in $[0, \bar{M}_i]$.[18] Observe that in any stationary Nash equilibrium $\sum_i \zeta_i(M) = M$. We show that $g(M) \equiv M - \sum_i \zeta_i(M)$ has a unique nonnegative root. Note that g is continuous, $g(0) < 0$ and $g(\max_i \bar{M}_i) > 0$, where the last claim follows from the bound on ζ_i. Then by the intermediate value theorem, g has a nonnegative root. Moreover, g is increasing and therefore the root is unique. □

Proof of Proposition 2 To prove the first statement, note that $m_k^{NE} = 0$ implies that (5) reduces to

$$[u(y) - u(\alpha_k y)]\, f(M^{NE}) \leq u'(y)$$

Since $\alpha_j \geq \alpha_k$, we have

$$\left[u(y) - u(\alpha_j y)\right] f(M^{NE}) \leq [u(y) - u(\alpha_k y)]\, f(M^{NE}).$$

Hence,

$$\left[u(y) - u(\alpha_j y)\right] f(M^{NE}) \leq u'(y)$$

and $m_j^{NE} = 0$.

To prove the second statement, note that abatement is positive and therefore the inequality in (5) holds:

$$\left[u(y - m_i^{NE}) - u(\alpha_i y)\right] f(M^{NE}) = u'(y - m_i^{NE}).$$

Given, that $u(\cdot)$ is strictly increasing, it is obvious that $m_k^{NE} = m_j^{NE} > 0$ if and only if $\alpha_j = \alpha_k$.

We prove the main claim by contradiction. Take two countries i and j such that $i < j$ (and therefore $\alpha_i > \alpha_j$) and suppose that $m_i^{NE} > m_j^{NE}$. Because of the first statement, we can focus on interior solutions without loss of generality. From (5), we get:

$$f(M^{NE}) = \frac{u'(y - m_i^{NE})}{u(y - m_i^{NE}) - u(\alpha_i y)} = \frac{u'(y - m_j^{NE})}{u(y - m_j^{NE}) - u(\alpha_j y)}.$$

Since $m_i^{NE} > m_j^{NE}$ and $u(\cdot)$ is concave, $u'(y - m_i^{NE}) > u'(y - m_j^{NE})$. This implies:

$$u(y - m_i^{NE}) - u(\alpha_i y) > u(y - m_j^{NE}) - u(\alpha_j y)$$
$$u(y - m_i^{NE}) - u(y - m_j^{NE}) > u(\alpha_i y) - u(\alpha_j y)$$

Note that the LHS of this inequality is negative and the RHS is positive. This contraction establishes that $m_i^{NE} \leq m_j^{NE}$. □

[18] We assume that $\max_i \bar{M}_i > 0$. Note that if $\max_i \bar{M}_i = 0$, then trivially there is a unique Nash equilibrium in which no country abates.

Proof of Proposition 3 First, suppose that every country abates a strictly positive amount, i.e., for all i, equation (4) holds with equality. Then summing (4) over i, we get

$$\delta p'(M^{NE})\left[\sum_i V_i^{NE} - u(\alpha_i y)\right] = (1-\delta)\sum_i u'(y - m_i^{NE}) \tag{10}$$

Let $\hat{m} = M^{NE}/n$ and let $\hat{V}_i = V_i(\hat{m}, M^{NE})$. Observe that $\sum_i \hat{V}_i > \sum_i V_i^{NE}$ since aggregate welfare increases when abatement is distributed more equally (for a given level of aggregate abatement) and, since $u''' \geq 0$, $\sum_i u'(y - m_i^{NE}) \geq nu'(y - \hat{m})$ by Jensen's inequality. From these observation and (10), we have

$$\delta p'(n\hat{m})\left[\sum_i \hat{V}_i - u(\alpha_i y)\right] \geq (1-\delta)nu'(y-\hat{m}). \tag{11}$$

In the social planner solution, we have

$$\delta p'(n\mu)\left[W - u(\alpha_i y)\right] = (1-\delta)nu'(y-\mu). \tag{12}$$

Now suppose, contrary to the claim of the proposition, that $\bar{m} \geq \mu$. Then

$$(1-\delta)nu'(y-\hat{m}) \geq (1-\delta)nu'(y-\mu) = \delta p'(n\mu)\left[W - u(\alpha_i y)\right],$$

where the equality follows from (12). Comparing this equation to (11), it must be that

$$\delta p'(n\hat{m})\left[\sum_i \hat{V}_i - u(\alpha_i y)\right] \geq \delta p'(n\mu)\left[W - u(\alpha_i y)\right]$$

Note that due to concavity of p, we have $p'(n\hat{m}) \leq p'(n\mu)$. Therefore, $\sum_i \hat{V}_i \geq W$, which contradicts the fact that W is defined as the (strict) maximum of total welfare. Hence, $\mu > \bar{m}$ and $M^{SP} > M^{NE}$.

Second, we examine boundary equilibria. Suppose that $m_1^{NE} = 0, \ldots, m_k^{NE} = 0$ and $m_{k+1}^{NE} > 0, \ldots, m_n^{NE} > 0$. If the social planner would only take into account the welfare of country $k+1$ up to n, then the aggregate level of abatement would be more than the aggregate level of abatement in the stationary Nash equilibrium. When it also takes into account the welfare of country 1 up to k, the social planner will increase the aggregate level of abatement. Hence, $M^{NE} < n\mu$ *a fortiori*. □

Proof of Lemma 1 Both the LHS and the RHS of IC_i are decreasing in m_i. The claim follows if we can show that the derivative of the LHS is strictly less than the derivative of the RHS. Suppose that in an abatement scheme country i has to abate m_i and aggregate abatement is M. Let m^* denote the optimal deviation from the abatement scheme. First we show that $m^* \leq m_i$.

Let m^* denote country i's optimal deviation. The aim is to show that $m^* \leq m_i$. Since country i's welfare from deviation is concave in m (cf. RHS of IC_i), it is sufficient to show that the derivate of welfare evaluated at m_i is negative:

$$\delta p'(M)[\tilde{V}_i - u(\alpha_i y)] \leq (1-\delta)u'(y - m_i)$$

Let $\sigma \equiv (u')^{-1}$. Therefore, the inequality can be rewritten as:

$$m_i \geq y - \sigma\left(\frac{\delta p'(M)[\tilde{V}_i - u(\alpha_i y)}{1-\delta}\right),$$

since σ is decreasing. In general, we need a minimal level of m_i to guarantee that the optimal deviation is downward. Unless

$$0 \geq y - \sigma\left(\frac{\delta p'(M)[\tilde{V}_i - u(\alpha_i y)]}{1-\delta}\right),$$

or equivalently

$$\delta p'(M)[\tilde{V}_i - u(\alpha_i y)] \leq (1-\delta)u'(y).$$

Observe that

$$\begin{aligned}\delta p'(M)[\tilde{V}_i - u(\alpha_i y)] &\leq \delta p'(M^{NE})[V_i^{NE} - u(\alpha_i y)] \leq (1-\delta)u'(y - m_i^{NE}) \\ &\leq (1-\delta)u'(y),\end{aligned}$$

where the first inequality follows from $M > M^{NE}$, concavity of p and the fact that $\tilde{V}_i < V^{NE}$, the second inequality from the definition of the stationary Nash equilibrium, and the final inequality from the concavity of u. Hence, all deviations are downward: $m^* \leq m_i$.

Then from the first-order condition, we have:

$$-(1-\delta)u'(y-m^*) + \delta\left[\tilde{V}_i - u(\alpha_i y)\right]p'(M_{-i}+m^*) \leq 0.$$

Consequently,

$$0 < \delta\left[\tilde{V}_i - u(\alpha_i y)\right]p'(M_{-i}+m^*) \leq (1-\delta)u'(y-m^*). \tag{13}$$

Remark that the derivative of the LHS of IC_i to m_i is

$$\frac{-(1-\delta)u'(y-m_i)}{1-\delta p(M)} < 0$$

and the derivative of the RHS of IC_i to m_i is

$$-\delta\left[\tilde{V}_i - u(\alpha_i y)\right]p'(M_{-i}+m^*) < 0$$

Using (13), we see that it suffices to show that

$$\frac{-(1-\delta)u'(y-m_i)}{1-\delta p(M)} < -(1-\delta)u'(y-m^*) \leq -\delta\left[\tilde{V}_i - u(\alpha_i y)\right]p'(M_{-i}+m^*) < 0$$

The only unproven inequality is

$$\frac{-(1-\delta)u'(y-m_i)}{1-\delta p(M)} < -(1-\delta)u'(y-m^*)$$

which follows directly from the fact that $1-\delta p(M) < 1$, concavity of the utility function and $m^* \leq m_i$. □

Proof of Lemma 2 It is obvious that $\widehat{m}$ is feasible for ε small enough. Note that conditional on M, maximizing $\sum_i V_i$ is equivalent to maximizing $\sum_i u(y-m_i)$. Therefore, we have to show that:

$$\sum_i u(y-\tilde{m}_i) < \sum_i u(y-\widehat{m}_i).$$

This is equivalent to showing that

$$u(y-\tilde{m}_j) + u(y-\tilde{m}_k) < u(y-\tilde{m}_j-\varepsilon) + u(y-\tilde{m}_k+\varepsilon)$$

Then using Taylor expansions, we get

$$u(y-\tilde{m}_j)+u(y-\tilde{m}_k) < u(y-\tilde{m}_j)-u'(y-\tilde{m}_j)\varepsilon+u(y-\tilde{m}_k)+u'(y-\tilde{m}_k)\varepsilon-\kappa_\varepsilon\varepsilon^2$$

for some $\kappa_\varepsilon \geq 0$ (since $u(\cdot)$ is concave). Therefore,

$$\kappa_\varepsilon\varepsilon < u'(y-\tilde{m}_k)-u'(y-\tilde{m}_j),$$

where the RHS is strictly positive by the strict concavity of $u(\cdot)$ and the assumption that $\tilde{m}_j < \tilde{m}_k$. Since $\lim_{\varepsilon\downarrow 0}\kappa_\varepsilon\varepsilon = 0$, there exists $\varepsilon > 0$ such that the inequality will hold. □

Proof of Proposition 4 Suppose γ is the proposed level of abatement for each country whose incentive constraints are satisfied if they abate at this level and the aggregate level of abatement is M. Let $U(\gamma)$ be the set of countries for which the upper boundary is binding at this level of abatement:

$$U(\gamma) = \{i \mid z_i \leq \gamma\}.$$

Note that U is a strict subset of $\{1,\ldots,n\}$ since $M < \sum_i z_i$ by assumption. Then, by Lemma 2, the abatement scheme proposed in the proposition is a welfare-maximizing outcome if

$$\gamma \sum_{i\notin U(\gamma)} 1 + \sum_{i\in U(\gamma)} z_i = M.$$

Define

$$\mathcal{F}(\gamma) = \gamma \sum_{i\notin U(\gamma)} 1 + \sum_{i\in U(\gamma)} z_i - M$$

Note that $\mathcal{F}(0) = -M < 0$, $\mathcal{F}(\max_i z_i) = \sum_i z_i - M > 0$ and $\mathcal{F}$ is increasing since U is a strict subset of $\{1,\ldots,n\}$. By the intermediate value theorem, we have that there is a unique value of $\gamma \in (0, \max_i z_i)$ such that $\mathcal{F}(\gamma) = 0$. □

References

1. Barrett S (1994) Self-enforcing international environmental agreements. Oxford Econ Papers 46:878–894
2. Barrett S (2003) Environment and statecraft: the strategy of environmental treaty-making. Oxford University Press, Oxford
3. Barrett S (2013) Climate treaties and approaching catastrophes. J Environ Econ Manag 66:235–250
4. Battaglini M, Harstad B (2016) Participation and duration of environmental agreements. J Polit Econ 124:160–204
5. Benchekroun H, Chaudhuri AR (2015) Cleaner technologies and the stability of international environmental agreements. J Public Econ Theory 17:887–915
6. Benchekroun H, Marrouch W, Chaudhuri AR (2011) Adaptation effectiveness and free-riding incentives in international environmental agreements. Technical report, CentER Discussion Paper
7. Breton M, Sbragia L, Zaccour G (2010) A dynamic model for international environmental agreements. Environ Resour Econ 45:25–48
8. Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Public Econ 52:309–328
9. Crépin A-S, Biggs R, Polasky S, Troell M, de Zeeuw A (2012) Regime shifts and management. Ecol Econ 84:15–22
10. d'Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark JA (1983) On the stability of collusive price leadership. Can J Econ 16:17–25
11. de Zeeuw A, Zemel A (2012) Regime shifts and uncertainty in pollution control. J Econ Dyn Control 36:939–950
12. Dutta P (1995) A folk theorem for stochastic games. J Econ Theory 66:1–32

13. Dutta P, Radner R (2009) A strategic analysis of global warming: theory and some numbers. J Econ Behav Organ 71:187–209
14. Eeckhoudt L, Gollier C (2005) The impact of prudence on optimal prevention. Econ Theor 26:989–994
15. Fuentes-Albero C, Rubio SJ (2010) Can international environmental cooperation be bought? Eur J Oper Res 202:255–264
16. Heijdra BJ, Heijnen P (2013) Environmental abatement and the macroeconomy in the presence of ecological thresholds. Environ Resour Econ 55:47–70
17. Hoel M (1992) International environment conventions: the case of uniform reductions of emissions. Environ Resour Econ 2:141–159
18. Karp L, Simon L (2013) Participation games and international environmental agreements: a non-parametric model. J Environ Econom Manag 65:326–344
19. Kolstad CD (2007) Systematic uncertainty in self-enforcing international environmental agreements. J Environ Econom Manag 53:68–79
20. Kolstad CD, Ulph A (2011) Uncertainty, learning and heterogeneity in international environmental agreements. Environ Resour Econ 50:389–403
21. Lemoine D, Traeger C (2014) Watch your step: optimal policy in a tipping climate. Am Econom J Econom Pol 6:137–166
22. Levine D (2000) The castle on the hill. Rev Econ Dyn 3:330–337
23. Mäler K-G, Xepapedeas A, de Zeeuw A (2003) The economics of shallow lakes. Environ Resour Econ 26:603–624
24. McGinty M (2007) International environmental agreements among asymmetric nations. Oxford Econom Papers 59:45–62
25. Nævdal E (2001) Optimal regulation of eutrophying lakes, fjords, and rivers in the presence of threshold effects. Am J Agric Econ 83:972–984
26. Nordhaus WD (2007) A review of the "Stern review on the economics of climate change". J Econ Lit 45:686–702
27. Polasky S, de Zeeuw A, Wagener F (2011) Optimal management with potential regime shifts. J Environ Econ Manag 62:229–240
28. Scheffer M, Carpenter S, Foley JA, Folke C, Walker B (2001) Catastrophic shifts in ecosystems. Nature 413:591–596
29. Stern N (ed) (2007) The economics of climate change: the Stern review. Cambridge University Press, Cambridge
30. Tahvonen O, Salo S (1996) Nonconvexities in optimal pollution accumulation. J Environ Econ Manag 31:160–177
31. van der Ploeg F, de Zeeuw A (2016) Non-cooperative and cooperative responses to climate catastrophes in the global economy: a north-south perspective. Environ Resour Econ 65:519–540
32. van der Ploeg F, de Zeeuw AJ (1992) International aspects of pollution control. Environ Resour Econ 2:117–139
33. Wagener F (2003) Skiba points and heteroclinic bifurcations, with applications to the shallow lake system. J Econom Dyn Control 27:1533–1561