Journal of Environmental Economics and Management ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

Climate treaties and approaching catastrophes ☆

Scott Barrett [a,b,*]

[a] School of International and Public Affairs & Earth Institute, Columbia University, New York, NY 10027, United States
[b] Princeton Institute for International and Regional Studies, Princeton University, Princeton, NJ 08544, United States

ARTICLE INFO

Article history:
Received 7 November 2011

Keywords:
Climate change
International environmental agreements
Catastrophe
Cooperation
Coordination
Uncertainty
Enforcement

ABSTRACT

If the threshold that triggers climate catastrophe is known with certainty, and the benefits of avoiding catastrophe are high relative to the costs, treaties can easily coordinate countries' behavior so as to avoid the threshold. Where the net benefits of avoiding catastrophe are lower, treaties typically fail to help countries cooperate to avoid catastrophe, sustaining only modest cuts in emissions. These results are unaffected by uncertainty about the impact of catastrophe. By contrast, uncertainty about the catastrophic threshold normally causes coordination to collapse. Whether the probability density function has "thin" or "fat" tails makes little difference.

1. Introduction

There is universal agreement, codified in the Framework Convention on Climate Change, that atmospheric concentrations of greenhouse gases should be stabilized "at a level that would prevent *dangerous* [my emphasis] anthropogenic interference with the climate system." The Kyoto Protocol tried to get countries to reduce their emissions without identifying a threshold for dangerous interference, but Kyoto failed, mainly for lack of an enforcement mechanism.[1] The recently negotiated Copenhagen and Cancun agreements also lack enforcement, and are legally non-binding besides, but unlike Kyoto these agreements identify a dangerous threshold—a 2° Celsius increase in average global temperature. In this paper I ask, Can the fear of crossing a catastrophic threshold overcome the enforcement challenge? Can it make climate treaties more effective?

The theory of international environmental agreements has generally offered a gloomy prognosis for cooperation, particularly on this issue.[2] To this point, however, the literature has considered only continuous abatement benefit functions. In this paper I take the benefit function to be *discontinuous* at a threshold. This simple change gives rise to a new result. If the threshold is known with *certainty*, and the loss from catastrophe vastly exceeds the costs of avoiding it, then the collective action problem changes fundamentally. Rather than *cooperate* to limit emissions, countries need only

☆ The idea for this paper came to me as I was thinking of how to respond to a question posed by William Nordhaus at a seminar I was giving at Yale a few years ago. I am grateful to Apurva Sanghi, Arthur Campbell, Astrid Dannenberg, Claude Henry, Erin Mansur, Geir Asheim, Kenneth Arrow, Mark Cane, Matthew Kotchen, Martin Weitzman, Robert Keohane, Robert Mendelsohn, Robert Socolow and Thomas Schelling for comments on a previous draft. I am also grateful to the editor, two anonymous referees, and participants at seminars given at Columbia, Harvard, Princeton, Stanford, the World Bank, and Yale (a later seminar) for their comments. An earlier version of this paper was given at the invitation of Michael Finus as a keynote lecture at the Environmental Protection and Sustainability Forum, University of Exeter, April 2011.

* Correspondence address: School of International and Public Affairs & Earth Institute, Columbia University, New York, NY 10027, United States. Fax: +1 212 854 4782.
E-mail address: sb3116@columbia.edu

[1] See Barrett [7].

[2] For surveys of the literature, see Michael Finus [13], Ulrich J. Wagner [26], and Barrett [6,7].

0095-0696/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jeem.2012.12.004

coordinate to avoid catastrophe. Under these circumstances, climate treaties can sustain the efficient outcome. Essentially, nature herself enforces an agreement to avoid catastrophe.

Here is why: when the net benefit of steering clear of the threshold is very high, avoiding catastrophe is a Nash equilibrium. (I mainly restrict attention in this paper to symmetric equilibria in pure strategies.) Since each country is too small to avert catastrophe on its own, however, there also exists a Pareto inferior Nash equilibrium in which the threshold is crossed. Under these circumstances, the challenge is for countries to coordinate on the mutually preferred equilibrium—a task for which treaties are exquisitely suited. When these circumstances do not apply, there exists a unique Nash equilibrium that sits on the wrong side of the threshold. Since there also exist other feasible, mutually preferred outcomes, countries will want to cooperate so as to improve on this inefficient equilibrium (as in a prisoners' dilemma game). A treaty can help here, too, but credible enforcement mechanisms for supplying this kind of global public good are generally weak.[3] Relative to the inefficient Nash equilibrium, which is common to both situations, coordination, when possible, almost always succeeds better than cooperation.

Weitzman [27] has drawn our attention to "uncertain catastrophes with tiny but highly unknown probabilities," showing that, if the probability of ever larger catastrophes does not fall faster than welfare losses increase as we venture deeper into the tails of the probability density function, then the expected gain from a policy to reduce emissions will be infinite. Under these circumstances, policy should do everything possible to reduce emissions as quickly as possible, even though doing so cannot guarantee that catastrophe will be avoided.[4] He calls this result, appropriately enough, "the dismal theorem."

Nordhaus [18]: 21 argues, to the contrary, that any conceivable catastrophic outcome can be avoided by policy, but that doing so requires "solving the global public goods problem by gathering most nations together to take collective action."[5] Weitzman does not address the collective action aspects of this challenge, but his theorem should carry through in a decentralized setting. If it pays the world as a whole always to devote more resources to reducing a threat that cannot be eliminated, no matter the cost, then it should pay individual countries (having the same preferences, and facing the same uncertainty as assumed by Weitzman) always to devote more resources to reducing the same existential threat, even when the benefits of doing so are widely dispersed. That is, under Weitzman's assumptions, coordination should be unnecessary. Every country should want to put its economy on a "climate war" footing, irrespective of whether other countries join them in this effort.

Underpinning the dismal theorem is the strong assumption that utility is unbounded.[6] If this assumption is relaxed, so that preferences are not very risk averse as consumption approaches zero, a different outcome emerges, one that is consistent with Weitzman's policy recommendation of "*a relatively more cautious approach* [emphasis added]" to reducing greenhouse gas emissions ([27]: 13). In this paper, for simplicity, I take utility to be linear. While this assumption rules out Weitzman's dismal theorem, I shall show that it nonetheless implies that countries may want to adopt a much more cautious approach as compared to one in which the threat of catastrophe can be ignored. Unfortunately, under these same circumstances, I also find that the uncertain prospect of approaching catastrophe has little if any effect on the non-cooperative outcome, or the ability of an international agreement to sustain collective action. The reason is that uncertainty (about the threshold for catastrophe) makes the expected damage function continuous, rendering coordination ineffective, and restoring the main results reported previously in the literature—that cooperation is needed, but difficult to sustain. In this model, whether the probability density function (pdf) has thin or fat tails makes little difference. Uncertainty about the magnitude of catastrophic damages is also relatively unimportant. It is uncertainty in the threshold that matters.

A certain threshold can be interpreted as a *discrete* uniform distribution having a single value. An interesting extension of this model, which ties my earlier analysis of certainty together with my later analysis of uncertainty, is the *continuous* uniform distribution. This distribution exhibits a critical discontinuity (in the distribution's right support), being fat-tailed throughout its range and zero-tailed elsewhere. I show that, for this special pdf, the result I obtained previously for the model of certainty extends to uncertainty—countries may be able to coordinate on the right-side discontinuity so as to guarantee avoiding catastrophe. However, as compared with the certainty case, I find that opportunities for coordination are extremely limited. Even under the favorable conditions of this special pdf, cooperation will almost certainly be needed and, as usual, difficult to sustain.

My overall conclusion is that fear of approaching catastrophes with uncertain thresholds strengthens the imperative to cooperate without improving the prospects for collective action.

[3] This is true even in a repeated game so long as treaties must be renegotiation-proof; see Barrett [5,6,7]. The one-shot game developed in this paper obeys an analogous property.

[4] If the conditions that give rise to this result are relaxed even slightly, then the threat of catastrophe raises numerous policy issues; see Kousky et al. [15].

[5] To Nordhaus [18], three conditions must hold for catastrophe to be cause for concern: policy failure, as just noted, high temperature sensitivity to concentrations, and extremely convex damages. In this paper, because temperature is subsumed in the analysis (damages are related to aggregate emissions), the latter two conditions will be satisfied when the abatement benefit function is discontinuous (or, in the uncertainty case, when the variance is very small) and the loss due to crossing the threshold (parameter X in this paper) is very large.

[6] Arrow [2] notes that, for individuals, the assumption implies that "the value of statistical life is infinite, a conclusion clearly contrary to all empirical evidence and to everyday observation."

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

Although the focus of this paper is on climate change, the model can be applied, with suitable modifications, to other situations. Consider the problem of space debris. If the volume of debris in low earth orbit exceeds a critical level, one more collision dramatically increases the probability of more collisions, creating a cascading effect that could render orbital space unusable (Kessler and Cour-Palais [14]). Here, catastrophe can be avoided by limiting additions to the stock of debris, or by designing satellites and rocket boosters so that they can be maneuvered to a "graveyard" or decaying orbit when they become obsolete. Another example is antibiotic resistance. Resistant strains of a pathogen sometimes have high evolutionary fitness in the presence of a drug treatment, but suffer a fitness cost relative to drug-sensitive strains when the drug is removed (Smith [23]). If use of the drug is low, the sensitive strains will win out. If use is high, the resistant strains will thrive. In between there exists a critical threshold for drug use, with the priority for public health being to stay on the good side of the threshold. Consider, finally, exploitation of a biological resource such as a species of fish for which there exists a strictly positive minimum viable population level. Economic overexploitation of the resource may be a problem even at higher stock levels, but exploitation below the critical level would be disastrous; crossing it would cause the species to go extinct. This may be a particularly important issue for fish populations such as tuna that aggregate, if school size is independent of stock abundance, and harvest economics depend on schooling (Clark and Mangel [11]).

Notwithstanding the importance of these and other applications, my main concern in this paper is with climate change. I shall introduce my model in Section 3, but it will help to begin by putting this model in the context of a previously published experiment, which was also concerned with climate change.

2. An experiment in catastrophe avoidance

The experiment by Milinski et al. [17] simulates the "collective-risk social dilemma" in preventing "dangerous climate change." In their experiment there are six players. Each is given €40. The game is played over ten periods. In each period, every player must choose to contribute €0, €2, or €4 without communicating. If, at the end of the game, at least €120 has been contributed, dangerous climate change is averted with certainty, and each player gets a payoff equal to the amount of money he has left (there are no refunds in this game). If less than €120 has been contributed, each player loses all the money she has left with probability 0.9. In their experiment, Milinski et al. [17] played the game with ten groups of students, only half of which succeeded in avoiding the threshold.

There are two symmetric pure strategy equilibria. In one, every player contributes €0 every period, giving each player an expected payoff of €4. In the other, every player contributes €2 every period, giving each a certain payoff of €20. (Of course, there also exist many asymmetric pure strategy equilibria in which different players contribute different amounts, possibly in different periods). The latter equilibrium is efficient; the former is not.

In contrast to the conventional representation of the climate change game, the game with catastrophic damages is pleading for coordination. Why, then, did half the groups in the Milinski et al. [17] experiment fail to coordinate? The main reason is almost surely that, by construction, the players were not allowed to communicate, let alone formulate a treaty to coordinate their contributions.[7]

The usual way of modeling an international environmental agreement is in three stages.[8] In stage one, countries choose independently whether to be a party or non-party to the agreement. In stage two, parties choose their actions (in this case, contributions) so as to maximize their collective payoff. Finally, in stage three, non-parties choose their actions with the aim of maximizing their individual payoffs. A treaty is self-enforcing if, given the treaty and participation level, non-participants do not want to change their behavior; if, given the participation level, parties to the treaty do not want to change the obligations expressed in the treaty; and if, given the participation decisions of other countries, each country does not want to change its decision of whether to be a party or non-party to the treaty.

Applying this notion of a self-enforcing treaty to the Milinski et al. [17] game, it will help to assume that contributions are made in a single period and that each player can contribute any amount up to his or her endowment. If participation were full, the treaty would then tell each country to contribute €20, netting each country a payoff of €20. Were a country to drop out of this agreement, the remaining parties, choosing collectively, would change their contributions. They would reason that, if they contributed an amount Y in total, then, taking this contribution as given, the non-party would contribute an amount $Z=€120-Y$ for $€120 \geq Y \geq €84$ and $Z=€0$ for $Y<€84$. Knowing this, the five remaining signatories could do no better than to contribute $Y=€84$ collectively (€16.80 each). This would net each of the five parties €23.20, whereas the sole non-signatory would get just €4. Recall that, were this country not to withdraw, it would get a payoff of €20. Obviously, with the treaty written in this way, no country has an incentive to withdraw, starting from a situation in

[7] A similar experiment by Tavoni et al. [25] confirms the importance of the assumption about communication.

[8] See Barrett [7]. This three-stage formulation imbues parties to the agreement with the special ability to commit to playing a collective abatement level—the so-called "Stackelberg leadership" assumption. If non-signatories have dominant strategies, this assumption is indistinguishable from the Nash assumption. In the Milinski et al. model, however, the payoff to avoiding catastrophe jumps at the threshold, implying that countries may not have dominant strategies. The main justification for the leadership assumption is the customary law principle, *pacta sunt servanda*: treaties are binding. Parties to an agreement are expected to fulfill their obligations, with custom being enforced outside the treaty (that is, outside the model). Note that this is a minimal commitment, since international law also says that countries are free to participate in treaties or not as they please—an assumption captured in Stage 1 of the above model. In this paper, the leadership assumption is also important for equilibrium selection—it ensures that, in a coordination game, the more efficient equilibrium will be supported by a self-enforcing agreement.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

which participation is full. Hence, a treaty comprising six signatories, each of which contributes €20, is self-enforcing. Moreover, while there exist two Nash equilibria in pure strategies in the underlying game, the self-enforcing agreement just noted is unique and supports the Pareto efficient equilibrium.

To sum up, while Milinski et al. [17] claim that countries may fail to avert catastrophe when doing so is feasible and efficient, simple theory suggest that, so long as countries are permitted to negotiate a treaty, catastrophe should be rather easy to avoid. In the remainder of this paper I inquire into whether this result could be expected to hold in richer (analytical) environments.

3. A climate change catastrophe game

Assume that countries have symmetric payoff functions. Denoting country i's abatement by q_i and aggregate abatement by Q, where $Q=\sum_{i=1}^{N} q_i$ and N is the number of countries, country i's payoff is assumed to be given by

$$\pi_i=\begin{cases} bQ-\frac{cq_i^2}{2} \text{ if } Q\geq \overline{Q} \\ bQ-X-\frac{cq_i^2}{2} \text{ if } Q<\overline{Q} \end{cases}. \tag{1}$$

In (1), parameter b stands for each country's marginal benefit of abating "gradual" climate change, cq_i represents country i's marginal abatement costs, and X denotes the damages each country suffers from "catastrophic" climate change. These damages are experienced if and only if global abatement falls short of $\overline{Q}$, the threshold.[9] If we let $X=0$, (1) collapses to the standard model in the literature on international environmental agreements.[10]

It is worth noting the important differences between (1) and Milinski et al. [17] model. First, in (1) contributions (abatement levels) are beneficial even if the threshold is exceeded; abatement reduces "gradual" as well as "catastrophic" climate change. Second, Milinski et al. vary X (more specifically, the expected value of X), but not the threshold, and (1) allows us to vary both.[11] Third, in (1), marginal abatement costs are increasing, whereas Milinksi et al. implicitly assume that they are constant. Fourth, Milinski and coauthors take abatement (contributions) to be discrete, and (1) allows abatement to be varied continuously. Finally, while (1) assumes that X is certain, and Milinksi et al. let X be uncertain, later in this paper I shall let both X *and* $\overline{Q}$ be uncertain.

In common with Milinski et al. [17] and Weitzman [27], model (1) abstracts away from the dynamics of climate change. It is, however, able to capture the key difference between "gradual" and "abrupt and catastrophic" climate change. It assumes that gradual climate change is so slow that marginal benefits are constant, and that abrupt climate change is so fast that transition is instantaneous.[12] The model can be interpreted as compressing perhaps a century of decision-making into a single period. Since the impacts of abrupt change will unfold more slowly than this (some of the impacts could take a millennium or more to play out), X should be interpreted as capturing the full consequences, including into the distant future, of decisions taken this century to avoid or exceed a threshold. The parameter b should be interpreted similarly for gradual climate change.

In the full cooperative outcome, the aggregate payoff will be

$$\Pi^{FC}=\begin{cases} bQN-\sum_i \frac{cq_i^2}{2} \text{ if } Q\geq \overline{Q} \\ bQN-XN-\sum_i \frac{cq_i^2}{2} \text{ if } Q<\overline{Q}. \end{cases} \tag{2}$$

If $X=0$, maximization of (2) yields $Q^{FC}=bN^2/c$. Assume $\overline{Q}>bN^2/c$ (avoiding the threshold requires abating more than is optimal for addressing only "gradual" climate change).[13] Then there are two possibilities. Either it will pay all countries collectively to meet the threshold ($Q^{FC}=\overline{Q}$), just, or it will not pay to meet the threshold ($Q^{FC}=bN^2/c$). Upon substituting

[9] Damages are normally related to temperature, temperature to concentrations, and concentrations to an emissions profile. However, there is strong evidence that temperature can be related directly to cumulative emissions (Allen et al. [1]; Zickfeld et al. [28]). In this paper, I take business as usual emissions as given. My focus is on reductions from this level: the level of abatement. The greater is the level of abatement, the smaller will be cumulative (total) emissions, and the lower will be temperature and, therefore, damages. Emissions will need to be reduced to a critical level, denoted here by $\overline{Q}$, if a dangerous threshold, such as the 2 °C change in temperature, is to be avoided.

[10] See especially Barrett [7].

[11] Milinski et al. [17] vary the expected value of X by varying the probability that damages will be "catastrophic" given that contributions fall short of the threshold.

[12] In a model with increasing marginal benefits, international cooperation still achieves very little [3], so it is not the constancy of marginal benefits that matters so much as continuity in the benefit function.

[13] This assumption implies that a focus only on "gradual" climate change could cause the world to cross a dangerous threshold. In the words of Stern [24], who warns of possible tipping points, "It is important to be clear that the "climate policy ramp" advocated by some economists involves a real possibility of devastating climatic changes." Of course, it is conceivable that the full cooperative abatement level could be determined solely by "gradual" climate change ($\overline{Q}<bN^2/c$). Suppose, then, that a treaty addressing only "gradual" climate change sustained the abatement level, Q^*. Then, if $\overline{Q}<Q^*\leq bN^2/c$, consideration of "catastrophic" climate change would have no effect on the usual analysis; abatement to avoid "gradual" climate change would avoid catastrophe in the bargain. If, to the contrary, $Q^*<\overline{Q}<bN^2/c$, consideration given to "catastrophic" climate change might improve the prospects for collective action, while still falling short of sustaining the full cooperative abatement level. In this case, the analysis would proceed in the same way as in this paper.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

these values into (2), it is easy to show that it will pay to meet the threshold if and only if

$$X \geq \frac{b^2N^2}{2c} - \left(b\overline{Q} - \frac{c\overline{Q}^2}{2N^2}\right). \tag{3}$$

Fig. 1 illustrates the relationship (which is non-linear due to marginal costs increasing). If the impact of catastrophic climate change, X, is "small," catastrophe is worth avoiding only if $\overline{Q}$, the threshold, is "small" (so that the costs of avoiding catastrophe are low). If $\overline{Q}$ is "large," catastrophe is worth avoiding only if X is "large."

In the non-cooperative outcome, each country i will maximize (1), taking as given the abatement choices of other countries. There are at most two symmetric Nash equilibria in pure strategies. In one, every country i plays $q_i = b/c$, and the threshold is exceeded. In the other, every country plays $q_i = \overline{Q}/N$, and the threshold is avoided, just. Suppose every country $j \neq i$ plays $q_j = \overline{Q}/N$. Then i will play either $q_i = \overline{Q}/N$ or $q_i = b/c$. Upon substituting these values in (1), it is easy to show that country i will prefer to play the former abatement level (in which case, avoiding catastrophe is a Nash equilibrium) rather than the latter if and only if

$$X \geq \frac{b^2}{2c} - \left(\frac{b\overline{Q}}{N} - \frac{c\overline{Q}^2}{2N^2}\right). \tag{4}$$

The top part of the Fig. 2 shows the space in which (3) and (4) both hold—that is, the space in which catastrophe avoidance is both (i) collectively optimal and (ii) a symmetric Nash equilibrium. The bottom part, unchanged from Fig. 1,

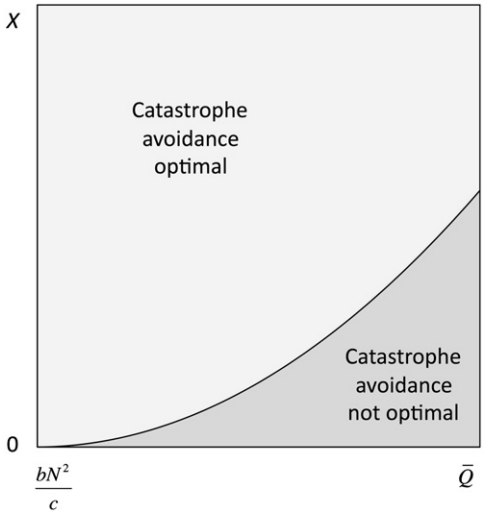

Fig. 1. Optimal catastrophe avoidance.

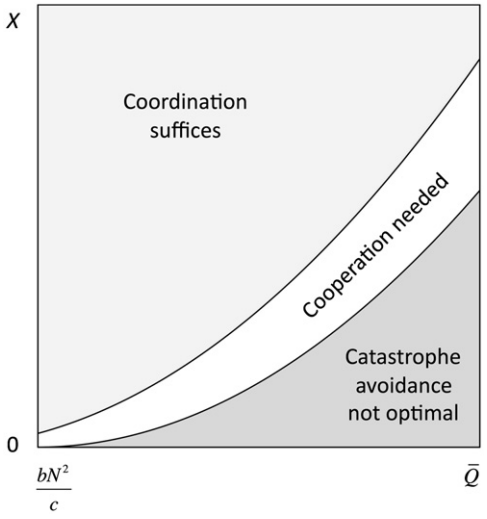

Fig. 2. When coordination suffices and cooperation is needed.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

shows the combinations of X and $\overline{Q}$ for which avoiding catastrophe is inefficient. Finally, sandwiched between these two spaces is a middle area in which (3) holds but (4) does not hold—that is, catastrophe avoidance is collectively optimal but cannot be supported as a symmetric Nash equilibrium. (Note that this middle area only exists because abatement reduces "gradual" as well as "catastrophic" climate change.) The figure shows that catastrophe avoidance cannot be sustained by coordination if X is extremely small, even if $\overline{Q}$ is close to the full cooperative abatement level for gradual climate change. Only if X is "big enough" will it be possible for countries to coordinate to avoid a "catastrophe."

In comparing (3) and (4), we see that coordination can always be relied upon to avoid catastrophe when doing so is optimal provided either $N=1$ (no externality) or $b=0$ (no benefit to abatement apart from avoiding catastrophe). The former result is to be expected. The latter result is surprising.[14] Intuitively, the incentives to avoid "catastrophic" climate change should be helped when the actions needed to do this also reduce "gradual" climate change. Imagine, however, that every country i plays $q_i=\overline{Q}/N$. Then, taking the abatement levels of other countries as given, a country that deviates unilaterally will suffer a smaller loss when the abatement by the other countries reduces gradual climate change. Ironically, the effect of abatement in reducing gradual climate change increases the incentive for a country to deviate from an agreement seeking to avert catastrophic climate change.

To summarize:

Proposition 1. *When inequality (3) does not hold, so that the impact of "catastrophic" climate change is low relative to the cost of avoiding catastrophe, the collective action problem is a prisoners' dilemma game for reducing "gradual" climate change. When (4) holds, so that the impact of catastrophic climate change is high relative to the cost of preventing catastrophe, the collective action problem is a coordination game for avoiding catastrophe. In between these situations, (3) holds but (4) does not hold, and the collective action problem is a prisoners' dilemma game for avoiding catastrophe.*[15]

4. Climate treaties for avoiding catastrophes

What role is there for climate treaties in averting an approaching catastrophe? Start from a situation in which every country participates in a treaty requiring that each play $q_i=\overline{Q}/N$. What conditions must hold for such a treaty to be self-enforcing?

Consider first the incentives for a country to withdraw from the treaty. Upon withdrawing, the deviating country will take as given the behavior of the remaining $N-1$ cooperating countries *as this is specified in the self-enforcing agreement*. Suppose that, once i withdraws, the agreement instructs these countries to play Q_{-i}. Then, upon withdrawing, i will either want to play $q_i=\overline{Q}-Q_{-i}$, to ensure that the threshold is met, or it will let the threshold slip and play $q_i=b/c$. Given that the other countries abate Q_{-i}, i will prefer to play the former level rather than the latter if

$$b\overline{Q}-\frac{c}{2}\left(\overline{Q}-Q_{-i}\right)^2\geq b\left(Q_{-i}+\frac{b}{c}\right)-X-\frac{c}{2}\left(\frac{b}{c}\right)^2 \tag{5}$$

Solving the quadratic, the minimum value of Q_{-i} which ensures that (5) is met is given by $\hat{Q}_{-i}=\overline{Q}-b/c-\sqrt{2X/c}$. Hence, off the equilibrium of full participation, if the $N-1$ cooperating countries play $\hat{Q}_{-i}$, then i will play $q_i=b/c+\sqrt{2X/c}$ and catastrophe will be avoided. If, to the contrary, they play $Q_{-i}<\hat{Q}_{-i}$, then i will play $q_i=b/c$, and the catastrophic threshold will be crossed.

How, then, will the $N-1$ remaining signatories want to play? If they play $Q_{-i}<\hat{Q}_{-i}$, then they cannot do better collectively than to play $Q_{-i}=b(N-1)^2/c$. Knowing how the deviant country will behave, they will prefer to abate $Q_{-i}=\hat{Q}_{-i}$ rather than $Q_{-i}=b(N-1)^2/c$ if and only if

$$b\overline{Q}-\frac{c}{2}\left(\frac{\overline{Q}-b/c-\sqrt{2X/c}}{N-1}\right)^2\geq b\left(\frac{b(N-1)^2}{c}+\frac{b}{c}\right)-X-\frac{c}{2}\left(\frac{b(N-1)}{c}\right)^2. \tag{6}$$

If (6) holds, and if i should withdraw from the agreement with full participation, the treaty will instruct the other $N-1$ countries to play (not $Q_{-i}=\overline{Q}(N-1)/N$, the Nash assumption, but) $Q_{-i}=\hat{Q}_{-i}$. Under these conditions, country i will not want to withdraw if

$$b\overline{Q}-\frac{c}{2}\left(\frac{\overline{Q}}{N}\right)^2\geq b\overline{Q}-\frac{c}{2}\left(\frac{b}{c}+\sqrt{\frac{2X}{c}}\right)^2. \tag{7}$$

Rewriting, we see that inequality (7) requires $b/c+\sqrt{2X/c}\geq\overline{Q}/N$; a country that withdraws is punished by being moved to abate *more* as a non-signatory than as a signatory (the opposite of free riding!).

[14] For Milinski et al. 's [17] model, there is no middle wedge; efficient catastrophe avoidance is always a Nash equilibrium.

[15] Obviously, these results apply specifically to model (1), which is linear-quadratic (linear benefits and quadratic costs). However, it is easy to see that these results also apply to the model discussed in Section 2, which is linear–linear with $b=0$, except that, in this model, there is no collective action problem for limiting gradual climate change. The paper by Barrett and Dannenberg [10], written after this paper was submitted, shows that Proposition 1 also holds for a model with linear benefits (with $b>0$) and stepwise-linear marginal costs (their model was developed for the purpose of conducting an experimental test of the theory presented in this paper); compare Fig. 1 in their paper with Fig. 2 in this one.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

If (6) and (7) hold, no country can gain by withdrawing unilaterally from the agreement that avoids catastrophe. It is easy to show (see the Appendix) that, if (7) holds, then so will (6) hold. Moreover, inequality (7) reduces to (4), which is the condition for when catastrophe avoidance can be sustained by coordination. Hence, we have:

Proposition 2. *When (4) holds, so that the impact of catastrophic climate change is high relative to the cost of preventing catastrophe, a self-enforcing international environmental agreement can sustain the full cooperative outcome, ensuring that catastrophe is avoided. Under these conditions, the treaty serves as a coordinating device.*

This analysis assumes full participation. It is well known that, in most cases, treaties can only support a lower level of participation. So, might collective action to avoid catastrophe be sustained even if (7) fails to hold?

Even if (7) were violated, inequality (6) could still hold. That is, the $N-1$ signatories could have an incentive to play $\overline{Q}-b/c-\sqrt{2X/c}$, knowing that the sole non-signatory would then abate $b/c+\sqrt{2X/c}<\overline{Q}/N$ (in which case the non-signatory would be a true free rider). Moreover, this agreement comprising $N-1$ countries would be self-enforcing, *provided it could be sustained by coordination*. However, it is easy to show that the conditions that enable $N-1$ countries to coordinate to avoid catastrophe also enable N countries to coordinate to sustain the full cooperative outcome (see the Appendix). Moreover, by extending this logic to other participation levels, it is clear that, if (7) fails to hold, it will not be possible to sustain coordination to avoid catastrophe by *any* number of countries ($\leq N$).

Of course, when (7) holds, it could be possible for fewer than N countries to coordinate so as to avoid catastrophe (this is just another way of saying that there may exist a large number of asymmetric Nash equilibria for avoiding catastrophe). However, coordination by fewer than N countries (with the others acting as free riders) offers no advantage over coordination by the grand coalition—only a disadvantage, since it is more costly for $N-1$ countries to sustain the threshold than for N countries to do so (with marginal costs increasing, cost-effective abatement requires that every country abate the same quantity). Since every country's expected payoff (with the expectation taken before an agreement is adopted, when no country knows if it will be a signatory or a non-signatory) is greater for an agreement comprising N countries than one comprising $N-1$ (or fewer) countries, provided both agreements are self-enforcing, in the negotiation stage all countries will favor a universal agreement. Moreover, once such an agreement has entered into force, if (7) holds, no country will have an incentive to withdraw.

Together, these results imply:

Proposition 3. *If there exists a self-enforcing international environmental agreement that avoids catastrophe by coordinating countries' abatement, then this agreement will be unique and sustain full participation.*

What happens if avoiding catastrophe is efficient and yet coordination fails? This would put us in the middle territory of Fig. 2. Countries could still try to *cooperate* to avoid catastrophe in this region, by using strategies to overcome the prisoners' dilemma, as in the usual model of an international environmental agreement. In this case, signatories would reduce their abatement in the belief that, by doing so, the threshold would be blown. However, the cooperation problem is now different from the usual one, which assumes only "gradual" climate change. For model (1), a deviation triggers a harsher punishment—the "catastrophic" loss due to crossing the threshold. It may, therefore, sustain greater cooperation.[16]

For it to be collectively rational for the $N-1$ signatories to play $Q_{-i}=b(N-1)^2/c$ after i has withdrawn, inequality (6) must not hold. Knowing this, country i will not want to withdraw if

$$X \geq \frac{b^2[2(N-1)^2+1]}{2c} - \left(b\overline{Q} - \frac{c\overline{Q}^2}{2N^2}\right). \tag{8}$$

Condition (8) is the same as (3) except for the intercept. It is straightforward to show that the intercept term in (8) is no greater than the intercept term in (3) for $N \leq 3$. That is, when coordination fails, a treaty can still be relied upon to sustain the full cooperative outcome—meaning, avoid catastrophe—if $N \leq 3$. It is well known that, with the functional forms assumed in (1), a self-enforcing international environmental agreement can sustain the full cooperative outcome even in the absence of catastrophe for $N \leq 3$. Consideration of catastrophe thus does nothing to change this result.

For $N > 3$, however, matters can be different. Cooperation may fail entirely to avoid catastrophe, or it may succeed in avoiding catastrophe but only within a portion of the middle wedge shown in Fig. 2.[17] In this region, inequality (8) holds but (6) does not hold.

Inequality (6) is awkward to work with, but an example can illustrate where and why cooperation can be effective within this middle wedge. Suppose $N=20$, $b=c=1$, and $\overline{Q}=600$.[18] Then the vertical distance of the cooperation wedge shown in Fig. 2, as defined by inequalities (3) and (4), requires $420.5 > X \geq 50$ (coordination succeeds for $X \geq 420.5$;

[16] "Punishment" appears in the one-shot model of an international environmental agreement when the treaty specifies that cooperating countries will lower their per-country abatement as the number of participating countries falls. We normally think of "punishment" in the context of a repeated game. But the stage game model of an international environmental agreement is particularly suited to exploring the enforcement of participation, whereas the repeated game model is better suited to exploring the enforcement of compliance. I have shown elsewhere that there is a natural correspondence between these two approaches [5–7].

[17] Using other functional forms, self-enforcing agreements can sustain full cooperation for $N > 3$ without considering the possibility of catastrophe. These include the quadratic-quadratic [3] and linear-linear [6] models. However, in both of these cases, a self-enforcing agreement can sustain the full cooperative outcome only when this outcome is nearly identical to the non-cooperative outcome.

[18] You can think of this assumption as suggesting that collective action by the G-20 group of countries could prevent catastrophic climate change.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

catastrophe avoidance is inefficient for $X < 50$). Condition (8) requires $X \geq 211.5$; the impact of catastrophe must be large enough to deter withdrawal. Of course, X must also be small enough that the N – 1 countries remaining in the agreement are willing to cross the threshold should i decline to cooperate. Upon substituting we find that (6) will not hold if $\sqrt{722(X+418.5)} < 599-\sqrt{2X}$, and it can be shown that there is *no* value for X within the relevant range ($420.5 > X \geq 211.5$) for which reversal of inequality (6) is satisfied. For these parameter values, cooperation to avoid catastrophe fails completely. Cooperation will have a better chance of avoiding catastrophe if $\overline{Q}$ is bigger. Raising the threshold helps because it makes credible the threat by the $N-1$ remaining signatories to lower their abatement dramatically should country i withdraw. For example, if the threshold is increased to 4000, the vertical range of the cooperation wedge becomes $19{,}800.5 > X \geq 16{,}200$ (coordination succeeds for $X \geq 19{,}800.5$; avoiding catastrophe is inefficient for $X < 16{,}200$), and cooperation can succeed in avoiding catastrophe—but only over a small portion of this range ($16{,}371.5 \geq X \geq 16{,}361.5$).

It can also be shown that if N countries cannot sustain cooperation to avoid catastrophe, then neither can fewer than N countries do so.[19] Of course, in the event that cooperation to avoid catastrophe fails, a self-enforcing international environmental agreement comprising three countries (for $N \geq 3$) could still succeed in reducing "gradual" climate change. However, such an agreement would reduce emissions just a little relative to the non-cooperative outcome; as is well known, each of the three signatories would abate $q_s = 3b/c$ and each non-signatory just b/c. Cooperation to limit "gradual" climate change would have only a modest impact on emissions.

To sum up:

Proposition 4. *When (3) holds but (4) does not hold, cooperation may succeed in averting catastrophe, but only under special circumstances. Moreover, even when these special circumstances apply, cooperation improves welfare relatively little.*[20] *Coordination, by contrast, succeeds spectacularly when the net benefits of avoiding catastrophe are particularly large.*

It should not surprise us that coordination makes all the difference. The Montreal Protocol succeeded partly because the strategic application of trade restrictions transformed protection of the ozone layer, ordinarily a prisoners' dilemma, into a coordination game [4,6]. Network externalities can play the same role in a treaty that strategically targets technology standards rather than emission limits [6,8]. In both of these cases, a self-enforcing treaty sustains coordination with full participation.[21] The novelty in the present paper is that here the opportunity for coordination is not strategic but rather a gift from Mother Nature.

Having established the preeminent role of coordination in averting catastrophe under conditions of certainty, in the remainder of this paper I shall limit my attention to the conditions under which coordination can reduce, if not eliminate, the probability of crossing a catastrophic threshold under conditions of uncertainty.

5. Impact and threshold uncertainty

[27] emphasizes fat-tailed uncertainty in climate sensitivity, compounded by uncertainties in translating temperature changes into welfare changes. Here I distinguish between uncertainty in the *impact* of catastrophe (X) and uncertainty about the *threshold* that triggers catastrophe $\left(\overline{Q}\right)$.

The two uncertainties are both very substantial, but also very different. One way to understand the nature of threshold uncertainty is to consider the recent attention-getting paper by Rockström et al. [19]. They identify a "planetary boundary" in terms of atmospheric CO_2 concentrations of 350 ppm by volume (ppmv) so as "to ensure the continued existence of the large polar ice sheets," while noting that the paleoclimatic record implies "that there is a critical threshold between 350 and 550 ppmv." Uncertainty about this threshold is purely scientific (and much of it is substantially irreducible). Uncertainties about the impacts associated with the loss of polar ice, by contrast, are both scientific and economic. Loss of the Greenland Ice Sheet, for example, would likely cause sea level to rise between 2 and 7 m over a period of 300 to over 1000 years (Lenton [16]). The value of X will obviously depend on the extent of sea level rise and its rate of change, but it will also depend on the *values* attached to these changes, including discounting. Much has been made in the economics literature of the importance of these values and their uncertainties, but I shall now show that, as regards the prospects for collective action, uncertainty about the impact of crossing a threshold is relatively unimportant. It is scientific uncertainty about the threshold that really matters.

Consider first uncertainty about the impact of catastrophe, X. It is straightforward to demonstrate that all the results shown thus far carry through if we substitute the expected value of X, $E(X)$, for X (this, again, is assuming that countries are

[19] For example, it would only pay $N-1$ countries to cooperate so as to ensure that the threshold was avoided when the other country, the free rider, reduced emissions by b/c, if $b\overline{Q}-(c/2)((\overline{Q}-b/c)/(N-1))^2 \geq b(b(N-1)^2/c+b/c)-X-(c/2)(b(N-1)/c)^2$. But if this condition applies, then (6) surely holds, which means we are in a coordination situation.

[20] Recall that the wedge for cooperation in Fig. 2 borders the region in which catastrophe avoidance is inefficient.

[21] In Barrett [4,6], trade restrictions sustain a first best. In Barrett [6,8], technology standards may sustain only a second best. In the present paper, the prospect of catastrophe may sustain a first best; it would sustain a second best only under the conditions described towards the end of footnote 13. With threshold uncertainty, as explained in the next section, a coordinating treaty can only sustain a second best, except for a special case discussed in Section 6.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

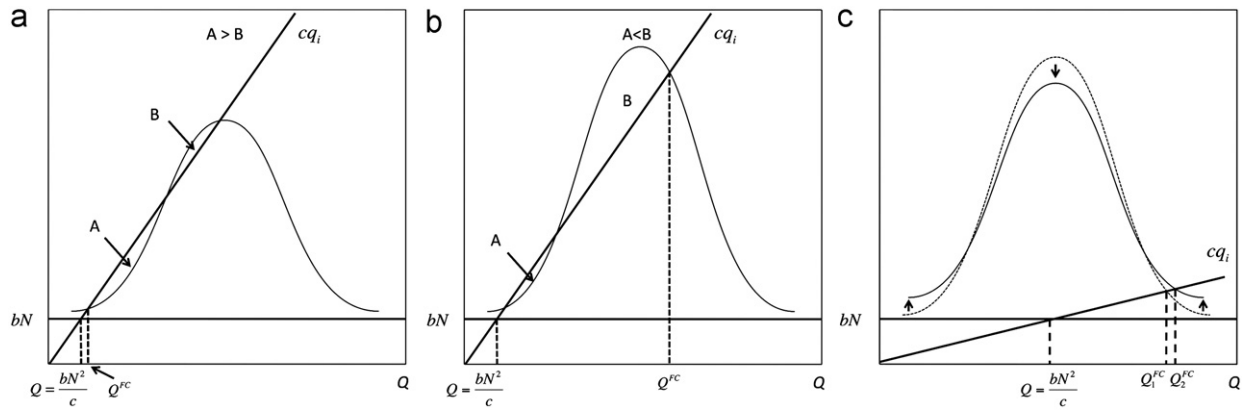

Fig. 3. (a) Prospect of catastrophe has little effect on the full cooperative outcome. (b) Prospect of catastrophe has profound effect on the full cooperative outcome. (c) Prospect of catastrophe with "thin" and "fat" tails.

risk neutral).[22] Uncertainty about the magnitude of catastrophic damages does not fundamentally alter the nature of the cooperation challenge. So long as the catastrophic threshold is certain, and $E(X)$ is large, countries will be able to coordinate so as to avert catastrophe.[23]

Consider now uncertainty about the threshold, $\overline{Q}$. In particular, let $\overline{Q}$ be a random variable with a continuous cumulative probability distribution $F(Q) = \Pr\left(\overline{Q} \le Q\right)$. Were countries to cooperate fully, they would now maximize

$$E(\Pi^c) = bQN - \sum_i \frac{cq_i^2}{2} - XN[1 - F(Q)], \tag{9}$$

which requires

$$bN - cq_i + XNf(Q) = 0, \tag{10}$$

where $f(Q)$ is the probability density function (pdf). Note that, in general, Eq. (10) will not be sufficient for a maximum.

The effect of uncertain catastrophe depends on the function f. If the pdf has infinite supports, f will always be positive, and the prospect of catastrophe will commend greater abatement, compared to a situation in which catastrophe is ignored. Recall that when catastrophe is certain, abatement in the full cooperative outcome may or may not be affected.

Fig. 3a illustrates Eq. (10) for a plausible pdf ($f(Q)$).[24] In the figure, the bell curve placed on top of the aggregate marginal benefit of avoiding "gradual" climate change represents $XNf(Q)$, the expected aggregate marginal benefit of catastrophe avoidance. For the pdf shown in this figure, concern about catastrophic climate change increases the full cooperative abatement level very slightly above bN^2/c, the optimal abatement level for gradual climate change ignoring catastrophic climate change. If abatement is increased beyond Q^{FC}, net benefits fall. They later increase before falling again, but so long as area A exceeds area B, as it does in Fig. 3a, the net benefits of increasing abatement beyond Q^{FC} will be negative, and it will be optimal to limit aggregate abatement to Q^{FC}.

Fig. 3b looks similar to 3a, but in this case area B exceeds area A, and the optimal aggregate abatement level increases dramatically, relative to bN^2/c, due to the uncertain prospect of catastrophe. Obviously, what matters in these figures is the size of area A relative to B. These relative values depend in turn on the values of b, X, and c as well as the pdf for the threshold.

Does the possibility of "fat tails" matter? As shown in Fig. 3c, a thickening of the tails of the pdf changes the full cooperative outcome very little, with the full cooperative abatement level increasing only slightly, from Q_1^{FC} to Q_2^{FC}.[25] Weitzman, of course, gets a very different result, but as noted in the introduction, he assumes that social preferences are strongly risk averse.

This brief discussion summarizes just a few of many possible constructions. My aim here is not to be comprehensive, but to illustrate how sensitive or insensitive policy recommendations can be to small changes in the constituent parts of Eq. (10).

[22] I am also assuming that $E(X)$ exists. For some fat-tailed distributions, expected value does not exist.

[23] This result should have been anticipated from my previous discussion of the Milinski et al. [17] experiment. Recall that, in this experiment, the value of X (and not $\overline{Q}$) was uncertain. An example of a situation in which the threshold was certain but the impacts were uncertain is the Millennium Bug or Y2K problem. Of course, this turned out not to be the catastrophe expected, though we shall never know whether this is because the action taken to avert a crisis succeeded or earlier fears of calamity were overblown. The theory presented here is consistent with both interpretations.

[24] All the pdfs drawn in this paper are symmetric, but my discussion and analysis does not depend on this assumption. There are reasons to believe the pdf for climate sensitivity exhibits positive skewness (Roe and Baker [20]).

[25] The diagram can be thought of as comparing the normal and t-distributions, the former being thin- and the latter thick-tailed.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

To sum up, we have:

Proposition 5. *Uncertainty about the impact of "catastrophic" climate change has no effect on the full cooperative abatement level. By contrast, uncertainty about the threshold may cause abatement in the full cooperative outcome to increase a lot relative to the full cooperative outcome for limiting "gradual" climate change.*

If countries fail to cooperate, each country i will choose q_i to maximize

$$E(\pi_i) = bQ - \frac{cq_i^2}{2} - X[1-F(Q)]. \tag{11}$$

The solution requires

$$b - cq_i + Xf(Q) = 0 \forall i. \tag{12}$$

What determines the gap between the full cooperative and non-cooperative outcomes? Look first at Fig. 4a. The top portion of the figure shows the full cooperative outcome. Though positioned somewhat differently, this portion of the figure is analogous to the situations depicted in Fig. 3a–c. The bottom portion of the figure shows the non-cooperative outcome. As drawn, Fig. 4a shows that the prospect of uncertain catastrophe has a substantial effect on the full cooperative outcome (raising it from bN^2/c to Q^{FC}), but virtually no effect on the non-cooperative abatement level (which, ignoring catastrophe, is bN/c, and which increases just a little to Q^{NC}, assuming that the pdf has an infinite support on the left side). Uncertainty about the threshold robs us of any chance for coordination on the full cooperative level of abatement. Indeed, compared to the certainty case (assuming that the certain threshold is equal to its expected value in Fig. 4a), uncertainty about the threshold makes countries want to abate *more*, and yet causes them to abate *less*.

Fig. 4b illustrates a situation in which the prospect of uncertain catastrophe increases abatement in both the full cooperative and non-cooperative outcomes (again, relative to a situation in which there was only gradual climate change), while at the same time narrowing the gap between the non-cooperative and full cooperative outcomes (ignoring catastrophe, this gap is $bN^2/c - bN/c$; taking catastrophe into account, it is $Q^{FC} - Q^{NC}$). In this case, even though the threshold is uncertain, the prospect of catastrophe creates an opportunity for coordination on the mutually preferred (but still inefficient) Nash equilibrium. It is obvious from the figure, however, that the circumstances that support this more cheerful situation are not necessarily to be expected. They include not only a high impact, X, but also a very low variance in the pdf and a small N—circumstances that are not favored by climate change.

Again, my aim here has not been to be comprehensive, but to indicate when collective action is likely to succeed or fail. The main results may be stated simply:

Proposition 6. *Compared to the certainty case, uncertainty about the impact of catastrophe has no effect on collective action. By contrast, uncertainty about the threshold is critical, being very likely to transform the collective action problem from a coordination game under threshold certainty to a prisoners' dilemma under threshold uncertainty.*

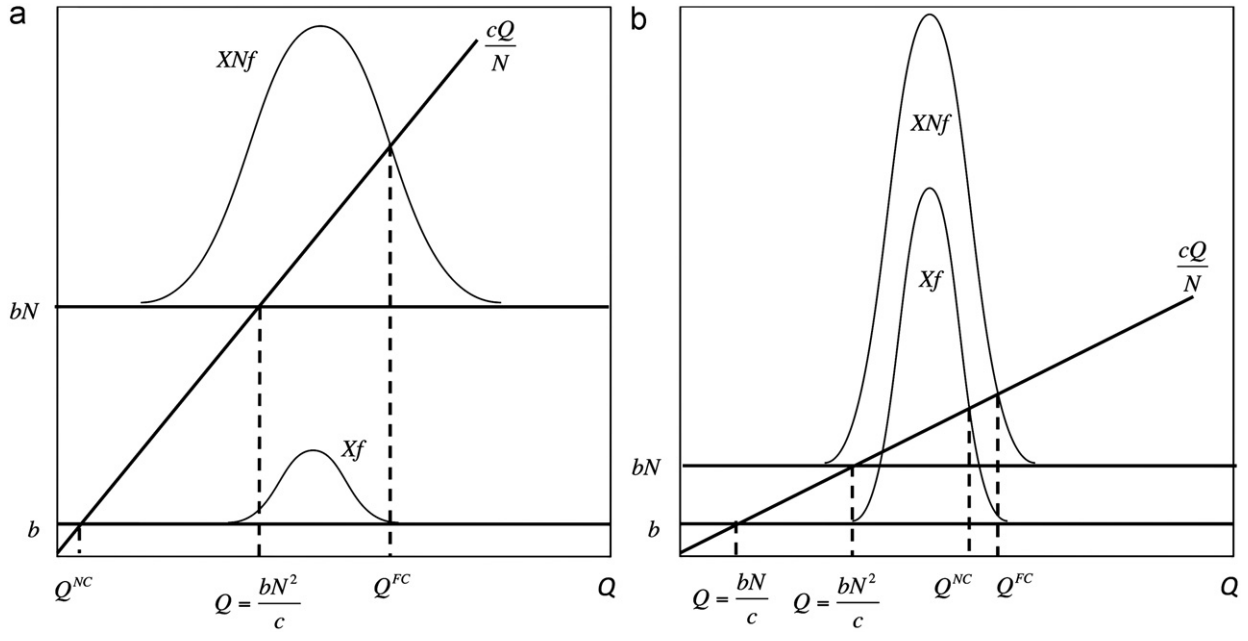

Fig. 4. (a) Uncertain prospect of catastrophe widens the gap between the non-cooperative and full cooperative outcomes. (b) Uncertain prospect of catastrophe narrows the gap between the non-cooperative and full cooperative outcomes.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

Proposition 7. *Compared to the certainty case, uncertainty about the threshold can make countries want to abate more in the full cooperative outcome, and yet cause them to abate less. Indeed, threshold uncertainty can cause countries to act as if the risk of catastrophe could be ignored, even when this risk is very great.*

To obtain more concrete results, we will need to choose a particular pdf. I do so in the next section.

6. Illustration for the uniform distribution

In this section I illustrate the points just made by assuming that the threshold obeys a continuous uniform distribution. An advantage of this distribution is that it offers a kind of blend of the two approaches emphasized in this paper. As in the previous section, there is threshold uncertainty. As in the earlier part of the paper, there is a discontinuity—in this case, at both of the distribution's (finite) supports. The discontinuity on the right side means that there may be a role for coordination in supporting a first best. As explained in the previous sub-section, this would not be the case for most distributions.

Assume, then, that the threshold concentration level is distributed uniformly with pdf

$$f(Q)=\begin{cases}0 & \text{for } Q<\overline{Q}_{\min}\\ \frac{1}{\overline{Q}_{\max}-\overline{Q}_{\min}} & \text{for } Q\in[\overline{Q}_{\min},\overline{Q}_{\max}]\\ 0 & \text{for } Q>\overline{Q}_{\max}.\end{cases} \tag{13}$$

This implies that we are certain about the *range* of values for the threshold, but uncertain about the *particular value*—that is, we have no reason to believe that any value within this range is more or less likely that any other value. In practical terms, the uniform distribution is consistent with Rockström et al.'s [19] representation of the climate problem, in which the paleoclimatic record suggests that the polar ice caps disappeared for concentrations between 350 and 550 ppmv, and preservation of the ice caps is taken to be the "planetary boundary," due presumably to the large impact of breaching it.[26]

The corresponding cumulative distribution function is

$$F(Q)=\begin{cases}0 & \text{for } Q<\overline{Q}_{\min}\\ \frac{Q-\overline{Q}_{\min}}{\overline{Q}_{\max}-\overline{Q}_{\min}} & \text{for } Q\in[\overline{Q}_{\min},\overline{Q}_{\max}]\\ 1 & \text{for } Q>\overline{Q}_{\max}.\end{cases} \tag{14}$$

If countries cooperate fully, they will maximize

$$E(\Pi^c)=\begin{cases}bQN-\sum_i\frac{cq_i^2}{2}-XN & \text{for } Q<\overline{Q}_{\min}\\ bQN-\sum_i\frac{cq_i^2}{2}-XN\left[1-\left(\frac{Q-\overline{Q}_{\min}}{\overline{Q}_{\max}-\overline{Q}_{\min}}\right)\right] & \text{for } Q\in\left[\overline{Q}_{\min},\overline{Q}_{\max}\right]\\ bQN-\sum_i\frac{cq_i^2}{2} & \text{for } Q>\overline{Q}_{\max}.\end{cases} \tag{15}$$

Assuming $Q_{\min}>bN^2/c$, countries will either play $Q=bN^2/c$ or they will play $Q\in\left[\overline{Q}_{\min},\overline{Q}_{\max}\right]$ (with this assumption, it will never pay to abate more than $\overline{Q}_{\max}$). In the latter case, full cooperation requires that countries abate the smaller of $\overline{Q}_{\max}$ or

$$Q^{FC}=\frac{bN^2}{c}+\frac{XN^2}{c\left(\overline{Q}_{\max}-\overline{Q}_{\min}\right)}. \tag{16}$$

Fig. 5a and b illustrate the two possibilities. In Fig. 5a, (16) holds; the solution is "interior." In Fig. 5b, a corner solution holds. (Note that, with $\overline{Q}_{\min}>bN^2/c$, it will never pay to abate $\overline{Q}_{\min}$.) Both figures are analogous to Fig. 3b.

Upon substituting $Q^{FC}=bN^2/c$ and (16) into (15), it is easy to show that countries should undertake some additional abatement (over and above the amount justified for addressing "gradual" climate change) to reduce the chance of catastrophe provided

$$X>\frac{2c(\overline{Q}_{\max}-\overline{Q}_{\min})}{N^2}\left(\overline{Q}_{\min}-\frac{bN^2}{c}\right). \tag{17a}$$

[26] I am simplifying Rockström et al.'s [19] arguments for 350 ppmv. In addition to preservation of the ice caps, they suggest that, after taking slow feedbacks into account, a doubling in CO_2 concentrations (presumably, to around 550 ppmv), would cause equilibrium temperature to increase about 6 °C, which would "severely challenge the viability of contemporary human societies."

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

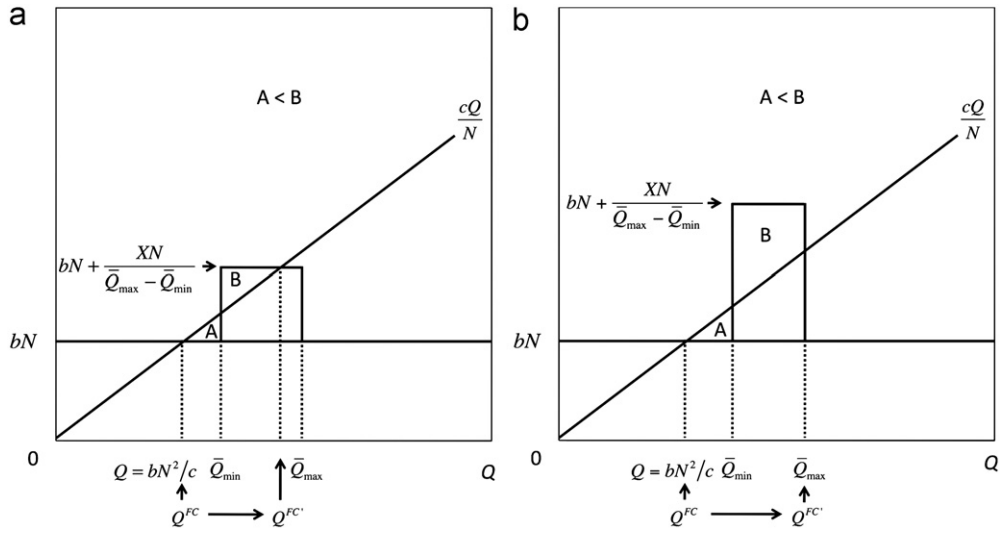

Fig. 5. (a) "Interior" solution for the full cooperative outcome with threshold uncertainty. (b) "Corner" solution for the full cooperative outcome with threshold uncertainty.

This is for an interior solution. For a corner solution the equivalent condition is

$$X > \frac{b^2N^2}{2c} - \left(b\overline{Q}_{\max} - \frac{c\overline{Q}_{\max}^2}{2N^2}\right). \tag{17b}$$

Fig. 5a and b again illustrate these conditions. In Fig. 5a, condition (17a) implies that triangle B is larger in area than triangle A. In Fig. 5b, condition (17b) implies that trapezoid B is larger in area than triangle A. As before, the values of areas A and B depend on b, c, X, and the pdf (in particular, $\overline{Q}_{\min}$ and $\overline{Q}_{\max}$).

If countries do not cooperate, each country i will maximize

$$E(\pi_i) = \begin{cases} bQ - \frac{cq_i^2}{2} - X & \text{for } Q < \overline{Q}_{\min} \\ bQ - \frac{cq_i^2}{2} - X\left[1 - \left(\frac{Q - \overline{Q}_{\min}}{\overline{Q}_{\max} - \overline{Q}_{\min}}\right)\right] & \text{for } Q \in [\overline{Q}_{\min}, \overline{Q}_{\max}]. \end{cases} \tag{18}$$

In a Nash equilibrium, country will either play $q_i = b/c$ or, solving (10),

$$q_i = \frac{b}{c} + \frac{X}{c\left(\overline{Q}_{\max} - \overline{Q}_{\min}\right)} \tag{19}$$

or $q_i = \overline{Q}_{\max}/N$, whichever of the latter two values is smaller.[27] Once again, the existence of two Nash equilibria (in pure strategies) makes abatement a coordination game. Upon comparing the payoffs, countries will abate so as to *reduce* the chance of catastrophe if

$$X > \frac{2c(\overline{Q}_{\max} - \overline{Q}_{\min})}{2N-1}\left(\overline{Q}_{\min} - \frac{b(2N-1)}{c}\right). \tag{20a}$$

They will abate so as to *eliminate* the possibility of catastrophe if

$$X > \left[\frac{b^2}{2c} - \left(\frac{b\overline{Q}_{\max}}{N} - \frac{c\overline{Q}_{\max}}{2N^2}\right)\right]. \tag{20b}$$

Conditions ((17a and b)) for the full cooperative outcome and (20a and b) for the non-cooperative outcome are identical only when $N=1$. As N increases, the gap between these outcomes widens. Fig. 6a (which resembles Fig. 4a) shows that the prospect of uncertain catastrophe can have a substantial effect on the full cooperative outcome, raising it from bN^2/c to $\overline{Q}_{\max}$, even as it has no effect at all on the non-cooperative outcome, which remains unchanged at $Q^{NC} = bN/c$. Fig. 6b (which is analogous to Fig. 4b) illustrates a situation in which complete avoidance of catastrophe can be sustained by coordination (only the bottom portion of the pdf for the aggregate benefit of avoiding catastrophe, drawn in broken lines, is shown in this figure). Unfortunately, this more pleasing result only holds under special circumstances. For example, X must be truly huge, especially if N is large.

[27] In this latter case I am limiting attention to the symmetric equilibrium.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

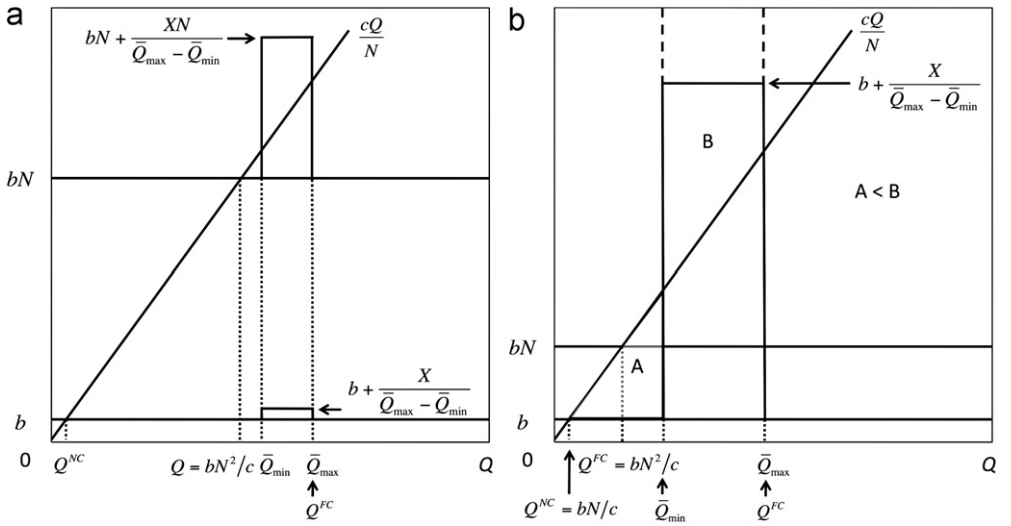

Fig. 6. (a) Uncertain prospect of catastrophe widens the gap between the non-cooperative and full cooperative outcomes. (b) Uncertain prospect of catastrophe closes the gap between the non-cooperative and full cooperative outcomes.

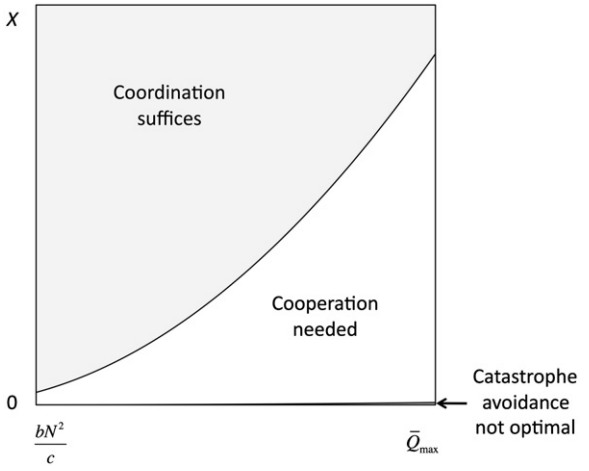

Fig. 7. When coordination suffices and cooperation is needed under threshold uncertainty.

Fig. 6a and b are both consistent with Rockström et al.'s [19] recommendation that the risk of catastrophe be eliminated, by holding concentrations down to 350 ppmv. Even without risk aversion, full cooperation can recommend a "precautionary" approach to catastrophe avoidance. However, the same conditions that support this conclusion may not enable countries to rally so as to achieve the collective action needed to sustain this outcome. If Fig. 6a described the world, countries would agree with Rockström et al.'s [19] recommendation, and yet fail to act so as to achieve it. Only if Fig. 6b portrayed the world could we expect actions to match aspirations.

Fig. 7 shows the range of values for X and $\overline{Q}_{max}$ for which catastrophe avoidance can be sustained by coordination (Fig. 7, constructed from (17b)and (20b) is thus akin to Fig. 2, constructed from (3) and (4)).[28] It is obvious from comparing Figs. 2 and 7 that uncertainty widens the area requiring cooperation (thereby shrinking the area for which coordination suffices). However, the extent of this widening is much greater than suggested by visual inspection of the figures. As can be seen by comparing inequalities (4) and (20b), the value of X that enables catastrophe to be avoided by coordination is N times greater in Fig. 7 than in Fig. 2. Uncertainty about the catastrophic threshold makes it very likely that avoiding "catastrophic" climate change is a prisoners' dilemma rather than a coordination game.

It will help again to work through an example. Recall from before that, for $b=c=1$, $N=20$, and $\overline{Q}=600$ (threshold certainty), full cooperation requires that catastrophe be avoided for $X \geq 50$, whereas coordination can avoid catastrophe only for $X \geq 420.5$. Table 1 shows these values, and others corresponding to them, for various scenarios of threshold

[28] The simulations producing these figures assume $b=c=1$ and $N=100$.

Table 1
Minimum values of X for various values of $\overline{Q}$ and $\overline{Q}_{min}-\overline{Q}_{max}$ (assuming $b=c=1$ and $N=20$).

Behavior	Consequence	$\overline{Q}$	$\overline{Q}_{min}-\overline{Q}_{max}$			
		600	590–610	550–650	500–700	450–750
Full cooperation	Reduce chance of catastrophe	–	38	75	100	75
	Eliminate chance of catastrophe	50	55	78	113	153
Coordination	Reduce chance of catastrophe	–	565	2364	4728	6323
	Eliminate chance of catastrophe	421	8703	9923	11,560	13,323

uncertainty. Reading across the top portion of the table we see a smooth transition for what is required for full cooperation as we move from threshold certainty to increasing levels of uncertainty. Notice that as the range of uncertainty increases, the size of the impact needed to justify elimination of the probability of catastrophe under full cooperation increases. This is because, holding the expected value of $\overline{Q}$ constant, a wider range for the pdf means a larger value for $\overline{Q}_{max}$; the benefit of completely avoiding catastrophe must increase to justify the higher cost of completely avoiding catastrophe. (See inequality (17b); in terms of Fig. 5b, as the range of the pdf widens, area A shrinks, while area B becomes both wider and shorter; for elimination of the risk to be justified, the value of X must increase so as to keep the northeast corner of trapezoid B above cQ/N.) By contrast, the minimum impact needed to justify reducing the chance of catastrophe below 100 percent increases and then decreases as the range for the distribution increases. Here there are two opposing forces. As the range of the distribution widens, the value of $f(Q)$ (represented by $1/(\overline{Q}_{max}-\overline{Q}_{min})$) falls, making an incremental reduction in the probability of catastrophe less valuable. (In (17a), the first multiplicand on the right increases; in Fig. 5a, the height of area B falls.) To compensate, the minimum X must increase. At the same time, a larger range causes the value of $\overline{Q}_{min}$ to fall, lowering the cost of reducing the probability of catastrophe. (In inequality (17a), the multiplicand on the far right shrinks; in Fig. 5a, area A shrinks and area B increases.) Now, coordination can be sustained for a lower minimum X.

Reading down the first column of numbers we can confirm the observation noted in Section 4 that the impact of catastrophe must be large for it to pay countries to coordinate so as to avoid the threshold (in the table, it must be nearly nine times as large, compared to the full cooperative outcome—421/50=8.4). More striking is how these requirements change as we move to the right in the table. The magnitude of the impact of catastrophe needed to give coordination a foothold increases dramatically, even for a very small range of uncertainty (in the table, even for the relatively tight range of 590–610, this value must be nearly 158 times as large compared to the full cooperative outcome—8703/55=158.2). Threshold uncertainty makes little difference to the full cooperative outcome (which increases from 50 to only 55 for these scenarios), but a huge difference to the ability of a self-enforcing agreement to sustain catastrophe avoidance by coordination. Recalling that the uniform distribution is particularly favorable to coordination, this is a disturbing finding.

7. Conclusions and implications

The standard model of a self-enforcing international environmental agreement predicts that collective action in reducing greenhouse gas emissions will be grossly inadequate. When this model is modified to incorporate a certain threshold with catastrophic damages, treaties can become highly effective. If the benefits of avoiding the threshold are high relative to the costs, the prospect of catastrophe transforms treaties into coordination devices.

Uncertainty about catastrophic damages is relatively unimportant to this calculus, whereas uncertainty about the threshold is critical. When the catastrophic threshold is certain, a slight relaxation in abatement from the threshold triggers a discontinuous change in damages—a huge deterrent to free riding. When the threshold is uncertain, a similar relaxation anywhere within the range of uncertainty for the threshold causes expected damages to increase just a little—a tiny disincentive to free ride. While the uncertain prospect of approaching catastrophes may commend substantially greater abatement in the full cooperative outcome, it may make little difference to non-cooperative behavior or to the ability of a climate treaty to sustain substantial cuts in global emissions.[29]

Multilateral efforts to reduce greenhouse gas emissions will be helped if threshold-uncertainty can be reduced very substantially (recall Figs. 4–6 especially). Though climate change is inherently uncertain, recent research suggests that signals of a generic nature may warn of an approaching threshold. A prime example is the phenomenon known as "critical slowing down," in which a system responds very slowly to perturbations as it nears a bifurcation [21]. Numerous abrupt climate shifts in the Earth's ancient history are consistent with this mathematical property [12], but this detection method is fallible; it may fail to identify some approaching thresholds, and predict others that are not actually there (Scheffer et al. [19]). As well, it only detects thresholds as we approach them, and by then the costs of stabilizing concentrations quickly so as to avoid catastrophe will be high. As shown here, under these circumstances, coordination will materialize only if

[29] Experiments relying on a modified version of the model developed here strongly support these conclusions. See Barrett and Dannenberg [10].

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

the impacts of breaching the threshold are truly huge. Of course, detection may also come too late for it to be possible to avoid catastrophe.

What, then, are the implications for policy? One implication, which I note with tongue in cheek, is that the collective action problem would be "solved" if we programmed a Doomsday Machine to set off a catastrophe with certainty once concentrations of greenhouse gases reached a specified level.[30] More seriously, negotiators should focus their attention on emission-reducing strategies that can promote collective action *without being conditional on climate thresholds*.[31] Finally, and most regrettably, we also need to prepare for the likelihood that we will someday cross a catastrophic threshold.[32]

Appendix A.1. ***Proof that (7) is a necessary and sufficient condition for a self-enforcing agreement to sustain the full cooperative outcome (avoiding catastrophe) when avoiding catastrophe is a Nash equilibrium.***

It will help to write out in long hand the condition for when avoiding catastrophe is a Nash equilibrium:

$$b\overline{Q}-\frac{c}{2}\left(\frac{\overline{Q}}{N}\right)^2 \geq b\left(\frac{\overline{Q}(N-1)}{N}+\frac{b}{c}\right)-X-\frac{c}{2}\left(\frac{b}{c}\right)^2. \tag{A.1}$$

Noting that inequality (7) requires $b/c+\sqrt{2X/c}\geq \overline{Q}/N$, the left hand side of (6) must be at least as large as the left hand side of (A.1). Now compare the right hand sides of these two inequalities. By assumption, $\overline{Q}/N > bN/c$, so we certainly have $\overline{Q}(N-1)/N > b(N-1)^2/c$, meaning that the first term on the right hand side of (6) must be smaller than the first term on the right hand side of (A.1). Finally, $b(N-1)/c > b/c$, so that the last term on the right hand side of (6) must be a larger negative value than its counterpart in (A.1). Taken together, this means that if (A.1) holds then (6) certainly holds; and, since (A.1) is the same as (7), this means that (7) is a sufficient condition for when catastrophe avoidance by N countries can be sustained by a self-enforcing agreement.

Appendix A.2. ***Proof that if $N-1$ countries can coordinate to avoid catastrophe, then so can N countries coordinate to avoid catastrophe.***

Suppose country i plays $q_i = b/c+\sqrt{2X/c}\geq \overline{Q}/N$ and each of the $j=1,\ldots,N-1$ other countries plays $q_j = (\overline{Q}-b/c-\sqrt{2X/c})/(N-1)$. This will be a Nash equilibrium provided that, if every other country plays these abatement levels, no country can gain by deviating. For i we require

$$b\overline{Q}-\frac{c}{2}\left(\frac{b}{c}+\sqrt{\frac{2X}{c}}\right)^2 \geq b\left(\overline{Q}-\frac{b}{c}-\sqrt{\frac{2X}{c}}+\frac{b}{c}\right)-X-\frac{c}{2}\left(\frac{b}{c}\right)^2 \tag{A.2}$$

and it is easy to confirm that the two sides of this expression are equal, making the weak inequality hold exactly.

For each country j we require

$$b\overline{Q}-\frac{c}{2}\left(\frac{\overline{Q}-b/c-\sqrt{2X/c}}{N-1}\right)^2 \geq b\left(\overline{Q}-\frac{(\overline{Q}-b/c-\sqrt{2X/c})}{(N-1)}+\frac{b}{c}\right)-X-\frac{c}{2}\left(\frac{b}{c}\right)^2,$$

which may be rewritten as

$$X\geq \frac{N}{(N-2)}\left\{\left[\frac{b^2}{2c}-\left(\frac{b\overline{Q}}{N}-\frac{c\overline{Q}^2}{2N^2}\right)\right]-\frac{\sqrt{2Xc}}{N^2}\left(\overline{Q}-\frac{bN}{c}\right)\right\}. \tag{A.3}$$

Constraint (A.3) lies above (4), the condition for coordination by N countries, provided

$$\left(\frac{\overline{Q}}{N}-\frac{b}{c}-\sqrt{\frac{2X}{c}}\right)\left(\overline{Q}-\frac{bN}{c}\right)\geq 0. \tag{A.4}$$

The second term on the left of (A.4) is positive by assumption and the first term is positive when (7) is violated. Hence, if it pays $N-1$ countries to coordinate to avoid catastrophe, it will also pay N countries to coordinate to avoid catastrophe.

[30] Tom Schelling told me in an email that my paper reminded him of something from fifty years ago: "There was some speculation, real or pretended, that the earth's atmosphere might have limited tolerance for nuclear explosions, and that beyond some threshold a nuclear blast might engulf the entire globe in a great flash of light that would exterminate all life on the surface. The thought didn't last long, but just long enough to stimulate the thought that if the threshold could be established with any certainty, it would be a great idea to detonate enough nuclear explosives to approach the threshold. Then nobody would dare to use a nuclear weapon. A real 'doomsday' proposal." A recent example of "strategic catastrophe-making" in the domestic context is the so-called "fiscal cliff," the combined effect of automatic budget cuts and tax hikes set to occur on midnight, December 31st, 2012 (the threshold), if the United States Congress does not pass new legislation to reduce US government debt. A problem with this strategy is that the impact of "catastrophe" is limited: the economy would not plunge immediately into recession. This is why the "fiscal cliff" has also been called the "fiscal slope."

[31] As noted previously, there may exist other ways in which treaties can coordinate behavior, including the use of trade restrictions [4,6] and the adoption of standards for technologies exhibiting network externalities [6,8].

[32] This is one reason why "geoengineering," especially the injection of reflective particles into the stratosphere, may need to be considered as a future option; see Schelling [22] and Barrett [9].

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004

References

[1] Myles R. Allen, David J. Frame, Chris Huntingford, Chris D. Jones, Jason A. Lowe, Malte Meinshausen, Nicolai Meinshausen, Warming caused by cumulative carbon emissions towards the trillionth tonne, Nature 458 (2009) 1163–1166.
[2] Kenneth J. Arrow, A note on uncertainty and discounting in models of economic growth, Journal of Risk and Uncertainty 38 (2009) 87–94.
[3] Scott Barrett, Self-enforcing international environmental agreements, Oxford Economic Papers 46 (1994) 878–894.
[4] Scott Barrett, The strategy of trade sanctions in international environmental agreements, Resource and Energy Economics 19 (1997) 345–361.
[5] Scott Barrett, A theory of full international cooperation, Journal of Theoretical Politics 11 (1999) 519–541.
[6] Scott Barrett, Environment and Statecraft: The Strategy of Environmental Treaty-Making, Oxford University Press, Oxford, 2003.
[7] Scott Barrett, The theory of international environmental agreements, in: Karl-Göran Mäler, Jeffrey Vincent (Eds.), Handbook of Environmental Economics, vol. 3, Elsevier, Amsterdam, 2005, pp. 1457–1516.
[8] Scott Barrett, Climate treaties and 'breakthrough' technologies, American Economic Review (Papers & Proceedings) 96 (2) (2006) 22–25.
[9] Scott Barrett, The incredible economics of geoengineering, Environmental and Resource Economics 39 (2008) 45–54.
[10] Scott Barrett, Astrid Dannenberg, Climate negotiations under scientific uncertainty, Proceedings of the National Academy of Sciences 109 (43) (2012) 17372–17376.
[11] Colin W. Clark, Marc Mangel, Aggregation and fishery dynamics: a theoretical study of schooling and the purse seine tuna fisheries, Fishery Bulletin 77 (2) (1979) 317–337.
[12] Vasilis Dakos, Marten Scheffer, Egbert H. van Nes, Victor Brovkin, Vladimir Petoukhov, Hermann Held, Slowing down as an early warning signal for abrupt climate change, Proceedings of the National Academy of Sciences 105 (38) (2008) 14308–14312.
[13] Michael Finus, Game Theory and International Environmental Cooperation, Edward Elgar, Cheltenham, 2001.
[14] Donald J. Kessler, Burton G. Cour-Palais, Frequency of artificial satellites: the creation of a debris belt, Journal of Geophysical Research 83 (A6) (1978) 2637–2646.
[15] Carolyn Kousky, Olga Rostapshova, Michael Toman, Richard Zeckhauser, Responding to Threats of Climate Change Mega-Catastrophes. Resources for the Future Discussion Paper 09–45, 2009.
[16] Timothy M. Lenton, Hermann Held, Elmar Kriegler, Jim W. Hal, Wolfgang Lucht, Stefan Rahmstorf, Hans Joachim Schellnhuber, Tipping elements in the earth's climate system, Proceedings of the National Academy of Sciences 105 (6) (2008) 1786–1793.
[17] Manfred Milinksi, Ralf D. Sommerfeld, Hans-Jürgen Krambeck, Floyd A. Reed, Jochem Marotzke, The collective-risk social dilemma and the prevention of simulated dangerous climate change, Proceedings of the National Academy of Sciences 105 (7) (2008) 2291–2294.
[18] William D. Nordhaus, An Analysis of the Dismal Theorem, Cowles Foundation Discussion Paper No. 1686, Yale University, 2009 January 16.
[19] Johan Rockström, Will Steffen, Kevin Noone, Åsa Persson, Stuart Chapin III, Eric F. Lambin, Timothy M. Lenton, Marten Scheffer, Carl Folke, Hans Joachim Schellnhuber, Björn Nykvist, Cynthia A. de Wit, Terry Hughes, Sander van der Leeuw, Henning Rodhe, Sverker Sörlin, Peter K. Snyder, Robert Costanza, Uno Svedin, Malin Falkenmark, Louise Karlberg, Robert W. Corell, Victoria J. Fabry, James Hansen, Brian Walker, Diana Liverman, Katherine Richardson, Paul Crutzen, Jonathan A. Foley, A safe operating space for humanity, Nature 461 (2009) 472–475.
[20] Gerald H. Roe, Marcia B. Baker, Why is climate sensitivity so unpredictable? Science 318 (2007) 629–632.
[21] Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, George Sugihara, Early-warning signals for critical transitions, Nature 461 (2009) 53–59.
[22] Thomas C. Schelling, The economic diplomacy of geoengineering, in: Thomas C. Schelling (Ed.), Strategies of Commitment, Harvard University Press, Cambridge, MA, 2006. 45–50.
[23] David L. Smith, The epidemiology of antibiotic resistance: policy levers, in: Ramanan Laxminarayan and Anup Malani; with David Howard and David L. Smith, Extending the Cure: Policy Responses to the Growing Threat of Antibiotic Resistance, Washington DC: Resources for the Future, 2007, pp. 39–68.
[24] Nicholas Stern, The economics of climate change, American Economic Review (Papers & Proceedings) 98 (2008) 2–37.
[25] Alessandro Tavoni, Astrid Dannenberg, Giorgos Kallis, Andreas Loeschel, Inequality, communication, and the avoidance of disastrous climate change in a public goods game, Proceedings of the National Academy of Sciences 108 (29) (2011) 11825–11829.
[26] Ulrich J. Wagner, The design of stable international environmental agreements: economic theory and political economy, Journal of Economic Surveys 15 (3) (2001) 377–411.
[27] Martin L. Weitzman, On modeling and interpreting the economics of catastrophic climate change, Review of Economics and Statistics 91 (1) (2009) 1–19.
[28] Zickfeld Kirsten, Michael Eby, H. Damon Matthews, Andrew J. Weaver, Setting cumulative emissions targets to reduce the risk of dangerous climate change, Proceedings of the National Academy of Sciences 106 (38) (2009) 16129–16134.

Please cite this article as: S. Barrett, Climate treaties and approaching catastrophes, Journal of Environmental Economics and Management (2013), http://dx.doi.org/10.1016/j.jeem.2012.12.004