Climate Treaties and Approaching Catastrophes

Scott Barrett*

Columbia University

Preliminary draft

24 March 2011

Abstract. Does the prospect of approaching "climate catastrophes" make international cooperation to limit emissions any easier? I find that, if the climate threshold is known with certainty, then the prospect of catastrophe fundamentally alters the cooperation challenge. Provided the costs of avoiding the threshold are low relative to the consequences of exceeding it, the challenge becomes one of *coordinating* to avoid the threshold, not *cooperating* to limit emissions. Where these conditions do not apply, international cooperation in cutting emissions, though efficient, is still difficult to enforce. The prospect of multiple catastrophic thresholds amplifies this gap between cooperation and non-cooperation. The availability of a backstop technology expands the space in which coordination is effective, but it also expands the space in which cooperation is needed. Uncertainty about the magnitude of catastrophic damages makes little difference. By contrast, uncertainty about the threshold that could trigger a catastrophe overturns the results for the certainty case. With threshold uncertainty, cooperation is still needed, and still difficult to enforce. Whether the probability density function has "thin" or "fat" tails makes little difference.

---

* Lenfest-Earth Institute Professor of Natural Resource Economics, Columbia University; sb3116@columbia.edu. The idea for this paper came to me as I was thinking of how to respond to a question posed by William Norhaus at a seminar at Yale. I am grateful to Mark Cane, Arthur Campbell, Claude Henry, Matthew Kotchen, Robert Mendelsohn, Erin Mansur, and Apurva Sanghi for commenting on an earlier version of this paper.

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* [emphasis added] anthropogenic interference with the climate system." The Kyoto Protocol was supposed to make a start toward meeting this goal, but, lacking an enforcement mechanism, it has failed to fulfill its promise.[1] The recently negotiated Copenhagen and Cancun agreements, being legally non-binding, are even weaker. Unlike Kyoto, however, these agreements identify a threshold for "dangerous" climate change—a 2 °C increase in average global temperature relative to pre-industrial levels.[2] In this paper I ask: Does the fear of crossing a dangerous threshold make climate treaties more effective? Does it ease the enforcement problem?

The theory of international environmental agreements has generally offered a gloomy prognosis for cooperation on this issue, due mainly to the difficulty of enforcement—a prognosis that, regrettably, two decades of negotiation effort has been unable to disprove. 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 some threshold. This change—the simplest but also the most extreme way of modeling catastrophes—gives rise to a very new result. If the threshold is *certain*, and the consequences of passing it are *profound*, the collective action problem changes fundamentally. Rather than *cooperate* to limit emissions, countries need only *coordinate* to avoid catastrophe. Under these circumstances, climate treaties can sustain full cooperation. Essentially, nature herself creates the conditions that allow international agreements to be effective.

---

[1] See Barrett (2005).

[2] This threshold for "dangerous" interference was first recognized by the Council of the European Union in 1996 and by the G8 in 2009.

Throughout this paper I allow both the threshold (which determines the costs of avoiding catastrophe) and the consequences of exceeding it to vary. This makes it possible to examine a multiple of catastrophe scenarios, some more harrowing than others, and some requiring a more demanding stabilization target than others. I show that it may not be worth avoiding every possible catastrophe scenario. Moreover, I find that coordination cannot avoid every catastrophe, even when doing so is feasible and globally optimal. For a range of values for the threshold and associated catastrophic damages, cooperation in cutting emissions is still needed, and still difficult to enforce. Coordination only helps to avoid the very worst outcomes. By implication, any new climate agreement would only improve on Kyoto if the threshold it identified were believed to be *truly* catastrophic.

Climatologists have identified not one threshold but many, each associated with a different catastrophe scenario. I find that, unless such thresholds are spaced far apart, full cooperation will require that the decision to avoid "the next" threshold recognize the knock-on benefit—if this one threshold is avoided, so will all the catastrophes that lie ahead. If countries act independently, however, and the thresholds on the horizon are not very tightly bunched together, countries will act as if to ignore these connections. The prospect of multiple catastrophes amplifies the free rider problem.

Limiting temperature change will probably require more than aggressive abatement. Given today's concentration level, the lagged response of temperature to concentrations, and inertia in transforming the global energy system, it is almost inconceivable that the 2 ˚C limit could be met without use of a new technology—"air capture." As defined here, air capture is an industrial process that removes CO_2 directly from the air, allowing the captured CO_2 to be stored underground or fixed in rock. Industrial air capture can potentially be scaled to any level, allowing us to "choose" a concentration target. It can also be located anywhere, separate from our energy and economic systems. The marginal cost of air capture is high, but to an approximation constant. Air capture is a true "backstop" technology.[3] I find that the availability of such a technology expands the space in which coordination alone can avoid catastrophe. However, because it also expands the space in which avoiding catastrophe is globally optimal, there still exist combinations of thresholds and catastrophic damages for which cooperation is needed.

Martin Weitzman (2009) has recently drawn our attention to "uncertain catastrophes with tiny but highly unknown probabilities." showing that, if the probability of ever larger catastrophes does not fall (as we venture deeper into the tails of the probability density function) faster than marginal utility rises (as the losses from catastrophe become ever greater), then policy should aim to do everything possible to reduce emissions as quickly as possible—and, even then, catastrophe may not be avoided.[4] He calls this result, appropriately enough, "the dismal theorem."

In a challenge to Weitzman, William Nordhaus (2009: 21) reasons that any conceivable catastrophic outcome can be avoided by policy—a non-dismal assessment, and one that I do not evaluate here—but that doing so requires, among other things, "solving the global public goods problem by gathering most nations together to take collective action."[5] Weitzman does not address this aspect of the challenge; his perspective is global. But his results 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 then it should pay individual countries (having the same preferences, and facing the same uncertainty

[3] See Barrett (2009a) for a discussion of the economics of a large number of options for mitigating climate change, including air capture; and Sarewitz and Nelson (2008) for an explanation for why air capture is especially likely to be adopted as compared with abatement.

[4] If one rejects this extreme position, then the threat of catastrophe raises numerous policy issues; for an overview of this perspective see Kousky *et al*. (2009).

[5] To Nordhaus, three conditions must hold for catastrophe to be avoided: policy failure, as just noted, high temperature sensitivity to concentrations, and extremely convex damages. In this paper, temperature is subsumed in the analysis (damages are related to aggregate emissions). Damages will be "extremely convex" if the abatement benefit function is discontinuous and the loss due to crossing the threshold, parameter "*X*" in this paper, is very large.

as assumed by Weitzman) always to devote more resources to reducing the same existential threat, however widely dispersed the benefits of those actions may be. The main real difference between this scenario and the certainty case examined in this paper is that, in Weitzman's formulation, coordination is unnecessary. Every country will want to put its economy on a "climate war" footing, irrespective of whether other countries join them in this effort.

As noted before, however, Weitzman puts all the focus on an extreme scenario—extreme as regards both the tail of the distribution and utility. If we relax either of these assumptions, the dismal theorem collapses. There is reason to do so. Weitzman's own practical advice for policy is restrained; to Weitzman (2009: 13), the dismal theorem "suggests as a policy response to climate change a relatively more cautious approach to GHG emissions…." Any analysis that recommends "a relatively more cautious approach" is one that admits some need to balance costs and benefits; and once the calculation is of this kind, we are likely to have to contend with free riding incentives.

Does the *uncertain* prospect of catastrophe help collective action or hinder it? I show that, if the catastrophic threshold is a random variable, then for very plausible probability density functions, the results for the certainty case change dramatically. Essentially, uncertainty makes the damage function continuous, restoring the main results reported previously in the literature—that cooperation is needed, but difficult to sustain (uncertainty makes coordination ineffective). Moreover, this is true even if the probability density function has "fat tails" (a necessary but not sufficient condition for Weitzman's dismal theorem). Interestingly, I find that uncertainty in the damages associated with crossing a threshold is of little significance. It is uncertainty in the threshold that matters.

To obtain this result, I make a number of simplifying assumptions, one of which is that utility is linear (social preferences are risk neutral). This explains why fat tails lose their power in my analysis, but in other respects the assumption matters relatively little. My analysis of full cooperation not only commends "a relatively more cautious approach," but in some situations favors a *dramatically* more cautious approach as compared to one in which the threat of catastrophe can be ignored. Unfortunately, under these same circumstances, I 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. This is a dismal conclusion of a different kind.

A certain threshold can be interpreted as a *discrete* uniform distribution having a single value (with probability 1). 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 is a decidedly thin-tailed distribution, but while Weitzman needs a thick-tailed distribution to support the dismal theorem, my analysis points to the need for discontinuity in the abatement benefit function to enable coordination. From this perspective, the continuous uniform distribution has the advantage of being discontinuous at each of its finite supports. 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 avert climate catastrophe. However, as compared with the certainty case, I find that the opportunities for coordination are very, very limited. Even under the favorable conditions of this special pdf, cooperation is needed and difficult to sustain.

My overall conclusion is thus that the prospect of approaching climate catastrophes strengthens the imperative to cooperate, but doesn't make preventive collective action any easier.

I shall develop all these arguments in due course, but first I want to motivate the modeling approach taken in this paper by describing an experiment in which catastrophe avoidance is played as a game.

2. Gambling for global public goods[6]

In a recent paper, Milinski *et al.* (2008) report the results of an experiment intended to simulate the "collective-risk social dilemma" in preventing "dangerous climate change." There are 6 players. Each is given €40. The game is played in 10 periods. In each period, every player must choose to contribute €0, €2, or €4. 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*. (2008) played the game with 10 groups of students, only half of which averted dangerous climate change.

There are two symmetric pure strategy equilibria. In one, every player contributes €0 every period. This gives each player an expected payoff of €4. In the other, every player contributes €2 every period. This gives every player 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.[7]). The latter equilibrium is efficient; the former is not.[8]

In contrast to the conventional representation of the climate change game, the game with catastrophic damages is pleading for coordination. We should not be surprised that half the groups in the Milinski *et al*. (2008) experiment failed to coordinate. By construction, the players were not allowed to communicate. They were also not allowed to formulate a treaty to coordinate their contributions.

[6] The title is from Dreber and Nowak's (2008) discussion of the paper by Milinski et al. (2008).
[7] For example, four players may contribute €2 each for each of the first 5 periods and €4 each for each of the last 5 periods, while the remaining two players contribute €O every period. In this case, the contributors each get €10 and the non-contributors each get €40.
[8] There also exists a symmetric mixed strategy equilibrium.

The usual way of modeling an international environmental agreement is in three stages (Barrett 2003). In stage 1, countries choose independently whether to be a party or non-party. In stage 2, parties choose their actions (in this case, contributions) so as to maximize their collective payoff. In stage 3, 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.

Let us now apply this notion of a self-enforcing treaty to the Milinski *et al*. (2008) game. To simplify matters, 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 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 ≥ Y ≥ €84 and Z = €0 for €84 > Y > €120. Knowing this, the 5 remaining signatories could do no better than to contribute Y = €84 collectively (€16.80 each). This would net each of the 5 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 which participation is full. The treaty comprising 6 signatories, each of which contributes €20, is self-enforcing.

To sum up, while Milinski *et al*. (2008) claim that countries may fail to avert catastrophe when doing so is feasible and efficient, I have shown that so long as countries are permitted to negotiate a treaty, catastrophe is rather easily avoided.[9] In the remainder of this paper I inquire into whether this result could be expected to hold in richer (analytical) environments.

3. The climate change catastrophe game

Let country *i*'s payoff be given by

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

If global abatement falls short of $\bar{Q}$, every country experiences "catastrophic" damages in the amount *X*.[10] As $X \to 0$, (1) resembles the standard model in the literature on international environmental agreements. For catastrophe to be a real prospect, *X* must be "large." Nordhaus (2009) has suggested that climate change would be "catastrophic" if it reduced per capita consumption by at least half, but in this paper I let *X* vary.

This representation differs from Milinski *et al.*'s (2008) model in several respects. First, contributions (abatement levels) are beneficial even if the threshold is not met; abatement reduces "gradual" as well as "catastrophic" climate change. Second, abatement is unbounded; negative emissions are feasible, using technologies like

---

[9] This implication of the theory is strongly supported by the results of the Miliniski *et al.* (2008) experiment. The groups that failed to avert dangerous climate change with certainty contributed on average €113, just €7 short of the full cooperative outcome. This average outcome gives each player an expected payoff of €2.12, which is lower than each player would get were none to contribute and substantially less than each would get if everyone contributed just a little bit more.

[10] 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.* 2009; Zickfeld *et al.* 2009). 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.

"air capture." Third, Milinski *et al*. (2008) vary *X* (more specifically, the *expected value* of X, by varying the probability that damages will be "catastrophic" given that contributions fall short of the threshold), but not the threshold, and (1) allows us to vary both. Fourth, in (1), marginal abatement costs are increasing, whereas Milinksi et al. (2008) implicitly assume that they are constant.[11] Finally, the loss due to passing the threshold is certain in (1) and uncertain in Milinski *et al*.'s model. Later I shall relax this assumption and show that it matters very little. I shall also allow the threshold to be uncertain—a further twist on the Milinski *et al*. model—and show that this assumption is critically important.

3.1 Full cooperation

In the full cooperative outcome, the aggregate payoff will be

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

If $X = 0$, maximization of (2) yields $Q^{FC} = bN^2/c$. Assume $\bar{Q} > bN^2/c$. Then there are two possibilities. Either it will pay all countries collectively to meet the threshold $(Q^{FC^*} = \bar{Q})$), just, or it will not pay to meet the threshold $(Q^{FC^*} = bN^2/c)$. It will pay to meet the threshold iff

$$b\bar{Q}N - \sum_i \frac{c}{2}\left(\frac{\bar{Q}}{N}\right)^2 \geq b\left(\frac{bN^2}{c}\right)N - XN - \sum_i \frac{c}{2}\left(\frac{bN}{c}\right)^2.$$

[11] Actually, Milinski *et al*. implicitly assume that marginal costs are constant up to a contribution level of €4, and infinitely high beyond that, since they do not allow larger per-period contributions.

Reducing gives

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

Figure 1a illustrates the effect of catastrophe on the full cooperative outcome. The full cooperative payoff for country *i* is given by the vertical distance between the benefit and cost curves. The "interior" maximum is at $Q = bN^2/c$. In Figure 1a, the "corner" maximum at $\bar{Q}$ yields a higher payoff. This value is thus the optimal abatement level, Q^{FC*}.

Figure 1b shows the marginal analysis coinciding with the situation depicted in Figure 1a. The triangle identified in the figure is the extra cost of avoiding the threshold. The benefit of doing so is *X*. If *X* is bigger than the value represented by this triangle, it will pay to abate $\bar{Q}$ rather than $Q = bN^2/c$.

This is for a given *X* and $\bar{Q}$. Figure 2 illustrates the combination of values for *X* and $\bar{Q}$ for which catastrophe avoidance is collectively optimal. (The relationship is non-linear due to marginal costs increasing.) If *X* is "small," catastrophe is worth avoiding only if $\bar{Q}$ is "small." If $\bar{Q}$ is "large," catastrophe is worth avoiding only if *X* is "large."

3.2 Non-cooperation

In the non-cooperative outcome, each country *i* will maximize (1), taking as given the abatement choices of other countries. There are two symmetric Nash equilibria in pure strategies. Either every country *i* plays $q_i = b/c$, and the threshold is missed, or every country plays $q_i = \bar{Q}/N$, and the threshold is just met. Suppose every country $j \neq i$ plays $q_j = \bar{Q}/N$. Then, if *i* plays $q_i = \bar{Q}/N$, *i* gets

$$\pi_i(\bar{Q}/N;\bar{Q}(N-1)/N) = b\bar{Q} - \frac{c}{2}\left(\frac{\bar{Q}}{N}\right)^2. \tag{4a}$$

Although *i* can ensure that the threshold is met, it may not have an incentive to do so. If *i* does not play $q_i = \bar{Q}/N$, it will play $q_i = b/c$, and get the payoff

$$\pi_i(b/c;\bar{Q}(N-1)/N) = b\left(\frac{\bar{Q}(N-1)}{N} + \frac{b}{c}\right) - X - \frac{c}{2}\left(\frac{b}{c}\right)^2. \tag{4b}$$

Given that every country $j \neq i$ plays $q_j = \bar{Q}/N$, *i* can do no better than to play $q_i = \bar{Q}/N$ provided (4a) exceeds (4b). Avoiding catastrophe is a symmetric Nash equilibrium iff

$$X \geq \frac{b^2}{2c} - \left(\frac{b\bar{Q}}{N} - \frac{c\bar{Q}^2}{2N^2}\right). \tag{5}$$

Figure 3 illustrates the non-cooperative outcome corresponding to the full cooperative outcome shown in Figure 1b. Notice that, in the non-cooperative outcome, the indicated triangle is larger. This is because countries that fail to cooperate ignore the benefit their abatement gives to others. Unless *X* is very large, there will be a gap between the full cooperative and non-cooperative outcomes.

Figure 4 shows this gap for combinations of values for *X* and $\bar{Q}$. The top part of the figure shows the space for which catastrophe avoidance is both (i) a symmetric Nash equilibrium and (ii) collectively optimal. Just below this is a space for which catastrophe avoidance is collectively optimal but not supportable as a symmetric Nash equilibrium. The figure shows that full cooperation cannot be sustained by coordination if X is very small, even if $\bar{Q}$ is close to the full cooperative level. Only if X is "big enough" can countries coordinate to avoid catastrophe.

In comparing (3) and (5), we see that the condition for avoiding catastrophe is identical if 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. 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 $q_i = \bar{Q}/N$. 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. The effect of abatement in reducing gradual climate change thus increases the incentive for a country to deviate from an agreement seeking to avert catastrophic climate change.

3.3 Multiple catastrophes

There may not be one catastrophic threshold but several. O'Neill and Oppenheimer (2004), for example, present three different catastrophe scenarios—loss of coral reefs, disintegration of the West Antarctic Ice Sheet, and shut down of the ocean circulation conveyor belt—with three different thresholds—1 ˚C, 2 ˚C, and 3 ˚C, respectively. How will the results presented thus far change if there are multiple approaching catastrophes?

Assume for simplicity that there are two thresholds. Threshold 1 is triggered for $Q < \bar{Q}_1$ and causes damage X_1; threshold 2 is triggered for $Q < \bar{Q}_2$ and causes damage X_2; and $\bar{Q}_1 > \bar{Q}_2$.[12] This means that, if we avoid the first threshold, we will also avoid the second. Assume $\bar{Q}_2 > bN^2/c$. Then full cooperation will require either

[12] Another possibility is that passing one threshold "tips" the system, causing us to pass the second threshold. In this case, the consequences of passing the first threshold would be $X_1 + X_2$.

avoiding both thresholds, avoiding the second threshold but not the first, or avoiding neither threshold. The payoffs in these cases are given by:

$$\Pi^{FC} = \begin{cases} bQN - \sum_i \frac{cq_i^2}{2} \text{ if } Q \geq \bar{Q}_1 \\ bQN - \sum_i \frac{cq_i^2}{2} - X_1 N \text{ if } \bar{Q}_1 > Q \geq \bar{Q}_2 \\ bQN - \sum_i \frac{cq_i^2}{2} - X_1 N - X_2 N \text{ if } \bar{Q}_2 > Q. \end{cases} \tag{6}$$

It is best to avoid both thresholds if and only if

$$X_1 + X_2 > \frac{b^2 N^2}{2c} - \left(b\bar{Q}_1 - \frac{c\bar{Q}_1^2}{2N^2} \right) \text{ and } X_1 > \left(b\bar{Q}_2 - \frac{c\bar{Q}_2^2}{2N^2} \right) - \left(b\bar{Q}_1 - \frac{c\bar{Q}_1^2}{2N^2} \right). \tag{7}$$

The first inequality in (7) says that avoiding both thresholds is better than avoiding neither. The second says that avoiding both is better than avoiding the second threshold only.

It is best to avoid the second threshold and not the first iff

$$X_2 > \frac{b^2 N^2}{2c} - \left(b\bar{Q}_2 - \frac{c\bar{Q}_2^2}{2N^2} \right) \text{ and } X_1 < \left(b\bar{Q}_2 - \frac{c\bar{Q}_2^2}{2N^2} \right) - \left(b\bar{Q}_1 - \frac{c\bar{Q}_1^2}{2N^2} \right). \tag{8}$$

The first inequality in (8) says that avoiding the second threshold only is better than avoiding neither. The second inequality says that avoiding the second threshold only is better than avoiding both thresholds.

The cost of avoiding the first threshold will be greater than the cost of avoiding only the second threshold. Of course, the benefits will be greater, too, since by avoiding the first threshold we will also avoid the second. But it is possible to avoid the second threshold and not the first—which is why, with two thresholds, two inequalities must hold to support full cooperation. (In general, for *n* thresholds, there will be *n* + 1 options, including the option of not avoiding any threshold; the full cooperative outcome will have to satisfy *n* conditions.)

As compared with the analysis of a single catastrophic threshold, the second conditions in (7) and (8) are novel. These conditions tell us that, if the two thresholds are relatively near one another, the extra cost of avoiding both catastrophes rather than one will be small, making it worthwhile either to avoid both catastrophes or neither. In this case, the value of X_2 is important in determining whether the first threshold should be avoided. By contrast, if the two thresholds are far apart, X_1 must be very large in order for it to be optimal to avoid the first threshold, with the value of X_2 having very little impact on this decision.

Analysis of the non-cooperative game is somewhat tricky, but an important insight emerges from very simple reasoning. Consider the first threshold. As before, play $\bar{Q}_1/N$ will be a Nash equilibrium only if every country *i* is worse off for deviating unilaterally. If *i* were to deviate, it could play b/c, just as in the case where there is only one threshold. Alternatively, it could play $\bar{Q}_2 - \bar{Q}_1(N-1)/N$ so as to avoid passing the second threshold. If inequality $b/c \geq \bar{Q}_2 - \bar{Q}_1(N-1)/N$ holds, *i* could do no better than to play b/c. It will never pay for *i* to abate less than this, and if the above inequality holds, it can't pay for *i* to abate more than this. This is because, by playing b/c, *i* is sure to avoid passing the second threshold when it deviates from the first threshold. If the above inequality does not hold, the analysis is more complicated. However, so long as *N* is "large" and $\bar{Q}_1$ and $\bar{Q}_2$ are not very, very close to one another, the above inequality will hold. Moreover, a similar result will hold for the second threshold. Overall, in a non-cooperative setting, countries are likely to behave as if each of a multiple of thresholds were the only threshold.

Taken together, these results imply that, with multiple catastrophes that are neither very far apart nor very, very close, the gap between full cooperation and non-cooperation will be amplified. There may exist situations in which full cooperation to avert a first catastrophe will be strengthened by the recognition that this will also avert a second catastrophe, but where these connections will be ignored by countries acting unilaterally.

3.4 Sustaining full cooperation by means of a self-enforcing agreement

Could a treaty sustain outcomes in the intermediate region of Figure 4? Let us see.

Start from a situation in which every country participates in a treaty requiring that each country play $q_i = \bar{Q}/N$. Will any country want to deviate unilaterally? Upon withdrawing from the agreement, 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 these countries play Q_{-i} when *i* does not participate. Then *i* will either want to play $q_i = \bar{Q} - Q_{-i}$, to ensure that the threshold is met, or it will let the threshold slip, and play $q_i = b/c$. It will prefer to play the former level rather than the latter if

$$b\bar{Q} - \frac{c}{2}(\bar{Q} - Q_{-i})^2 \geq b\left(Q_{-i} + \frac{b}{c}\right) - X - \frac{c}{2}\left(\frac{b}{c}\right)^2 \tag{9}$$

Solving the quadratic, the minimum value of Q_{-i} which ensures that (9) is met is given by $\hat{Q}_{-i} = \bar{Q} - b/c - \sqrt{2X/c}$.

Given that a country has deviated, the *N* – 1 cooperating countries will either want to play $\hat{Q}_{-i}$ precisely or they will want to abandon the effort to avoid catastrophe. Suppose they play $\hat{Q}_{-i}$ precisely. Then it will pay country *i* not to deviate if

$$b\bar{Q}-\frac{c}{2}\left(\frac{\bar{Q}}{N}\right)^2 \geq b\bar{Q}-\frac{c}{2}\left(\frac{b}{c}+\sqrt{\frac{2X}{c}}\right)^2 \tag{10}$$

Inequality (10) reduces to (5). In this case, a treaty is able to sustain full cooperation only if full cooperation can be sustained by coordination. Treaties in this case are pure coordination devices.

The other possibility noted above is that the defection causes all countries to abandon the effort to avoid catastrophe. In this case, the withdrawal will cause the other countries to play $Q_{-i}=b(N-1)^2/c$; *i*, of course, will play $q_i=b/c$. Under these circumstances, *i* will prefer not to deviate provided

$$b\bar{Q}-\frac{c}{2}\left(\frac{\bar{Q}}{N}\right)^2 \geq b\left(\frac{b(N-1)^2}{c}+\frac{b}{c}\right)-X-\frac{c}{2}\left(\frac{b}{c}\right)^2 \tag{11}$$

Rearranging gives

$$X \geq \frac{b^2[2(N-1)^2+1]}{2c}-\left(b\bar{Q}-\frac{c\bar{Q}^2}{2N^2}\right) \tag{12}$$

Condition (12) is the same as (3) except for the intercept. It is straightforward to prove that, should a unilateral withdrawal cause other countries to play $Q_{-i}=b(N-1)^2/c$, full cooperation can be sustained by a treaty only if $N \leq 3$. It is also easy to show that, when $X = 0$, full cooperation can be sustained by a treaty only when N ≤ 3.[13] Hence, when coordination alone is unable to sustain full cooperation, the prospect of catastrophe fails to help a self-enforcing agreement sustain any additional cooperation.[14]

---

[13] For surveys highlighting this result, see Finus (2001) and Barrett (2005).

[14] Note that I have not solved for the conditions under which parties to a treaty would choose to respond to a unilateral defection by playing $Q_{-i}=b(N-1)^2/c$. The expression is easy to derive, but it is a complicated expression. Note as well that I have not explored here the prospects for partial cooperation. Doing so would require a slightly different approach. Should there be two non-signatories, we would need to decide how these countries would choose to play. It seems likely that they would play mixed strategies.

4. Avoiding catastrophe with backstop technologies

Our greatest concern must lie with the area in Figure 3 in which $\bar{Q}$ and X are both large, and yet coordination alone cannot steer us clear of catastrophe. However, a key assumption is that marginal abatement costs are increasing. I now suppose there exists a backstop technology with constant marginal costs. As noted in the introduction, the only available backstop technology for climate change is "air capture."[15]

Let country i's payoff now be given by

$$\pi_i = \begin{cases} b(Q+Z) - \dfrac{cq_i^2}{2} - \gamma z_i \text{ if } Q+Z \geq \bar{Q} \\ b(Q+Z) - X - \dfrac{cq_i^2}{2} - \gamma z_i \text{ if } Q+Z < \bar{Q}, \end{cases} \tag{13}$$

where z_i is i's level of air capture and $Z = \sum_{i=1}^{N} z_i$.

If countries cooperate fully they will maximize

$$\Pi^{FC} = \begin{cases} bN(Q+Z) - \sum_i \dfrac{cq_i^2}{2} - \sum_i \gamma z_i \text{ if } Q+Z \geq \bar{Q} \\ bN(Q+Z) - XN - \sum_i \dfrac{cq_i^2}{2} - \sum_i \gamma z_i \text{ if } Q+Z < \bar{Q}. \end{cases} \tag{14}$$

Assuming $\gamma > bN$, full cooperation requires $q^{FC} = bN/c$ and $z^{FC} = 0$ for X = 0. Hence, the backstop technology will only be used to avert catastrophe. In the full cooperative outcome, abatement must be cost-effective. This means that countries

[15] For a analysis of the role this technology can play in addressing "gradual" climate change, see Barrett (2010). Another technology that may be used to avert catastrophic climate change is "geoengineering." See Weitzman (2009), Kousky *et al*. (2009), and Barrett (2009), Note, however, that while this technology has a number of advantages over air capture (especially, cost), it would not limit atmospheric concentrations of greenhouse gases; it is not a true backstop technology.

will abate up to a level $q_i = \gamma/c$ and rely on air capture for larger levels of mitigation. Hence, for $\bar{Q} \leq \gamma N/c$, the previous analysis is unchanged. For $\bar{Q} > \gamma N/c$, it will pay to avert catastrophe if and only if

$$b\bar{Q}N - \frac{c}{2}\left(\frac{\gamma}{c}\right)^2 N - \gamma\left(\frac{\bar{Q}}{N} - \frac{\gamma}{c}\right)N \geq b\left(\frac{bN^2}{c}\right)N - XN - \frac{c}{2}\left(\frac{bN}{c}\right)^2 N \tag{15}$$

or

$$X \geq \frac{-(\gamma - bN)(\gamma + bN)}{2c} + \frac{(\gamma - bN)\bar{Q}}{N} \tag{16}$$

Suppose now that every country $j \neq i$ plays $q_j = \gamma/c$ and $z_j = \left(\bar{Q}/N - \gamma/c\right)$. Then, if country i plays $q_i = \gamma/c$ and $z_i = \left(\bar{Q}/N - \gamma/c\right)$, i gets

$$b\bar{Q} - \frac{c}{2}\left(\frac{\gamma}{c}\right)^2 - \gamma\left(\frac{\bar{Q}}{N} - \frac{\gamma}{c}\right) \tag{17}$$

If i deviates it will play $q_i = b/c$ and get the same payoff as in (4b). Country i will not deviate if

$$X \geq -\frac{(\gamma + b)(\gamma - b)}{2c} + \frac{(\gamma - b)\bar{Q}}{N} \tag{18}$$

Condition (18) holds for $\bar{Q} \geq \gamma N/c$ and it is easy to show that (18) and (5) are equivalent for $\bar{Q} = \gamma N/c$.

Figure 5 illustrates the new result. The availability of a backstop technology expands the space in which coordination suffices to sustain an efficient outcome. However, it also expands the space in which full cooperation commends avoiding catastrophe.

5. Uncertain catastrophic damages

Weitzman (2009) emphasizes fat-tailed uncertainty in climate sensitivity, saying that this would be compounded by uncertainties in translating temperature changes into welfare changes. Here I distinguish between uncertainty in the loss due to catastrophe (*X*) and uncertainty about the threshold that would trigger catastrophe ($\bar{Q}$).

How does uncertainty about the loss, *X*, affect the above results? It is straightforward to demonstrate that all the results shown thus far carry through if we substitute $E(X)$ for *X*.[16] 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.[17] As I shall now show, however, this same result does *not* hold as regards uncertainty about the threshold, $\bar{Q}$.

6. Uncertain catastrophic thresholds

Consider now the case where $\bar{Q}$ is a random variable with a continuous cumulative probability distribution $F(Q) = \Pr\left(\bar{Q} \leq Q\right)$. Again, we can think of $\bar{Q}$ as representing the total amount of abatement (reductions in emissions from business as usual) needed to keep temperature below some threshold value—say, 2 °C.

A possible representation of *F*(*Q*) is shown in Figure 6. In the figure, $P(\bar{Q} \leq 0) > 0$, implying that there is a positive probability that catastrophe will never materialize, even if we do nothing to limit emissions. In addition, $P(\bar{Q} \leq Q^{\max}) < 1$, implying that,

[16] This assumes that *E*(*X*) exists. For some fat-tailed distributions, expected value does not exist. In these cases, decision-making would need to be guided by a different criterion.
[17] This result should have been anticipated from my previous discussion of the Milinski *et al.* (2008) experiment. Recall that, in this experiment, the value of *X* (and not $\bar{Q}$) was uncertain.

if abatement has an upper limit, $Q^{\max}$, it may not be possible to avoid catastrophe with certainty. (A backstop technology may ease this constraint.) Of course, other representations of the cumulative probability distribution are possible, but Figure 6 looks very similar to the relations derived from climate models.[18]

6.1 Full cooperation

Were countries to cooperate fully, they would now maximize

$$E(\Pi^c) = bQN - \sum_j \frac{cq_j^2}{2} - XN\left[1 - F(Q)\right] \tag{19}$$

which requires

$$bN - cq_i + XNf(Q) = 0 \tag{20}$$

where $f(Q)$ is the probability density function. Note that, in general, eq. (20) 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.

Figure 7a illustrates eq. (20) for a plausible pdf, one that is consistent with the cumulative probability distribution shown in Figure 6. For this pdf, concern about

[18] See, in particular, Zickfeld (2009), Figures 3a-c. In these figures, the horizontal axis represents total emissions, not abatement; and the vertical axis represents the probability of temperature exceeding a threshold, not of temperature falling short of a threshold. After adjusting for these differences, Zickfeld's figures look very much like Figure 6.

catastrophic climate change increases the full cooperative abatement level just a little.

There are other possibilities. In Figure 7b, the calculus for choosing an abatement level is changed dramatically by the uncertain prospect of catastrophe. In Figure 7c, the pdf is the same as in Figure 7a, but the value of *X* is increased, and the prospect of catastrophe now has a huge effect on the full cooperative outcome. What matters in all of these figures is the size of areas *A* and *B*.

These figures help to illuminate an important policy question. Suppose that it is universally agreed that temperature change ought to be limited to 2 ˚C. Given uncertainty in climate sensitivity, what concentration target (and, therefore, cumulative emission level or abatement level from business as usual) should we aim for? If we were willing to meet the target with probability 18 percent, we would limit concentrations to 550 ppm CO_2e, but if we wanted to meet the goal with probability 93 percent, we would need to limit concentrations to 350 ppm CO_2e (Anderson and Bows 2008). This framing of the problem implies that choice of a concentration target depends on risk aversion. However, the analysis developed here points to other considerations. It shows that the former target would make sense if the important relationships were consistent with Figure 7a, and that the latter target would be commended if these relationships resembled Figure 7b or 7c. In short, the optimal level of atmospheric concentrations depends on the values of *b*, *X,* and *c* as well as the pdf for the threshold, *in addition to preferences towards risk*.

Does the possibility of "fat tails" matter? As shown in Figure 8, a thickening of the tails of the pdf changes the full cooperative outcome very little.[19] Weitzman, of course, gets a very different result. But in contrast to Weitzman, I am assuming constant marginal utility. This makes it important to consider the marginal costs of

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

abatement, which I take to be increasing and which Weitzman does not (need to) model.

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. (20).

6.2 Non-cooperation

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

$$E(\pi_i) = bQ - \frac{cq_i^2}{2} - X\left[1 - F(Q)\right] \tag{21}$$

The solution requires

$$b - cq_i + Xf(Q) = 0 \forall i \tag{22}$$

What determines the gap between the full cooperative and non-cooperative outcomes? In the example shown in Figure 9a, the prospect of uncertain catastrophe has a substantial effect on the full cooperative outcome but virtually no effect on the non-cooperative abatement level.

Figure 9b illustrates a situation in which the prospect of uncertain catastrophe narrows the gap between the non-cooperative and full cooperative outcomes. It is obvious from the figure that the circumstances that support this more cheerful outcome aren't necessarily to be expected. For example, uncertainty is more likely to narrow the gap when *N* is small. But when *N* is small cooperation is easier to sustain in any event—and for the climate problem, *N* isn't small.

6.3 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 above, there is uncertainty. Like the earlier part of the paper, there is a discontinuity—in this case, at both of the distribution's supports. The discontinuity on the right side means that there may be a role for coordination. 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

$$f(Q) = \begin{cases} 0 & \text{for } Q < \bar{Q}_{\min} \\ \dfrac{1}{\bar{Q}_{\max} - \bar{Q}_{\min}} & \text{for } Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}] \\ 0 & \text{for } Q > \bar{Q}_{\max}. \end{cases} \tag{23}$$

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). The corresponding cumulative distribution function is

$$F(Q) = \begin{cases} 0 & \text{for } Q < \bar{Q}_{\min} \\ \dfrac{Q - \bar{Q}_{\min}}{\bar{Q}_{\max} - \bar{Q}_{\min}} & \text{for } Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}] \\ 1 & \text{for } Q > \bar{Q}_{\max}. \end{cases} \tag{24}$$

If countries cooperate fully, they will maximize

$$E(\Pi^c) = \begin{cases} bQN - \sum_j \frac{cq_j^2}{2} - XN & \text{for } Q < \bar{Q}_{\min} \\ bQN - \sum_j \frac{cq_j^2}{2} - XN\left[1 - \frac{(Q - \bar{Q}_{\min})}{(\bar{Q}_{\max} - \bar{Q}_{\min})}\right] & \text{for } Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}] \\ bQN - \sum_j \frac{cq_j^2}{2} & \text{for } Q > \bar{Q}_{\max}. \end{cases} \quad (25)$$

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

$$Q^{FC} = \frac{bN^2}{c} + \frac{XN^2}{c(\bar{Q}_{\max} - \bar{Q}_{\min})}. \quad (26)$$

Figures 10a and 10b illustrate the two possibilities. In Figure 10a, (26) holds; the solution is "interior." In Figure 10b, a corner solution holds. (Note that, with $\bar{Q}_{\min} > bN^2/c$ it will never pay to abate $\bar{Q}_{\min}$.)

Upon substituting $Q^{FC} = bN^2/c$ and (26) into (25), it is easy to show that countries should undertake some additional abatement to reduce the chance of catastrophe provided

$$X > \frac{2c(\bar{Q}_{\max} - \bar{Q}_{\min})}{N^2}\left(\bar{Q}_{min} - \frac{bN^2}{c}\right). \quad (27a)$$

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

$$X > \frac{c}{2N^2}\left(\bar{Q}_{\max} - \frac{bN^2}{c}\right)^2. \quad (27b)$$

Figures 10a-10b illustrate these conditions. In Figure 10a, condition (27a) implies that the triangle B is larger in area than triangle A. In Figure 10b, condition (27b) implies that trapezoid B is larger in area than triangle A.

Figures 11a-11e show how these conditions relate to the parameters. Figure 11a shows that an increase in X can increase the full cooperative abatement level dramatically. Figure 11b shows that a reduction in $\bar{Q}_{\max}$, holding $\bar{Q}_{\min}$ constant, can also increase abatement in the full cooperative outcome. In this case, the pdf is essentially squeezed and pushed up; uncertainty is reduced. In Figure 11c, uncertainty is unchanged, but the range is shifted. Avoiding uncertain catastrophe is cheaper, and full cooperation now requires an increase in abatement. Figure 11d shows that an increase in b can also increase the full cooperative abatement level. Finally, Figure 11d shows that a reduction in the marginal costs of reducing emissions can cause the full cooperative abatement level to increase.

If countries do not cooperate, each country i will maximize

$$E(\pi_i) = \begin{cases} bQ - \dfrac{cq_i^2}{2} - X & \text{for } Q < \bar{Q}_{\min} \\ bQ - \dfrac{cq_i^2}{2} - X\left[1 - \dfrac{(Q - \bar{Q}_{\min})}{(\bar{Q}_{\max} - \bar{Q}_{\min})}\right] & \text{for } Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}] \end{cases} \tag{28}$$

In the Nash equilibrium, countries will either play $q = b/c$ or, solving (19),

$$q_i^* = \frac{b}{c} + \frac{X}{c(\bar{Q}_{\max} - \bar{Q}_{\min})} \tag{29}$$

or $q_i^* = \bar{Q}_{\max}/N$, whichever is smaller.[20] Upon comparing the payoffs, countries will invest effort in reducing the chance of catastrophe if

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

$$X > \frac{2c(\bar{Q}_{\max} - \bar{Q}_{\min})}{(2N-1)}\left(\bar{Q}_{\min} - \frac{b(2N-1)}{c}\right). \tag{30a}$$

They will avoid the possibility of catastrophe if

$$X > \frac{c}{2N}\left(\bar{Q}_{\max} - \frac{bN}{c}\right)^2. \tag{30b}$$

In comparing (27) and (30), it is clear that the conditions needed to prevent catastrophe are identical only when $N = 1$. As *N* increases, the gap widens. Figure 12a (which, of course, resembles Figure 9a) shows that the prospect of uncertain catastrophe can have a substantial effect on the full cooperative outcome even as it has no effect at all on the non-cooperative outcome. Figure 12b illustrates a situation in which avoiding catastrophe can be sustained by coordination. Unfortunately, to obtain this more cheerful result, *X* must be truly huge, especially if *N* is large.

Figure 13 shows the range of values for X and $\bar{Q}_{\max}$ for which catastrophe avoidance can be sustained by coordination. It is obvious from comparing Figure 4 and 13 that uncertainty widens the area requiring cooperation. However, the extent of this widening is greater than indicated by the figures. At the horizontal intercept, where $\bar{Q}$ and $\bar{Q}_{\max}$ are just a little bit bigger than bN^2/c, the value of X that enables catastrophe to be avoided using coordination is 100 times greater in Figure 13 than in Figure 4.[21] Uncertainty about the catastrophic threshold implies that coordination is unlikely to suffice; cooperation is still needed, and still difficult to sustain.

To sum up, even assuming a uniform distribution with a discontinuity in the far right support, it is very unlikely that coordination will suffice to enforce a treaty limiting greenhouse gas emissions.

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

7. Conclusions

Under some but not all circumstances, the prospect of approaching catastrophes fundamentally alters the challenge of getting countries to cooperate to limit global climate change. The standard model of a self-enforcing international environmental agreement predicts that, for a problem like climate change, cooperation in reducing emissions will be grossly inadequate. When this model is modified to incorporate a certain threshold with catastrophic damages, full cooperation can be sustained in many cases, including the worst imaginable cases. Essentially, the prospect of catastrophe transforms the cooperation problem into one requiring only coordination.

Uncertainty about the threshold overturns this result. When the catastrophe threshold is certain, a slight relaxation in abatement triggers a discontinuous change in damages—a huge deterrent to free riding. When the threshold is uncertain, this same policy change causes expected damages to increase just a little, making countries more inclined to let abatement slip (that is, 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.

These results are consistent with how the climate negotiations have played out so far. Efforts to identify catastrophic thresholds have had little impact on cooperation; they have only made the failure to cooperate seem more alarming.

References

Allen, Myles R., David J. Frame, Chris Huntingford, Chris D. Jones, Jason A. Lowe, Malte Meinshausen, and Nicolai Meinshausen (2009). "Warming Caused by Cumulative Carbon Emissions Towards the Trillionth Tonne," *Nature* 458: 1163-1166.

Anderson, Kevin and Alice Bows (2008). "Reframing the Climate Change Challenge in Light of Post-2000 Emission Trends." *Philosophical Transactions of the Royal Society A* 366: 3863-3882.

Barrett, Scott (2003). *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford: Oxford University Press.

Barrett, Scott (2005). "The Theory of International Environmental Agreements," in K.-G. Mäler and J. Vincent (eds.), *Handbook of Environmental Economics*, Vol. 3, Amsterdam: Elsevier, pp. 1457-1516.

Barrett, Scott (2009). "The Coming Global Climate-Technology Revolution." *Journal of Economic Perspectives* 23(2): 53-75.

Barrett, Scott (2010). "Climate Treaties and Backstop Technologies." *CESifo Ecoomic Studies*, forthcoming.

Dreber, Anna and Martin A. Nowak (2008). "Gambling for Global Goods." *Proceedings of the National Academy of Sciences* 105(7): 2261-2262.

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

Kousky, Carolyn, Olga Rostapshova, Michael Toman, and Richard Zeckhauser (2009). "Responding to Threats of Climate Change Mega-Catastrophes." Resources for the Future Discussion Paper 09-45.

Milinksi, Manfred, Ralf D. Sommerfeld, Hans-Jürgen Krambeck, Floyd A. Reed, and Jochem Marotzke (2008). "The Collective-Risk Social Dilemma and the Prevention of Simulated Dangerous Climate Change." *Proceedings of the National Academy of Sciences* 105(7): 2291-2294.

Nordhaus, William (2009). "An Analysis of the Dismal Theorem," mimeo, Yale University, January 16.

O'Neill, Brian C. and Michael Oppenheimer (2002). "Dangerous Climate Impacts and the Kyoto Protocol." *Science* 296(5575): 1971-1972.

Sarewitz, Daniel and Richard Nelson (2008). “Three Rules for Technological Fixes.” *Nature* 456: 871-872.

Weitzman, Martin L. (2009). “On Modeling and Interpreting the Economics of Catastrophic Climate Change,” *Review of Economics and Statistics*, 91(1): 1-19.

Zickfeld, Kirsten, Michael Eby, H. Damon Matthews, and Andrew J. Weaver (2009). “Setting Cumulative Emissions Targets to Reduce the Risk of Dangerous Climate Change.” *Proceedings of the National Academies*, 106(38): 16129-16134.

FIGURE 1a
Effect of catastrophe on the full cooperative outcome

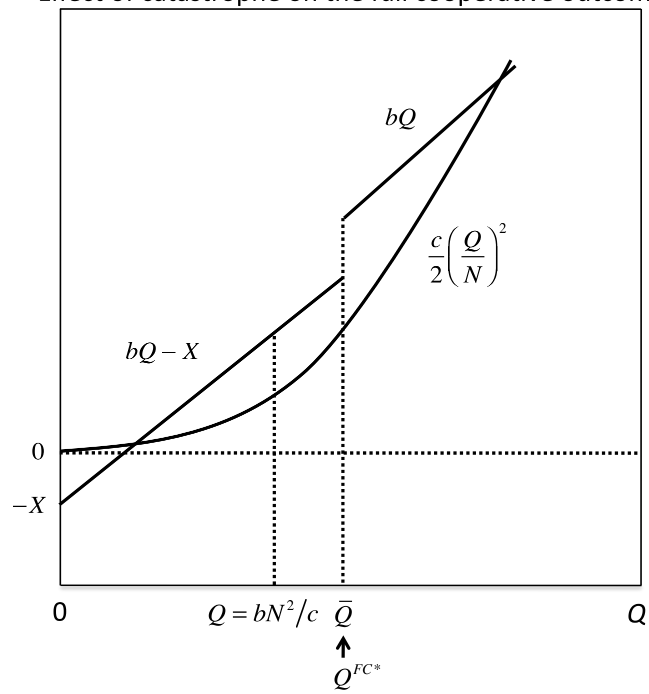

FIGURE 1b
Effect of catastrophe on the full cooperative outcome

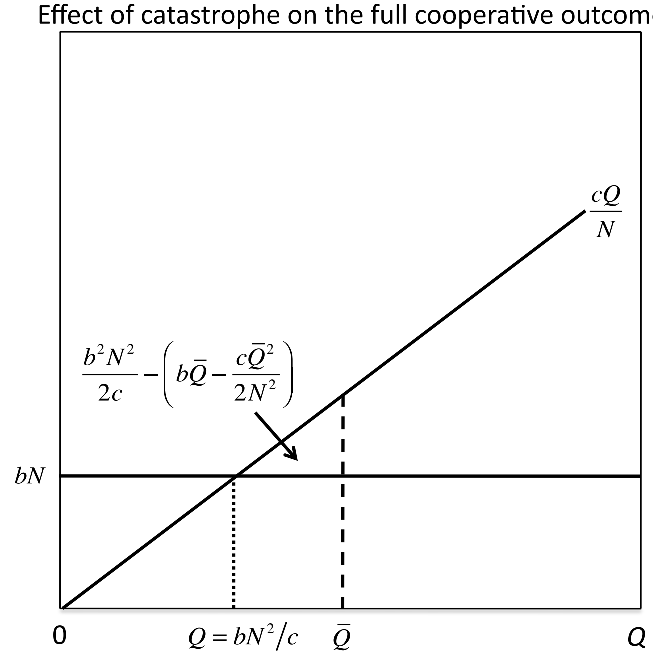

FIGURE 2

Optimal catastrophe avoidance

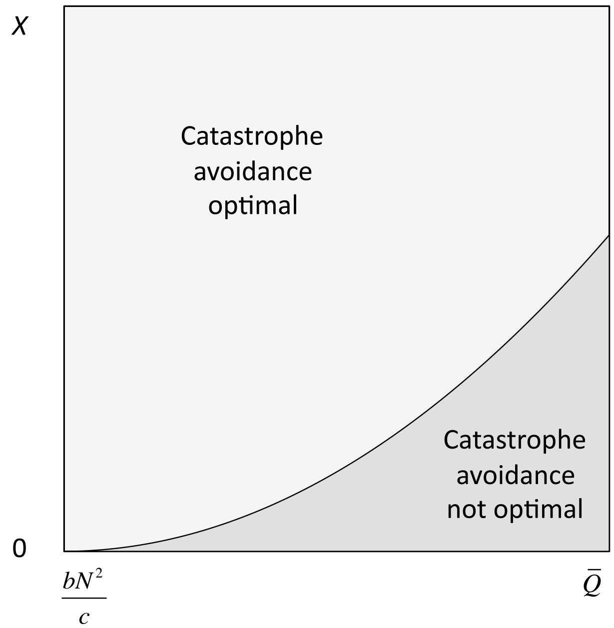

FIGURE 3

Effect of catastrophe on the non-cooperative outcome

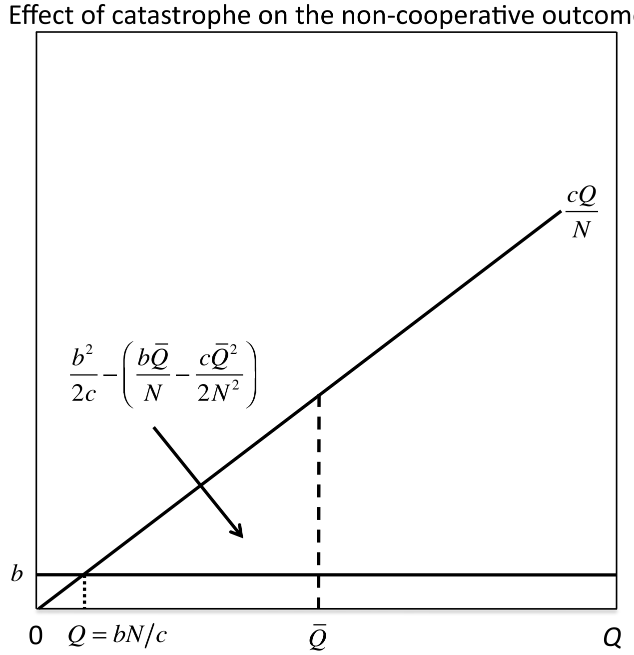

FIGURE 4
When coordination suffices and cooperation is needed

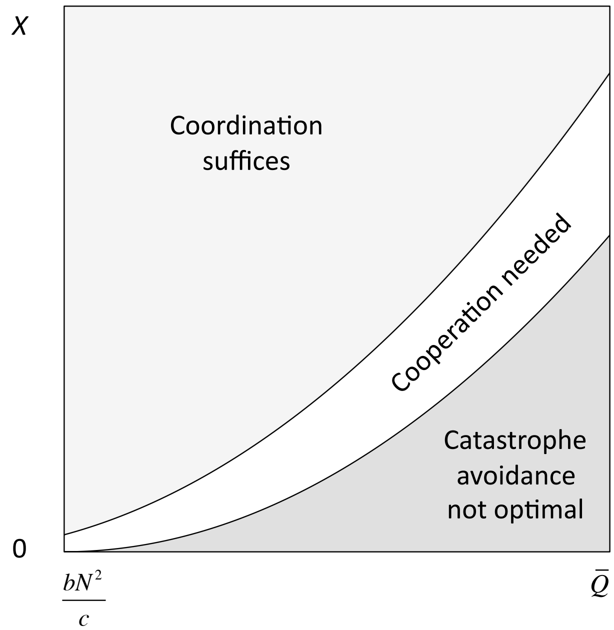

FIGURE 5
Avoiding catastrophe with a backstop technology

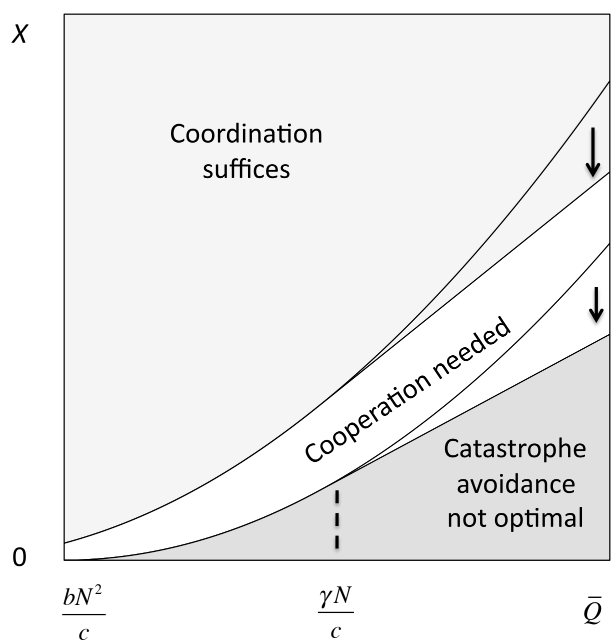

FIGURE 6
Possible cumulative probability function

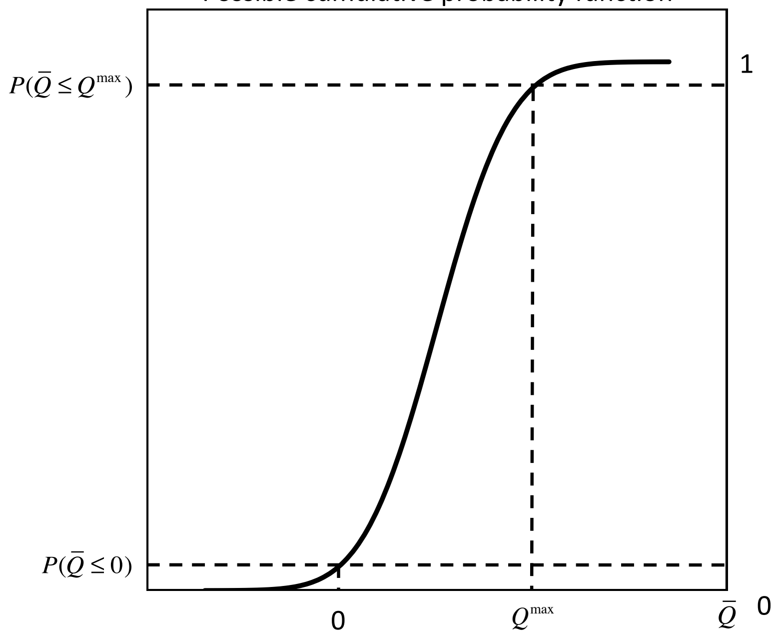

FIGURE 7a
Prospect of catastrophe has little effect on full cooperative outcome

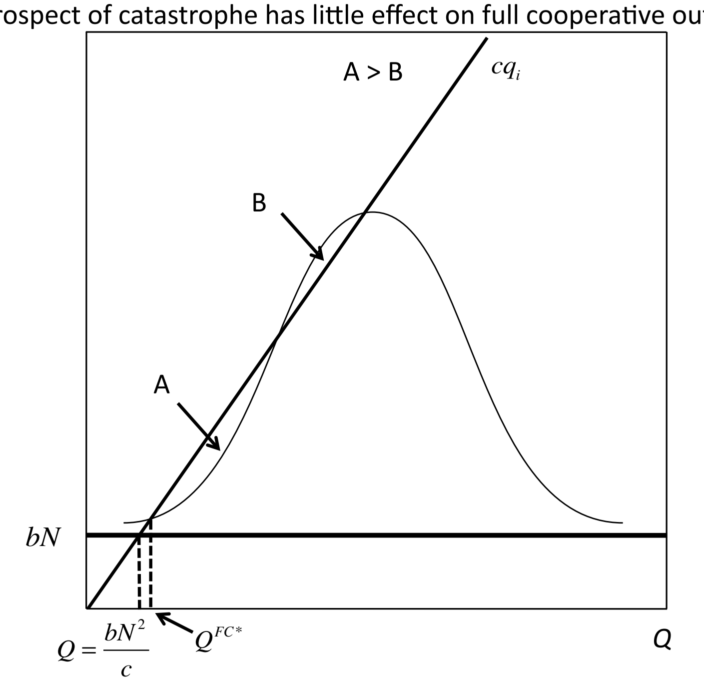

FIGURE 7b
Prospect of catastrophe has profound effect on full cooperative outcome

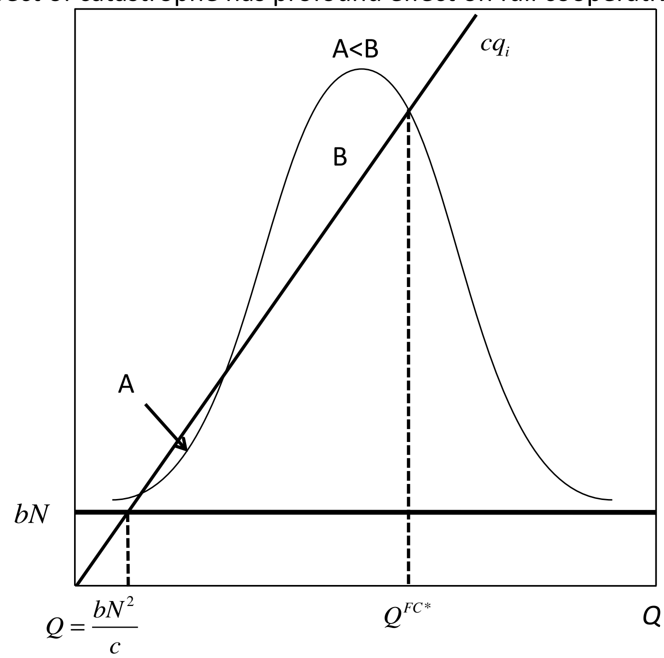

FIGURE 7c
Prospect of catastrophe has substantial effect on full cooperative outcome

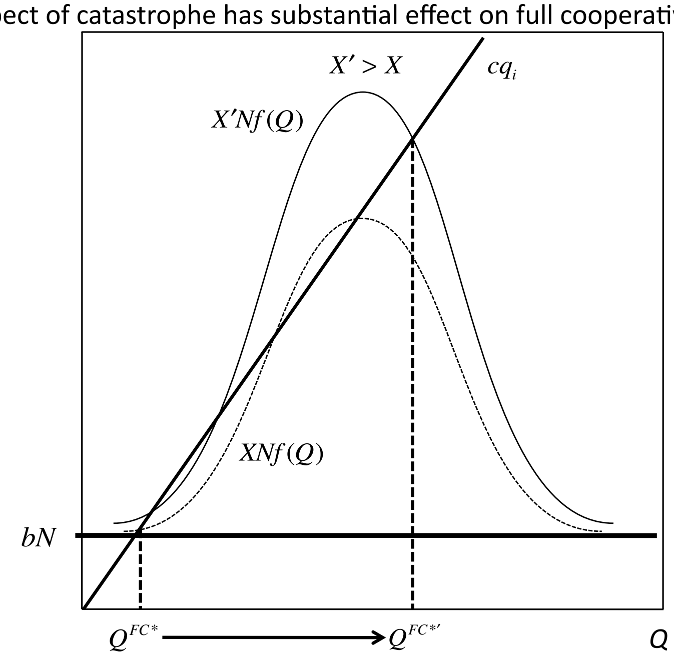

FIGURE 8
Prospect of catastrophe with "thin" and "fat" tails

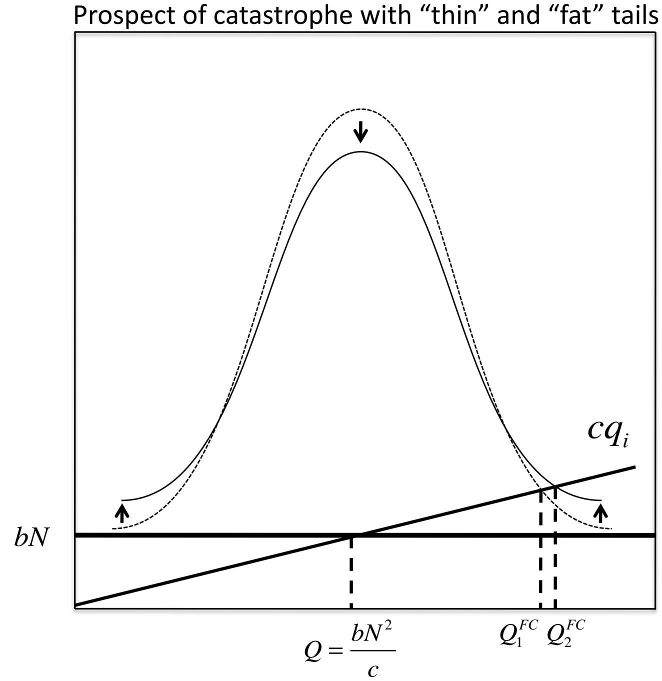

FIGURE 9a
The non-cooperative and full cooperative outcomes

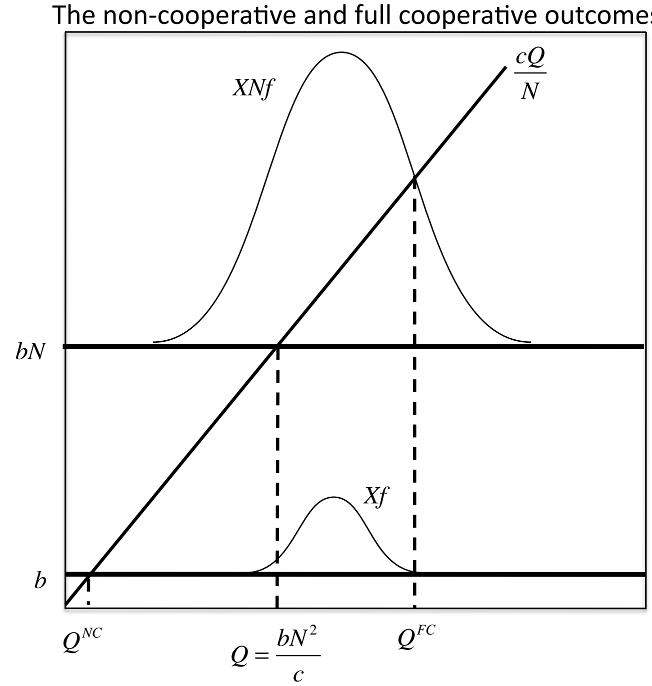

FIGURE 9b
The non-cooperative and full cooperative outcomes

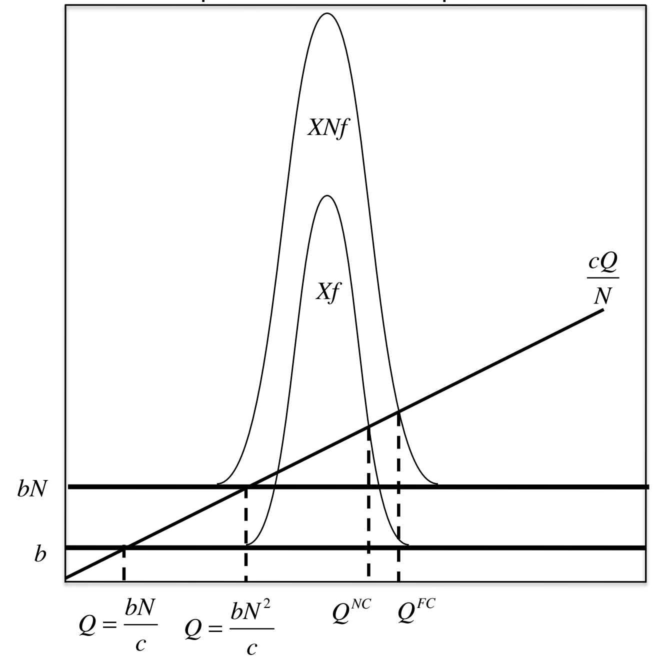

FIGURE 10a
"Interior" full cooperative outcome

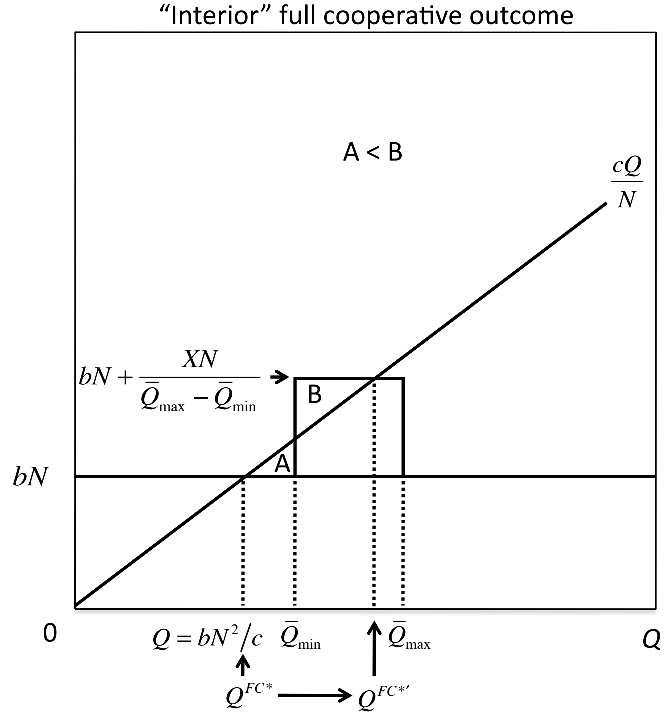

FIGURE 10b

"Corner" full cooperative outcome

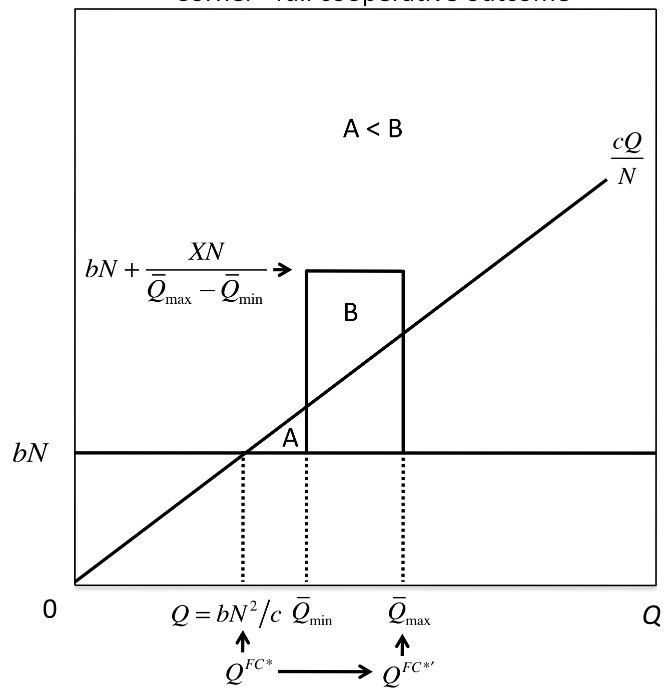

FIGURE 11a

Effect of *X* on full cooperative outcome

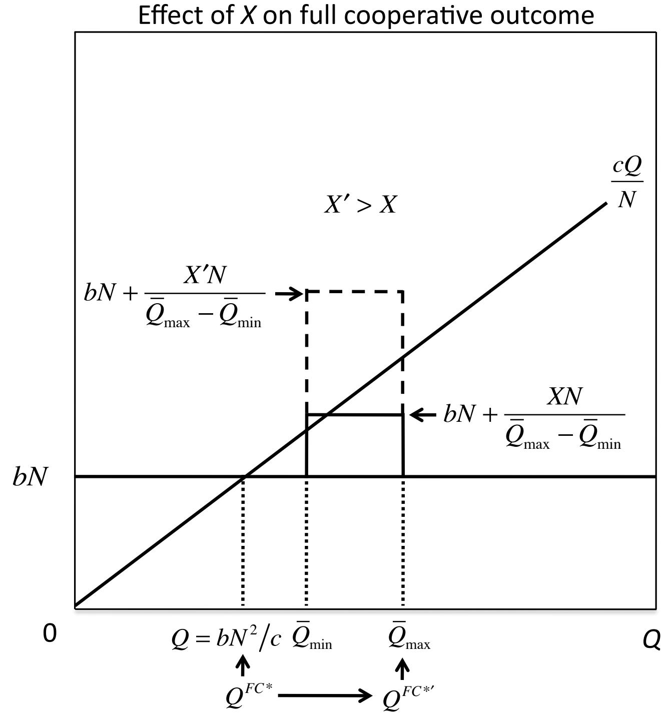

FIGURE 11b

Effect of $\bar{Q}_{\max}$ on full cooperative outcome

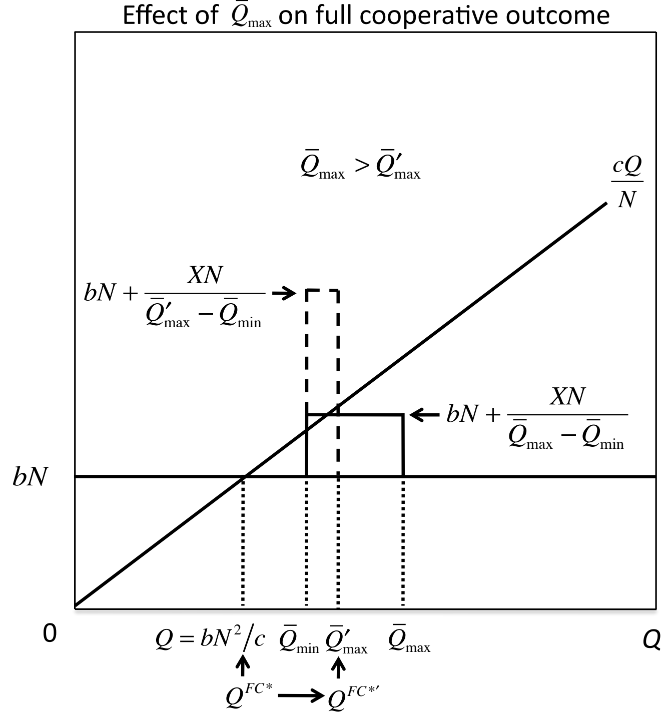

FIGURE 11c

Effect of $\bar{Q}_{\min}$ on full cooperative outcome

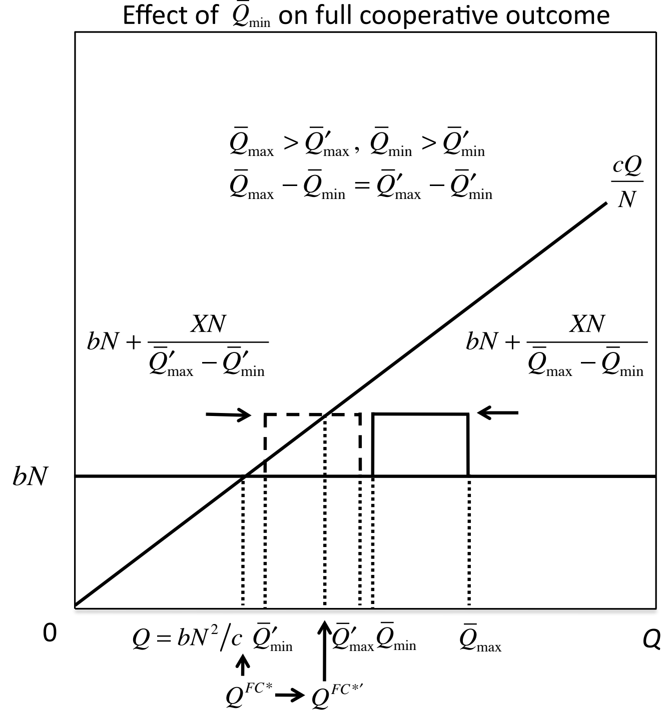

FIGURE 11d

Effect of b on full cooperative outcome

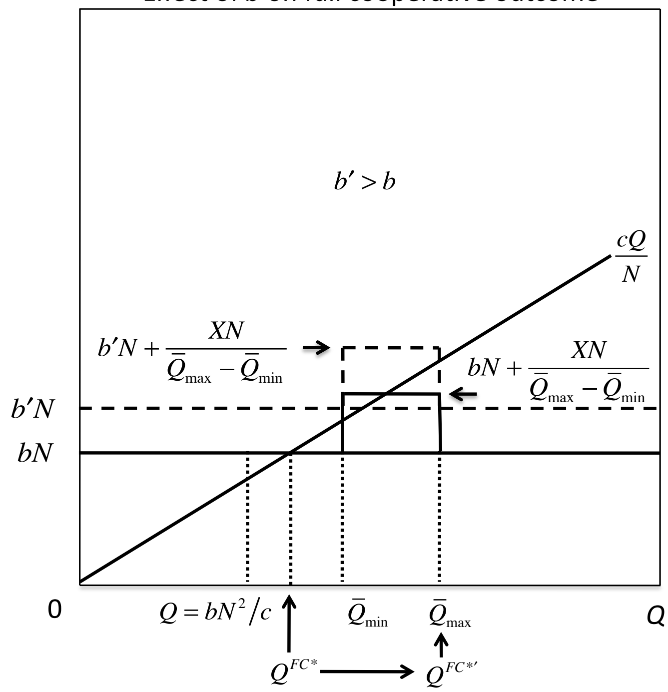

FIGURE 11e

Effect of c on full cooperative outcome

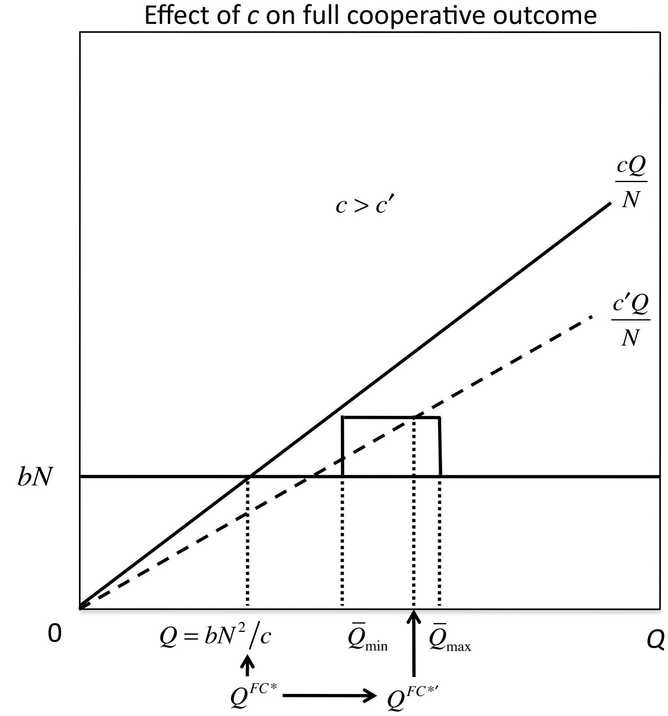

FIGURE 12a

The non-cooperative and full cooperative outcomes

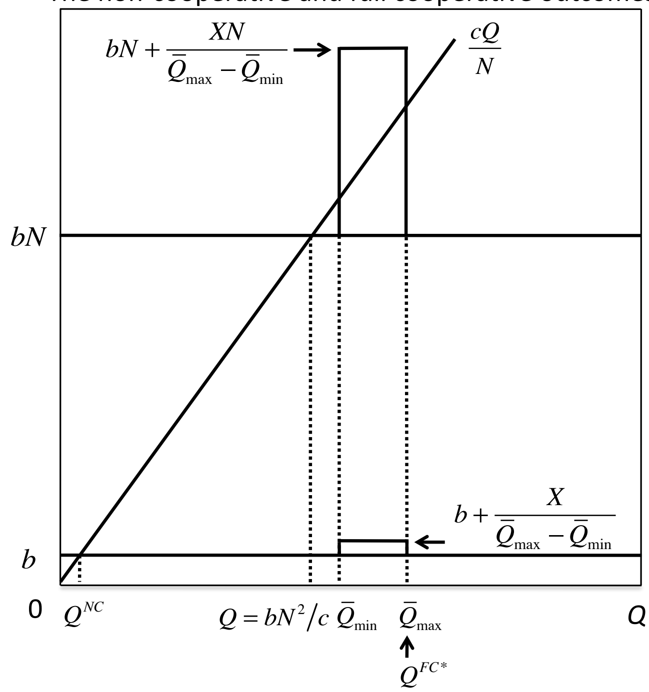

FIGURE 12b

The non-cooperative and full cooperative outcomes

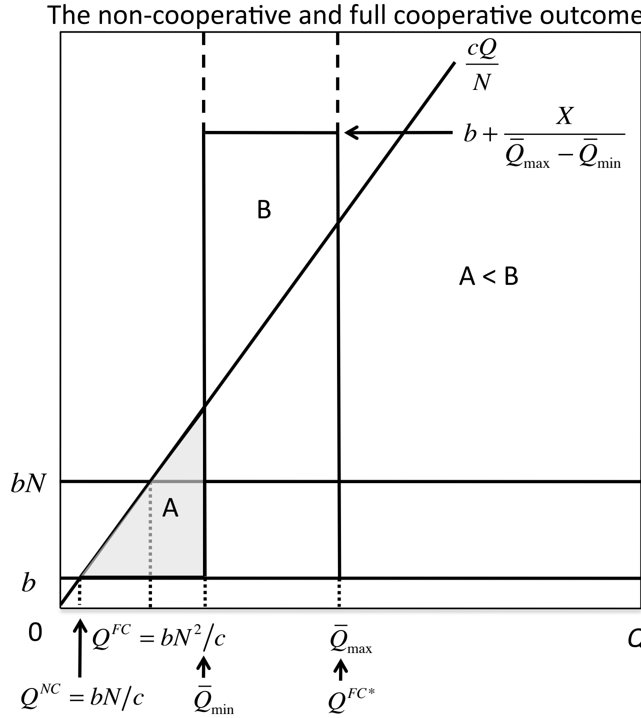

FIGURE 13
When coordination suffices and cooperation is needed