



Working Papers

www.cesifo.org/wp

On the Sensitivity of Collective Action to Uncertainty about Climate Tipping Points

Scott Barrett
Astrid Dannenberg

CESIFO WORKING PAPER NO. 4643
CATEGORY 10: ENERGY AND CLIMATE ECONOMICS
FEBRUARY 2014

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

On the Sensitivity of Collective Action to Uncertainty about Climate Tipping Points

Abstract

Previous research shows that collective action to avoid a catastrophic threshold, such as a climate “tipping point,” is unaffected by uncertainty about the impact of crossing the threshold but that collective action collapses if the location of the threshold is uncertain. Theory suggests that behavior should differ dramatically either side of a dividing line for threshold uncertainty. Inside the dividing line, where uncertainty is small, collective action should succeed. Outside the dividing line, where uncertainty is large, collective action should fail. We test this prediction in the experimental lab. Our results strongly support the prediction: behavior is highly sensitive to uncertainty around the dividing line.

JEL-Code: C720, F510, H410, H870, Q540.

Keywords: collective action, prisoners’ dilemma, coordination game, threshold uncertainty, negotiations, tipping points, climate change, experimental economics.

Scott Barrett
Earth Institute and School of International
and Public Affairs
Columbia University
USA – New York, NY 10027
sb3116@columbia.edu

Astrid Dannenberg
Earth Institute
Columbia University
USA – New York, NY 10027
ad2901@columbia.edu

On the Sensitivity of Collective Action to Uncertainty about Climate Tipping Points

Scott Barrett^{1*} & Astrid Dannenberg²

*¹Earth Institute and School of International and Public Affairs, Columbia University, New York, NY 10027; Princeton Institute for International and Regional Studies, Princeton University, Princeton, NJ 08544. *Email: sb3116@columbia.edu.*

²Earth Institute, Columbia University, New York, NY 10027; University of Gothenburg, 405 30 Gothenburg, Sweden.

Despite more than two decades of diplomatic effort, concentrations of greenhouse gases continue to trend upwards, creating the risk that we may someday cross a threshold for “dangerous” climate change¹⁻³. While climate thresholds are very uncertain, new research is trying to devise “early warning signals” of an approaching tipping point⁴⁻¹¹. This research offers a tantalizing promise: while collective action fails when threshold uncertainty is large, reductions in this uncertainty may bring about the behavioral change needed to avert a climate “catastrophe”⁵. Here we present the results of an experiment, rooted in a game-theoretic model, showing that behavior differs dramatically either side of a dividing line for threshold uncertainty. On one side of the dividing line, where threshold uncertainty is relatively large, free riding proves irresistible and trust illusive, making it virtually inevitable that the tipping point will be crossed. On the other side, where threshold uncertainty is small, the incentive to coordinate is strong and trust more robust, often leading the players to avoid crossing the tipping point. Our results show that uncertainty must be

reduced to this “good” side of the dividing line in order to stimulate the behavioral shift needed to avoid ”dangerous” climate change.

Our approach can be applied to a variety of situations, from the collapse of a fishery to sudden transitions of ecological^{6,12-15} or other complex systems^{6,16}, but our focus in this paper is on averting “dangerous” climate change¹⁻¹¹, perhaps the greatest challenge for collective action ever. We begin by presenting a game-theoretic model of countries’ decisions to limit their greenhouse gas emissions and avoid crossing a critical tipping point. In our model, avoiding the tipping point is feasible and collectively optimal, but individual optimization by countries sustains this desirable outcome only when uncertainty about the location of the threshold is sufficiently small; when uncertainty is larger, our model predicts that individual behavior, motivated by self-interest, will push countries over the tipping point, resulting in “catastrophe.” Our main contribution is to test this prediction of a behavioral regime shift in the lab.

Our game-theoretic model assumes that there are N identical countries, each able to reduce emissions by up to q_{\max}^A units using technology A and by up to q_{\max}^B units using technology B . The per-unit costs of reducing emissions using these technologies are constant but different, with $c^A < c^B$. Technology A can be thought of as low-cost “ordinary abatement,” and B as a high-cost technology for removing carbon dioxide from the atmosphere¹⁷.

Let Q denote the total reduction in emissions by all countries using both technologies, and let b represent the marginal benefit to an individual country of

avoiding “gradual” climate change. Assuming $c^B > bN > c^A > b$ gives the classical prisoners’ dilemma. For these parameter values, self-interest impels each country to abate zero, whereas collectively all countries are better off when each abates q_{\max}^A units using technology A . Technology B would never be used to address “gradual” climate change in this model, but is essential for avoiding the tipping point.

Since climate thresholds can be related to cumulative emissions^{18,19}, threshold avoidance can be expressed in terms of abatement from business as usual. Denote the threshold by \bar{Q} , and assume that \bar{Q} is a random variable distributed uniformly such that the probability of avoiding the threshold is 0 for $Q < \bar{Q}_{\min}$, $(Q - \bar{Q}_{\min}) / (\bar{Q}_{\max} - \bar{Q}_{\min})$ for $Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$, and 1 for $Q > \bar{Q}_{\max}$. We assume that avoidance of the threshold is technically feasible but requires using technology B in addition to A . Abatement short of \bar{Q} results in loss of value X . Theory²⁰ and experimental evidence²¹ suggest that uncertainty about the impact of crossing the threshold, X , should not affect behavior and so we assume that the value of X is certain.

We next solve two different optimization problems (see the Methods section below in addition to the *SI Methodological Details*). We first show that all countries *collectively* will want to abate \bar{Q}_{\max} so long as $X \geq (c^B - bN)(\bar{Q}_{\max} - Nq_{\max}^A) / N$. In this paper we assume that this condition is always satisfied, making the consequences of crossing the tipping point truly “catastrophic.”

We next show that if every other country abates \bar{Q}_{\max}/N , each will want to abate \bar{Q}_{\max}/N , making the avoidance of “catastrophe” a Nash equilibrium, provided $X \geq (c^B - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$. Of course, in this game zero abatement is also a Nash equilibrium. Hence, when this second condition holds, the collective action problem is for countries to *coordinate* their abatement so as to sustain the “safe” Nash equilibrium. When this condition does not hold, the players are trapped in a prisoners’ dilemma; they would rather stay on the safe side of the tipping point, but free rider incentives draw them inexorably towards the unique but “catastrophic” equilibrium.

Rearranging this second condition, we can define $\phi = X/(c^B - b)$ to be the *dividing line* for the range of threshold uncertainty. Assuming that all countries abate the same amount, our model thus predicts that countries will play $q_i^A = q_{\max}^A$, $q_i^B = \bar{Q}_{\max}/N - q_{\max}^A$ to the left of the dividing line, where $\bar{Q}_{\max} - \bar{Q}_{\min} \leq \phi$, and $q_i^A = q_i^B = 0$ to the right of the dividing line, where $\bar{Q}_{\max} - \bar{Q}_{\min} > \phi$. (Note that, in our experiment, because abatement is expressed in discrete units, the efficient outcome may not be exactly symmetric; see *SI Methodological Details*).

In the experiment, the game is played by groups of ten players. At the start of each game, every participant was given €11, distributed between Accounts A (€1) and B (€10). Contributions to the public good consisted of poker chips (abatement) purchased from these accounts. Chips purchased from Account A cost €0.10 each ($c^A = 0.1$), and there were 10 chips ($q_{\max}^A = 10$). Chips paid for out of Account B cost €1.00 each ($c^B = 1$), and again there were 10 chips ($q_{\max}^B = 10$). Participants were also given an “endowment fund” of €20, allocated to Account C. This fund could not be

used to contribute to the public good; it was included only to ensure that no player could be left out of pocket. At the conclusion of the game, each participant received a payoff equal to the amount of money left in his or her accounts, after making two adjustments: First, each subject was given €0.05 for every poker chip contributed by the group ($b = 0.05$) regardless of who had contributed the chips and from which account they had been taken. Second, each subject's payoff was reduced by €15 ($X = 15$) unless \bar{Q} or more chips were contributed by the group. In the baseline treatment, the threshold was certain and set equal to $\bar{Q} = 150$ (chips). In the other treatments, the threshold was a random variable, distributed uniformly over a range of values: *145/155*, *140/160*, *135/165*, and *100/200*. Thus all treatments share the same expected value (150) but differ in the size of the range of possible threshold values. Notice that, for the above parameter values, $\phi = 15.8$. Our model thus predicts that players should avoid “catastrophe” in the *150* and *145/155* treatments but not in the *140/160*, *135/165*, and *100/200* treatments.

In total, 500 students participated in the computerized experiment, 100 per treatment (10 groups \times 10 players/group). The games were played in stages. In the first stage, every participant proposed a contribution target for their group and pledged an amount they would contribute individually. It was common knowledge that these declarations were non-binding but would be communicated to the group. After these declarations were revealed, the participants chose their actual contributions in the second stage. All the participants were then informed about everyone's decisions and asked to complete a short questionnaire, giving a picture of their reasoning and emotions during the game. Finally, a computerized “spinning wheel” was activated to determine the actual value for the threshold.

The experimental results strongly support the hypotheses arising from our theoretical model (Figure 1). In the *150* treatment, “catastrophe” is avoided eight out of 10 times. In *145/155*, “catastrophe” is avoided four out of 10 times with probability 100% and in the other six cases with probability of at least 30%. The difference between *150* and *145/155* is statistically insignificant (Mann-Whitney-Wilcoxon (MWW) test, $n = 20$, $p = .23$). By contrast, in *140/160*, “catastrophe” is *never* avoided, despite the closeness of this uncertainty range to *145/155*. “Catastrophe” also occurs every time in the *135/165* treatment. In *100/200*, one group out of 10 managed to reduce the probability of “catastrophe” seven percent. Again, the differences among *140/160*, *135/165*, and *100/200* are insignificant ($n = 20$, $p > .05$ each). But the differences between *150* and *145/155* on the one hand and *140/160*, *135/165*, and *100/200* on the other hand are all highly significant ($n = 20$, $p < .01$ each). Consistent with the theory, there is a qualitative change in behavior either side of the dividing line.

As predicted, contributions in the treatments *150* and *145/155* do not differ significantly from the full cooperative levels of 150 and 155, respectively (see Table 1 and Figure 2; *T*-test, $n = 10$, $p = .72$ for *150* and $p = .25$ for *145/155*). Moreover, the average contribution in *145/155* is higher than in *150*, though the difference lacks statistical significance (MWW test, $n = 20$, $p = .85$).

Consistent with our theoretical model, contributions in the treatments *140/160*, *135/165*, and *100/200* are significantly different from their full cooperative levels—160, 165, and 200, respectively (*T*-test, $n = 10$, $p = .00$ each). However, contributions exceed the predicted value of zero (one-sided *T*-test, $n = 10$, $p = .00$ each). We should

not be surprised. To the right of the dividing line, the players face a prisoners' dilemma, and it is a common finding in the experimental literature that groups playing this game contribute somewhere between the full cooperative and non-cooperative levels²² (*SI Literature*). In our experiment, the tendency to contribute above the non-cooperative level is especially strong because contributions from Account A are very cheap.

To the left of the dividing line, average proposals and pledges closely track their full cooperative levels; to the right, they fall relative to full cooperation as the uncertainty range increases (Table 1 and Figure 2). Our ex post questionnaire (*SI Empirical Analysis*) reveals that, to the left of the dividing line, proposals are mainly chosen to maximize the collective payoff of the group, pledges to signal intended contributions. To the right of the dividing line, in contrast, proposals and pledges are mainly chosen to stimulate contributions by others.

Figure 3 reveals the effect of uncertainty on individual pledges and contributions. To the left of the dividing line, contributions are tightly bunched near the full cooperative levels. To the right, contributions move progressively closer to zero and 10, the cheap-chips level, as the range of uncertainty increases. When the uncertainty range reaches *100/200*, the full cooperative level has lost its attraction, with the majority of players contributing zero or 10, and with only 2 players contributing 20. To the left of the dividing line, the players had incentives to be trustworthy and trusting, and most players contributed at least as much as they pledged (98% in *150* and 80% in *145/155*). To the right of the dividing line, the incentives were different, and far fewer

players contributed the amounts they pledged (55% in *140/160*, 46% in *135/165*, and only 18% in *100/200*).

The contribution of poker chips in our experiment is best thought of as a metaphor for the things countries need to do to prevent “dangerous” climate change. Rockström *et al.*, for example, identify an atmospheric CO₂ concentration level of 350 ppmv as “safe,” based on paleoclimatic evidence suggesting that the polar ice sheets “tipped” previously somewhere between 350 and 550 ppmv²³. Assuming that these values represent the range of the distribution for a critical threshold, our model can be interpreted, roughly, as suggesting that countries would do no better collectively than to limit concentrations to 350 ppmv, provided our first condition ($X \geq (c^B - bN)(\bar{Q}_{\max} - Nq_{\max}^A)/N$) held, but that they could be expected to behave so as to allow concentrations to top 550 ppmv unless our second condition ($X \geq (c^B - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$) also held. To give our analysis empirical real-world relevance, more sophisticated versions of these relations could be developed and estimated, but this will require further research²⁴.

Improved climate models²⁵⁻²⁷ and early warning signals are valuable, not least as an aid to adaptation. By reducing threshold uncertainty, they may also stimulate emergency measures to limit emissions. However, early warning signals may fail completely^{8,28}; or, being prone to false positives and false negatives⁹, may reduce uncertainty by too little to prevent a critical threshold from being breached. Making matters worse, early warning signals arrive late. Even if early warning changed the

incentives to act, the *ability* to act may be severely circumscribed, leaving few immediate options besides adaptation and geoengineering.

The impact of crossing a tipping point can be interpreted as a punishment imposed by Mother Nature. When uncertainty about a tipping point is small, the fear of crossing it serves as an effective deterrent. When threshold uncertainty is large, however, this punishment fails as a deterrent, and *strategic* enforcement mechanisms are needed to deter players from straying into the climate “danger zone”^{29,30}. A good example of strategic enforcement is the use of trade restrictions in the Montreal Protocol, which has successfully protected the ozone layer. Similar mechanisms, our research implies, are needed to avoid “dangerous” climate change.

Methods

The two key conditions of our theoretical model result from two different optimization exercises. If countries cooperate fully, they will choose their individual abatement levels so as to maximize their expected aggregate payoff, $E(\Pi) = bQN - \sum_i c^A q_i^A - \sum_i c^B q_i^B - XN \left[1 - (Q - \bar{Q}_{\min}) / (\bar{Q}_{\max} - \bar{Q}_{\min}) \right]$ for

$Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$. Under our assumptions, it will pay all countries collectively to abate \bar{Q}_{\max} in the aggregate, just, so long as $XN \geq (c^B - bN)(\bar{Q}_{\max} - Nq_{\max}^A)$ (*SI Methodological Details*). We assume that this condition is always satisfied. If countries choose their abatement levels independently, every country i will maximize its own expected payoff, $E(\pi_i) = bQ - c^A q_i^A - c^B q_i^B - X \left[1 - (Q - \bar{Q}_{\min}) / (\bar{Q}_{\max} - \bar{Q}_{\min}) \right]$ for $Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$, taking the abatement levels of other countries as given. Zero abatement is always a Nash equilibrium, but per-country abatement in the amount

\bar{Q}_{\max}/N is also a Nash equilibrium when $X \geq (c^B - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$ (*SI Methodological Details*). This last condition is satisfied in two of our experimental treatments (150 and 145/155); it is not satisfied for the three remaining treatments (140/160, 135/165, and 100/200).

The experimental sessions were conducted in a computer laboratory at the University of Magdeburg, Germany, using students recruited from the general student population. In each session, subjects were seated randomly at linked computers. A set of written instructions including several numerical examples and control questions was handed out. The control questions tested subjects' understanding of the game to ensure that they were aware of the available strategies and the implications of making different choices. At the beginning of each session, subjects were assigned randomly to 10-person groups and played five practice rounds, with the membership of groups changing after each round. After a final reshuffling of members, each group played the game for real. Note that there is no significant correlation between the average contributions made in the practice rounds and the contributions made in the real round (in every case, the p values were insignificant at the 10% level). To ensure anonymity, each member of a group was identified by a different letter (A to J). At the end of each session, after the actual threshold value was determined by the "spinning wheel," students were paid their earnings in cash (for more details see the *SI Methodological Details*).

References:

1. Alley, R. B. *et al.* Abrupt climate change. *Science* **299**, 2005-2010 (2003).

2. Lenton, T. M. *et al.* Tipping elements in the earth's climate system. *Proc. Natl. Acad. Sci. USA* **105**, 1786-1793 (2008).
3. Kriegler, E., Hall, J. W., Held, H., Dawson, R. & Schellnhuber, H. J. Imprecise probability assessment of tipping points in the climate system. *Proc. Natl. Acad. Sci. USA* **106**, 5041-5046 (2009).
4. Dakos, V. *et al.* Slowing down as an early warning signal for abrupt climate change. *Proc. Natl. Acad. Sci. USA* **105**, 14308-14312 (2008).
5. Biggs, R., Carpenter, S. R. & Brock, W. A. Turning back from the brink: detecting an impending regime shift in time to avert it. *Proc. Natl. Acad. Sci. USA* **106**, 826-831 (2009).
6. Scheffer, M. *Critical Transitions in Nature and Society* (Princeton Univ. Press, Princeton NJ, 2009).
7. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53-59 (2009).
8. Ditlevsen, P. D. & Johnsen, S. J. Tipping points: early warning and wishful thinking. *Geophys. Res. Lett.* **37**, L19703 (2010).
9. Lenton, T. M. Early warning of climate tipping points. *Nature Clim. Change* **1**, 201-209 (2011).
10. Scheffer, M. *et al.* Anticipating Critical Transitions. *Science* **338**, 344-348 (2012).
11. Lenton, T. M., Livina, V. N., Dakos, V., van Nes, E. H. & Scheffer, M. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. *Phil. Trans. R. Soc. A.* **370**, 1185-1204 (2012).
12. Carpenter, S. R. *et al.* Early warnings of regime shifts: a whole-ecosystem experiment. *Science* **332**, 1079-1082 (2011).

13. Wang, R. *et al.* Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature* **492**, 419-422 (2012).
14. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591-596 (2001).
15. Dai, L., Vorselen, D., Korolev, K. S. & Gore, J. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* **336**, 1175-1177 (2012).
16. May, R. M., Levin, S. A. & Sugihara, G. Complex systems: ecology for bankers. *Nature* **451**, 893-895 (2008).
17. Lackner, K. S. *et al.* The urgency of the development of CO₂ capture from ambient air. *Proc. Natl. Acad. Sci. USA* **109**, 13156-13162 (2012).
18. Allen M. R. *et al.* Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163-1166 (2009).
19. Zickfeld, K., Eby, M., Matthews, H. D. & Weaver, A. J. Setting cumulative emissions targets to reduce the risk of dangerous climate change. *Proc. Natl. Acad. Sci. USA* **106**, 16129-16134 (2009).
20. Barrett, S. Climate treaties and approaching catastrophes. *J. Environ. Econ. Manage.* <http://dx.doi.org/10.1016/j.jeem.2012.12.004> (2013).
21. Barrett, S. & Dannenberg, A. Climate negotiations under scientific uncertainty. *Proc. Natl. Acad. Sci. USA* **109**, 17372-17376 (2012).
22. Ledyard, J. O. in *Handbook of Experimental Economics* (eds Kagel, J. H. & Roth, A. E.) 111-194 (Princeton Univ. Press, Princeton, 1995).
23. Rockström, J. *et al.* A Safe Operating Safe for Humanity. *Nature* **461**: 472-475 (2009).

24. Lenton, T.M. & Ciscar, J.-C. Integrating tipping points into climate impact assessments. *Clim. Change* **117**, 585-597.
25. Robinson, A., Calov, R. & Ganopolski, A. Multistability and critical thresholds of the Greenland ice sheet. *Nature Clim. Change* **2**, 429-432 (2012).
26. Hawkins, E. *et al.* Bistability of the Atlantic overturning circulation in a global climate model and links to ocean freshwater transport. *Geophys. Res. Lett.* **38**, L10605 (2011).
27. Drijfhout, S.S., Weber, S.L., & van der Swaluw, E. The stability of the MOC as diagnosed from model projections for pre-industrial, present and future climates. *Clim. Dynam.* doi:10.1007/s00382-010-930-z (2010).
28. Hastings, A. & Wysham, D. B. Regime shifts in ecological systems can occur with no warning. *Ecol. Lett.* **13**, 464-472 (2010).
29. Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge Univ. Press, New York, 1990).
30. Barrett, S. *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (Oxford Univ. Press, Oxford, 2003).

Correspondence and requests for materials should be addressed to S.B (email: sb3116@columbia.edu).

Acknowledgments: We thank James Rising for programming our “spinning wheel,” the MaXLab team at Magdeburg University for use of their lab, and Princeton University’s Communicating Uncertainty group for funding our experiments.

Author contributions: SB and AD contributed equally to this work. They both designed and performed the research and analysed the data and wrote the paper.

Additional information:

The authors declare no competing financial interests. Supplementary information accompanies this paper on www.nature.com/natureclimatechange. Reprints and permissions information are available online at <http://www.nature.com/reprints>.

Figure legends:

Figure 1 Probability of “catastrophe” by treatment. In *150*, “catastrophe” is avoided eight out of 10 times. In *145/155*, “catastrophe” is avoided four out of 10 times with probability 100% and in the other six cases with probability between 30% and 80%. In *140/160* and *135/165*, “catastrophe” is never avoided. In *100/200*, “catastrophe” occurs nine out of 10 times with probability 100% and once with probability 93%.

Figure 2 Treatment means versus predicted values. Mean contributions are consistent with the predicted values to the left of the dividing line. To the right of the dividing line, mean contributions lie between the full cooperative and the predicted (non-cooperative) values. Mean proposals and mean pledges match the full cooperative values to the left of the dividing line; to the right, a wedge opens up between these values as the uncertainty range widens.

Figure 3 Individual pledges and contributions by treatment. To the left of the dividing line, pledges and contributions are tightly bunched, more so in *150* than in *145/155*. To the right of the dividing line, in *140/160*, *135/165*, and *100/200*, values vary widely, with contributions increasingly falling short of pledges with higher uncertainty. A series of Spearman’s correlation tests gives: $n = 100$, $\rho = .38$, $p = .00$ in *150*; $\rho = .59$, $p = .00$ in *145/155*; $\rho = .33$, $p = .00$ in *140/160*; $\rho = -.01$, $p = .93$ in *135/165*; $\rho = .10$, $p = .34$ in *100/200*. A small noise (2%) has been inserted to make all data points visible.

Tables:

Table 1 Summary statistics by treatment.

Treatment	Proposal		Pledge		Contribution		Group contribution	
	Mean (Std dev)	Mode (%)	Mean (Std dev)	Mode (%)	Mean (Std dev)	Mode (%)	Mean (Std dev)	Min / max
<i>150</i>	151.9 (1.57)	150 (83%)	14.7 (0.51)	15 (74%)	15.1 (0.77)	15 (56%)	150.9 (7.69)	136 / 159
<i>145/155</i>	158.0 (1.40)	160 (48%)	15.1 (0.62)	16 (53%)	15.4 (0.38)	16 (45%)	153.5 (3.84)	148 / 160
<i>140/160</i>	161.0 (2.64)	160 (69%)	15.1 (0.83)	16 (64%)	11.7 (1.69)	16 (33%)	117.4 (16.85)	80 / 139
<i>135/165</i>	163.3 (8.75)	170 (41%)	15.4 (1.10)	17 (36%)	11.3 (1.98)	10 (33%)	112.9 (19.84)	68 / 130
<i>100/200</i>	166.3 (9.85)	200 (29%)	15.8 (1.69)	20 (32%)	7.7 (1.67)	10 (36%)	77.2 (16.67)	55 / 107

To the left of the dividing line for threshold uncertainty (the area shaded in blue), actual contributions closely follow the proposals and pledges. To the right of the dividing line (the area shaded in red), contributions fall short of the proposals and pledges. Here, mean and especially modal proposals and pledges increase with the uncertainty range (that is, with the full cooperative contribution level), while mean and modal contributions decrease. The treatments *150* and *100/200* are taken from ref 21.

Figure 2

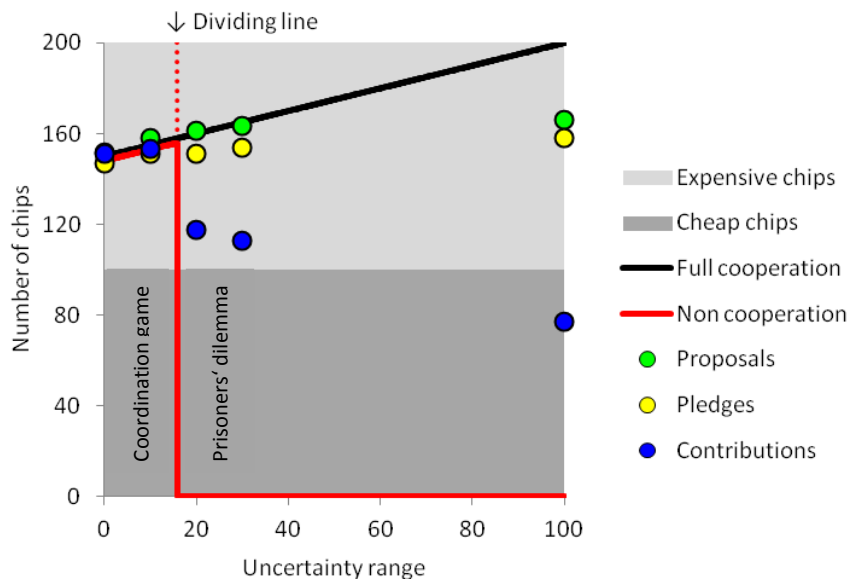
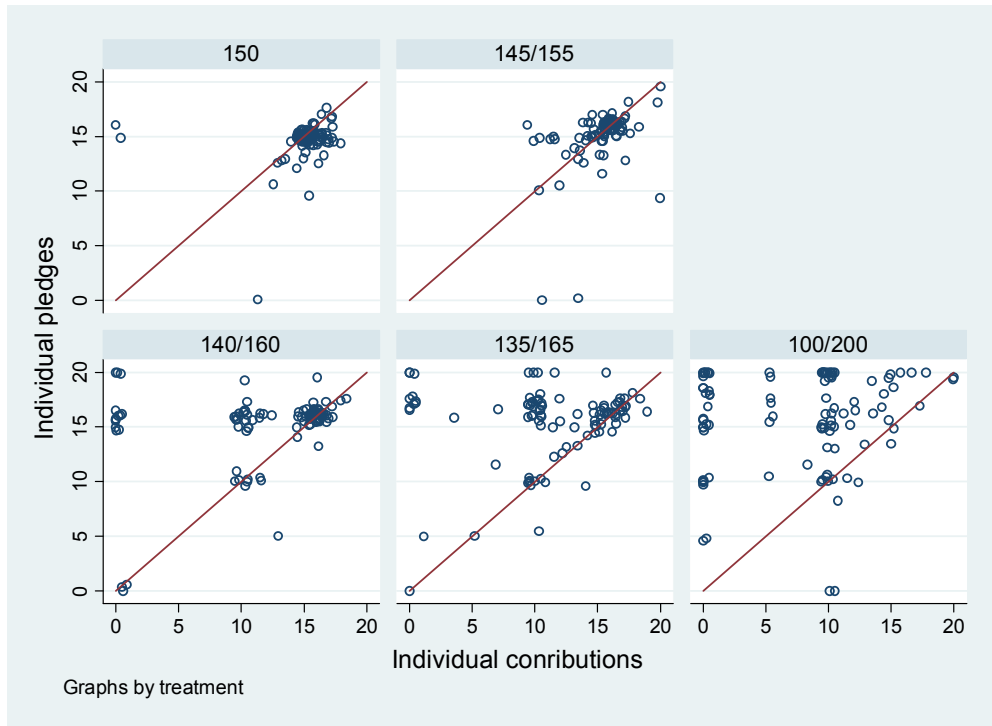


Figure 3



Supplementary Information for

On the Sensitivity of Collective Action to Uncertainty about Climate Tipping Points

Scott Barrett & Astrid Dannenberg

This file contains Supplementary Text, Supplementary Tables S1-S6, Supplementary Figures S1-S4, and additional references.

1. Supplementary Methodological Details

Derivations of main results from theoretical model

In this section we provide details behind the key theoretical results described in the paper. Our analysis here assumes that abatement levels (contributions) are continuous, but the results do not change in a significant way when these values must be integers, as in our experiment (see, for example, the caption to Figure S1). All the variables and parameters used in this model are described in Table S1.

Lemma 1. *Full cooperation requires eliminating the chance of catastrophe if and only if $XN \geq (c^B - bN)(\bar{Q}_{\max} - Nq_{\max}^A)$.*

Proof. Full cooperation requires that countries maximize their expected collective payoff

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

subject to $q_i^A \in [0, q_{\max}^A]$ and $q_i^B \in [0, q_{\max}^B] \forall i$. By assumption, $bN > c^A$; and so maximization of $E(\Pi)$ requires $q_i^A = q_{\max}^A$ for all i irrespective of the value of Q . Also by assumption, $c^B > bN$, so that full cooperation requires setting $q_i^B = 0$ for all i unless this abatement can reduce the probability of catastrophe—that is, unless $Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$. Upon maximizing the middle expression in eq. (1) by choice of q_i^B for $Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$, a necessary condition for setting $q_i^B > 0$ is $XN \geq (c^B - bN)(\bar{Q}_{\max} - \bar{Q}_{\min})$. Given the linear nature of this model, it will pay to set $q_i^B = \bar{Q}_{\max}/N - q_{\max}^A$ if this weak inequality holds (and abatement is assumed symmetric) and $q_i^B = 0$ if the inequality does not hold. The expected collective payoffs corresponding to these choices are

$$E(\Pi; q_i^A = q_{\max}^A, q_i^B = 0) = bq_{\max}^A N^2 - c^A q_{\max}^A N - XN$$

$$E(\Pi; q_i^A = q_{\max}^A, q_i^B = \bar{Q}_{\max}/N - q_{\max}^A) = b\bar{Q}_{\max} N - c^A q_{\max}^A N - c^B (\bar{Q}_{\max}/N - q_{\max}^A) N$$

The latter payoff is at least as great as the former (that is, full cooperation requires eliminating the chance of catastrophe) if and only if

$$XN \geq (c^B - bN)(\bar{Q}_{\max} - q_{\max}^A N). \quad (2)$$

As noted before, we also require $XN \geq (c^B - bN)(\bar{Q}_{\max} - \bar{Q}_{\min})$. By assumption, $\bar{Q}_{\min} \geq Nq_{\max}^A$. Hence, this second condition will hold so long as (2) holds, making inequality (2) a necessary and sufficient condition for full cooperation to require eliminating the chance of catastrophe. ■

Lemma 2. *Play $q_i^A = 0$, $q_i^B = 0$ for all i is a Nash equilibrium.*

Proof. Taking the abatement of every other country as given, each country i is assumed to maximize its expected payoff

$$E(\pi_i) = \begin{cases} bQ - c^A q_i^A - c^B q_i^B - X & \text{for } Q < \bar{Q}_{\min} \\ bQ - c^A q_i^A - c^B q_i^B - X \left[1 - \frac{(Q - \bar{Q}_{\min})}{(\bar{Q}_{\max} - \bar{Q}_{\min})} \right] & \text{for } Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}] \\ bQ - c^A q_i^A - c^B q_i^B & \text{for } Q > \bar{Q}_{\max} \end{cases} \quad (3)$$

subject to $q_i^A \in [0, q_{\max}^A]$, $q_i^B \in [0, q_{\max}^B]$, and the abatement by countries other than i being given. In this model, only the aggregate level of abatement by other countries matters; from i 's perspective this amount is given by $Q_{-i} = \sum_{j \neq i} (q_j^A + q_j^B)$.

So, how will i choose to play? Suppose that $Q_{-i} \in [0, \bar{Q}_{\min} - q_{\max}^A - q_{\max}^B]$. Then, no matter what i does, aggregate abatement cannot exceed \bar{Q}_{\min} . Moreover, since $c^A > b$ by assumption, i won't even want to undertake any ordinary abatement to reduce gradual climate change (since $c^B > c^A$, i will also not undertake any abatement using technology B). Since countries are symmetric, this will be true for every country i , making play $q_i^A = q_i^B = 0 \forall i$ a Nash equilibrium. ■

Lemma 3. *Play $q_i^A = q_{\max}^A$, $q_i^B = \bar{Q}_{\max}/N - q_{\max}^A$ for all i is a Nash equilibrium if $X \geq (c^B - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$.*

Proof. Provided $Q_{-i} \in (\bar{Q}_{\min} - q_{\max}^A - q_{\max}^B, \bar{Q}_{\max})$, abatement by country i may be in i 's self-interest. This is because, by reducing its own emissions, i can reduce the probability of catastrophe. By assumption, $\bar{Q}_{\min} \geq Nq_{\max}^A$. Hence, in a symmetric outcome, every country must abate using technology B in addition to using A in order to reduce the probability of catastrophe. From eq. (3), for $Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$, maximization of $E(\pi_i)$ requires choosing: (1) $q_i^B > 0$ if $b + X/(\bar{Q}_{\max} - \bar{Q}_{\min}) - c^B \geq 0$ or, upon rearranging, $X \geq (c^B - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$, and (2) $q_i^A > 0$ if $X \geq (c^A - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$. By assumption, $c^B > c^A$; and so the second condition will hold so long as the first inequality holds. Since technology A is cheaper than B, if it pays i to abate one unit of emissions using B, it will also pay i to abate $q_i^A = q_{\max}^A$.

Moreover, since the first condition holds for every unit of abatement using technology B for $Q \in [\bar{Q}_{\min}, \bar{Q}_{\max}]$, in a symmetric equilibrium it will pay every country i to abate $q_i^B = \bar{Q}_{\max}/N - q_{\max}^A$ so that the probability of catastrophe is reduced to zero. In summary, provided $X \geq (c^B - b)(\bar{Q}_{\max} - \bar{Q}_{\min})$, there exists a Nash equilibrium in which it pays every country i to abate $q_i^A = q_{\max}^A$ and $q_i^B = \bar{Q}_{\max}/N - q_{\max}^A$. ■

Our main result is easily proved using these three lemmas:

Proposition. *Let $\phi = X/(c^B - b)$ represent the dividing line for the uncertainty range $\bar{Q}_{\max} - \bar{Q}_{\min}$, and assume that the condition specified in Lemma 1 holds so that it is collectively optimal for every player to play $q_i^A = q_{\max}^A$ and $q_i^B = \bar{Q}_{\max}/N - q_{\max}^A$. Then, for $\bar{Q}_{\max} - \bar{Q}_{\min} \leq \phi$, sustaining the full cooperative outcome and thus avoiding catastrophe is a Nash equilibrium, achievable by coordination, whereas for $\bar{Q}_{\max} - \bar{Q}_{\min} > \phi$ the Nash equilibrium is unique and requires that every country play $q_i^A = q_i^B = 0$, making abatement a prisoners' dilemma.*

Figures S1 and S2 illustrate the situations that pertain, respectively, either side of the dividing line, $\phi = X/(c^B - b)$. Table S2 summarizes the experimental treatments and the corresponding theoretical hypotheses. In our experiment, the condition specified in Lemma 1 requires $\bar{Q}_{\max} \leq 400$, which obviously holds for all treatments.

Experimental Design and Procedure

As is conventional practice in the literature, our experiment involved students as participants (all but one of the twelve studies listed in Table S3 used students). The experimental sessions were held in a computer lab at the University of Magdeburg, Germany, with 500 students recruited from the general student population (using the recruiting software Orsee; S1). The five treatments were run in a between-subjects design, meaning that each student took part in one treatment only. One hundred subjects were randomly assigned to each treatment. In each session, 20 subjects were seated randomly at linked computers (using the game software Ztree; S2). Before the subjects got to know the game, they were invited to play and become familiar with the computerized “spinning wheel” (see Figure S3). Probabilities must be communicated with care (S3), which is why we used this novel way of demonstrating a uniform probability distribution; it facilitated the participants’ understanding of the risk and its implications for the game. Written instructions, including several numerical examples and control questions, were then handed out. The instructions assumed a neutral frame for the context and language of the experiment in order to avoid any potential bias (there was no mention of “climate change”).

Experimental instructions

(The instructions below are for the 150 treatment, translated from German.)

Welcome to our experiment!

1. *General information*

In our experiment you can earn money. How much you earn depends on the gameplay, or more precisely on the decisions you and your fellow co-players make. You will receive €20 for your participation, but note that a loss during the experiment will be deducted from that amount while a gain will be added. For a successful run of this experiment, it is essential that you not talk to other participants. Now, read the following rules of the game carefully. If you have any questions, please give us a hand signal. It is important that you read up to the STOP sign only. Please wait when you get there, as we will give you a brief oral presentation before we continue.

2. *Game rules*

There are 10 symmetric players in the game, meaning you and 9 other players. Each player is faced with the same decision problem. All decisions in the experiment are anonymous. For the purpose of anonymity, you will be identified by a letter (between A and J), which you will see in the lower left corner of your display.

At the beginning of the game, you will receive 20 poker chips, which are credited to two personal accounts, Account A and Account B. During the experiment, you can use the poker chips in your accounts to contribute to a joint project or you can leave these chips untouched. Chips from Account A are cheap; they cost €0.10 each. But you can contribute no more than 10 chips from Account A. Chips from Account B are more expensive; they cost €1.00 each. You can contribute at most 10 chips from Account B. So, overall you can contribute any integer amount of chips between 0 and 20 to the joint project: 10 chips from Account A and 10 chips from Account B.

At the end of the game, the amount of chips you have left in Accounts A and B will be paid to you in cash: €0.10 for each chip in Account A and €1.00 for each chip in Account B. There are two further adjustments: First, you will get €0.05 for every poker chip contributed to the joint project, irrespective of who contributed the chip and whether the chip was purchased from Account A or B. Second, if the group as a whole contributes fewer than 150 poker chips to the joint project, every player will lose €15. If the group contributes 150 or more poker chips to the joint project, no player will lose any money.

Before you and the other players decide how many chips to contribute, everyone will be given an opportunity to make two non-binding announcements: First, each player will make a proposal for how many chips the group as a whole should contribute to the joint project. Second, each player will make a pledge for how many chips he or she intends to contribute to the joint project. All the proposals and pledges made by the players will be displayed before you and the others decide how much to contribute.

The game will be played and paid out only once. You should think carefully about how to decide in the game. Before playing “for real,” 5 trial rounds will be played so that you and the other players can become familiar with the game. The people you will play with in the 5 trial rounds and the real round will always change, so that you will never play more than one round with the same group of people.

3. *Example*

Here, you can see a (hypothetical) example of the decisions made by the 10 players.

[Screenshot]

The leftmost column (“Proposals”) shows every player’s proposal for the collective contribution target. The column next to it (“Pledges”) shows each player’s pledge for his or her own contribution. The next column (“Contributions”) displays each player’s actual contribution. The rightmost column (“Payoffs”) shows the corresponding payoff levels. When you have played the game, you will see this information displayed.

[STOP sign]

Please wait for the oral description of the game.

4. Control questions

Please answer the following control questions.

a. Take a look at the hypothetical example in part 3 above. Are the collective contributions of the group as a whole sufficient to avoid the loss?

Yes No

b. How many chips does each player have to contribute on average, if the group were to contribute 150 chips in total?

0 5 10 15 20

c. Assume that the group as a whole (including yourself) has contributed 0 chips to the joint account. What would be your payoff (excluding the participation fee)?

–€4.00 €0.00 €10.00 €12.50 €17.50

d. Assume that the group as a whole (including yourself) has contributed 100 chips to the joint account. Assume also that you have contributed 10 chips from Account A and 0 chips from Account B. What would be your payoff (excluding the participation fee)?

–€4.00 €0.00 €10.00 €12.50 €17.50

e. Assume that the group as a whole (including yourself) has contributed 150 chips to the joint account. Assume also that you have contributed 10 chips from Account A and 5 chips from Account B. What would be your payoff (excluding the participation fee)?

–€4.00 €0.00 €10.00 €12.50 €17.50

f. Assume that the group as a whole (including yourself) has contributed 150 chips to the joint account. Assume also that you have contributed 10 chips from Account A and 0 chips from Account B. What would be your payoff (excluding the participation fee)?

–€4.00 €0.00 €10.00 €12.50 €17.50

g. Assume that the group as a whole (including yourself) has contributed 200 chips to the joint account. Assume also that you have contributed 10 chips

from Account A and 10 chips from Account B. What would be your payoff (excluding the participation fee)?

- €4.00 €0.00 €10.00 €12.50 €17.50

Please use the pocket calculator to calculate other examples! Give us a hand signal after you have answered all the control questions. We will come to you and check that you have answered all the questions. The game will begin after we have checked the answers of all the players and answered any questions you may have. Good luck!

2. Supplementary Literature

Public goods games have been studied extensively in experimental settings. Many researchers have employed the linear public goods game with a unique Pareto inefficient equilibrium to study human cooperation. The results show that a full range of behavior exists; some contribute everything (the Pareto-efficient amount), some contribute zero (the free riding amount), and some choose a middle course. Average contributions are typically halfway between the Pareto-efficient level and the free riding level and, when the game is repeated finitely, decline steadily with repetition (for reviews, see S4 and S5). In threshold public goods games, the public good is provided only if the sum of contributions reaches a predetermined threshold, and there exist two sets of Nash equilibria (in pure strategies), one of which is efficient. The players' task in these games is to correlate their action (coordination) rather than to suppress free riding incentives (cooperation). The provision of these public goods often but not always succeeds. Increases in thresholds increase contributions but also increase the likelihood that the threshold will not be reached (for reviews, see S4 and S6).

Threshold public goods experiments with varying uncertainty about the threshold have been conducted in the context of discrete public goods and common pool resources; for an overview see Table S3 (and S7-S18). In the latter literature, players are allowed to claim any amount of a collectively owned resource, but they are unaware of the precise resource size and receive a payoff of zero if the total quantity claimed exceeds the resource stock. A key result of this literature is that increasing resource uncertainty causes players to overestimate the resource size and, as a consequence, request more. They also expect their co-players to request more. These studies generally refer to cognitive biases in the perception of uncertainty to explain the results, e.g. an erroneous belief that the variability and the central tendency of a probability distribution are positively correlated (S16). Other explanations include an optimism bias (S15) and an egoism bias (S7). As can be seen in Table S3, the typical experimental protocol involves five players in a group. The games are played over several rounds and every participant plays all uncertainty ranges (within-subjects test). The players are not allowed to communicate with each other and typically they don't receive information about others' behavior or the true threshold until the end of the experiment. A closer look at the literature reveals that increasing uncertainty about the threshold decreases contributions to the public good (or increases requests of the common pool resource) only if the increase in the uncertainty range is very large. The effects are insignificant when the increase in the uncertainty range is moderate.

Our paper departs from this literature in a number of ways. First, we study threshold uncertainty in the context of climate change; contributions in our model alleviate

“gradual” climate change and “abrupt and catastrophic” climate change; players had two ways of contributing, rather than one, each achieved at a different marginal cost. Second, our games were played by 10 players—significantly more than in previous studies (which refrained from deception). Third, we used a between-subjects design, i.e. every participant played one uncertainty range only. Together with the use of the “spinning wheel” to communicate the uncertainty, we think this design facilitates the participants’ understanding of the games. Finally, we placed great emphasis on communication—a possibility ignored by all previous studies. Specifically, we allow players not only to pledge their own intended contribution, but also to propose a target for the group contribution. The experimental literature on communication has shown that this form of restricted and anonymous communication among participants can improve coordination but works much less reliably for cooperation (for reviews see S6 and S19-S21, see also S22-S23).

3. Supplementary Empirical Analyses

Table S4 shows the significance of differences in proposals, pledges, and contributions between treatments. As predicted, the average contribution for 145/155 is higher than for 150 but, given the tightness of this uncertainty range, the difference lacks statistical significance. Note that the results for 145/155 are all the more remarkable considering that the players in this treatment lack an obvious focal point (S24). For 150, it is fairly obvious that each player should contribute 15. For 145/155, however, there does not exist an efficient symmetric equilibrium (given that contributions must be in integer values). To support the full cooperative outcome, some players must contribute more than others, and determining which individuals should contribute more and which less requires extraordinary coordination.

Consistent with theoretical predictions, contributions in 150 and 145/155 do differ significantly from the contributions made in each treatment to the right of the dividing line, 140/160, 135/165, and 100/200. Table S4 also shows that contributions made under large uncertainty (100/200) differ significantly from those made under intermediate uncertainty (140/160 and 135/165). Although this observation is not predicted by the theory, it confirms a finding from previous experiments on threshold uncertainty (see Table S3).

A similar picture evolves for the variances in contributions. There are no significant differences in standard deviations within the coordination zone to the left of the dividing line (between 150 and 145/155) or within the prisoners’ dilemma zone to the right of the dividing line (between 140/160, 135/165, and 100/200). But comparing the treatments on either side of the dividing line reveals almost always statistically significant differences.

In a symmetric coordination game, with communication, the players all know what to do; above all, they want to reach the efficient equilibrium and avoid catastrophe. Each treatment in this territory thus has a single attractor: the full cooperative contribution level. This is why aggregate contributions in these games are relatively close to the full cooperative level and show little variation. In a prisoners’ dilemma, by contrast, the players are ambivalent. Their collective interest makes them want to contribute so as to avoid catastrophe; but their self-interest makes them want to lower their contributions to save costs. In addition to the full cooperative contribution level, there are two more attractors in this game: zero, the Nash contribution level, and 10, the

cheap-chips level. The multiplicity of attractors leads to a much larger variation in aggregate contributions. As the uncertainty range increases, the full cooperative level loses its attraction; more and more people contribute either zero or 10. The percentages of people making these contributions increase from 38% in 140/160 to 44% in 136/165 and 66% in 100/200.

Figure S4 shows to what extent players' contributions deviated from their pledges depending on what the group as a whole had pledged. The vertical axis shows the aggregate pledges (the sum of the individual pledges) and the horizontal axis shows the gap between players' individual pledges and their actual contributions. In the 150 treatment, the aggregate pledges cluster around 150, the obvious focal point. For aggregate pledges equal or greater than 150, the gap between individual pledges and contributions is very small; players had no reason to deviate from their pledge. For aggregate pledges below 150, the gap is negative, indicating that players increased their contributions when they noticed that their pledges would not suffice to avoid catastrophe. The lower the aggregate pledge the larger is the gap. There are only two players who gave less than they pledged. Things are a little more complicated in 145/155. There are two clusters for the aggregate pledges; one around 155 and another one just below 150. For aggregate pledges around 155, the gap between individual pledges and contributions is relatively small. Though, the gap is not as small as in 150. When the aggregate pledge is below 150, the gap is larger. Again, the lower the aggregate pledge the larger is the gap. But this time the gap widens in both directions. Players' reactions to a low aggregate pledge varied; some increased their contributions (moved to the left of the vertical line at zero) while others decreased their contributions (moved to the right). We interpret this as saying that a sufficiently high aggregate pledge affirmed players' trust and made them want to keep their word while a low aggregate pledge eroded trust and made them unsure how to play the game.

The picture for the treatments within the prisoners' dilemma zone (140/160, 135/165, and 100/200) is very different. The deviations between individual pledges and contributions do not follow a certain pattern; players lacked the incentive to keep their word regardless of what the group as a whole had pledged. Also, with a larger uncertainty range more and more players moved to the right of the vertical line at zero, meaning that the gap between what people say and what they do became larger.

Table S5 presents participants' responses to our follow-up questionnaire. In general, whenever the questions are about general attitudes, responses vary little across treatments (questions 7, 8, and 9). The same is true for risk aversion (question 12). Subjects' risk aversion does not significantly differ between treatments (Chi-square test, $n = 200$, $p > .05$ each). There is no significant correlation between individual risk aversion and individual contributions (Spearman's correlation test, $n = 100$, $p > .05$ each) or between the number of risk averse members in a group and group contributions ($n = 10$, $p > .05$ each). Thus, we cannot reject the hypothesis that risk aversion and behavior in the game are independent. The average estimations of the threshold (question 13) are not significantly different from the expected value 150 (T-test, $n = 100$, $p > .05$ each) and they do not significantly differ between treatments (MWW test, $n = 200$, $p > .05$ each). Not surprisingly, participants in all treatments did not feel very confident about their estimations (question 14). However, the confidence level does not significantly differ between treatments (Chi-square test, $n = 200$, $p > .05$ each). This suggests that the different contribution levels we observe in the different experimental treatments are not the consequence of a misperception of risk

or different levels of confidence. There is no significant correlation between individual confidence levels and contributions in the game (Spearman's correlation test, $n = 100$, $p > .05$ each).

By contrast, responses to the questions about subjects' reasoning and emotions during the games vary considerably between treatments. Subjects playing in *150* and *145/155* were more contented with the game's outcome and less regretful than subjects playing in *140/160*, *135/165*, and *100/200* (questions 1 and 2). Fairness and trust were more important in the coordination games (questions 3, 4, and 10). For example, when asked, "Did you trust the other players to make the contributions they pledged?" 47% answered "Very much" in *145/155*, whereas only 27% gave this answer in *140/160*. The results differ on either side of the dividing line, but also reflect a trend. When asked the above question, only 10% of the people playing *100/200* answered "Very much," whereas the percentage answering "Not at all" increased from 2% in *150* to 41% in *100/200*. Similarly, proposals and pledges were found to be more helpful for coordination (questions 5 and 6). When asked, "Do you agree with the statement that the exchange of pledges was helpful?" 50% of respondents playing *145/155* responded "Very much," whereas only 26% of the respondents playing *140/160* gave this answer. There is again a clear difference either side of the dividing line. There is also a background trend. The percentage of respondents answering "Very much" declines from 68% in *150* to 10% in *100/200*. By contrast, the percentage answering "Not at all" increases from 3% in *150* to 33% in *100/200*. Interestingly, when asked if they feel betrayed by others (question 11), responses form an inverse U-shape across treatments. The share of participants answering this question with "Very much" rose from 10% in *150* to 30% in *140/160* and then declined to 16% in *100/200*, arguably because the players in the latter treatment didn't expect a better outcome. Overall, we find a general trend: greater uncertainty makes mutual trust and cooperation increasingly difficult. Around the dividing line, however, there is a qualitative and profound change in behavior and motives.

Table S6 presents subjects' responses to the open questions about their motivation for making their proposals, pledges, and contributions, respectively. The responses were classified and assigned to certain response categories. The comparison of responses between the treatments can be summarized as follows: As a motivation to propose a certain group target, joint payoff maximization became less important with increasing uncertainty while the stimulation of others' contributions and suggesting a realistic target became more important. Fairness and safety tended to be of less relevance under higher uncertainty. As a motivation for the pledges, truthful signaling of intended contributions and the creation of trust clearly became less important with increasing uncertainty while the stimulation of others' contributions became more important. Finally, for the actual contribution decision, the wish to contribute the fair share of the burden and fulfill the own pledge became less important with increasing uncertainty. The willingness to compensate for potentially missing contributions decreased. Instead, own payoff maximization, resignation and distrust became more important. Also, with increasing uncertainty, more and more people stated that they threw in the cheap chips to compromise between group interest and own interest.

4. Supplementary References

- S1. Greiner, B. in *Forschung und wissenschaftliches Rechnen* (eds. Kremer, K. & Macho, V.) 79-93 (GWDG Bericht 63, Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen 2004).
- S2. Fischbacher, U. Z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* **10**, 171-178 (2007).
- S3. Spiegelhalter, D., Pearson, M. & Short, I. Visualizing uncertainty about the future. *Science* **333**, 1393-1400 (2011).
- S4. 21. Ledyard, J. O. in *Handbook of Experimental Economics* (eds. Kagel, J. H. & Roth, A. E.) 111-194 (Princeton Univ. Press, Princeton, 1995).
- S5. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics* **14**, 47-83 (2011).
- S6. Croson, R. T. A. & Marks, M. B. Step Returns in Threshold Public Goods: A Meta- and Experimental Analysis. *Experimental Economics* **2**, 239-259 (2000).
- S7. Gustafsson, M., Biel, A. & Gärling, T. Egoism Bias in Social Dilemmas with Resource Uncertainty. *Group Processes & Intergroup Relations* **3**, 351-365 (2000).
- S8. McBride, M. Threshold uncertainty in discrete public good games: an experimental study. *Economics of Governance* **11**, 77-99 (2010).
- S9. Suleiman, R., Budescu, D. V. & Rapoport, A. Provision of Step-Level Public Goods with Uncertain Provision Threshold and Continuous Contribution. *Group Decision and Negotiation* **10**, 253-274 (2001).
- S10. Wit, A. & Wilke, H. Public good provision under environmental and social uncertainty. *European Journal of Social Psychology* **28**, 249-256 (1998).
- S11. Budescu, D. V., Rapoport, A. & Suleiman, R. Resource Dilemmas with Environmental Uncertainty and Asymmetric Players. *European Journal of Social Psychology* **20**, 475-487 (1990).
- S12. Budescu, D. V., Rapoport, A. & Suleiman, R. Common Pool Resource Dilemmas under Uncertainty: Qualitative Tests of Equilibrium Solutions. *Games and Economic Behavior* **10**, 171-201 (1995a).
- S13. Budescu, D. V., Suleiman, R. & Rapoport, A. Positional Order and Group Size Effects in Resource Dilemmas with Uncertain Resources. *Organizational Behavior and Human Decision Processes* **61**, 225-238 (1995b).
- S14. Gustafsson, M., Biel, A. & Gärling, T. Overharvesting of resources of unknown size. *Acta Psychol. (Amst.)* **103**, 47-64 (1999a).
- S15. Gustafsson, M., Biel, A. & Gärling, T. Outcome-Desirability in Resource Management Problems. *Thinking & Reasoning* **5**, 327-337 (1999b).
- S16. Rapoport, A., Budescu, D. V., Suleiman, R. & Weg E. in *Social dilemmas: Theoretical issues and research findings* (eds. Liebrand, W. B. G., Messick, D. M. & Wilke, H. A. M.) 43-57 (Pergamon Press, Oxford 1992).
- S17. Rapoport, A., Budescu, D. V. & Suleiman, R. Sequential Request from Randomly Distributed Resources. *Journal of Mathematical Psychology* **37**, 241-265 (1993).

- S18. Suleiman, R., Rapoport, A. & Budescu, D. V. Fixed Position and Property Rights in Sequential Resource Dilemmas under Uncertainty. *Acta Psychol. (Amst.)* **93**, 229-245 (1996).
- S19. Balliet, D. Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. *Journal of Conflict Resolution* **54**, 39-57 (2010).
- S20. Bicchieri, C. & Lev-On, A. Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy & Economics* **6**, 139-168 (2007).
- S21. Crawford, V. A Survey of Experiments on Communication via Cheap Talk. *Journal of Economic Theory* **78**, 286-298 (1998).
- S22. Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A. & Marotzke, J. The collective-risk social dilemma and the prevention of dangerous climate change. *Proc. Natl. Acad. Sci. USA* **105**, 2291-2294 (2008).
- S23. Tavoni, A., Dannenberg, A., Kallis, G. & Loeschel, A. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci. USA* **108**, 11825-11829 (2011).
- S24. Schelling, T. C. *The Strategy of Conflict* (Harvard Univ Press, Cambridge, 1960).
- S25. Barrett, S. & Dannenberg, A. Climate negotiations under scientific uncertainty. *Proc. Natl. Acad. Sci. USA* **109**, 17372-17376 (2012).

Table S1. Description of parameters and variables

Description	Symbol	Value
Number of players	N	10
Marginal benefit of reducing gradual climate change	b	0.05
Marginal cost of abatement using technology A	c^A	0.10
Marginal cost of abatement using technology B	c^B	1.00
Maximum abatement using technology A	q_{\max}^A	10
Maximum abatement using technology B	q_{\max}^B	10
Cost of catastrophe	X	15
Lower threshold bound	\bar{Q}_{\min}	Varies by treatment*
Upper threshold bound	\bar{Q}_{\max}	Varies by treatment**
Player i 's abatement using technology A	q_i^A	
Player i 's abatement using technology B	q_i^B	
Aggregate abatement	$Q = \sum_i (q_i^A + q_i^B)$	
Aggregate abatement except for player i	$Q_{-i} = \sum_{j, j \neq i} (q_j^A + q_j^B)$	
Player i 's expected payoff	$E(\pi_i)$	
Expected collective payoff	$E(\Pi)$	

* 150 for the 150 treatment; 145 for 145/155, 140 for 140/160, 135 for 135/165, and 100 for 100/200. ** 150 for the 150 treatment; 155 for 145/155, 160 for 140/160, 165 for 135/165, and 200 for 100/200.

Table S2. Experimental treatments and hypotheses.

Treatment	Threshold	Game	Full cooperation	Non cooperation	Catastrophe avoided?
<i>150</i>	150	Coordination	150	0, 150	Yes
<i>145/155</i>	[145, 155]	Coordination	155	0, 155	Yes
<i>140/160</i>	[140, 160]	Prisoners' Dilemma	160	0	No
<i>135/165</i>	[135, 165]	Prisoners' Dilemma	165	0	No
<i>100/200</i>	[100, 200]	Prisoners' Dilemma	200	0	No

The treatments *150* and *145/155* represent coordination games to the left of the dividing line for threshold uncertainty; the remaining treatments *140/160*, *135/165*, and *100/200* represent prisoners' dilemma games to the right of the dividing line. Assuming that players use the proposals and pledges to coordinate to the efficient equilibrium, contributions in *150* and *145/155* are predicted to reach the full cooperative level and avoid catastrophe. Since the Nash equilibrium in the prisoners' dilemma games is unique, contributions are predicted to fall to zero. The treatments *150* and *100/200* are taken from (S25).

Table S3. Cooperation experiments with varying threshold uncertainty.

Study (S7-S18)	Contribution decision	Uncertainty test and rounds	Benefits	Marginal costs	N	Threshold distribution	Threshold ranges	Communication	Effect of increasing uncertainty on contributions
<i>Climate change games</i>									
This paper	Simultaneous; discrete [0, 20]	Between-subject test; one-shot	Continuous and discontinuous	Stepwise constant	10	Continuous uniform	150 [145, 155] [140, 160] [135, 165] [100, 200]	Yes; proposals and pledges	Negative unless the increase in uncertainty is very small
<i>Discrete public goods games</i>									
Gustafsson et al. 2000 *	Simultaneous; [0, 2000]	Within-subject test; 2 x 8 rounds	Discontinuous	Constant	5	Sequences of numbers with same mean (5000) and different variances represent prior thresholds	Std dev = 625.2 or 1068.9	No	Negative
McBride 2010	Simultaneous; binary {0, 1}	Within-subject test; 2 x 15 rounds	Discontinuous	Constant	5	Discrete uniform	3 {2, 3, 4} {1, 2, 3, 4, 5}	No	Tends to be negative for low public good value
Suleiman et al. 2001	Simultaneous; discrete [0, 5]	Within-subject test; 3 x 8 rounds	Discontinuous	Constant	5	Discrete uniform	10 [8, 12] [6, 14] 15 [13, 17] [11, 19]	No	Positive for low threshold mean
Wit / Wilke 1998 *	Sequential; discrete [0, 15]	Between-subject test; one-shot	Discontinuous	Constant	100	Continuous uniform	[0, 1000] [0, 1400]	No	Negative, especially when combined with high strategic uncertainty
<i>Common pool resource games</i>									
Budescu et al. 1990	Simultaneous; any request	Within-subject test; 3 x 10 rounds	Discontinuous	Constant	5	Continuous uniform	500 [250, 750] [0, 1000]	No	Negative
Budescu et al. 1995a	Simultaneous and sequential; any request	Within-subject test; 5 x 4 rounds	Discontinuous	Constant	5	Continuous uniform	[495, 505] [465, 535] [400, 600] [310, 690] [220, 780]	No	Negative for large uncertainty
Budescu et al. 1995b	Simultaneous and sequential; any request	Within-subject test; 3 x 10 rounds	Discontinuous	Constant	5	Continuous uniform	500 [250, 750] [0, 1000]	No	Negative
					3	Continuous uniform	500 [480, 520] [420, 580] [120, 880] [0, 1000]	No	Negative for large uncertainty
					2	Continuous uniform	500 [480, 520] [450, 550] [420, 580] [240, 760] [120, 880] [60, 940] [0, 1000]	No	Negative for large uncertainty
Gustafsson et al. 1999a *	Simultaneous; any request	Within-subject test; 6 x 4 rounds	Discontinuous	Constant	5	Continuous uniform	[250, 750] [0, 1000] [750, 1250] [500, 1500] [1250, 1750] [1000, 2000]	No	Tends to be negative
Gustafsson et al. 1999b *	Simultaneous; any request	Within-subject test; 2 x 8 rounds	Discontinuous	Constant	5	Sequences of numbers with same mean (5000) and different variances represent prior resource sizes	Std dev = 625.2 or 1068.9	No	Negative
Rapoport et al. 1992	Simultaneous; any request	Within-subject test; 5 x 4 rounds	Discontinuous	Constant	5	Continuous uniform	500 [375, 625] [250, 750] [125, 875] [0, 1000]	No	Negative for large uncertainty
Rapoport et al. 1993	Sequential; any request	Within-subject test; 3 x 10 rounds	Discontinuous	Constant	5	Continuous uniform	500 [250, 750] [0, 1000]	No	Negative
Suleiman et al. 1996	Sequential; any request	Within-subject test; 3 x 5 rounds	Discontinuous	Constant	5	Continuous uniform	500 [250, 750] [0, 1000]	No	Negative
* Hypothetical or deceptive experiments in which participants were given false information about the game and their payoff.									

Table S4. Significance of differences between treatments.

	Proposal	Pledge	Contribution	Proposal	Pledge	Contribution	Proposal	Pledge	Contribution	Proposal	Pledge	Contribution
<i>145/155</i>	.0002 (.5289)	.0878 (.3823)	.8490 (.1770)									
<i>140/160</i>	.0002 (.5606)	.3250 (.0311)	.0004 (.1188)	.0027 (.3470)	.9095 (.1612)	.0002 (.0260)						
<i>135/165</i>	.0032 (.0145)	.1120 (.0116)	.0002 (.0425)	.0232 (.0108)	.2721 (.0470)	.0002 (.0090)	.1112 (.0290)	.5702 (.3365)	.8498 (.5993)			
<i>100/200</i>	.0002 (.0052)	.0638 (.0170)	.0002 (.0137)	.0961 (.0039)	.1502 (.0375)	.0002 (.0008)	.2113 (.0107)	.1726 (.1285)	.0009 (.6730)	.5449 (.6707)	.4268 (.3484)	.0015 (.8428)
Treatment	<i>150</i>			<i>145/155</i>			<i>140/160</i>			<i>135/165</i>		

P-values from a Mann-Whitney-Wilcoxon rank-sum test of treatment differences in mean values; in parentheses, *p*-values from a Levene test of treatment differences in variances.

Table S5. Responses to the ex post questionnaire.

Question	Response	150	145/155	140/160	135/165	100/200
1) Were you generally satisfied with the game's outcome?	Very much	63	32	6	5	10
	Somewhat	18	40	22	36	31
	Little	5	17	32	21	26
	Not at all	14	11	40	38	33
2) Knowing how the game was played, with the benefit of hindsight, do you wish you had made a different contribution?	Very much	2	6	11	16	11
	Somewhat	19	19	24	21	17
	Little	27	24	19	33	22
	Not at all	52	51	46	30	50
3) Did fairness play a role for your contribution decision?	Very much	61	48	44	36	24
	Somewhat	16	32	24	30	10
	Little	11	8	18	16	21
	Not at all	12	12	14	18	45
4) Did trust play a role for your contribution decision?	Very much	58	44	43	39	18
	Somewhat	22	36	28	28	12
	Little	9	13	17	19	23
	Not at all	11	7	12	14	47
5) Do you agree with the statement that the exchange of proposals was helpful?	Very much	49	37	20	17	6
	Somewhat	27	32	38	40	28
	Little	13	21	29	29	34
	Not at all	11	10	13	14	32
6) Do you agree with the statement that the exchange of pledges was helpful?	Very much	68	50	26	25	10
	Somewhat	24	34	42	27	30
	Little	5	13	19	32	27
	Not at all	3	3	13	16	33
7) Generally speaking, do you trust other people?	Very much	25	29	23	27	21
	Somewhat	60	51	55	63	60
	Little	13	20	19	9	17
	Not at all	2	0	3	1	2
8) Generally speaking, do you agree with the statement that, if a person fails to keep his or her word, they deserve another chance?	Very much	24	39	35	39	41
	Somewhat	54	46	44	47	45
	Little	18	14	18	13	14
	Not at all	4	1	3	1	0
9) Generally speaking, do you try to keep your word?	Always	56	54	55	44	36
	Often	41	45	43	54	60
	Sometimes	1	1	2	2	4
	Rarely	1	0	0	0	0
	Never	1	0	0	0	0
10) Did you trust the other players to make the contributions they pledged?	Very much	47	47	27	29	10
	Somewhat	43	38	41	32	23
	Little	8	12	24	19	26
	Not at all	2	3	8	20	41
11) Knowing how the game was played, with the benefit of hindsight, do you feel, that some of the other players betrayed your trust in them?	Very much	10	14	30	28	16
	Somewhat	12	39	28	20	21
	Little	37	33	23	28	23
	Not at all	41	14	19	24	40
12) Please imagine the following situation in another unrelated experiment: You have an initial endowment of €40. There is a 50% possibility that you will lose your €40. However, you can avoid this loss by paying €20 up front. Would you rather pay this amount and get €20 for certain or would you rather accept the risk of losing the €40 with probability 50%?	€40 uncertain	15	15	18	12	25
	Indifferent	27	14	14	22	13
	€20 certain	58	71	68	66	62
13) The contribution threshold will soon be determined by the "spinning wheel". What single value do you estimate for the threshold?	Mean		149.87	149.68	149.05	150.71
14) How confident are you about your estimate of the contribution threshold?	Very much		10	5	13	6
	Somewhat		33	41	32	39
	Little		35	30	32	35
	Not at all		22	24	23	20

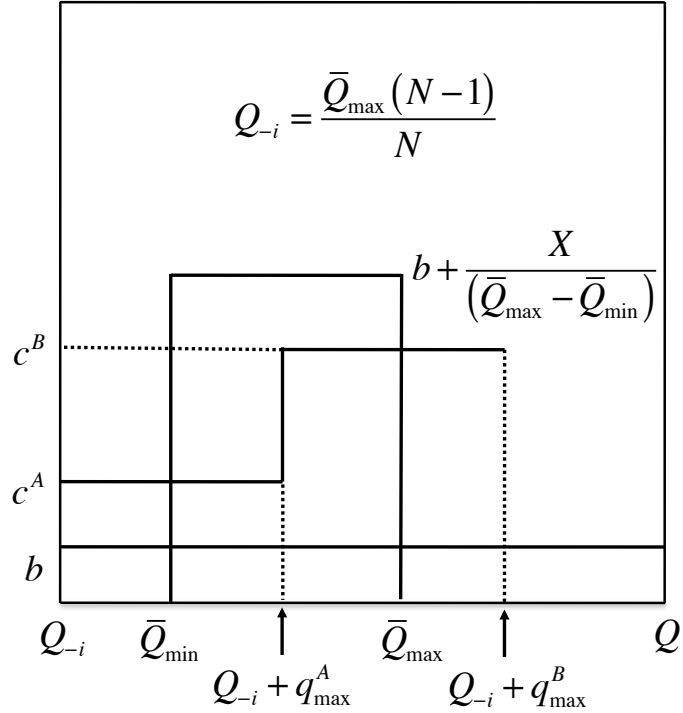
Numbers are percentages of subjects per treatment (except for question 13, which shows mean values).

Table S6. Responses to the ex post open-ended questions.

Question	Response	150	145/155	140/160	135/165	100/200
1) What was the most important reason for your proposal for the group contribution?	Joint payoff maximization	82	77	63	42	22
	Fairness	3	8	5	4	1
	Safety	8	3	6	4	0
	Stimulation of others' contributions	2	5	10	18	31
	Realistic target	0	5	12	27	39
	Other reason	5	2	4	5	7
2) What was the most important reason for your pledge for your own intended contribution?	Signaling of intended contribution / creation of trust	71	72	50	29	24
	Stimulation of others' contributions	17	19	33	46	66
	Safety	5	5	8	15	4
	Other reason	7	4	9	10	6
3) What was the most important reason for your contribution?	Fair share to reach target / own pledge	56	64	47	34	12
	Compensation of potentially missing contributions / safety	33	17	5	4	0
	Own payoff maximization	10	15	16	21	24
	Resignation / distrust	0	1	17	21	30
	Cheap chips / compromise between group and own interest	0	1	11	19	33
	Other reason	1	2	4	1	1

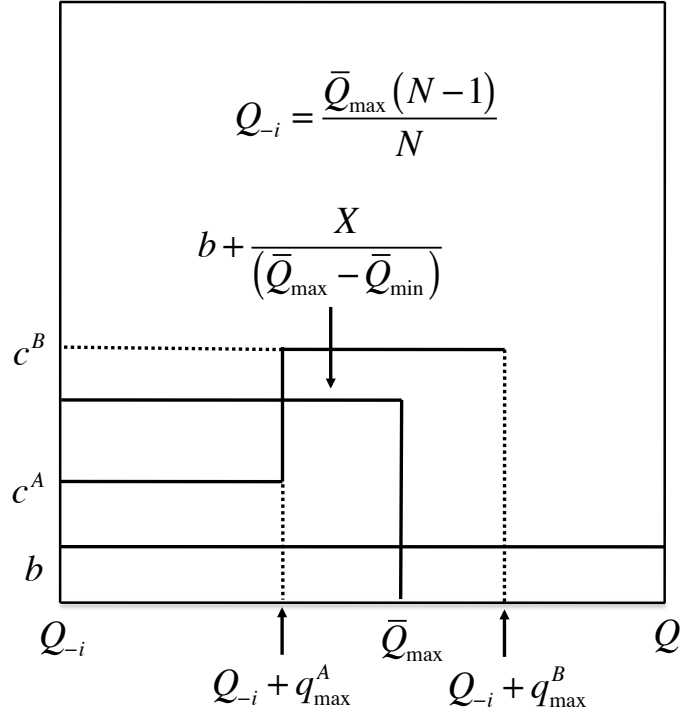
Subjects' responses were classified by keyword search. Numbers are percentages of subjects per treatment.

Figure S1. Avoiding catastrophe a Nash equilibrium.



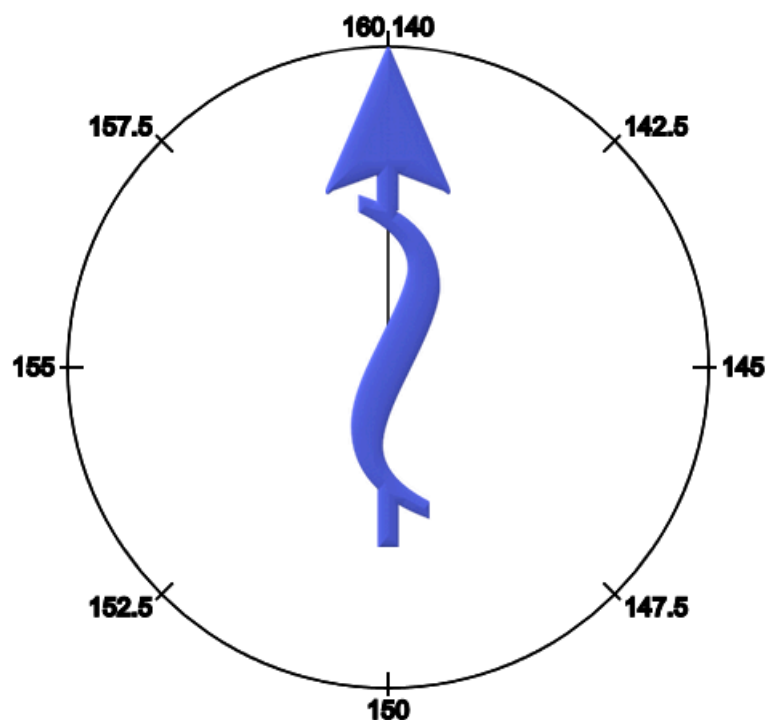
This figure describes qualitatively the incentives for every country i to abate so as to eliminate any chance of catastrophe, given that all other countries abate the same amount. The figure is consistent with the treatment 145/155. In this case, we have $N = 10$, $b = 0.05$, $c^A = 0.10$, $c^B = 1$, $X = 15$, $Q_{-i} = 139.5$, $\bar{Q}_{\max} = 155$, $\bar{Q}_{\min} = 145$, $q_{\max}^A = 10$, and $q_{\max}^B = 10$. Hence, the dividing line is $\phi = X / (c^B - b) = 15.8$; and, since $\bar{Q}_{\max} - \bar{Q}_{\min} = 10$, the game is played to the *left* of the dividing line. Using the above values we find $b + X / (\bar{Q}_{\max} - \bar{Q}_{\min}) = 1.55$. Since this value exceeds $c^B = 1$, it will pay i to abate $q_i^A = q_{\max}^A = 10$, and $q_i^B = \bar{Q}_{\max} / N - q_{\max}^A = 5.5$. This assumes continuous abatement. In our experiment, abatement must be in integer units. It is easy to confirm that play $q^A = 10$, $q^B = 5$ by half the players and play $q^A = 10$, $q^B = 6$ by the other half is a Nash equilibrium. Of course, there are also many more asymmetric Nash equilibria, but in our set up, contributions that are approximately symmetric are particularly focal.

Figure S2. Avoiding catastrophe not a Nash equilibrium.



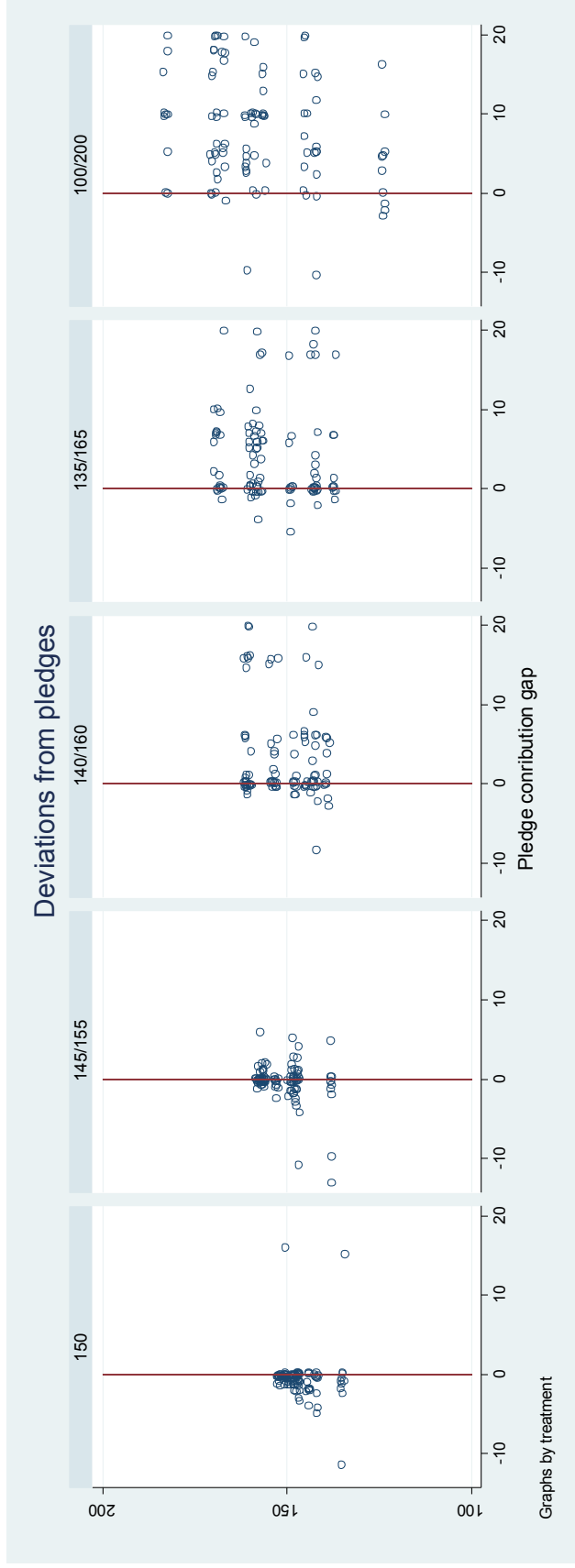
This figure describes qualitatively the absence of incentives for abatement by countries acting independently. The figure is consistent with the treatments *140/160*, *135/165*, and *100/200*. Let's focus on *140/160*. In this case, we have $N = 10$, $b = 0.05$, $c^A = 0.10$, $c^B = 1$, $X = 15$, $Q_{-i} = 144$, $\bar{Q}_{\max} = 160$, $\bar{Q}_{\min} = 140$, $q_{\max}^A = 10$, and $q_{\max}^B = 10$. The dividing line is $\phi = X / (c^B - b) = 15.8$; and, since $\bar{Q}_{\max} - \bar{Q}_{\min} = 20$, the game is played to the *right* of the dividing line. Using the above values we find $b + X / (\bar{Q}_{\max} - \bar{Q}_{\min}) = 0.8$. Since this value is less than $c^B = 1$, it will pay i to abate $q_i^B = 0$. In other words, if every country j , $j \neq i$ were initially to play $q_j = \bar{Q}_{\max} / N$, it will pay i to play $q_i^B = 0$. Hence, play $q_i^A = q_{\max}^A$, $q_i^B = \bar{Q}_{\max} / N - q_{\max}^A$ cannot be a Nash equilibrium.

Figure S3. The “spinning wheel.”



At the end of the experiment, a volunteer was invited to activate a computerized “spinning wheel,” with the “ends” of the wheel at 12 o’clock representing the minimum and maximum values of the uncertainty range (*145/155*, *140/160* shown here, *135/165*, and *100/200*). Every subject was able to observe the wheel being spun, and see where the arrow came to rest, determining the value for the threshold (the selected value was also displayed next to the wheel). The wheel generated random numbers (rounded to two decimal places), i.e. it was not possible to manipulate the outcome. All participants had 15 minutes to play and become familiar with the wheel before they got to know the game. To spin the spinner, players had to click on the spinner and swipe to the side, away from the spinner.

Figure S4. Deviations between pledges and actual contributions.



The vertical axis shows the aggregate pledges (the sum of the individual pledges) and the horizontal axis shows the gap between players' individual pledges and their actual contributions. A series of Spearman's correlation tests between the aggregate pledges and the absolute gap between individual pledges and actual contributions gives: $n = 100$, $\rho = -.32$, $p = .00$ for 150; $\rho = -.29$, $p = .00$ for 145/155; $\rho = -.02$, $p = .81$ for 140/160; $\rho = -.16$, $p = .12$ for 135/165; $\rho = -.16$, $p = .12$ for 100/200.