LETTER:

Threshold uncertainty, early-warning signals, and the prevention of dangerous climate change

Mark J. Hurlstone[1*] and Ben R. Newell[2]

[1]School of Psychological Science, University of Western Australia, Perth, Australia
[2]School of Psychology, UNSW, Sydney, Australia
[*]mark.hurlstone@uwa.edu.au

The goal of the Paris Climate Agreement is to keep global temperature rise well below 2°C. In this agreement—and its antecedents negotiated in Copenhagen and Cancun—the fear of crossing a dangerous climate threshold is supposed to serve as the catalyst for cooperation amongst countries. However, there are deep uncertainties about the location of the threshold for dangerous climate change,[1–3] and recent evidence indicates that this threshold uncertainty is a major impediment to collective action.[4–7] Early-warning signals of approaching climate thresholds[8–11] are a potential remedy to this threshold uncertainty problem, and initial experimental evidence suggests that such early-detection systems may help improve the prospects of cooperation.[6] Here, we provide a direct experimental assessment of this early-warning signal hypothesis. Using a catastrophe avoidance cooperation game, we show that large initial—and subsequently unabated—threshold uncertainty undermines cooperation, consistent with earlier studies,[5–7] but additionally that a marked subsequent reduction in threshold uncertainty—mimicking an early-warning signal—does little to improve the prospects of avoiding crossing a dangerous threshold. Regrettably, our findings suggest early-warning signals indicating that a critical climate threshold is approaching are unlikely to offer the leverage necessary to motivate countries to take the necessary action to avert catastrophe.

The goal of the United Nations Framework Convention on Climate Change (UNFCCC) is to achieve "stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system". But what constitutes dangerous interference? In 2009, the signatories of the Copenhagen Accord reached an agreed definition, namely that in accordance with "the scientific view the increase in global temperature should be below 2 degrees Celsius". However, contrary to the Copenhagen Accord, there is no scientific view that 2°C is the threshold for dangerous anthropogenic interference. Although there is a consensus regarding the existence of dangerous climate thresholds, the location of those thresholds is highly uncertain.[1–3] Political actors and climate negotiators are not oblivious to this scientific uncertainty. Indeed, no sooner had the signatories of the Copenhagen Accord agreed upon the 2-degree-target than a year later in Cancun discussions were raised regarding the possibility of adopting a more conservative 1.5°C target. This uncertainty is enshrined in the new Paris Agreement which—in addition to reaffirming the 2-degree-target—underscores the desirability of "pursuing efforts to limit the temperature increase to 1.5°C".

Recently, an experimental literature has emerged to examine the consequences for the climate negotiations of uncertainty about climate thresholds. Within this literature, the problem of avoiding dangerous climate change has been simulated using laboratory cooperation experiments.[12,13] In these experiments, groups of players must cooperate by investing money from a personal operating fund into hypothetical emission abatement to avoid crossing a dangerous threshold, which if breached triggers catastrophic economic losses for all (see Supplementary Literature). This literature has revealed that when the threshold is known with certainty, groups can effectively coordinate their efforts to remain on the safe side of the dangerous threshold, but when the threshold is uncertain, coordination collapses and catastrophe is all but guaranteed.[5–7,14] Under threshold certainty, there is a strong relationship between what groups propose to do; what they pledge to contribute; and what they actually contribute.[6] Under threshold uncertainty, groups propose to do less than is needed to avert catastrophe; pledge to contribute less than is required to meet their proposals; and contribute less than their pledges.[6,15]

The parallels with the real climate negotiations are striking and sobering. Under the Paris Agreement, countries

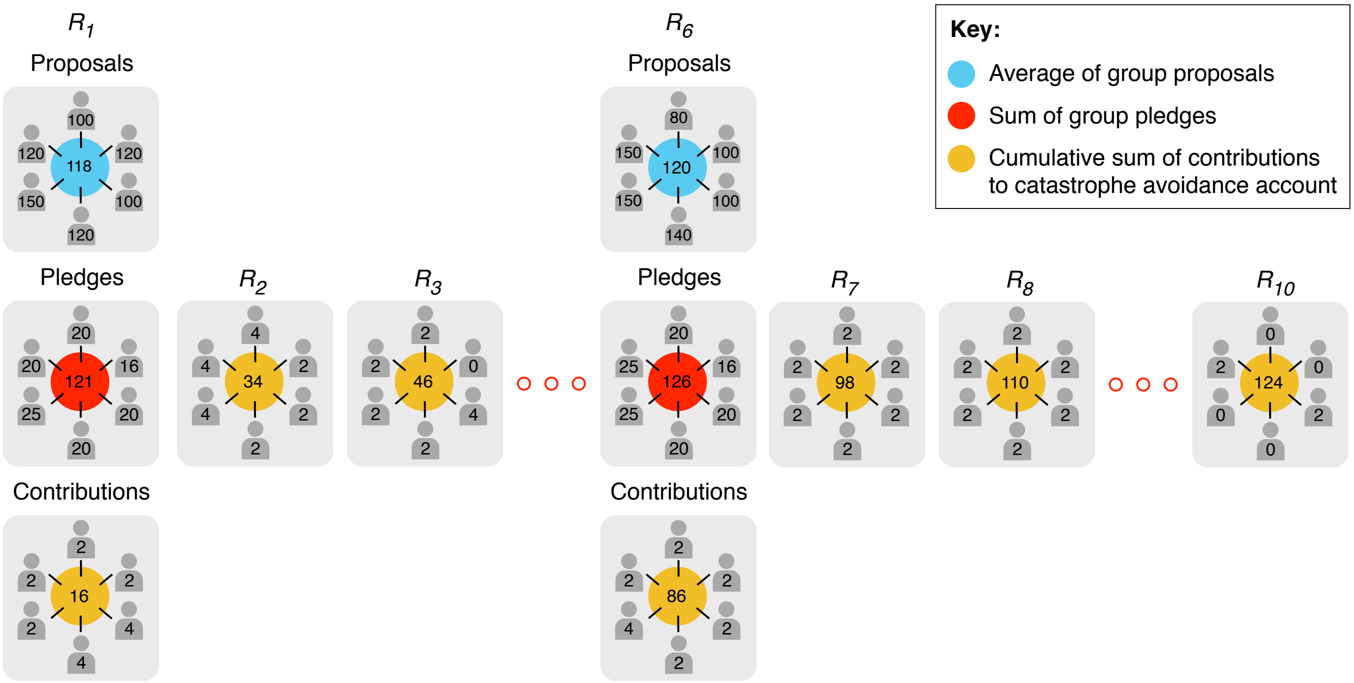

Figure | 1 An illustration of the structure of the catastrophe avoidance cooperation game. At the start of the game, \$40 is credited to the personal account of each player ($N = 6$). In treatment T_1, players are instructed that the threshold is \$120, whereas in treatments T_2–T_4, players are told the threshold is a uniform random value between \$0–\$240, but they will not know the actual value of the threshold until the end of the game. On each of 10 rounds, R_{1-10}, each player must decide simultaneously and independently whether to contribute \$0, \$2, or \$4 from their personal account into a damage prevention account. At the start of round 1—and again on round 6—players simultaneously and independently submit two non-binding announcements before making their contribution decision. First, each player submits a ‘proposal’ regarding the target level of contributions the group should aim for by round 10, and the average of these proposals becomes the agreed collective target. Next, each player submits a ‘pledge’ regarding the total amount that they will personally contribute across the 10 rounds toward reaching the agreed collective target. In treatments T_3 and T_4, before players submit their second set of non-binding proposals on round 6, they are instructed that the uncertainty about the threshold has reduced, and that the threshold is now a uniform random value between \$84–\$156 (T_3) or \$108–\$132 (T_4). At the end of the game, the contributions in the damage prevention account are compared with the known (T_1) or randomly chosen (T_2–T_4) threshold. In the uncertainty treatments, the computer determines the exact threshold amount by drawing a random number from a uniform distribution either over the interval [0, 240] (T_2), [84, 156] (T_3), or [108, 132] (T_4). If the total contributions equal or exceed the threshold, then the damage is avoided and players get to keep the remaining contents of their personal accounts, otherwise they lose 90% of their remaining funds.

have proposed to do less than is required to limit the risk of catastrophe (the agreement aims to restrict warming to 2°C but recognises that a 1.5°C goal is probably required) and they have pledged to contribute much less than is required to reach the collective goal.[16–18] Laboratory cooperation experiments suggest that countries actual contributions will be less than their pledges, leaving little hope of staying below the 2°C limit.

If the uncertainty surrounding the location of the dangerous threshold could be reduced sufficiently, this might provide the leverage necessary to transform the climate negotiations. Uncertainty about the location of a dangerous threshold can be reduced through the detection of early-warning signals of approaching climate transitions.[8–11] Such early detection systems could provide the reduction in threshold uncertainty needed to help countries avert catastrophe. That such signals might facilitate cooperation was demonstrated in a recent experiment that revealed groups can often surmount the threshold uncertainty problem, provided that uncertainty is confined to a narrow range.[6] However, this experiment involved a one-shot interaction in which groups always faced the same level of threshold uncertainty,

Table 1 | Overview of the design of the experiment

Treatment	τ Rounds 1-10		Expected Value	N Participants
T_1 (Certainty)	\$120		\$120	$10 \times 6 = 60$
T_2 (Uncertainty)	[\$0, \$240]		E(τ) = \$120	$10 \times 6 = 60$
	τ Rounds 1-5	τ Rounds 6-10		
T_3 (Warning-Wide)	[\$0, \$240]	[\$84, \$156]	E(τ) = \$120	$10 \times 6 = 60$
T_4 (Warning-Narrow)	[\$0, \$240]	[\$108, \$132]	E(τ) = \$120	$10 \times 6 = 60$

τ, threshold for catastrophe.

whereas an appropriate test of the early-warning signal hypothesis requires a repeated interaction scenario, wherein groups face large threshold uncertainty initially, followed by a reduction in that uncertainty as the threshold is approached to mimic the action of an early-warning signal.

Here, we present the results of an experiment that meets this requirement. Our experiment involved 240 participants who were randomly allocated to six-player groups to play a catastrophe avoidance game (Fig. 1).[7,19,20] At the start of the game, each player was given a \$40 endowment. On each of ten rounds, players decided whether to contribute \$0, \$2, or \$4 of their endowment into a damage avoidance account. The players knew that if the total amount contributed by the end of the game equalled or exceeded a threshold amount, then they would get to keep their remaining endowment, otherwise they would lose 90% of this remaining amount. Before the contribution decisions on rounds 1 and 6, each player submitted two non-binding communications: (1) a proposal regarding how much the group should collectively contribute over the 10 rounds, and (2) a pledge regarding how much they personally intended to contribute toward reaching this collective goal. At the end of each round, the players were informed about everyone's current round contribution and cumulative contribution across all rounds played so far.

The experiment involved four treatments T_1–T_4, each comprising 10 groups (Table 1). In T_1, the threshold was certain, whereas in treatments T_2–T_4 it was uncertain. In the former treatment, groups were told the threshold was \$120, whereas in the latter treatments they were informed it was a uniform random value between \$0 and \$240 but the exact amount would not be determined and announced until the conclusion of the game. Treatments T_3 and T_4 differed from T_2 in that on round 6—before the second set of non-binding proposals and pledges—unexpectedly groups received an early-warning that the uncertainty surrounding the threshold had been reduced. Specifically, in T_3, groups were instructed that the threshold was now a random amount between \$84 and \$156 (70% reduction in threshold uncertainty), whereas in T_4, they were instructed that the threshold was now a random amount between \$108 and \$132 (90% reduction in threshold uncertainty). The comparison between T_1 and T_2 indexes whether threshold uncertainty undermines cooperation, whereas the comparison between T_2 and either T_3 or T_4 indexes whether an early-warning signal benefits cooperation under threshold uncertainty, and whether the size of any such benefit is dependent upon the magnitude of the reduction in threshold uncertainty.

For data analysis, we begin by examining the probability of avoiding catastrophe according to treatment based on comparison of group contributions with the known (T_1) or randomly chosen (T_2–T_4) threshold (Fig. 2a; see Supplementary Statistical Analyses for a kindred analysis based on various different hypothetical thresholds). In T_1, where the threshold is certain, 90% of groups succeeded in avoiding catastrophe. In T_2, where the threshold is deeply uncertain, only 50% of groups succeeded in avoiding catastrophe. In T_3 and T_4, where the threshold is deeply uncertain initially, but then reduces by 70% and 90% mid-game, only 30% and 50% of groups succeeded in avoiding catastrophe, respectively. The difference between T_1 and T_2 is marginally significant (Fisher exact, $P = .070$), whereas the difference between T_1 and $T_2 + T_3 + T_4$ is highly significant (Fisher exact, $P = .013$). Thus, threshold uncertainty

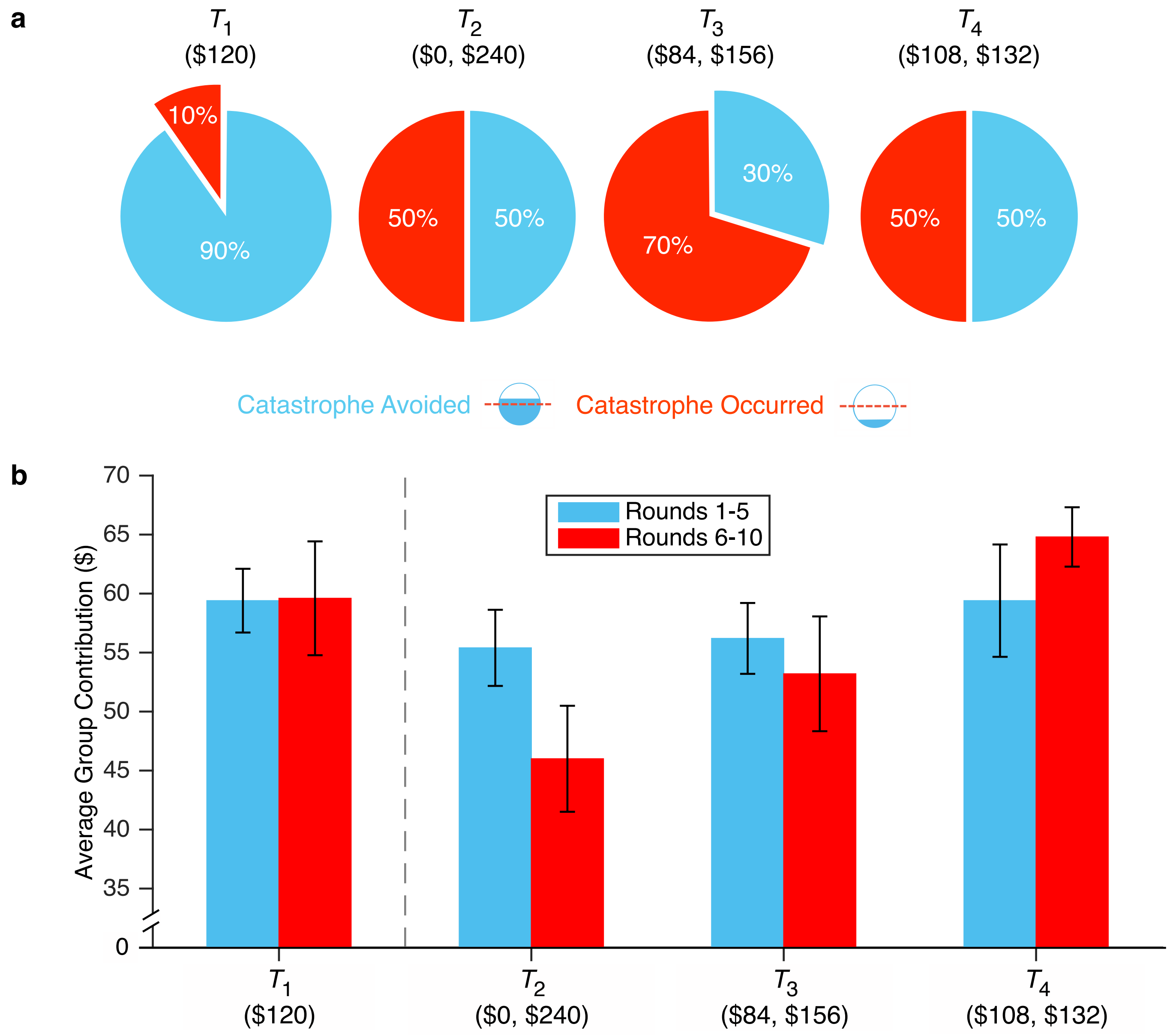

Figure | 2 **Early-warning signals fail to stave off the impediment of threshold uncertainty. a,** The percentage of groups in which catastrophe occurred and was avoided according to treatment. Whereas most groups averted catastrophe in T_1, only 50% of groups did so in T_2, and the provision of an early-warning signal in T_3 and T_4 is ineffectual in overturning the handicap of large initial threshold uncertainty. **b,** Average group contributions, according to treatment, over rounds 1–5 and 6–10. Group contributions do not differ by treatment over rounds 1-5, but over rounds 6-10 contributions are lower in both T_2 and T_3 than in T_1. In contrast, contributions are higher in T_4 than in T_1–T_3 but not by enough to increase the probability of avoiding catastrophe.

significantly reduced the probability of avoiding catastrophe. Critically, the differences between T_2 and T_3 (Fisher exact, $P = .650$), and between T_2 and T_4 (Fisher exact, $P = 1.000$) were both nonsignificant. Thus, the provision of an early-warning signal did not overturn the impediment of threshold uncertainty.

Fig. 2b shows shows the ex ante (rounds 1–5) and ex post (rounds 6–10) early-warning signal group contributions by treatment. Ex ante contributions do not differ significantly by treatment (Kruskal-Wallis, $\chi^2_{df=3} = 0.72$, $P = .869$), whereas ex post contributions do (Kruskal-Wallis, $\chi^2_{df=3} = 10.95$, $P = .012$). Ex post contributions are significantly lower in T_2 than T_1 (Mann-Whitney = 80.00, $P = .025$), confirming that threshold uncertainty not only reduced the

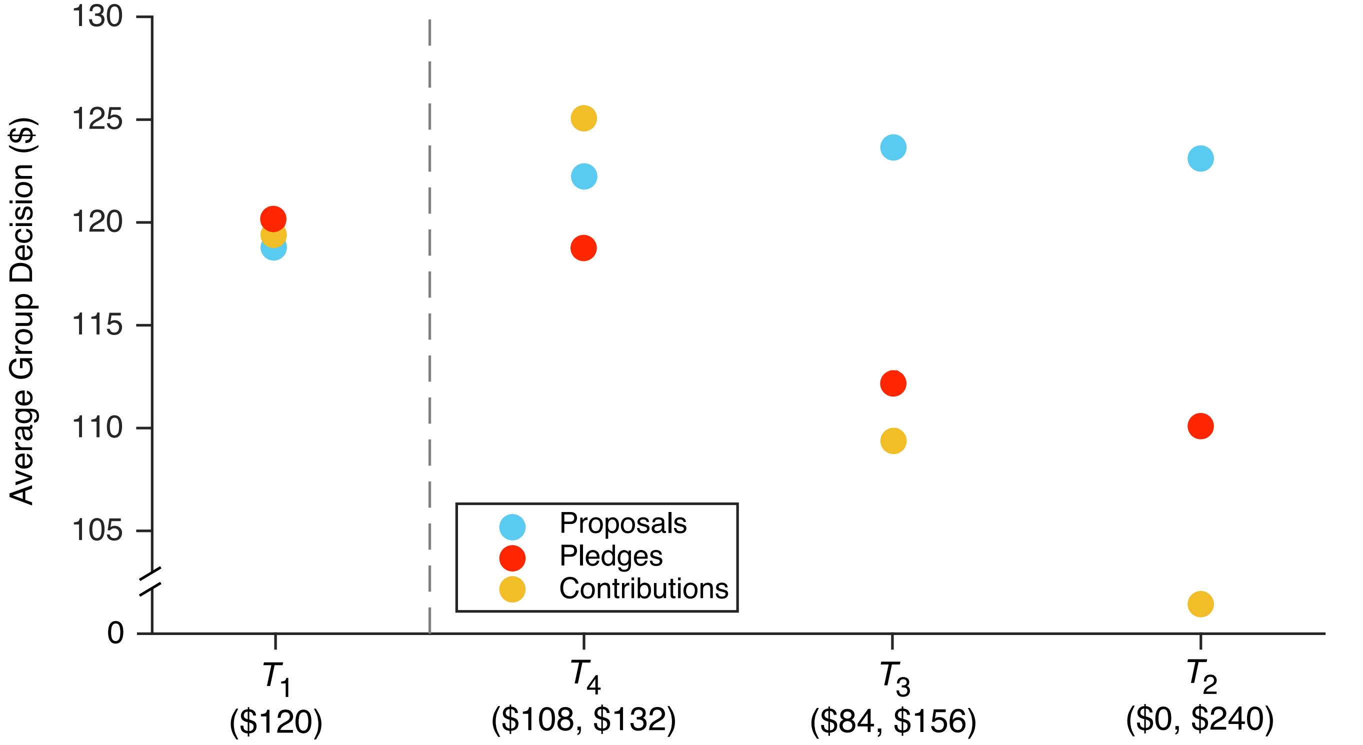

Figure | 3 Average group proposals, pledges, and contributions as a function of the four treatments. As threshold uncertainty increases, so too does the gap between what groups agree to do, commit to do, and actually do.

probability of averting catastrophe, but also reduced group contributions. Critically, whereas ex post contributions do not differ significantly between T_3 and T_2 (Mann-Whitney = 33.500, $P = .224$), ex post contributions are significantly higher in T_4 than T_2 (Mann-Whitney = 9.00, $P = .002$)—inspection of the round-by-round dynamics of contributions (see Supplementary Statistical Analyses) reveals that that this was largely due to a punctuated peak in contributions on round 6, which subsequently tailed-off. Thus, whilst an early-warning signalling a 70% reduction in threshold uncertainty did nothing to stimulate contributions, an early-warning signalling a 90% reduction increased contributions to a level comparable to that observed in T_1.

Next, we compared group proposals, pledges, and contributions across treatments. Since group proposals and pledges on round 1 did not differ appreciably from those on round 6 (see Supplementary Statistical Analyses), for simplicity, we combined each into a single measure by averaging the group proposals and pledges on the two rounds. The results are shown in Fig. 3, where the treatments have been organised, from left to right, in order of increasing threshold uncertainty. Thus, T_4 is now presented adjacent to T_1, whereas T_2 is presented on the far right. Consistent with previous work,[6] as threshold uncertainty increases, so too does the gap between what groups agree to do, pledge to do, and actually do. In T_1 and T_4, group proposals, pledges, and contributions fall closely in line. Indeed, in T_4 contributions are numerically higher than proposals and pledges. By contrast, in T_3 and T_2, pledges are less than proposals, and contributions, in turn, are less than pledges. Fig. 3 also clarifies why, despite stimulating contributions, the early-warning signal in T_4 did not increase the probability of avoiding catastrophe. This is because the reduction in threshold uncertainty had no effect on group proposals. Thus, despite the fact groups could avert catastrophe with certainty with a collective contribution of $132, they continued to choose a sub-optimal collective target around $120, which, if reached, still leaves a 50% chance of catastrophe occurring.

Consistent with earlier work,[5–7] our results suggest threshold uncertainty is a serious impediment to collective action. Early-warning signals offer the tantalising prospect that as a threshold is approached, uncertainty about its location may be reduced sufficiently to provide the incentive necessary to transform behaviour. However, the best way to reduce uncertainty about a threshold is to get closer to it, but by then it may already be too late to take emergency measures to avoid crossing it. There is also the risk that an early-warning signal may go undetected, meaning we may not know about the location of the threshold until it has already been breached. What our results show is that even if an early-warning signal is detected, it is unlikely to spur the level of cooperation required to steer clear of the threshold. Although a 90% reduction in threshold uncertainty in our treatment T_4 stimulated group contributions, it nevertheless failed to increase the probability of avoiding crossing the dangerous threshold. This suggests for an early-warning signal to be effective, the reduction in threshold uncertainty needs to be substantial, which would mean getting as close to the threshold as possible, which may leave too little time to take evasive action. Even then, our results imply that any residual uncertainty about the location of the threshold may still result in countries choosing a sub-optimal collective target that leaves open the possibility the threshold will be crossed.

The contemporary climate agreements rely on the fear of crossing a dangerous threshold to enforce cooperation. However, because of scientific uncertainty, this approach is unlikely to spur collective action to prevent dangerous climate change.[4–7] To these results, our findings indicate that even the prospect of early-warning signals is unlikely to offer the leverage necessary to avert catastrophe, although they may serve as an aid to pre-emptive adaptation.[8] This suggests that strategic enforcement mechanisms need to be built into these agreements that recreate—through strategic treaty design—the conditions that exist when the threshold is certain.[21–23] Trade sanctions, the removal of essential permits, or expulsion from cherished markets are examples of strategic enforcement mechanisms. The challenge for climate negotiators is to identify the incentives to persuade countries to commit to an enforceable climate treaty that leaves them vulnerable to being punished by other countries. However, as the window closes on our opportunity to prevent a climate emergency, some experts have suggested it is time to entertain other more radical solutions, such as solar geoengineering, which is cheap and can reduce global temperatures rapidly, potentially providing additional time for a more effective climate regime to emerge.[24–27] Of course, solar geoengineering is beset by known and unknown risks[25] but public citizens appear to be willing to invest in such risky technology.[28,29] Moreover, if collective efforts to reduce global emissions continue to fail, then not using solar geoengineering could represent an even greater risk.[30]

Methods

The experiment employed the collective-risk social dilemma catastrophe avoidance public goods game developed by Milinski and colleagues.[19] We used an augmented version of the basic game that permitted communication between players in the form of nonbinding proposals and pledges, a design feature introduced in experiments by Dannenberg and colleagues.[7,20] Ethical approval to conduct the experiment was granted by the Human Ethics office at the University of Western Australia (UWA) (RA/4/1/6996: Committing to the public good).

Participants

Two hundred and forty members of the campus community at UWA participated in the experiment (mean age = 24.37 years; SD = 7.30; range = 17–56; females = 60%). Participants were recruited using the Online Recruitment System for Experimental Economics (ORSEE),[31] an open source web-based recruitment platform used by the Behavioural Economics Laboratory (http://bel-uwa.github.io) at UWA. The ORSEE database contains a pool of over 1,500 UWA staff and students from a range of academic disciplines. Participants were recruited by issuing electronic invitations to randomly selected individuals in the ORSEE database to attend the experimental sessions.

Design

The experiment employed a 4 (treatment: T_1 vs. T_2 vs. T_3 vs. T_4) $\times$ 10 (round: 1–10) mixed design: treatment was a between-groups factor, whereas round was a within-groups factor. Participants were tested in groups of six players (ten groups per treatment). We commenced testing with the uncertainty treatments T_2, T_3, and T_4—randomly allocating each six-person group to one of the three treatments—before collecting the data for the certainty treatment T_1.

Apparatus, materials, and procedure

Experimental sessions were conducted in the Behavioural Economics Laboratory, a computerised laboratory for running economic experiments at UWA, in the presence of two experimenters. At the start of a session, players were randomly seated at interconnected computer terminals running the Zurich Toolbox for Readymade Economic Experiments (z-Tree),[32] which was used to register and communicate their decisions during the experiment. The computer terminals were separated by privacy blinds to prevent player collusion. Participants read an information sheet and provided written informed consent initially, after which they read the experimental instructions, and answered a series of control questions (see Supplementary Experimental Instructions) to ensure they understood the rules of play. The experiment did not commence until the experimenters had verified that all players had answered the control questions correctly. To ensure anonymity, each player was assigned a pseudonym before the game commenced (Ananke, Telesto, Despina, Japetus, Kallisto, or Metis). During the game, each player's decisions were communicated to the other players under his or her designated pseudonym.

At the start of the game, each player was given a \$40 endowment. On each of ten rounds, players decided simultaneously and independently whether to contribute \$0, \$2, or \$4 of their endowment into an account for damage prevention. The players knew that the total amount invested in the damage prevention account by the end of the game must equal or exceed a threshold amount, otherwise each player would lose 90% of her or his remaining endowment. In treatment T_1, the instructions emphasised that the threshold amount to-be-reached by the end of the game was \$120. By contrast, in the uncertainty treatments T_2, T_3, and T_4 the instructions emphasised that the threshold amount was a random amount between \$0 and \$240, with each whole dollar amount having an equal probability of being selected, but the exact amount would not be determined and declared until the conclusion of the game.

At the start of rounds 1 and 6, each player simultaneously and independently submitted two non-binding announcements. First, each player submitted a proposal regarding how much the group should contribute in total over the ten rounds. After each player had registered her or his proposal, the proposals of all players, as well as the group average, were displayed on all computers simultaneously. The players knew that the average group proposal would serve as the agreed collective target. Second, each player submitted a pledge regarding how much money she or he would personally contribute in total over the ten rounds. Once each player had registered her or his pledge, the pledges of all players, as well as the group total, were displayed on all computers simultaneously along with the group proposals to facilitate comparison.

At the end of each round, the contribution decisions of all six players, their cumulative contributions across all rounds played so far, and their proposals and pledges were displayed on all computers simultaneously (in addition to the total current round contributions, total contributions across all rounds played so far, average group proposal, and total group pledges). In this way, as the game progressed, each player was able to gauge whether her or his group members were adhering to their pledges, and whether the group contributions were consistent with achieving the agreed (average) group proposal.

At the start of round 6, before the second set of non-binding announcements, groups in treatments T_3 and T_4 were given an on-screen warning informing them that the uncertainty surrounding the location of the threshold had now been reduced. Specifically, in T_3, groups were informed that the threshold amount was now a random amount between \$84–\$156 (equivalent to a 70% reduction in threshold uncertainty), whereas in T_4, groups were informed that the threshold amount was now a random amount between \$108–\$132 (equivalent to a 90% reduction in threshold uncertainty). In T_1 and T_2, the known threshold (\$120) and uncertain threshold range (\$0–\$240), respectively, remained the same as specified at the outset, and groups in these treatments did not therefore receive any additional information about the threshold. Instead, at the start of round 6, groups in these treatments proceeded directly to submit their second set of non-binding announcements.

At the end of the game, the threshold amount and the contents of the damage prevention account were communicated to the group. In the uncertainty treatments, the computer determined the exact threshold amount by drawing a random number from a uniform distribution either over the interval [0, 240] (T_2), [84, 156] (T_3), or [108, 132] (T_4). Once this information had been communicated to the group, participants completed a brief economic preferences questionnaire comprising single-item self-report measure of risk aversion, loss aversion, trust, fairness, altruism, and temporal discounting (see Supplementary Statistical Analyses). Participants were then paid in cash either the full remainder of their endowment (if the group contributions reached or exceeded the threshold amount) or 10% of the leftovers of their endowment (if the group failed to reach the threshold amount), in addition to a \$10 attendance fee. The average payout was \$20.15 (inclusive of attendance fee). The cash was concealed in envelopes to protect the anonymity of players.

Acknowledgements

This research was funded by a grant from the Climate Adaptation Flagship of the Commonwealth Scientific and Industrial Research Organisation awarded to MJH.

Author contributions

MJH and BN conceived and designed the experiment; MJH programmed the experiment, collected the data, analysed the results, and wrote the paper; BN reviewed and edited the paper.

Data availability

All raw data associated with this study, along with the computer programs used to execute the experimental treatments, have been deposited in a publicly accessible GitHub repository at https://github.com/mark-hurlstone/Hurlstone-Newell-2019.

References

1. Lenton, T. M. *et al.* Tipping Elements in the Earths Climate System. *Proc. Natl. Acad. Sci. USA* **105**(6), 1786–1793 (2008).
2. Kriegler, E. *et al.* Imprecise probability assessment of tipping points in the climate system. *Proc. Natl. Acad. Sci. USA* **106**(13), 5041–5046 (2009).
3. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
4. Barrett, S. Climate treaties and approaching catastrophes. *J. Environ. Econ. Manage.* **66**, 235–250 (2013).
5. Barrett, S. & Dannenberg, A. Climate negotiations under scientific uncertainty. *Proc. Natl. Acad. Sci. USA* **109**(43), 17372–17376 (2012).

6. Barrett, S. & Dannenberg, A. Sensitivity of collective action to uncertainty about climate tipping points. *Nat. Clim. Change.* **4**, 36–39 (2014).
7. Dannenberg, A. *et al.* On the provision of public goods with probabilistic and ambiguous thresholds. *Environ. Resource. Econ.* **61**, 365–383 (2015).
8. Lenton, T. Early warning of climate tipping points. *Nat. Clim. Change.* **1**, 201–209 (2011).
9. Lenton, T. M. *et al.* Early warning of climate tipping points from critical slowing down: Comparing methods to improve robustness. *Phil. Trans. R. Soc. A.* **370**, 1185–1204 (2012).
10. Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
11. Scheffer, M. *et al.* Anticipating critical transitions. *Science* **338**, 344–338 (2012).
12. Dannenberg, A., & Tavoni, A. Collective Action in Dangerous Climate Change Games. In Ariel Dinar, Felix Munoz-Garcia, Ana Espinola-Arredondo, Richard M. Matthew, Tony Bryant, Anabela Botelho (eds.), WSPC References of Natural Resources and Environmental Policy in the Era of Global Change, Vol. 4, Experimental Economics (2017).
13. Hurlstone, M. J. *et al.* Cooperation studies of catastrophe avoidance: Implications for climate negotiations. *Climatic Change* **140**(2), 119–133 (2017).
14. Brown, T. C., & Kroll, S. Avoiding an uncertain catastrophe: Climate change mitigation under risk and wealth heterogeneity. *Climatic Change* **141**, 155–166 (2017).
15. Barrett, S. & Dannenberg, A. An experimental investigation into 'pledge and review' in climate negotiations. *Climatic Change* **138**, 339–351 (2016).
16. Rogelj. J. *et al.* Paris Agreement climate proposals need a boost to keep warming well below 2°C. *Nature* **534**, 631–639 (2016).
17. Robiou du Pont, Y. *et al.* Equitable mitigation to achieve the Paris Agreement goals. *Nat. Clim. Change.* **7**, 38–43 (2017).
18. UNFCCC. *Synthesis Report on the Aggregate Effect of the Intended Nationally Determined Contributions.* Report No. FCCC/CP (UNFCCC, 2015).
19. Milinski, M. *et al.* The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc. Natl. Acad. Sci. USA* **105**(7), 2291–2294 (2008).
20. Tavoni, A. *et al.* Inequality, communication and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci. USA* **108**(29), 11825–11829 (2011).
21. Barrett, S. *Environment and statecraft: The strategy of environmental treaty-making* (Oxford Univ. Press, 2003).
22. Barrett, S. Collective action to avoid catastrophe: When countries succeed, when they fail, and why. *Glob. Policy.* **7**, 45–55, (2016).
23. Barrett, S., & Danneberg, A. Negotiating to avoid 'gradual' vs 'dangerous' climate change: An experimental test of two prisoners' dilemmas', in T. Cherry, J. Hovi and D. M. McEvoy (eds) *Towards a new climate agreement: Conflict, resolution, and governance*. London: Routledge, pp. 61–75 (2014).
24. Barrett, S. Solar geoengineering's brave new world: Thoughts on the governance of an unprecedented technology. *Rev. Env. Econ. Policy.* **8**, 249–269, (2014).
25. Barrett, S. *et al.* Climate engineering reconsidered. *Nat. Clim. Change.* **4**, 527–529 (2014).
26. Lenton, T. M. Can emergency geoengineering really prevent climate tipping points? In *Geoengineering our Climate?* (pp. 67–70). Routledge, 2018.
27. Schelling, T. C. Norms, conventions, and institutions to cope with climate change. In *Climate change and common sense: Essays in honour of Tom Schelling*. Oxford University Press, 2012.

28. Andrews, T. M. *et al.* High-risk high-reward investments to mitigate climate change. *Nat. Clim. Change.* **8**, 890–894 (2018).

29. Gosnell, G. Climate policy: A risk-seeking future. *Nat. Clim. Change.* **8**, 855–860 (2018).

30. Barrett, S. Choices in the climate commons. *Science* **362**, 1217 (2018).

31. Greiner, B. Subject pool recruitment procedures: Organizing experiments with ORSEE. *J. Econ. Sci. Assoc* **1**(1), 114–125 (2015).

32. Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ* **10**, 171–178 (2007).