SUPPLEMENTARY MATERIALS:

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

This supplementary document reports additional details about the study conducted by[1] examining the impact of early warning signals on cooperation under threshold uncertainty in a dangerous climate change game. The document includes the instructions and control questions given to participants, a game-theoretic model of the experiment, and ancillary statistical analyses of the experimental results. Note that this document is not meant to be self-explanatory—please consult *authors removed for masked review*[1] for further information.

1 Supplementary Experimental Instructions (adapted from[2])

Certainty Treatment

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 players make. Regardless of the gameplay, you will receive $10 for your participation. For a successful experiment, it is necessary that you do not talk to other participants or do not communicate in any other way. Now please read the following rules of the game carefully. If you have any questions, please raise your hand.

2. Game Rules

There are six players in the game, meaning you and five other players. Each player is faced with the same decision problem. In the beginning of the experiment, you receive a starting capital of $40, which is credited to your personal account. During the experiment, you can use the money in your account or let it be. In the end, your current account balance is paid to you in cash. Your decisions are anonymous. For the purpose of anonymity, you will be allocated a pseudonym which will be used for the whole duration of the game. The pseudonyms are chosen from the names of moons in the Solar System (Ananke, Telesto, Despina, Japetus, Kallisto or Metis). Once the game begins you will be able to see your pseudonym in the lower left corner of your display.

The experiment has exactly ten rounds. In each round, you can invest your money in order to try and prevent damage. The damage will have a considerable negative financial impact on all players. In each round of the game, all six players are asked the following question at the same time:

“How much do you want to invest to prevent damage?”

You can answer with $0, $2, or $4. After each player has made her or his decision, the six decisions are displayed at the same time. After that, all money paid by the players is assigned to a special account for damage prevention.

At the end of the game (after exactly ten rounds), the computer calculates the total investments made by all players of the group. If the total investments are equal to or greater than a threshold amount, the damage is prevented and each player is paid the money remaining in her or his account, meaning the $40 starting capital minus the money the player has invested in preventing damage over the course of the game. However, if the total investments are lower than the threshold amount, the damage occurs: All players lose 90% of the remaining money in their personal accounts. The threshold amount to be reached in order to prevent damage is $120.

At the end of the game all players together must have invested at least $120 to prevent the damage. If a single player has invested, say, a total of $10 in damage prevention after ten rounds, he or she has a credit of $30 on his or her personal account. If the group of players as a whole has invested $120 or more in damage prevention, the damage will not occur and this player will receive $30 from the game. However, if the group has invested less than $120, the damage will occur and the player will receive $3 (10% of $30) from the game.

Please note the following feature of the game: Before the players decide how much they want to invest into preventing damage, they make two non-binding announcements. First, each player makes a proposal for how much the group as a whole should invest into preventing damage over the total of ten rounds. Second, each player makes a pledge for how much money they intend to invest in total over the ten rounds into preventing damage. After these two non-binding announcements, the proposals and pledges made by all players (and an average and total value from all proposals and pledges, respectively) will be shown on the monitor. At the end of round 5, all players can make a new non-binding proposal for the total investments to be made by the group over the ten rounds, and a new non-binding pledge for how much money they intend to invest in total over the ten rounds.

3. An Example

Here, you can see an example of the decisions made by the six players in one round (round 3). Please direct your attention first to the far right of the graphic.

Proposals Rounds 1-10		Pledges Rounds 1-10		Investments Rounds 1-3		Investments Round 3	
Ananke	100	Ananke	10	Ananke	6	Ananke	2
Telesto	80	Telesto	12	Telesto	6	Telesto	4
Despina	120	Despina	20	Despina	4	Despina	0
Japetus	100	Japetus	12	Japetus	10	Japetus	4
Kallisto	110	Kallisto	22	Kallisto	4	Kallisto	2
Metis	140	Metis	24	Metis	4	Metis	0
Average	108	Total	100	Total	34	Total Round 3	12

The fourth column shows the investments made in the current round (round 3). The players Ananke and Kallisto have invested $2 each, the players Telesto and Japetus have invested $4 each and Despina and Metis have not made any investments. In total, $12 were invested in this round. The third column shows the cumulative investments made by each player from the first to the current round (rounds 1–3). The players Ananke and Telesto have each invested $6 in the first three rounds. Despina, Kallisto and Metis have each invested $4 and Japetus has invested $10 in the first three rounds. In total, $34 were invested in the first three rounds.

The first column shows the proposals made by each player regarding how much the group as a whole should invest into preventing damage over the ten rounds in total. For example, Metis suggests that the group should invest $140. The average of all proposals is $108. The second column shows the pledges made by each player regarding how much they will personally invest in the damage prevention account over the ten rounds in total. For example, over the ten rounds Kallisto has pledged to personally invest $22 in total. The total of all pledges is $100. In the game, you will see this information after each round.

4. Control Questions

1. How much does a player have to invest on average over the course of ten rounds, if the group was to invest $120 in total?

☐ $10 ☐ $12 ☐ $20 ☐ $30 ☐ $60

2. Assume the group has invested the threshold amount to prevent damage, and that you have invested $16 in total. How much cash do you get at the end of the game (excluding the $10 participation fee)?

I get $ ____________.

3. Take a look at the table in part 3 of the instructions.

(a) How much did Ananke and Kallisto propose the group should invest in damage prevention over the ten rounds?

Ananke proposed $____________. Kallisto proposed $____________.

(b) How much did Japetus and Metis pledge to invest in damage prevention over the ten rounds?

Japetus pledged $____________. Metis pledged $____________.

(c) How much money do Despina and Japetus have in their personal accounts after round 3?

Despina has $____________ in her account. Japetus has $____________ in his account.

4. True or false? At the start of the game, and once again at the end of round 5, each player makes: (I) a non-binding proposal of how much the group should collectively invest in damage prevention over the ten rounds, and (II) a non-binding pledge of how much they will personally invest in damage prevention over the ten rounds.

☐ True ☐ False

5. Assume you invested a total of $20 over the ten rounds and the threshold amount was not reached by your group. How much cash do you get at the end of the game (excluding the $10 participation fee)?

☐ $0 ☐ $2 ☐ $4 ☐ $10 ☐ $20

6. Assume that the group has invested a total of $100 over the ten rounds. Does the damage occur in this case? (please tick the correct answer).

☐ Yes ☐ No

7. Assume that the group has invested a total of $125 over the ten rounds. Does the damage occur in this case? (please tick the correct answer).

☐ Yes ☐ No

Please raise your hand after you have answered all control questions. We will come to you and check the answers. The game will begin after we have checked the answers of all players and answered any questions you may have. Good luck!

Uncertainty, Warning-Wide, and Warning-Narrow Treatments

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 players make. Regardless of the gameplay, you will receive $10 for your participation. For a successful experiment, it is necessary that you do not talk to other participants or do not communicate in any other way. Now please read the following rules of the game carefully. If you have any questions, please raise your hand.

2. Game Rules

There are six players in the game, meaning you and five other players. Each player is faced with the same decision problem. In the beginning of the experiment, you receive a starting capital of $40, which is credited to your personal account. During the experiment, you can use the money in your account or let it be. In the end, your current account balance is paid to you in cash. Your decisions are anonymous. For the purpose of anonymity, you will be allocated a pseudonym which will be used for the whole duration of the game. The pseudonyms are chosen from the names of moons in the Solar System (Ananke, Telesto, Despina, Japetus, Kallisto or Metis). Once the game begins you will be able to see your pseudonym in the lower left corner of your display.

The experiment has exactly ten rounds. In each round, you can invest your money in order to try and prevent damage. The damage will have a considerable negative financial impact on all players. In each round of the game, all six players are asked the following question at the same time:

"How much do you want to invest to prevent damage?"

You can answer with $0, $2, or $4. After each player has made her or his decision, the six decisions are displayed at the same time. After that, all money paid by the players is assigned to a special account for damage prevention.

At the end of the game (after exactly ten rounds), the computer calculates the total investments made by all players of the group. If the total investments are equal to or greater than a threshold amount, the damage is prevented and each player is paid the money remaining in her or his account, meaning the $40 starting capital minus the money the player has invested in preventing damage over the course of the game. However, if the total investments are lower than the threshold amount, the damage occurs: All players lose 90% of the remaining money in their personal accounts. The threshold amount to be reached in order to prevent damage is some amount between $0 and $240, but you will not know the exact amount until the conclusion of the game. At the end of the experiment, the exact threshold amount will be drawn randomly by the computer. The draw is programmed so that each whole dollar amount between $0 and $240 has an equal probability of being selected.

Suppose at the end of the game that the randomly drawn threshold amount is $100. All players together must have invested at least $100 to prevent the damage. If a single player has invested, say, a total of $10 in damage prevention after ten rounds, he or she has a credit of $30 on his or her personal account. If the group of players as a whole has invested $100 or more in damage prevention, the damage will not occur and this player will receive $30 from the game. However, if the group has invested less than $100, the damage will occur and the player will receive $3 (10% of $30) from the game.

Please note the following feature of the game: Before the players decide how much they want to invest into preventing damage, they make two non-binding announcements. First, each player makes a proposal for how much the group as a whole should invest into preventing damage over the total of ten rounds. Second, each player makes a pledge for how much money they intend to invest in total over the ten rounds into preventing damage. After these two non-binding announcements, the proposals and pledges made by all players (and an average and total value from all proposals and pledges, respectively) will be shown on the monitor. At the end of round 5, all players can make a new non-binding proposal for the total investments to be made by the group over the ten rounds, and a new non-binding pledge for how much money they intend to invest in total over the ten rounds.

3. An Example

Here, you can see an example of the decisions made by the six players in one round (round 3). Please direct your attention first to the far right of the graphic.

The fourth column shows the investments made in the current round (round 3). The players Ananke and Kallisto have invested $2 each, the players Telesto and Japetus have invested $4 each and Despina and Metis have not made any investments. In total, $12 were invested in this round. The third column shows the cumulative investments made by each player from the first to the

Proposals Rounds 1-10	
Ananke	100
Telesto	80
Despina	120
Japetus	100
Kallisto	110
Metis	140
Average	108

Pledges Rounds 1-10	
Ananke	10
Telesto	12
Despina	20
Japetus	12
Kallisto	22
Metis	24
Total	100

Investments Rounds 1-3	
Ananke	6
Telesto	6
Despina	4
Japetus	10
Kallisto	4
Metis	4
Total	34

Investments Round 3	
Ananke	2
Telesto	4
Despina	0
Japetus	4
Kallisto	2
Metis	0
Total Round 3	12

current round (rounds 1–3). The players Ananke and Telesto have each invested $6 in the first three rounds. Despina, Kallisto and Metis have each invested $4 and Japetus has invested $10 in the first three rounds. In total, $34 were invested in the first three rounds.

The first column shows the proposals made by each player regarding how much the group as a whole should invest into preventing damage over the ten rounds in total. For example, Metis suggests that the group should invest $140. The average of all proposals is $108. The second column shows the pledges made by each player regarding how much they will personally invest in the damage prevention account over the ten rounds in total. For example, over the ten rounds Kallisto has pledged to personally invest $22 in total. The total of all pledges is $100. In the game, you will see this information after each round.

4. Control Questions

1. How much does a player have to invest on average over the course of ten rounds, if the group was to invest $60 in total?

☐ $10 ☐ $12 ☐ $20 ☐ $30 ☐ $60

2. How much does a player have to invest on average over the course of ten rounds, if the group was to invest $180 in total?

☐ $10 ☐ $12 ☐ $20 ☐ $30 ☐ $60

3. Assume the group has invested the threshold amount to prevent damage, and that you have invested $16 in total. How much cash do you get at the end of the game (excluding the $10 participation fee)?

I get $ ____________.

4. Take a look at the table in part 3 of the instructions.

(a) How much did Ananke and Kallisto propose the group should invest in damage prevention over the ten rounds?

Ananke proposed $____________. Kallisto proposed $____________.

(b) How much did Japetus and Metis pledge to invest in damage prevention over the ten rounds?

Japetus pledged $____________. Metis pledged $____________.

(c) How much money do Despina and Japetus have in their personal accounts after round 3?

Despina has $___________ in her account. Japetus has $___________ in his account.

5. True or false? At the start of the game, and once again at the end of round 5, each player makes: (I) a non-binding proposal of how much the group should collectively invest in damage prevention over the ten rounds, and (II) a non-binding pledge of how much they will personally invest in damage prevention over the ten rounds.

☐ True ☐ False

6. True or false? In the random draw to determine the threshold amount at the end of the game, each whole dollar amount between $0 and $240 has the same probability of being selected.

☐ True ☐ False

7. Assume you invested a total of $20 over the ten rounds and the threshold amount was not reached by your group. How much cash do you get at the end of the game (excluding the $10 participation fee)?

☐ $0 ☐ $2 ☐ $4 ☐ $10 ☐ $20

8. Assume that the group has invested a total of $100 over the ten rounds. The draw shows that the threshold amount to avoid damage is $160. Does the damage occur in this case? (please tick the correct answer).

☐ Yes ☐ No

9. Assume that the group has invested a total of $80 over the ten rounds. The draw shows that the threshold amount to avoid damage is $20. Does the damage occur in this case? (please tick the correct answer).

☐ Yes ☐ No

10. What is the probability of the threshold amount to prevent damage being greater than $60? ___________.

11. What is the probability of the threshold amount to prevent damage being greater than $180? ___________.

Please raise your hand after you have answered all control questions. We will come to you and check the answers. The game will begin after we have checked the answers of all players and answered any questions you may have. Good luck!

2 Analysis of Experimental Model

The imperfect information, repeated and multiple-player structure of the experimental game allows for multiple Nash equilibria, and this complexity precludes a full equilibrium analysis. Therefore, in this section, we analyse the game under a set of simplifying assumptions and focus on two solutions—the internal cooperative equilibrium and Nash equilibrium. This is possible because the game has a single pay-off period at the end of the game and can therefore be partially analysed as an equivalent one-shot static game. Barrett and Dannenberg[3] provide a similar analysis of such a game. The range of Nash responses is explored diagrammatically in the final section. In addition to the internal solution, there are Nash equilibria where all players contribute zero. There are also Nash responses where, on the one hand, other players contribute just less than the Nash equilibrium, and a player has an incentive to contribute more than the others; on the other hand, there are Nash responses where other players contribute more than the Nash equilibrium, and a player has an incentive to reduce their contribution and "free-ride".

Recall from the experimental design that the threshold does not change during the course of the game in the certainty ($120) and uncertainty ($0-$240) treatments, whereas in the early warning treatments the threshold range is initially the same as in the uncertainty treatment, but is then reduced by 70% in the warning-wide treatment ($84-$156) and 90% in the warning-narrow treatment ($108-$132) at the start of round 6. According to our model, it is the final state of the threshold that matters in terms of total group contributions. Accordingly, the analyses that follow are based on the new threshold ranges announced at the mid-point of the experimental game in the warning-wide and warning-narrow treatments. We consider how expected contributions should differ before and after the announcement of the new threshold range in *authors removed for masked review*.[1]

2.1 Model simplifying assumptions

The model simplifying assumptions are as follows: (1) all N players (countries) are identical and risk-neutral; (2) in each period players contribute q_{it} which is the same in each round; and (3) the total contribution (abatement) of N countries over T rounds i is:

$$Q_T = \sum_{i=1}^{N} \sum_{t=1}^{T} q_{it}$$

2.2 Model structure

The structure of the model is as follows. The climate tipping point is distributed $\tilde{Q} \sim f(\bar{Q}, \varepsilon)$, where the probability density is given by $f(\bar{Q}, \varepsilon)$, $\bar{Q}$ is the mean of the distribution and ε the dispersal around the mean in a uniform distribution. The probability of avoiding a tipping point is given by the cumulative distribution:

$$\mathrm{P}(Q \leq \sum_{i=1}^{N} Q_{iT}) = F(\bar{Q}, \varepsilon, \sum_{i=1}^{N} Q_{iT})$$

The payoff depends upon whether total contributions exceed the realised threshold:

$$J_i(Q_{iT}|Q_{-iT}, \tilde{Q}) = \begin{cases} \tau(X_{i0} - Q_{iT}) & Q_{iT} + Q_{-iT} < \tilde{Q} \\ X_{i0} - Q_{iT} & Q_{iT} + Q_{-iT} \geq \tilde{Q} \end{cases}$$

Relating this to the notation for a uniform distribution:

$$E[J_i(Q_{iT}|Q_{-iT}] = \begin{cases} \tau(X_{i0} - Q_{iT}) & Q_{iT} + Q_{-iT} < Q_{min} \\ (X_{i0} - Q_{iT})((1-\tau)F(\bar{Q}, \varepsilon, Q_{iT} + Q_{-iT}) + \tau) & Q_{iT} + Q_{-iT} \in [Q_{min}, Q_{max}] \\ (X_{i0} - Q_{iT}) & Q_{iT} + Q_{-iT} > Q_{max} \end{cases}$$

The payoff is restricted by contribution constraints in each period and in total. These limit the overall contribution and the rate at which the player can respond in a period. Thus, following Barrett and Dannenberg,[3] $q_{it} \in [0, q_{max}]$. Over the planning horizon, the maximum contribution by a player is Tq_{max}. There is a further implicit constraint imposed in the game that $Q_{iT} \in [0, X_{i0}]$, that is, the total contribution cannot exceed the initial endowment.

The pay-off for player i at the end of the planning horizon is the initial endowment X_{i0} less the cumulative contributions up to the end of period T, that is Q_{iT}. If the total contributions from all players are more than or equal to the threshold, the payoff is $X_{i0} - Q_{iT}$; if less than the threshold $\tau(X_{i0} - Q_{iT})$, where $0 \leq \tau \leq 1$ determines the penalty related to not achieving the realised threshold. The expected payoff for a player is given by:

$$E[J_i(Q_{it}|Q_{-it}, \tilde{Q})] = (X_{i0} - Q_{iT})F(\bar{Q}, \varepsilon, Q_{it} + Q_{-it}) + \tau(X_{i0} - Q_{iT})(1 - F(\bar{Q}, \varepsilon, Q_{it} + Q_{-it}))$$

This equation can be rearranged so that the expected benefits term and expected costs are separated:

$$E[J_i(Q_{it}|Q_{-it},\tilde{Q})] = X_{i0}(\tau + (1-\tau)F(\bar{Q},\varepsilon,Q_{iT}+Q_{-iT})) - Q_{iT}(\tau+(1-\tau)F(\bar{Q},\varepsilon,Q_{iT}+Q_{-iT}))$$

The cooperative solution is a special case where the above equation is maximised for a group of players that act as a single entity, and the initial endowment is defined by $X_0 = X_{i0}N$:

$$E[J_i(Q_{iT}|Q)] = X_0(\tau+(1-\tau)F(\bar{Q},\varepsilon,Q_{iT})) - Q_{iT}(\tau+(1-\tau)F(\bar{Q},\varepsilon,Q_{iT}))$$

2.3 Case 1: Internal Cooperative Solution and Nash equilibria

Taking derivatives of the above equation with respect to Q_{iT} and assuming an internal solution yields a marginal condition:

$$X_0(1-\tau)f(\bar{Q},\varepsilon,Q_{iT}) = (\tau+(1-\tau)F(\bar{Q},\varepsilon,Q_{iT})) + Q_{iT}(\tau+(1-\tau)f(\bar{Q},\varepsilon,Q_{iT}))$$

If we substitute in a cumulative uniform distribution with a lower bound $Q_{min} = \bar{Q}-\varepsilon$ and an upper bound $Q_{max} = \bar{Q}+\varepsilon$ the optimal cooperative solution is:

$$Q_T^C = \frac{1}{2}\left(\frac{(Q_{min}-\tau Q_{max})}{(1-\tau)} + X_0\right) = \frac{1}{2}\left(\frac{((\bar{Q}-\varepsilon)-\tau(\bar{Q}+\varepsilon))}{(1-\tau)} + X_0\right) \quad \text{(S1)}$$

$$Q_T^C = \frac{1}{2}(G(\bar{Q},\varepsilon,\tau)+X_0); \text{where } G(\bar{Q},\varepsilon,\tau) = \frac{((\bar{Q}-\varepsilon)-\tau(\bar{Q}+\varepsilon))}{(1-\tau)}$$

The cooperative solution scales as the number of players varies to give $Q_{iT}^C = Q_T^C/N$.

The Nash equilibrium contribution of a single player is:

$$Q_{iT}^N = \frac{(Q_{min}-\tau Q_{max})}{(1+N)(1-\tau)} + \frac{X_{i0}}{(1+N)}$$

$$Q_{iT}^N = \frac{1}{(1+N)}\left(G(\bar{Q},\varepsilon,\tau)+X_{i0}\right)$$

This result is multiplied by N to give the total contribution of all players; we substitute $X_0 = NX_{i0}$:

$$Q_T^N = NQ_{iT}^N = \frac{N}{(1+N)}\left(G(\bar{Q},\varepsilon,\tau)+X_{i0}\right) = \frac{N}{(1+N)}G(\bar{Q},\varepsilon,\tau) + \frac{X_0}{(1+N)} \quad \text{(S2)}$$

To show that: $Q_T^C - Q_T^N \geq 0$

We solve (S1) and (S2) for $\left(G(\bar{Q},\varepsilon,\tau)+X_{i0}\right)$ and equate:

$$Q_T^C = \frac{1+N}{2}Q_T^N - \frac{(N-1)}{2}G(\bar{Q},\varepsilon,\tau)$$

Assuming an internal solution, where $0 < Q_T^C < Q_{max}$ and $0 < Q_T^N < Q_{max}$, $Q_T^C - Q_T^N \geq 0$ holds unambiguously in the case where $N \geq 2$ and $G(\bar{Q},\varepsilon,\tau) < 0$.

For instance, in the uncertainty treatment:

$G(\bar{Q},\varepsilon,\tau) = (0-(0.1\times 240))/(1-0.1) = -26.67$

$Q_T^C = \frac{7}{2}Q_T^N - \frac{5}{2}(-26.67)$

$Q_T^N = 11.43$

$Q_T^C = 106.67$

2.4 Case 2: A deterministic threshold

A deterministic threshold, where $\varepsilon = 0$ and the threshold is $\bar{Q}$, results in two "corner point" solutions. The cooperative solution, where the countries contribute their fair share to avoid crossing the threshold, is $Q_{iT}^C = \bar{Q}/N$. There are two focal Nash equilibria, either $Q_{iT}^N = 0$ or $Q_{iT}^N = \bar{Q} - Q_{iT}$. The last equilibrium arises when the contributions of the other players is high enough for the player to "top-up" the contributions of the other players to reach the deterministic threshold.

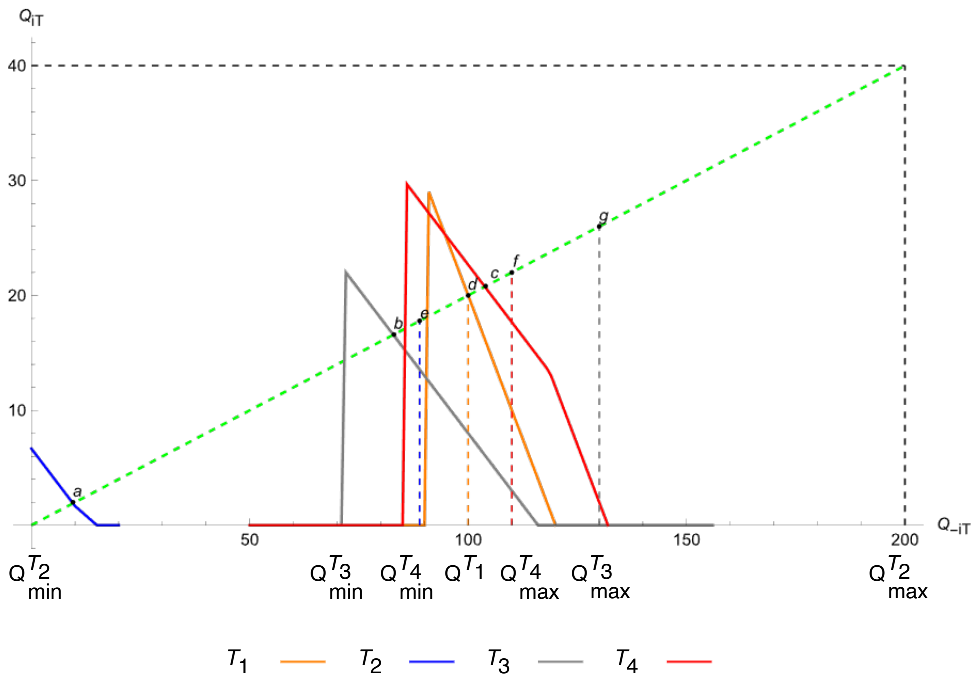

Figure S1 | Nash response curves for the four treatments.

2.5 Case 3: Corner point solutions

There are two corner point solutions in this problem, with one Nash equilibrium at the lower bound and the other at the zero-risk point. This is illustrated in Table 2 of *authors removed for masked review*[1] and Fig. S1 (these are specific solutions relating to the treatments applied). The horizontal axis in Fig. S1 gives the contributions of all players except player *i*, whereas the vertical axis gives the contribution of player *i*. The parameters Q^d_{min} and Q^d_{max} ($d = T_1$ [certainty], T_2 [uncertainty], T_3 [warning wide], T_4 [warning narrow]) give the ranges for the uniform distribution adjusted on the basis that player *i* contributes their share. For the deterministic (certainty) treatment, the parameter Q^{T_1} gives the deterministic target of 100 for the other players.

For each treatment, the Nash response of the player follows a "sawtooth" pattern. Below the internal Nash equilibria (*a*, *b*, *c* and *d*), where all players contribute the same positive amount, this point is found where the Nash response curves cross the green dashed line. At very low contributions, the Nash response is to contribute zero. All treatments have a Nash equilibrium where all players contribute zero.

As contributions rise from other players, there is a threshold where player *i* contributes more than their share, as defined by the Nash equilibrium. As contributions increase further, then player *i*'s contribution declines and reaches the Nash equilibrium. If the contribution of the other players continues to increase, player *i*'s contribution is zero at some point.

The blue Nash response curve is the optimal response by player *i* to a range of contributions by the other players when the range is from \$0 to \$240. Point *a* is a Nash equilibrium where all players contribute the same amount. This should be compared with the cooperative solution at *e*. For the warning-wide treatment (T_3), the grey line gives the Nash response, *b* is the Nash equilibrium, and *g* is the cooperative solution which, in this case, eliminates risk as the total contributions = \$156. Similarly, for the warning-narrow treatment (T_4), the red line gives the Nash response, *c* is the Nash equilibrium, and *f* is the cooperative solution. For the certainty treatment (T_1), the fixed target has a Nash equilibrium and cooperative solution that coincide at *d*.

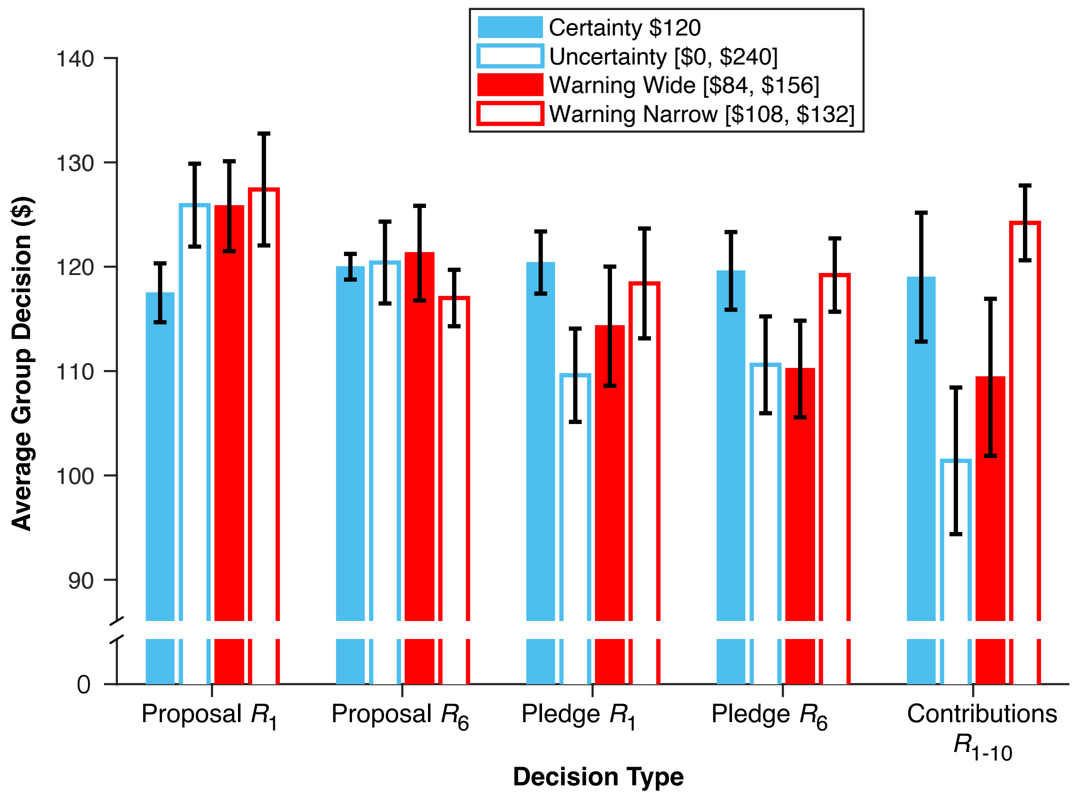

Figure S2 | Group proposals, pledges, and contributions as a function of treatment. Error bars represent standard errors.

3 Supplementary Statistical Analyses

3.1 Proposals, pledges, and contributions

Fig. S2 shows the average group proposals and pledges in rounds 1 and 6 and group contributions (collapsed over the ten rounds) by treatment. Group proposals in rounds 1 and 6 hovered closely around the $120 mark in all instances, confirming that this was the focal[4] threshold value in all treatments. In the certainty and warning-narrow treatments, group pledges are generally consistent with the proposed group amounts, except that in the warning-narrow treatment, in round 1, group pledges are lower than group proposals, whereas, in the uncertainty and warning-wide treatments, group pledges are lower than group proposals in both rounds. There were no significant differences between treatments for group proposals in round 1 (Kruskal-Wallis, $\chi^2_{df=3}$ = 5.84, P = .119) or round 6 (Kruskal-Wallis, $\chi^2_{df=3}$ = 1.98, P = .576). However, collapsing across treatments, group proposals in round 6 (119.7 $\pm$ 1.63) were slightly, but significantly, lower than in round 1 (124.15 $\pm$ 2.11) (Wilcoxon Signed-Rank, W = 409.00, P = .022). There were no significant differences between treatments for group pledges in round 1 (Kruskal-Wallis, $\chi^2_{df=3}$ = 3.74, P = .291) or round 6 (Kruskal-Wallis, $\chi^2_{df=3}$ = 4.31, P = .230). Collapsing across treatments, there was no significant difference between group pledges in round 1 (115.68 $\pm$ 2.36) and round 6 (114.90 $\pm$ 2.12) (Wilcoxon Signed-Rank, W = 280.00, P = .537). Finally, there was a significant difference in group contributions as a function of treatment (Kruskal-Wallis, $\chi^2_{df=3}$ = 8.00, P = .046). Contributions were significantly lower in the uncertainty treatment than the certainty treatment (Mann-Whitney, 80.50, P = .023), and although contributions did not differ significantly between the uncertainty and warning-wide treatments (Mann-Whitney, 39.00, P = .427), contributions were significantly higher in the warning-narrow treatment than the uncertainty treatment (Mann-Whitney, 15.50, P = .010). Thus, an early warning signal that reduced uncertainty to within 10% of the threshold value spurred group contributions, whereas an early warning signal that reduced uncertainty to within 30% of the threshold value exerted no effect on group contributions.

To determine if proposals and pledges were consequential with respect to actual contributions, we conducted a linear regression, with group contributions as the dependent measure and proposals and pledges in rounds 1 and 6 as predictors. The resulting model was significant, compared to a constant-only model, $F(4, 35) = 9.48$, $P < .001$. The results of the analysis are shown in Table S1, from which it can be seen that only group pledges in round 6 ($P < .001$) were a reliable signal of actual group contributions.

Table S1 | Linear regression predicting group contributions

	Unstandardised β	**Standard Error**	**Standardised β**	***t***	***p***
Intercept	-42.47	38.20		-1.11	0.274
Proposal R_1	0.05	0.21	0.03	0.24	0.810
Proposal R_6	0.10	0.25	0.05	0.40	0.694
Pledge R_1	0.26	0.21	0.19	1.25	0.222
Pledge R_6	0.93	0.21	0.60	4.39	$< .001$

3.2 Economic preferences and contributions

To provide a further window into the factors that influenced individual contributions in the catastrophe avoidance game, participants completed an individual differences questionnaire at the end of the game which measured their risk, time, and social preferences.[5] Table S2 shows the average responses to each of the six economic preference items, which measured risk aversion, loss aversion, fairness, trust, altruism, and temporal discounting, as a function of the four treatments. Responses did not differ significantly across treatments for either of the economic preference items: (Kruskal-Wallis, $\chi^2_{df=3} = 4.55$, $P = .208$) for risk aversion, (Kruskal-Wallis, $\chi^2_{df=3} = 5.87$, $P = .118$) for loss aversion, (Kruskal-Wallis, $\chi^2_{df=3} = 0.56$, $P = .906$) for fairness, (Kruskal-Wallis, $\chi^2_{df=3} = 1.14$, $P = .768$) for trust, (Kruskal-Wallis, $\chi^2_{df=3} = 4.46$, $P = .216$) for altruism, and (Kruskal-Wallis, $\chi^2_{df=3} = 5.09$, $P = .165$) for temporal discounting. To examine whether economic preferences influenced player contributions in the catastrophe avoidance game, we conducted a linear regression with individual player contributions as the dependent measure and responses on each of the six economic preference items as predictors. The model was significant, relative to a constant-only model, $F(6, 233) = 3.21$, $P = .005$. Table S3 summarises the results for each of the six predictors, from which it can be seen that only altruism was a significant predictor ($P = .005$), with higher levels of self-reported altruism being positively associated with contributions in the catastrophe avoidance game.

Table S2 | Mean responses on the post-game economic preferences questionnaire as a function of treatment

Construct	Question	Certainty	Uncertainty	Warning Wide	Warning Narrow
Risk aversion	How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?	6.12 (2.26)	5.48 (2.39)	5.88 (2.34)	5.30 (2.17)
Loss aversion	How well does the following statement describe you as a person? I generally hate to lose something more than I like to gain something.	6.68 (2.21)	6.10 (2.06)	5.88 (2.23)	6.40 (2.52)
Fairness	Please indicate your level of agreement with the following statement: when a group of people must work toward a common goal, it is important that each group member contributes an equal amount of effort.	8.22 (2.26)	8.12 (2.36)	8.47 (1.82)	8.32 (2.27)
Trust	How well does the following statement describe you as a person? As long as I am not convinced otherwise, I assume that people have only the best intentions.	6.05 (2.53)	5.57 (2.45)	5.63 (2.69)	5.75 (2.80)
Altruism	How willing are you to help others without expecting anything in return?	7.60 (1.98)	7.23 (1.92)	6.93 (2.02)	7.28 (2.12)
Temporal discounting	How willing are you to give up something today in order to benefit from doing so in the future?	7.85 (1.64)	7.48 (1.56)	7.22 (1.68)	7.28 (1.86)

All items required a response on an eleven point scale. For the risk aversion item, participants were asked to: "Please use a scale from 0 to 10, where 0 means you are completely unwilling to take risks and 10 means you are very willing to take risks"; for the loss aversion and trust items participants were asked to: "Please use a scale from 0 to 10, where 0 means does not describe me at all and 10 means describes me perfectly"; for the fairness item participants were asked to: "Please use a scale from 0 to 10, where 0 means strongly disagree and 10 means strongly agree"; for the altruism item participants were asked to: "Please use a scale from 0 to 10, where 0 means you are completely unwilling to help others and 10 means you are very willing to help others"; for the temporal discounting item participants were asked to: "Please use a scale from 0 to 10, where 0 means you are completely unwilling to give up something today and 10 means you are very willing to give up something today".

Table S3 | Linear regression predicting group contributions

	Unstandardised *β*	Standard Error	Standardised *β*	*t*	*p*
Intercept	8.06	2.97		2.72	0.007
Risk aversion	-0.01	0.18	-0.00	-0.05	0.957
Loss aversion	0.28	0.18	0.10	1.51	0.132
Fairness	0.18	0.20	0.06	0.94	0.347
Trust	0.07	0.17	0.03	0.41	0.680
Altruism	0.64	0.23	0.20	2.83	0.005
Temporal discounting	0.35	0.24	0.09	1.42	0.157

References

1. Hurlstone, M. J. *et al.* Threshold uncertainty, early warning signals, and the prevention of dangerous climate change. *Manuscript submitted for publication* (2024).
2. Dannenberg, A. *et al.* On the provision of public goods with probabilistic and ambiguous thresholds. *Environ. Resource. Econ.* **61**, 365–383 (2015).
3. Barrett, S. & Dannenberg, A. Sensitivity of collective action to uncertainty about climate tipping points. *Nat. Clim. Change.* **4**, 36–39 (2014).
4. Schelling, T. C. *The strategy of confict.* Harvard University Press (1960).
5. Falk, A. *et al.* The preference survey module: A validated instrument for measuring risk, time, and social preferences. *SSRN Electronic Journal*, January (2016).