

Avoiding an uncertain catastrophe: climate change mitigation under risk and wealth heterogeneity

Thomas C. Brown¹ · Stephan Kroll²

Received: 22 April 2016 / Accepted: 21 December 2016 / Published online: 3 January 2017
© Springer Science+Business Media Dordrecht (outside the USA) 2017

Abstract For environmental problems such as climate change, uncertainty about future conditions makes it difficult to know what the goal of mitigation efforts should be, and inequality among the affected parties makes it hard for them to know how much they each should do toward reaching the goal. We examine the effects of scientific uncertainty and wealth inequality in experiments where subjects decide how much to contribute toward reducing a common threat. We also explore how the framing of uncertainty affects collective action. Our results suggest that uncertainty lowers contributions, but contributions remain surprisingly high even in treatments with a variable loss probability, where such behavior is individually suboptimal (and where the underlying game is a prisoner's dilemma). Further, we find that the characterization of uncertainty is crucial and that inequality need not lower contributions at all.

1 Introduction

Efforts to avoid irreversible damage to common-pool resources tend to coalesce around a critical threshold beyond which catastrophe is likely to be unavoidable, such as the 2 °C limit on global temperature rise identified in the 2009 Copenhagen Accord. Although critical tipping points, such as the temperature rise beyond which the Greenland ice sheet will melt, are envisioned, the locations of such tipping points are not known precisely. For example, the IPCC states regarding the Greenland ice sheet, “Current estimates indicate that the [tipping point] is greater than about 1°C ... but less than about 4°C ... global mean warming with

Electronic supplementary material The online version of this article (doi:10.1007/s10584-016-1889-5) contains supplementary material, which is available to authorized users.

✉ Thomas C. Brown
thomas.brown@colostate.edu

¹ Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, CO 80526, USA

² Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO 80523, USA

respect to pre-industrial” (IPCC, 2013, p. 29). The Paris Climate Accord in December 2015 essentially acknowledges that temperature thresholds are only vaguely related to such tipping points in stating that the goal is to hold “the increase in the global average temperature to well below 2°C above pre-industrial levels” and to “pursue efforts to limit the temperature increase to 1.5°C.”

Uncertainty about the threshold level of effort needed to avoid a critical tipping point can be characterized in two ways: the exact location of the threshold might be unknown, or the probability of facing a catastrophic outcome as result of mitigation efforts not reaching the threshold might not be fixed (e.g., missing the threshold by just a bit might make a collapse less likely than missing it by a lot). These two types of uncertainty could undo the positive effect of having a threshold on the likelihood of avoiding the tipping point. When the threshold is unknown, or when the extent of the catastrophe depends on how far agents’ efforts fall short of the threshold, then the effect of the threshold as a coordination device becomes weaker. The logic is simple: if the missing threshold has only a small marginal effect instead of a stark discrete effect, then the costs for each individual contributing a little less than what is needed to get to the threshold might be smaller than the benefits of reaching the threshold.

The effect of uncertainty could be even worse if the agents are heterogeneous in income. Income heterogeneity makes coordination more difficult than with a known threshold because what constitutes a fair individual contribution is more difficult to discern. Such income disparities may yet hamper success in reaching the emission targets agreed in 2015 in Paris.

2 Literature review and research questions

Public-good experiments are one way to explore the ramifications of alternative incentive structures on people’s willingness to cooperate in achieving a common goal. We examine the effects of scientific uncertainty and income inequality in the context of the dynamic threshold public-good experiment pioneered by Milinski et al. (2008). In their basic setup, subjects, in groups of six, each start with an equal endowment, of which they can contribute up to 10% to the public good in each of ten rounds. Subjects face a certain probability of losing all of their remaining funds if the group’s aggregate contribution after the ten rounds does not reach a known threshold. We use variations of this setup to address three research questions.

First, we compare the effect of a certain threshold and known probability of loss, as in Milinski et al. (2008), with that of an uncertain threshold. These two situations are fundamentally different. Introducing uncertainty about the location of a threshold can turn a coordination game back into a prisoner’s dilemma game (Barrett 2013). The theoretical prediction of different behavior in those two situations has been demonstrated empirically by Barrett and Dannenberg (2012) who found that the vast majority of groups facing a certain threshold coordinated on the “safe” equilibrium, in which aggregate contributions reached the threshold, whereas contributions of most groups in the treatment with an uncertain threshold were so low that the catastrophic outcome was inevitable. Barrett and Dannenberg (2014) extended the analysis by varying the uncertainty range and observed that large uncertainty resulted almost always in catastrophic outcomes, whereas low uncertainty about the threshold nearly always triggered high enough contributions to avoid catastrophes. Finally, Dannenberg et al. (2015) compared contributions when the threshold was known with treatments where it was uncertain but known to vary uniformly within a given range and where it was known only to fall

within that range (with no information about the distribution). They found not only that introducing uncertainty about the threshold lowered contributions but also that lack of knowledge about the distribution lowered contributions most of all. We retest the effect of uncertainty. In particular, our first research question concerns the effect of moving from a discrete chance of loss (loss occurs for sure if the threshold is not reached) to a continuous change in the chance of a loss (additional contributions incrementally lower the chance of loss).

Second, we model the uncertainty faced by subjects in two apparently quite different but, as it turns out, equivalent ways. Our effort here is motivated in part by the finding that the way scientific uncertainty is presented in the context of future climate change can influence how the public perceives the actual problem and responds to it (Shackley and Wynne 1996). In one case, the threshold is known and fixed, as in the certain threshold treatment, but the loss probability is a function of the contribution level—aggregate contributions close to the threshold result in a small loss probability, whereas low contribution levels make a loss more likely. In the other case, we keep the loss probability constant at 1—subjects will lose their remaining funds with certainty if the group does not reach the threshold—but the threshold itself is uncertain (it is a random variable, uniformly distributed over a range from 0 to a known maximum). As implemented, these different frames yield identical payoff structures. However, the former framing makes the probability of loss uncertain whereas the latter makes the location of the threshold uncertain. Our second research question is then whether the framing has an impact on subjects' and groups' behavior. We expect that when the threshold itself is uncertain, as opposed to being known, the lack of a specific goal toward which participants can aim will result in lower contributions. Goal specificity is recognized, in the workplace and elsewhere, as an important determinant of successful performance (Locke and Latham 2002).

Third, since income inequality is of critical importance in climate negotiations, we examine for our last research question whether heterogeneity in endowments has a positive, negative, or no impact at all on contributions to public goods. Income inequality has been of considerable interest in experimental economics: some linear public good experiments found a negative effect of income inequality (e.g., Cherry et al. 2005; van Dijk et al. 2002), others found a positive effect (e.g., Chan et al. 1996; Reuben and Riedl 2013), and another found no effect (Chan et al. 1999). A few papers also examined the effect of heterogeneous endowments in threshold public-good games and found no or only a negative effect (e.g., Croson and Marks 2001; Rapoport 1988; Rapoport et al. 1989; Rapoport and Suleiman 1993). Further, the few papers that examined endowment heterogeneity in collective-risk social dilemmas arrived at different conclusions: one (Tavoni et al. 2011) observed that heterogeneity lowered the rate at which groups reached a known threshold, but two others (Burton-Chellew et al. 2013; Milinski et al. 2011) did not find a significant effect.

3 Experimental design and expectations

To examine these three research questions, we implement six treatments of a threshold public-good game. Among the six treatments we cross two factors, the *Endowment* that players begin with (either equal or unequal) and the *Threat* they face. Our three threat configurations are when the threshold is known and the probability of loss, if the threshold is not reached, is fixed (H1), when the threshold is known but the loss probability is variable (H2), and when the threshold is unknown but the loss probability is fixed (H3) (Table 1).

Table 1 Experimental design

Treatment	Number of groups	Endowment (\$)	Threat name	Threat configuration	
				Threshold (\$)	Probability of loss ^a
T1	10	Equal (40)	H1	Known (120)	Fixed (0.5)
T2	12	Equal (40)	H2	Known (120)	Variable (0–1) ^c
T3	10	Equal (40)	H3	Unknown (0–120) ^b	Fixed (1.0)
T4	10	Unequal (30 or 50)	H1	Known (120)	Fixed (0.5)
T5	11	Unequal (30 or 50)	H2	Known (120)	Variable (0–1) ^c
T6	10	Unequal (30 or 50)	H3	Unknown (0–120) ^b	Fixed (1.0)

^a Probability of loss if the threshold is not reached

^b Chosen randomly from a uniform distribution

^c Decreases linearly in C_{10} up to $C_{10} = 120$; constant at 0 for $C_{10} > 120$

Each treatment consists of three independent and identical rounds, with ten periods per round, administered to groups of six subjects. At the end of each period, contributions of all players are displayed along with the group's total contribution up to that point. Subjects know that if the sum of their contributions to the group account at the end of period 10, C_{10} , does not reach the threshold X , they will face a probability p of losing all of their remaining funds. With c_{i10} as the sum of player i 's contributions to the public good over the ten periods and E as the original endowment, player i 's payoff is

$$EV(\pi_i) = \begin{cases} (1-p)(E-c_{i10}) & \text{if } C_{10} < X \\ E - c_{i10} & \text{if } C_{10} \geq X \end{cases} \quad (1)$$

This setup is designed to imitate the real-world situation where countries face major potential loss and have the opportunity to avoid the loss, but only at a significant cost to most, if not all, of them.

We first describe treatments T1–T3, in which all endowments are equal, and then indicate how treatments T4–T6 (unequal endowments) differ from T1–T3 (Table 1). In treatments T1–T3, all six subjects in a group start each round with an endowment of \$40, of which they can contribute \$0, \$2, or \$4 to the group account in each of the ten periods (all currency amounts are in U.S. dollars). Treatment T1 replicates the Milinski et al. (2008) experiment, with a certain threshold of $X = \$120$ and a certain loss probability of $p = 0.5$ with which all group members lose their remaining funds if the group's contributions do not reach the threshold. Thus, in T1, player i 's expected payoff (π_i) is

$$EV(\pi_i) = \begin{cases} .5(40 - c_{i10}) & \text{if } C_{10} < 120 \\ 40 - c_{i10} & \text{if } C_{10} \geq 120 \end{cases} \quad (2)$$

As usual in threshold public-good games of this kind, there are potentially many Nash equilibria. In T1, one Nash equilibrium is for every subject to contribute 0 over the entire round, because unless the group contributions are sufficient to reach the threshold any individual contribution lowers the subject's expected payoff. With the given parameters, a second kind of Nash equilibrium exists, in which the group's total contribution exactly reaches the threshold of \$120. For risk-neutral subjects, there is only one such equilibrium—everybody contributes \$20—since risk-neutral subjects would always prefer to keep \$40 with a loss probability of 0.5 than to keep an amount less than \$20 with certainty. For risk-averse subjects,

however, there are also other equilibria with individual contributions other than \$20 as long as the group's total contribution reaches \$120. In all cases, risk-averse subjects prefer the certain outcome with the group's contributions reaching the threshold over the 50-50 gamble. However, that preference alone is insufficient incentive to keep a subject contributing in all periods within a round. Once it becomes clear to a subject that the group is unlikely to reach the threshold, even a risk-averse subject may abandon the cause.

In treatments T2 and T3, uncertainty about the loss is more complex than in T1. In T2, the threshold is still \$120, which is known to subjects, but the loss probability is no longer fixed at 0.5; rather, it is a linear function of the group's total contribution, decreasing from 1 for a C_{10} of \$0 to 0 for a C_{10} of \$120. To be precise, the loss probability (p) is

$$p = \begin{cases} \frac{120 - C_{10}}{120} & \text{if } C_{10} < 120 \\ 0 & \text{if } C_{10} \geq 120 \end{cases} . \quad (3)$$

Therefore, player i 's expected payoff is:

$$EV(\pi_i) = \begin{cases} (1-p)(40 - c_{i10}) & \text{if } C_{10} < 120 \\ 40 - c_{i10} & \text{if } C_{10} \geq 120 \end{cases} . \quad (4)$$

In T3, the loss probability if the group fails to reach the threshold is fixed at $p = 1$ (so loss is certain if the threshold is not reached), but this time the threshold itself (X) is uncertain and randomly drawn from a uniform distribution between 0 and 120. Given an aggregate contribution C_{10} smaller than \$120, the probability of it being larger than X is $C_{10}/120$, and thus, the effective probability of loss when $C_{10} < 120$ is $1 - C_{10}/120$, such that

$$p = \begin{cases} 1 - \frac{C_{10}}{120} & \text{if } C_{10} < 120 \\ 0 & \text{if } C_{10} \geq 120 \end{cases} , \quad (5)$$

which is identical to the loss probability p of T2 shown in Eq. 3. Thus, the expected payoff of T3 for any given combination of individual contributions is identical to the expected payoff for T2 for the same combination—treatments T2 and T3, although differently framed, are identical in effect.

In both treatments T2 and T3, \$120 still constitutes a particular threshold, since the group is “safe” with certainty if their aggregate contribution reaches \$120, as in T1. T2 and T3 differ, however, from T1 in that the cost of C_{10} being below \$120 is less dramatic and more gradual than it is in T1 and thus more similar to many real-world situations. That the cost of not reaching the threshold is smaller seems to be good news for subjects. At least theoretically, however, it is not, because the incentive for the group to reach \$120 (and maximize the group total payoff) is smaller. Indeed, it easily can be shown that for risk-neutral subjects everyone contributing \$20 is no longer a Nash equilibrium. Maximizing the individual payoff function (Eq. 4) with respect to c_{i10} yields the best-response function:

$$c_{i10}(d) = \begin{cases} \frac{40 - d}{2} & \text{if } d \leq 40 \\ 0 & \text{if } d > 40 \end{cases} , \quad (6)$$

where $d = C_{10} - c_{i10}$ (the sum of the contributions of the other players). In equilibrium $c_{i10}^* = 40/\gamma \approx 5.71, \forall i$. Given the discrete nature of contributions in our game and that subjects can only make even-numbered contributions, $c_{i10}^* = 6$ and $C_{10} = 36$. Obviously, individual

contributions of \$6 (which would result in expected payoffs of 10.2 for each group member) are not enough for the group to reach the threshold of \$120 and for each group member to receive a certain payoff of \$20. If the other five group members each contribute \$20 for a total of \$100, then subject i maximizes her expected payoff by contributing 0; that is, everybody contributing \$20 is no longer an equilibrium.

Everybody contributing \$20, however, while not an equilibrium, is clearly the socially optimal situation now, as any increase of an individual contribution (until the aggregate contribution level reaches \$120), independent of what others do, has a positive net impact on the sum of payoffs. This is not true in T1, where only contributions that put the group's aggregate level above the threshold have a positive impact on the sum of payoffs. Contrary to the coordination game in T1, the games in T2 and T3 are therefore prisoner's dilemmas again, which are similar to traditional linear public-good games without a threshold (see Barrett 2013 on this distinction).

Based on these theoretical predictions, and on results from similar experiments, we expect regarding *question 1* that the majority of groups in T1 will try to coordinate on the "contribution equilibrium." In comparison, the predicted success rates in T2 and T3 are unambiguously less optimistic about reaching the \$120 threshold. Contrary to a high success rate in T1, we expect zero success rates in T2 and T3 (with success defined as reaching the "safe threshold" of \$120). Further, although we expect contributions to be positive in T2 or T3, we expect them to be lower than in T1.

Turning to *question 2*, although T2 and T3 are fundamentally the same, they differ in framing—T2 stresses the loss probability while T3 focuses on threshold uncertainty. If subjects focus squarely on the incentives, we expect to find no difference in contributions between T2 and T3. Alternatively, if framing of the risk matters in this case, we would expect contributions in T2 to exceed those of T3 due to the reduced specificity of the goal in T3, as described above.

Note that T2 and T3 differ from each other in two ways, which each would tend to push contributions in a different direction. Making the threshold location variable between 0 and 120 (as in T3), as opposed to fixed at 120 (as in T2), puts less pressure on groups to reach the 120 threshold, whereas fixing the loss probability at 1 if the threshold is not reached (T3), as opposed to being variable between 0 and 1 (T2), puts more pressure on groups to reach the threshold. Because the two effects work in opposite directions, it becomes an empirical question: which one (if either) is stronger? If fewer T3 groups than T2 groups reach 120, we would attribute that drop in cooperation to making the threshold location variable. Conversely, if fewer T2 groups than T3 groups reach 120, we would attribute the drop in cooperation to making the loss probability variable.

Treatments T4–T6 replicate T1–T3, respectively, except that in T4–T6 endowments are unequal (Table 1). Three subjects are randomly assigned to be "rich" (endowment of \$50) and the other three are "poor" (endowment of \$30). The rich subjects can contribute in each period \$0, \$2, \$3, or \$5, and the poor can contribute \$0, \$1, \$2, or \$3. The four values subjects may enter were chosen in order to give them the opportunity to contribute nothing, contribute everything, make equal absolute contributions to assure $C_{10} = 120$ (everybody gives \$2 each period), or make contributions that assure the threshold is reached and that payoffs are equal (the poor give \$1 in each period and the rich give \$3, so that everybody keeps \$2 for themselves). Moreover, while equal proportional contributions (\$1.50 for poor and \$2.50 for rich) were not possible in a given period, poor subjects going back and forth between \$1 and \$2 and rich subjects between \$2 and \$3 results in equal proportional contributions within a round.

As in T1, there are two equilibria for risk-neutral subjects in T4: everyone gives \$0 and everyone gives half of their endowment, which is enough to bring the aggregate contribution to \$120 ($3 \cdot 15 + 3 \cdot 25 = 120$). In the latter equilibrium, subjects earn \$15 and \$25, respectively, with certainty, the same as their expected earnings would be in the zero-contribution equilibrium. Risk-averse subjects have, depending on their levels of risk aversion, additional equilibria in which the aggregate contribution equals \$120. In treatments T5 and T6, the best-response function is identical to that of T2 and T3 and is shown in Eq. 6 except that for poor subjects, the 40 in Eq. 6 is replaced with 30 and for rich subjects the 40 is replaced with 50. The response function yields a corner - solution equilibrium in which the poor subjects contribute 0 and the rich subjects contribute 12.5 for an aggregate contribution of 37.5 and expected payoffs of 9.4 and 12.2 for poor and rich subjects, respectively. With discrete contributions, the poor contribute 0 and the rich contribute 12 or 13 in equilibrium, with total contributions of 36 to 38 and expected payoffs in the range of 9 to 9.5 for the poor and 11.4 to 12 for the rich. Given that equilibrium predictions of T4–T6 are very similar to those of T1–T3, along with the lack of consistent evidence in the literature of an effect of income inequality, regarding *question 3* we cannot expect differences in success rates or contribution levels between equal and unequal endowments.

4 Methods

Our six treatments (T1–T6) engaged a total of 378 subjects in 63 groups of six subjects each. At least ten groups participated in each treatment. Each subject provided 30 responses, ten in each of three identical rounds. At the end of the experiment, one of the three rounds was randomly chosen to determine the payoffs. Subjects were recruited in introductory psychology classes at Colorado State University, listed over 40 different academic majors (with psychology being the most prevalent at about 18%), were about 55% female, and had a mean age of about 20. No subject participated in more than one session. Subjects were individually seated at computer terminals, did not converse among themselves, and were not aware of who among the other participants bid a specific amount. The experiment was implemented using the z-Tree software (Fischbacher 2007). On average, a session, including reading the instructions, lasted 50 min. The instructions carefully explained the opportunity to earn a cash payoff and the risk that the potential payoff could be lost. The average payoff was about \$23 for subjects in groups that kept their remaining funds and, of course, \$0 for subjects in groups that lost their remaining funds. Across all subjects the average payout was \$18.4. See the [Supplementary material](#) for experimental scripts.

Analysis of the three research questions was performed using data from all six treatments in a single generalized linear mixed model (implemented with the GLIMMIX procedure of SAS) with Threat, Endowment, and Round as class variables. The model allows comparisons of the different levels of Threat, Endowment, and Round. The Tukey multiple range test is used to adjust for the effect of having performed multiple tests. Two dependent variables are analyzed: success rate is entered as a binary variable (the group either reached the threshold or not), and total contribution is entered as Gaussian after adjustment with the box-cox transformation to approximate a normal distribution. A p value cutoff of 0.05 is used for all tests. The null hypothesis in all statistical tests was of no difference between the two treatments or conditions being compared. All of the expectations regarding research questions 1 and 2 apply as well to treatments T4–T6, respectively, so that we examine those expectations in terms of H1–H3.

Our main statistical tests are based on data from all three rounds of each treatment. Round was found to not be significant in the analyses of success rate ($p = 0.483$) or total contribution ($p = 0.088$). Thus, we summarize results by combining across rounds. See the [Supplementary material](#) for more, including round-specific results in Tables S1 and S2 and treatment-level statistical test results in Table S3.

5 Results

Question 1: on the effect of a discrete versus incremental change in the chance of loss

The success rate when the threshold is known and loss probability is fixed (H1, Table 1) is somewhat, though not significantly (at the $p = 0.05$ level), higher than with threat H2 ($p = 0.082$) and is significantly higher than with threat H3 ($p = 0.001$). Further, the success rates given H2 are obviously positive, but with H3 they are not. For example, in T1, 60% (18 out of 30) of all group opportunities managed to reach the threshold, and in T2, 25% made it, but in T3, not a single group made it (Fig. 1).

The generally lower success rates of H2 and H3 compared to those of H1 are not surprising in light of the fact that reaching the threshold does not have the same stark effect in H2 and H3 as it does in H1. A group in H2 and H3 narrowly missing the threshold would face only a slightly higher probability of losing the remaining endowment, while in H1 that probability abruptly drops from 0.5 to 0 when the threshold is reached.

Total group contributions are higher in H1 than in H2 ($p = 0.037$) or H3 ($p < 0.001$). For example, average contributions are \$113, \$110, and \$94 in T1, T2, and T3, respectively (Fig. 2). Recall that theory predicts total contribution for risk-neutral subjects to be \$120 for H1 (if they settle on the “better” Nash equilibrium) but only \$36 for H2 and H3. Even the lowest individual group total contribution across all rounds of treatments with H2 and H3 threats, \$64, is much higher than the theoretically predicted contribution of \$36–38 for risk-neutral subjects.

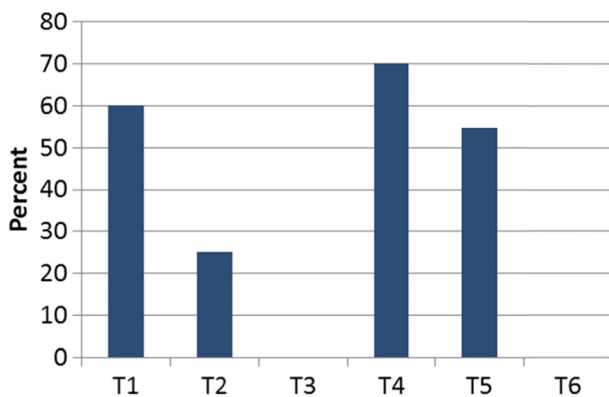


Fig. 1 Success rate (percent of groups reaching \$120 total contribution). Given a known threshold and fixed probability of loss if the threshold is not reached, the threshold was reached in 18 (T1) and 21 (T4) of the 30 (ten groups $\times$ three rounds) opportunities. With a known threshold and variable probability of loss, the “safe” threshold was reached in 9 of 36 (T2) and 18 of 33 (T5) opportunities. With an unknown threshold but fixed probability of loss (T3 and T6), the “safe” threshold was never reached

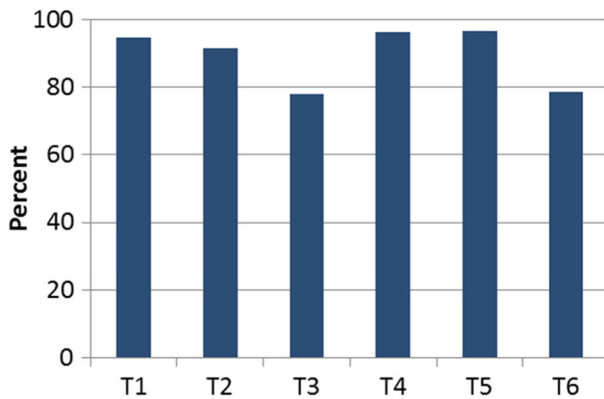


Fig. 2 Average percent of the safe threshold (\$120) contribution. Absent uncertainty about the threshold or loss probability, groups contributed on average \$113 (T1) and \$116 (T4), or about 95% of the threshold amount of \$120. With a known threshold but variable loss probability, they contributed on average \$110 (T2) and \$116 (T5) (again roughly 95% of \$120). With an unknown threshold but fixed probability of loss they contributed on average \$88 (T3) and \$97 (T6) (roughly 79% of \$120)

The high contribution levels in treatments with H2 and H3 threats are unexpected. Having a threshold in place, even if the loss in the case of not reaching the threshold is not as stark or if the threshold is uncertain, still seems to motivate subjects more than theory based on self-interest and risk neutrality would predict—which is good news for climate change policies. Results for these four treatments show that in situations where contributions potentially have a positive impact, most subjects made significant contributions toward the group goal (66% of all contributions across these four treatments were positive, and only 3% of the subjects contributed \$0 in all periods).

Recall that Dannenberg et al. (2015) also investigated the effect of an uncertain threshold using very similar treatments to ours. Their “Certainty” treatment matches our T1 and their “Risk” treatment matches our T3, the principal difference being that subjects in their experiment made non-binding proposals before periods 1 and 6, whereas in our experiment no such inter-subject communication occurred. Because they implemented only one round of their experiment, we compare their results to our round 1 results. Mean total contributions in their Certainty and Risk treatments were 121 and 101 (for a difference of 20), whereas in our T1 and T3 treatments, total contributions were 115 and 96 (for a difference of 19). The overall difference between their results and ours could possibly be due to the non-binding proposals that their subjects announced. Most notable, however, is the similarity of the differences between Certainty/T1 (where the threshold is known) and Risk/T3 (where the threshold is unknown and varies uniformly between 0 and 120).

Observation of individual group behavior affords an interesting view of a key difference between the H1 and H2/H3 incentive structures. Three groups in H1 (two in T1 and one in T4) contributed a total of only about \$80 in round 1, considerably lower than any other H1 groups (see Fig. S1). Two of these three groups contributed significantly less in the following two rounds (about \$72 and \$36 in rounds 2 and 3), suggesting that subjects in those groups were becoming progressively less optimistic about reaching the threshold, and the third group contributed significantly more in the following two rounds (\$102 and \$118 in rounds 2 and 3), indicating an optimistic attempt to reach the threshold despite the poor (round 1) beginning. Such significant downward or upward trends in total contribution following a low initial total

contribution are absent from all groups in the H2 and H3 treatments, where groups that achieved a low total contribution in round 1 tended to repeat that level in the subsequent two rounds (Fig. S1). The likely explanation for the difference between the H1 and H2/H3 responses to low total contributions in round 1 is that H1 contributions have no effect on the risk of loss unless the threshold is reached (so that contributing is counter-productive unless the goal is in sight), whereas with H2/H3 all contributions (until the safe threshold is reached) serve to reduce the risk of loss.

Question 2: on framing of the risk If subjects were solely motivated by the monetary incentives incorporated in the two frames compared here (H2 vs. H3), success rates and contributions would not differ between frames, for the incentives are identical. However, the success rate is higher in H2 than in H3 ($p=0.038$) (Fig. 1) and so is the total contribution ($p<0.001$) (Fig. 2). Having the threshold as a target, even if the loss probability is unknown, more strongly motivates groups to cooperate and get out of their prisoner's dilemma situation than knowing the loss probability and not knowing the exact threshold. This finding suggests that a threshold provides a clear goal, known to all participants, toward which they can aim. When the location of the threshold is unknown the threshold apparently loses some salience.

Question 3: on the effect of unequal endowments Perusal of Figs. 1 and 2 has already indicated that in our setup, income heterogeneity has only a positive effect, if it has any. We find that success rates do not differ between T1 and T4 ($p=0.322$) or between T3 and T6 (both rates are 0), but do between T2 and T5 ($p=0.002$). Consistent with success rates, contributions do not differ between T1 and T4 ($p=0.509$) or between T3 and T6 ($p=0.861$) but do differ between T2 and T5 ($p=0.016$). We cannot offer a good explanation for this difference in the effect of income heterogeneity between the H2 threat configuration on the one hand and the H1 and H3 configurations on the other hand, and we suspect that it may be an anomaly (see the [Supplementary material](#) for more on this). In any case, the general finding that income heterogeneity does not have a negative effect on cooperation again is good news for climate negotiations, all else equal.

6 Discussion

We add realism to a dynamic public-good-with-threshold game often used to model behavior in combating climate change, by introducing income heterogeneity and two forms of scientific uncertainty surrounding the threshold and loss probabilities. Our results suggest the following. (1) When the threshold is precisely known and failure to reach it entails a huge risk, groups overwhelmingly try to reach the threshold level of contribution. Scientific uncertainty about loss probability or the location of the threshold, such that the chance of loss drops incrementally as total contributions rise, lowers contribution levels, but does not have the dramatic negative effect on contributions that theory, given risk neutrality, would predict. Even when the location of the threshold was unknown—and thus where contributing is individually suboptimal (and the underlying game is a prisoner's dilemma)—aggregate contributions on average dropped by less than 20% compared to when the threshold was known with certainty. (2) Framing of the risk situation matters. Groups facing a fixed threshold and variable loss probability cooperate more than groups facing a variable threshold and a fixed loss probability,

suggesting that the presence of a threshold matters, even if the risk of loss if it is not reached is uncertain. (3) These first two results apply even if actors are heterogeneous in income.

These findings confirm that lack of a precise, known threshold increases the difficulty of protecting resources from irreversible harm, but show that it need not severely diminish the effort, even in the absence of negotiation. In part, this occurs because even when we are uncertain about the science we can assume that the greater our effort the lower is the likelihood of a future catastrophe. Further, the finding that framing matters suggests that the international community has been wise to emphasize a threshold level of effort in climate change mitigation talks, for even if the risks we face are uncertain, the focus on a threshold helps focus the joint effort. Finally, our findings indicate that income heterogeneity is not by itself a barrier to avoiding a shared disaster.

Acknowledgements We thank Paul Bell, Gretchen Nurse Rainbolt, Christopher Goemans, Nicholas Janusch, Alexander Maas, and seminar participants at the CU Environmental and Resource Economics Workshop, Beaver Creek, CO; Workshop on Behavioral and Experimental Economics, Florence, Italy; International Symposium on Society and Resource Management, Estes Park, CO; and Potsdam Institute for Climate Impact Research, Potsdam, Germany.

Author contributions TCB and SK shared equally in all aspects of the research and writing.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest

References

- Barrett S (2013) Climate treaties and approaching catastrophes. *J Environ Econ Manag* 66:235–250
- Barrett S, Dannenberg A (2012) Climate negotiations under scientific uncertainty. *Proc Natl Acad Sci U S A* 109: 17372–17376
- Barrett S, Dannenberg A (2014) Sensitivity of collective action to uncertainty about climate tipping points. *Nat Clim Change* 4:36–39
- Burton-Chellew MN, May RM, West SA (2013) Combined inequality in wealth and risk leads to disaster in the climate change game. *Clim Change* 120:815–830
- Chan KS, Mestelman S, Moir R, Muller RA (1996) The voluntary provision of public goods under varying income distributions. *Can J Economics* 29:54–69
- Chan K, Mestelman S, Moir R, Muller RA (1999) Heterogeneity and the voluntary provision of public goods. *Exp Econ* 2:5–30
- Cherry TL, Kroll S, Shogren JF (2005) The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab. *J Econ Behav Organ* 57:357–365
- Crosen R, Marks M (2001) The effect of recommended contributions in the voluntary provision of public goods. *Econ Inq* 39:238–249
- Dannenberg A, Löschel A, Paolacci G, Reif C, Tavoni A (2015) On the provision of public goods with probabilistic and ambiguous thresholds. *Environ Resour Econ* 61:365–383
- Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 10:171–178
- IPCC (2013) Summary for policymakers. In: Stocker TF Q, Plattner D, Tignor G-K, Allen M, Boschung SK, Nauels J, Xia A, Bex Y, Midgley V, Midgley PM (eds) *Climate change 2013: The physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, p 1535
- Locke EA, Latham GP (2002) Building a practically useful theory of goal setting and task motivation. *Am Psychol* 57: 705–717
- Milinski M, Sommerfeld RD, Krambeck H-J, Reed FA, Marotzke J (2008) The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc Natl Acad Sci U S A* 105:2291–2294

- Milinski M, Röhl T, Marotzke J (2011) Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Clim Change* 109:807–814
- Rapoport A (1988) Provision of step-level public goods: effects of inequality in resources. *J Pers Soc Psychol* 54: 432–440
- Rapoport A, Suleiman R (1993) Incremental contribution in step-level public goods games with asymmetric players. *Organ Behav Hum Decis Process* 55:171–194
- Rapoport A, Bornstein G, Erev I (1989) Intergroup competition for public goods: effects of unequal resources and relative group size. *J Pers Soc Psychol* 56:748–756
- Reuben E, Riedl A (2013) Enforcement of contribution norms in public goods games with heterogeneous populations. *Game Econ Behav* 77:122–137
- Shackley S, Wynne B (1996) Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority. *Sci Technol Hum Val* 21:275–302
- Tavoni A, Dannenberg A, Kallis G, Löschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc Natl Acad Sci U S A* 108:11825–11829
- van Dijk F, Sonnemans J, van Winden F (2002) Social ties in a public good experiment. *J Public Econ* 85:275–299