Climatic Change
DOI 10.1007/s10584-016-1838-3

REVIEW ARTICLE

Cooperation studies of catastrophe avoidance: implications for climate negotiations

Mark J. Hurlstone[1,2] · Susie Wang[1,2] · Annabel Price[1] · Zoe Leviston[1,2] · Iain Walker[1,2,3]

Received: 27 October 2015 / Accepted: 8 October 2016

Abstract The landmark agreement recently negotiated in Paris represents an ambitious plan to combat climate change. Nevertheless, countries' current climate pledges are insufficient to achieve the agreement's goal of keeping global mean temperature rise "well below" 2 °C. It is apparent that climate negotiators need to be equipped with additional strategies for fostering cooperation if a climate catastrophe is to be averted. We review the results arising from an emerging literature in which the problem of avoiding dangerous climate change has been simulated using cooperation experiments in which individuals play a game requiring collective action to avert a catastrophe. This literature has uncovered five key variables that influence the likelihood of avoiding disaster: (1) the perceived risk of collective failure, (2) inequalities in historical responsibility, wealth, and risk exposure, (3) uncertainty surrounding the threshold for catastrophe, (4) intergenerational discounting, and (5) the prospect of reward or punishment based on reputation. Along with the results of a recent experimental assessment of the key instruments of the Paris Agreement, we consider how knowledge of the effects of these variables might be harnessed by climate negotiators to improve the prospects of reaching a solution to global climate change.

Keywords Climate change · Cooperation · Climate negotiations · Global public good · Collective-risk social dilemma

✉ Mark J. Hurlstone
mark.hurlstone@uwa.edu.au; mark.hurlstone@github.io

[1] School of Psychology, University of Western Australia, Crawley 6009 WA, Australia

[2] Commonwealth Scientific and Industrial Research Organisation, Queensland, Australia

[3] University of Canberra, ACT 2617, Australia

Published online: 27 October 2016

1 Introduction

Human greenhouse gas emissions are continuing to rise at unprecedented rates, posing the risk that one day mean global temperature will exceed a dangerous threshold (Alley et al. 2003; Lenton et al. 2008; Schellnhuber et al. 2006)—defined in the Copenhagen Accord as 2 °C. Catastrophe avoidance requires the collective action of all nations to reduce their emissions of greenhouse gases by 50 % by 2050 so that atmospheric concentrations can be stabilized (Meinshausen et al. 2009; Peters et al. 2013; Roeckner et al. 2011). The protection of the global climate is a global public good that confers benefits that no country can be excluded from receiving. However, because climate protection requires economically costly emission abatement, each country has an incentive to act selfishly by refusing to cooperate in the hope that others will carry the burden of supplying the global public good—the so-called free rider problem. Climate protection is therefore an example of a more general and pervasive problem known as the "tragedy of the commons" (Hardin 1968), wherein individuals, groups, or countries acting in their self interest overexploit a common resource to the detriment of their collective long-term best interests.

As the protection of the global climate depends upon the aggregate effort of all nations, international climate negotiations are crucial so that countries can coordinate their efforts. However, until the landmark agreement recently negotiated in Paris under the United Nations Framework Convention on Climate Change (UNFCCC), more than twenty years of climate negotiations had failed to produce a meaningful climate regime. The Paris Agreement is a major milestone and represents an ambitious plan to combat climate change that compels most countries to reduce their emissions. Although it retains the 2 °C target of the Copenhagen Accord, it also underscores the desirability of restricting warming to 1.5 °C. Meeting this goal will not be easy, given that countries' current climate pledges—known as Intended Nationally Determined Contributions—are only sufficient to limit warming to around 2.7 °C,[1] rather than 2 °C (let alone 1.5 °C). The Paris Agreement is therefore the first step in a long journey and it will need to be improved upon and complemented by parallel approaches. In pursuit of this goal, it would be helpful if climate negotiators were equipped with strategies to facilitate this process. In this article, we consider the lessons for climate negotiations that might be gleaned from an emerging literature in which the problem of avoiding dangerous climate change has been simulated using laboratory cooperation experiments.

2 A climate catastrophe avoidance game

Milinski and colleagues (Milinski et al. 2008) have devised a laboratory cooperation game known as the "collective-risk social dilemma" (CRSD) for simulating the problem of avoiding dangerous climate change. The game involves groups of six players. Each is given an operating fund of €40 and must decide whether to contribute €0, €2, or €4 in each of 10 rounds to a climate account without communicating. At the end of each round, the contributions of each group member are made public. If at least €120 has been contributed by the end of the game then catastrophic climate change is averted with certainty and players get to keep the leftovers of their operating fund. However, if the group fails to reach the threshold then catastrophic climate change occurs with a pre-specified probability (e.g., 90 %)

[1] Climate Action Tracker; http://climateactiontracker.org

that what remains of each player's operating fund will be lost. The €120 target can be construed as a temperature threshold—such as the 2 °C goal—whilst the player contributions are a metaphor for the level of investment of countries in emission reductions.

The CRSD is a coordination game—where players must coordinate strategies for their mutual benefit—with two symmetrical pure strategy Nash equilibria. One is a "dangerous" equilibrium where each player contributes €0, whereas the other is a "safe" equilibrium where each player contributes €20 (there are also several "safe" asymmetric pure strategy Nash equilibria where different players contribute different amounts). Coordination games usually possess a "focal point" (Schelling 1960) with salient characteristics that facilitates coordination. In the CRSD, fear of catastrophe makes the €120 contribution level salient, rendering it a natural focal point.

3 Factors affecting cooperation

We now review five major findings observed using the CRSD—and kindred climate change cooperation games—that cast light on the factors that spur and inhibit collective efforts to avert catastrophe. To aid interpretation, Table 1 provides a summary of the key attributes of the experiments that are the focus of our review.

3.1 Perception of risk

To avert a catastrophe, countries must be convinced they will be seriously adversely affected by climate change. This was shown in the original experiment using the CRSD by Milinski et al. (2008) in which groups were allocated to one of three different collective-risk conditions. In the low-risk condition, the probability of catastrophe if group members failed to reach the threshold was 10 %; in the moderate-risk condition it was 50 %; whilst in the high-risk condition it was 90 %. When the risk was low or moderate, catastrophe was virtually assured—under low-risk, all groups failed to reach the threshold, whilst under moderate-risk only 10 % of groups were successful. By contrast, under high-risk 50 % of groups reached the threshold and the remaining groups came close.[2] This experiment suggests that the perception of a high risk of danger is a necessary—but not sufficient—condition to stave off catastrophe. These results mesh well with game theoretic analyses of the CRSD showing that coordination for a global good is best realized when the perception of risk is high (Pacheco et al. 2014; Santos & Pacheco 2011; Vasconcelos et al. 2014).

The Milinski et al. (2008) experiment raises two important implications for climate negotiations. First, the framing of risks matters greatly. The climate negotiations rely upon the assessment reports of the Intergovernmental Panel on Climate Change (IPCC), which use probabilistic statements (e.g., Very Likely) to communicate risks. However, Budescu et al. (2012, 2014) have shown that people persistently misinterpret these statements. Notably, they tend to underestimate high probabilities, which is especially worrying given the Milinski et al. (2008) demonstration that the perception of a high risk of catastrophe is necessary to drive cooperation. Fortunately, this tendency can be mitigated by supplementing probabilistic statements with numerical ranges (Budescu et al. 2012, 2014). The IPCC

[2] See Barrett (2011) for a discussion of how the incentives to cooperate vary with the risk level in this experiment.

Table 1 Overview of basic attributes of key climate change related cooperation experiments. See main text for further details

Reference	Cooperation game	Variables manipulated	Group size	N rounds	Risk of catastrophe	Communication
Milinski et al. (2008)	CRSD[a]	Risk of catastrophe (low- vs. moderate- vs. high-risk)	6	10	10 %, 50 %, or 90 % (corresponding to low-, moderate-, and high-risk conditions, respectively)	No
Tavoni et al. (2011)	CRSD	Inherited inequality (equal vs. unequal) and communication (no-pledge vs. with-pledge)	6	10 (3 passive + 7 active)	50 %	Non-binding pledges in with-pledge condition only
Milinski et al. (2011)	CRSD	Wealth inequality (rich groups vs. poor groups vs. mixed groups) and intermediate target (no intermediate threshold vs. intermediate threshold)	6	10	90 %	No
Burton-Chellew et al. (2013)	CRSD	Combined inequalities (egalitarian vs. unequal-wealth vs. rich-suffer vs. poor-suffer)	6	10	70 % or 80 % in the egalitarian and unequal-wealth conditions; 90 % or 95 % for rich players and 50 % or 65 % for poor players in rich-suffer condition (with reverse mapping of risk to player wealth in poor-suffer condition)	No
Barrett & Dannenberg (2012)	Variant of CRSD	Uncertainty (certainty vs. impact uncertainty vs. threshold uncertainty vs. impact-and-threshold uncertainty)	10	1	100 %	Non-binding pledges in all conditions

Table 1 (continued)

Reference	Cooperation game	Variables manipulated	Group size	N rounds	Risk of catastrophe	Communication
Barrett & Dannenberg (2014a)	Variant of CRSD	Threshold uncertainty (€150 vs. €145–€155 vs. €140–160 vs. €135–165 vs. €100–200; corresponding to certain threshold and low, moderate, large, and extreme threshold uncertainty, respectively)	10	1	100 %	Non-binding pledges in all conditions
Jacquet et al. (2013)	CRSD	Temporal discounting (short-delay vs. long-delay vs. intergenerational)	6	10	90 %	No
Milinski et al. (2006)	Non-threshold climate cooperation game alternated with indirect reciprocity game	Anonymity of investments (public vs. anonymous) in climate cooperation game[b]	6	20[c]	Not applicable	No
Barrett & Dannenberg (2016)	Variant of CRSD	Peer-review (no-review vs. ex-ante review vs. mid-review vs. ex-post review) with threshold uncertainty in all conditions	5	1	100 %	Non-binding pledges in all conditions and social approval ratings in review conditions

[a] Collective-risk social dilemma

[b] Milinski et al. (2006) also manipulated the presence or absence of information about climate change risks but we do not consider this ancillary aspect of their experiment.

[c] viz. 10 rounds of the climate cooperation game and 10 rounds of the indirect reciprocity game played in alternation.

would be wise to transition to this dual-scale risk communication strategy to give climate negotiators a more accurate perception of risks. Negotiators would also benefit from knowledge of strategies for communicating risks most efficiently.

The second implication relates to the possibility that the scientific community may have understated the risks of climate change. Evidence for this directional bias comes from the observation of systematic under-predictions by climate scientists of key attributes of climate change (Brysse et al. 2013) and undue conservatism regarding climate risks in the fourth assessment report of the IPCC (Freudenburg & Muselli 2010). Lewandowsky et al. (2015) attribute this tendency to "err on the side of least drama" to the influence of contrarian talking points from public discourse about climate change into scientific thinking—a phenomenon they refer to as "seepage". The experiment of Milinski et al. (2008) suggests it is important not to downplay the severity of the climate problem, since a reduced perception of risk could hamper collective efforts to avert catastrophe. Accordingly, climate negotiators and the IPCC need to be able to detect and avoid seepage, and Lewandowsky et al. (2015) provide recommendations for how this might be accomplished.

3.2 Inequalities

The Milinski et al. (2008) experiment assumed that all players are created equal. However, this is not an accurate reflection of the real game of climate change where inequalities exist in terms of historical responsibility, wealth, and risk exposure. Several studies have added greater ecological validity to the CRSD by incorporating such inequalities.

Tavoni et al. (2011) examined the impact of inherited inequality using an augmented CRSD divided into "passive" and "active" phase components, with a 50 % risk of catastrophe. In the passive phase (rounds 1-3), the computer determined the contributions made by each player. In the equal condition, all group members were forced to allocate €2 per round, whereas in the unequal condition, half the players were forced to contribute €4 per round, whilst the other half were forced to contribute €0 per round. Given a starting operating fund of €40, this meant that by the end of the passive phase, all players in the equal condition had a remaining operating fund of €34 each, whereas in the unequal condition, the poor and rich players had a remaining operating fund of €28 and €40 each, respectively. In the active phase (rounds 4-10), players decided for themselves how much to contribute on each round.

The authors also incorporated a communication manipulation. In the with-pledge condition, at the end of rounds 3 and 7, players were able to announce how much they intended to contribute to the climate account over the next three rounds via the submission of non-binding pledges. These pledges were then revealed to each member of the group so they could ascertain whether the threshold would be reached if everyone adhered to their commitments. Contributions in this condition were contrasted with those in a no-pledge condition where communication between players was prohibited.

In the absence of pledges, inequality was an impediment to cooperation—50 % of groups reached the threshold in the equal condition, compared to 20 % in the unequal condition. However, allowing players to submit pledges enhanced cooperation in both conditions and almost neutralized the impediment of inequality—70 % of groups in the equal condition and 60 % in the unequal condition reached the threshold. Communication attenuated the handicap of inequality because rich players were able to signal to poor players early on their willingness to compensate for the latter's lesser resource capacity, and the poor players in turn were willing to trust that the rich players would fulfill their pledges. These findings are encouraging given that the risk of catastrophe was only 50 %—a condition under which free rider incentives are strong. As noted by Barrett (2011), they suggest that had communication been permitted in the Milinski et al. (2008) study, most groups in the 90 % risk condition would have averted catastrophe.

The handicap of wealth inequality can also be attenuated through the introduction of an intermediate climate target, as demonstrated by Milinski et al. (2011). In their experiment, players received an operating fund—the leftovers of which they could definitely keep—as well as an endowment that they would lose with 90 % probability if they did not reach the €120 threshold. The experimenters created rich and poor players by varying the size of their operating funds and endowments—rich players received a €40 operating fund and a €60 endowment, whereas poor players received a €20 operating fund and a €30 endowment. The game was played using rich groups, poor groups, and mixed groups (3 rich players + 3 poor players). In one condition, groups were informed that in addition to reaching the €120 threshold they also had to reach an intermediate threshold of €60 by the end of round 5. If they failed to reach the intermediate threshold then in rounds 6-10 there was a 20 % probability in each round that an intermediate climate event would occur, the consequence of which was a 10 % reduction of each player's operating fund and endowment. Cooperation under this scenario was contrasted with that in another scenario where no intermediate threshold was provided.

Without an intermediate threshold, all rich groups, no poor groups, and 60 % of mixed groups reached the final threshold. When an intermediate threshold was provided rich groups once again always averted catastrophe, but the key result was that 33 % of poor groups and 67 % of mixed groups now reached the final threshold. The increase in cooperation in mixed groups with an intermediate threshold arose because rich players compensated for the lower contributions of poor players.

The cooperation problem is rendered more difficult when inequalities exist in terms of both wealth and risk exposure, as shown by Burton-Chellew et al. (2013). In their experiment, groups were assigned to one of four conditions. In the egalitarian condition, all six group members were given €40 in operating funds and faced the same risk of catastrophe (viz. 70 % or 80 %) if they failed to reach the threshold. In the unequal-wealth condition, group members also faced the same risks, but differed in their wealth, with two rich players receiving operating funds of €80 each, and four poor players receiving €20 each. In the rich-suffer and poor-suffer conditions, wealth heterogeneity was induced in the same way as in the unequal-wealth condition, but the risks of catastrophe for rich and poor players were also varied. In the rich-suffer condition, the risk was higher for rich than for poor players (viz. 90 % or 95 % vs. 50 % or 65 %, respectively), whereas the reverse was true in the poor-suffer condition.

Cooperation was highest in the egalitarian condition, with 88 % of groups reaching the threshold, whilst cooperation was only slightly reduced in the unequal-wealth and rich-suffer conditions, with 63 % and 75 % of groups, respectively, reaching the threshold. That wealth inequality did not exert a greater impact on cooperation in these latter conditions occurred because when rich players faced the same or greater risks as poor players, they were willing to contribute more to avert catastrophe. The crucial question is whether this assistance would be provided when poor players were more at risk. Regrettably, the answer is no—in the poor-suffer condition cooperation collapsed, with only 13 % of groups reaching the threshold.

These experiments suggest that equity considerations must take centre stage at climate negotiations. Early leadership from powerful countries is a crucial ingredient for collective success, since the poor are unable to compensate for the inertia of the rich. Effective coordination mechanisms are therefore required to facilitate this process. The pledge-and-review mechanism in the Paris Agreement operates in a similar way to the communication instrument used by Tavoni et al. (2011) and may help promote equitable burden-sharing. However, the pledges submitted by countries before Paris were not based on a common metric, rendering it difficult to assess comparability of effort and deduce whether countries are pledging their fair share. For pledge-and-review to be an effective coordination mechanism, improved methods for assessing and comparing mitigation efforts are needed (Aldy & Pizer 2014). The experiment of Milinski et al. (2011) suggests that improved coordination might also be achieved by highlighting near-term climate threats and providing intermediate targets to mitigate these risks. As well as encouraging early redistribution of wealth, such near-term targets may reduce the likelihood of overshooting the long-term goal. Such coordinating mechanisms are all the more important given the marked and detrimental effect of combined inequalities in wealth and risk shown by Burton-Chellew et al. (2013). This result is especially concerning in light of recent evidence that rich countries perceive climate change as less threatening than poor countries (Lo 2015; Lo & Chow 2015). This reduced perception of risk may be misguided, however. Although the poor will suffer most, this does not imply that the rich will not suffer. Indeed, recent modeling suggests that the economic impacts of climate change will be far greater than previously anticipated (Moore & Diaz 2015). A necessary aspect of successful climate negotiations therefore may be to convince rich countries that climate change threatens ruin to all, not merely the poor.

3.3 Threshold and impact uncertainty

Another difference between the true game of climate change and the experiments examined so far is that the latter assume that the location of the dangerous threshold and the impact of crossing it are known with certainty. However, in the true game, uncertainty pervades these and other aspects of the climate change problem. Although the Paris Agreement identifies a target of 2 °C, in reality there is considerable scientific uncertainty regarding the location of the dangerous climate threshold (Lenton et al. 2008). Similarly, estimates of the expected damages resulting from dangerous climate change differ widely (Lenton et al. 2008). It therefore follows that any adequate laboratory analogue of the problem must incorporate such uncertainties. This has been done in a sequence of experiments by Barrett and Dannenberg who have shown experimentally (Barrett & Dannenberg 2014b) and theoretically (Barrett, 2013) that the existence of a dangerous climate threshold spurs cooperation relative to a scenario based on gradual climate change alone. However, uncertainty surrounding the threshold causes cooperation to collapse (Barrett & Dannenberg 2012, 2014a; Dannenberg et al. 2015; see also Hasson et al. 2012), whereas uncertainty about the impact of crossing the threshold has no effect on behavior (Barrett & Dannenberg 2012).

Barrett and Dannenberg (2012) examined the influence of uncertainty about the threshold for catastrophe and the consequences of crossing it on cooperation. In their variant of the CRSD, groups consist of ten players who are each allocated €31, which is divided into an operating fund of €11 and an endowment of €20. The operating fund can be used to invest in "weak" or "strong" emission abatement by purchasing chips (max = 10 of each type) at a cost of €0.10 or €1.00, respectively. The game is played over a single round which is divided into two stages—a communication stage, where each player proposes a contribution target for the group and pledges an amount they will contribute individually, followed by a contribution stage where each player chooses their actual contributions. The goal is to reach an investment threshold T, otherwise a cost C is deducted from each player's endowment with certainty (viz. 100 % risk of catastrophe).

Barrett and Dannenberg's (2012) experiment contained four conditions. In the certainty condition, the dangerous threshold and the impact of crossing it were both known with certainty (viz. T = €150 and C = 15). In the impact uncertainty condition, the dangerous threshold was known with certainty, but the impact of crossing it was not (viz. T = €150 and C was uniformly distributed between €10–€20). In the threshold uncertainty condition, the impact of crossing the dangerous threshold was known with certainty, but the location of the threshold was not (viz. T was uniformly distributed between €100–€200, and C = 15). Finally, in the impact-and-threshold uncertainty condition, both the dangerous threshold and the impact of crossing it were uncertain (viz. T was uniformly distributed between €100–€200, and C was uniformly distributed between €10–€20).

When the threshold and damages were known with certainty, 80 % of groups averted catastrophe and this figure rose to 100 % in the impact uncertainty condition. However, only 10 % and 30 % of groups in the threshold uncertainty and impact-and-threshold uncertainty conditions, respectively, averted catastrophe. These results demonstrate that uncertainty surrounding the threshold for dangerous climate change is a major handicap to cooperation, whereas uncertainty about the damages is ineffectual.

How much must uncertainty about the threshold be reduced? In a subsequent experiment, Barrett and Dannenberg (2014a) varied the size of the window of uncertainty surrounding the threshold. In the certainty condition, T = €150, whereas in four threshold uncertainty conditions T was uniformly distributed between either: (1) €145–€155, (2) €140–€160, (3) €135–€165, or (4) €100–€200 (C = €15 in all conditions). When the threshold was known with certainty, 80 % of groups averted catastrophe, but even with a very narrow window of uncertainty surrounding the threshold (viz. €145–€155), this figure dropped to 40 %. When the window of uncertainty was increased, no groups averted catastrophe. Thus, even a small degree of uncertainty about the threshold is a major impediment to collective action.

These experiments suggest that uncertainty about climate damages is relatively inconsequential, whereas uncertainty about the threshold is crucial. Unless the threshold is known with certainty, it is unlikely that countries will be able to avert catastrophe. Uncertainty about the threshold undermines cooperation because it reduces the credibility of Nature's threat to tip the climate system into chaos if the threshold is breached, which is the incentive for collective action (Barrett 2014). It turns the climate coordination game under a certain threshold into a prisoners' dilemma. The key feature of a prisoners' dilemma is that no matter what other countries do, each country has an incentive to do nothing, even though the collective best outcome arises when all countries cooperate. There is only one Nash equilibrium in a prisoners' dilemma—all parties defect.

A clear implication of these results for climate negotiations is that if the science of climate change were more certain and a red line for danger could be identified then fear of crossing it would discipline behavior. However, irreducible uncertainties surrounding the location of the critical threshold render this prospect unlikely. Accordingly, strategic enforcement mechanisms are required that can re-create the incentive to cooperate that exists when the threshold is known with certainty (Barrett & Dannenberg 2014b)—if Nature cannot provide the free rider deterrent, then countries must do so themselves. Examples of viable enforcement mechanisms include trade sanctions, exclusion from a cherished market, or the removal of an essential license. The key to an effective enforcement mechanism is that the punishment imposed must be severe and credible, and it must be perceived as such by countries (Barrett 2003). Crucially, enforcement mechanisms are purely strategic devices—their purpose is not to be used but to provide the deterrent necessary to transform behavior. Well-designed enforcement mechanisms are effective because they allow countries to escape the prisoners' dilemma by turning it into a coordination game.

3.4 Intergenerational discounting

The problem of avoiding dangerous climate change results not merely from a conflict between self and collective interest (tragedy of the commons), but also because the costs of failing to cooperate to avert it will be felt by future generations, leaving current actors with little incentive to fix the problem (tragedy of the time horizon). Thus, a defining feature of climate change is its intergenerational nature, which involves trade-offs between the self and future others. It is well-known that temporal discounting—the tendency to prefer immediate over delayed rewards, and delayed over immediate costs—influences individual decision making (Frederick et al. 2002; Loewenstein et al. 2003). Temporal discounting over short-term time horizons is known as intragenerational discounting, whereas temporal discounting over extremely long-term time horizons is known as intergenerational discounting. In individual decision making, both forms of discounting are known impediments to climate action (Lorenzoni & Pidgeon 2006; Spence et al. 2012; Weber 2010). Do intra- and inter-generational discounting also manifest in a group setting resembling climate negotiations?

This question was addressed using the CRSD in an experiment by Jacquet et al. (2013). In their experiment, group-members were given an operating fund of €40 and an endowment of €45. Players always received the leftovers of their operating funds immediately after the end of the game, whereas there was a 90 % chance that the endowment would be wiped out if the threshold was not reached. The experiment involved three conditions that differed in terms of the temporal delay between the end of the game and the endowment being awarded, and the number of beneficiaries of the endowment. In the short-delay condition, the endowment was paid to group members 1 day after the experiment; in the long-delay condition, it was paid to group members 7 weeks after the experiment; in the intergenerational condition, it was invested in a reforestation project to sequester CO_2, the beneficiaries of which would be future generations (a temporal delay of several decades, with a much wider range of beneficiaries). The difference in cooperation between the short- and long-delay conditions can be used as an index of intragenerational discounting, whereas the difference in cooperation between the short- or long-delay and intergenerational conditions can be used as an index of intergenerational discounting.

Temporal discounting caused a marked decrease in cooperation—in the short-delay condition, 70 % of groups reached the threshold, whereas only 36 % of groups did so in the long-delay condition, and no groups reached the threshold in the intergenerational condition. Thus, the experiment provided evidence for the operation of both intra- and intergenerational discounting, but intergenerational discounting was the stronger impediment to cooperation of the two.

This experiment suggests that the intergenerational nature of climate change is a major obstacle to cooperation in climate negotiations—we cannot depend on the intergenerational altruism of countries to solve the climate problem. Accordingly, climate negotiations will be significantly undermined if defection is the only means by which countries can reap immediate benefits. To counteract the adverse effect of intergenerational discounting on cooperation, it will be necessary to institute strategic mechanisms that facilitate the restructuring of incentives so that the short-term gains of cooperation exceed those of defection. In the next section, we consider how reputation might be used as a tool to realize this goal.

3.5 Reputation

In the experiments considered so far, the reputation of players cannot be used as a tool to spur cooperation. However, if a positive reputation can be used as a currency to obtain rewards and avoid punishments, stable cooperation can be achieved. This was demonstrated in an experiment by Milinski et al. (2006) using a non-threshold climate cooperation game in which groups of six players were given a €12 operating fund and could invest €0, €1, or €2 into a climate account over 10 rounds. Any money invested in the climate account was doubled in value and used to fund a press advertisement on climate protection.

Rounds of this climate cooperation game were alternated with rounds of an indirect reciprocity game. In each round of the latter game, players adopted the role of "donor" or "receiver" once. When afforded the role of donor, a player had to decide whether or not to give a reward of €3 to another player at a cost of €1.50 to themselves; conversely, when given the role of receiver, a player could potentially receive a €3 reward from another player. Importantly, if a player acted as a potential donor to a second player, then the second player could not subsequently serve as the potential donor to the first.

The critical manipulation was that on "odd" rounds of the climate cooperation game, investments in the climate account were made public, whereas on "even" rounds they were made anonymous. Thus, on public rounds, there was an incentive for players to cooperate by investing in the public good because this affords a positive reputation, which should be rewarded in the indirect reciprocity game, whereas failure to cooperate affords a negative reputation, which may incur punishment. By contrast, on anonymous rounds, the incentive to cooperate and the disincentive to free ride is removed because reputation cannot be used as a basis for obtaining benefits in the indirect reciprocity game.

The key finding was that investments in the climate cooperation game were markedly higher on public than anonymous rounds. Moreover, public investments were not just higher, they were also more stable, whereas anonymous investments decreased monotonically over rounds. Investments were higher on public rounds because cooperation was subsequently rewarded in the indirect reciprocity game, whereas defection was punished. These results identify reputation as a powerful tool for leveraging cooperation.

Is it possible to harness the power of reputation in the real climate change game by specifying a second cooperative scenario in which a good reputation can be used to obtain rewards and stave off punishments? It transpires that the Milinski et al. (2006) experiment provides a proof of concept for the recently proposed mechanism of "climate clubs" (Nordhaus 2015; Stewart et al. 2013; Victor 2015). The idea behind such clubs is that a small and powerful "coalition of the willing" establishes a voluntary group that produces tangible benefits with public good characteristics. These "club goods" might be, for example, development of a new technology, pooled finances, or common technological standards that confer non-climate economic or non-economic benefits to club members. Joining the club requires would-be-members meet certain emission reduction objectives—that is, entry into the club is contingent on one having a good reputation for climate protection, as is sustaining club membership. The club serves the primary function of conferring benefits to its members—emission reductions are achieved as a by-product. If the benefits of club membership are sufficiently attractive this offers the leverage necessary to get free riders to invest in climate protection so they too can access these benefits. Non-members are punished by denying them the benefits of the club.

An advantage of the club mechanism is that by starting initially with a small group of members it is possible to build cooperation gradually from the ground-up, thereby attenuating the free-rider problem, which is more prevalent in larger groups (Pacheco et al. 2014; Santos & Pacheco 2011; Vasconcelos et al. 2014) like the UNFCCC negotiating forums. Furthermore, by transforming the incentives so that the short-term benefits of cooperation may outweigh those of defection, climate clubs offer a mechanism to surmount the intergenerational discounting problem.

4 An experimental assessment of the paris agreement

The key feature of the Paris Agreement is its pledge-and-review mechanism. One of the goals of incorporating this mechanism is to provide countries with the opportunity to express their approval or disapproval of other countries' pledges and contributions, thereby casting a spotlight on role models and free riders.

Before concluding, we consider a very recent experiment by Barrett and Dannenberg (2016) that tested the likely consequences of adoption of this mechanism in climate negotiations. Their basic experimental set-up was similar to that used in Barrett and Dannenberg (2012; see earlier) but with groups of five—as opposed to ten—players and uncertainty about the threshold only. The game was played in a sequence of stages. In the first stage, each player submitted proposals for the target investments the group should aim for; in the second stage, each player pledged how much they would contribute toward reaching the proposed target; in the third stage players submitted their actual contributions; in the fourth stage players had the opportunity—having witnessed the contributions of their co-players—to revise their contribution.

In the no-review baseline condition, the game was conducted as described above, whereas in the review conditions players were afforded the opportunity to indicate their level of approval or disapproval of their group members' submissions. There were three different review conditions that differed according to the locus of the review component—ex-ante (after the second stage involving pledges), mid-review (between the third and fourth stages involving contributions), and ex-post (after the final stage involving contributions).

Barrett and Dannenberg (2016) found that irrespective of the condition to which they were assigned, groups proposed collective targets that were less than required to prevent catastrophe with certainty; pledged to contribute less than those targets; and in turn contributed less than their pledges (see also Barrett & Dannenberg 2012, 2014a). Peer review increased proposals, pledges, and—to a lesser extent—contributions in all three review conditions, but only slightly so.

The implications of these results for climate negotiations are clear. The Paris pledges will at best limit warming to 2.7 °C, so as in Barrett and Dannenberg's (2016) experiment countries have already pledged to do less than is required to reach the collective target. Their experiment suggests that even with a pledge-and-review mechanism, countries' actual emission reductions are likely to be less than their pledges. This suggests that complementary strategies running in parallel with the Paris Agreement will be required to achieve the level of cooperation needed to fulfill its objective.

A potentially effective strategy would be to exploit linkage mechanisms that harness the power of existing non-climate institutions for international cooperation in the service of climate protection. For example, despite being introduced to solve a different problem, the Montreal Protocol on Substances that Deplete the Ozone Layer has served an important function in protecting the climate by phasing out the use of ozone-depleting substances that are also greenhouse gases that contribute to climate change (Velders et al. 2007). Prior to the Paris negotiations an amendment to the Montreal Protocol was negotiated to phase down the use of global-warming-inducing Hydrofluorocarbons. This amendment should be successful because the Montreal Protocol contains an enforcement mechanism in the form of trade restrictions to ensure compliance amongst its parties. Climate negotiators would be wise to pursue future actions under the Montreal Protocol to phase out other climate forcing substances to achieve additional climate benefits.

5 Summary

Our foregoing analysis has identified several potential avenues to more successful climate negotiations. Specifically, the observation that the perception of risk is a key driver of collective action suggests that efforts must be undertaken to convince countries of the expected high-risks of dangerous climate change, especially rich countries who perceive it as a less serious threat. Improved methods are required for communicating climate risks and identifying and preventing seepage of contrarian arguments into climate negotiations. Greater transparency and improved methods for determining comparability of effort are also needed so that the Paris pledge-and-review mechanism can help negotiators achieve a fair climate deal. The institution of other coordinating mechanisms—such as near-term climate targets—may also foster more equitable burden-sharing between countries. To overcome the significant impediment of uncertainty about the threshold for catastrophe, climate negotiators must pursue efforts to incorporate a credible enforcement mechanism into the Paris Agreement or develop a new agreement with this feature. In the meantime, opportunities exist through linkage mechanisms for climate negotiators to co-opt existing non-climate institutions for transnational cooperation that already contain such mechanisms to achieve climate benefits. The intergenerational nature of climate change is another major handicap to collective action. However, the observation that reputation can facilitate climate cooperation when it generates benefits in other contexts suggests that this impediment might be tamed through the formation of climate clubs that yield short-term benefits that exceed the immediate rewards of free riding. An experimental assessment of the key instruments of the Paris Agreement suggests it will fail to limit global temperature rise to below 2 °C. As the window closes on humankind's opportunity to meet this goal, serious consideration should be given by climate negotiators to the strategies outlined here.

References

Aldy JE, Pizer WA (2014) Alternative metrics for comparing domestic climate change mitigation efforts and the emerging international climate policy architecture. REEP 10:3–24

Alley RB, Marotzke J, Nordhaus WD, Overpeck JT, Peteet DM, Pielke RA, Pierrehumbert RT, Rhines PB, Stocker TF, Talley LD, Wallace JM (2003) Abrupt climate change. Science 299:2005–2010

Barrett S (2003) Environment and statecraft: The strategy of environmental treaty-making

Barrett S (2011) Avoiding disastrous climate change is possible but not inevitable. P Natl Acad Sci USA 108:11733–11734

Barrett S (2013) Climate treaties and approaching catastrophies. J Environ Econ Manag 66:235–250

Barrett S (2014) Why have climate negotiations proved so disappointing? Sustainable Humanity, Sustainable Nature: Our Responsibility

Barrett S, Dannenberg A (2012) Climate negotiations under scientific certainty. P Natl Acad Sci USA 109:17372–17376

Barrett S, Dannenberg A (2014a) Sensitivity of collective action to uncertainty about climate tipping points. Nature Clim Chang 4:36–39

Barrett S, Dannenberg A (2014b) Negotiating to avoid 'gradual' versus 'dangerous' climate change: An experimental test of two prisoners dilemmas. In: Cherry T, Hovi J, McEvoy DM (eds) Towards a New Climate Agreement: Conflict, Resolution, and Governance. Routledge, London

Barrett S, Dannenberg A (2016) An experimental investigation into 'pledge and review' in climate negotiations. Climatic Change

Brysse K, Oreskes N, O'Reilly J, Oppenheimer M (2013) Climate change prediction: Erring on the side of least drama? Global Environ Chang 23:327–337

Budescu DV, Por H, Broomell SB (2012) Effective communication of uncertainty in the IPCC reports: A nationally representative survey. Clim Chang 113:181–200

Budescu DV, Por H, Broomell SB, Smithson M (2014) The interpretation of IPCC probabilistic statements around the world. Nat Clim Chang 4:508–512

Burton-Chellew MN, May RM, West SA (2013) Combined inequality in wealth and risk leads to disaster in the climate change game. Clim Chang 120:815–830

Dannenberg A, Löschel A, Paolacci G, Reif C, Tavoni A (2015) On the provision of public goods with probabilistic and ambiguous thresholds. Environ Res Econ 61:365–383

Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: A critical review. J Econ Lit 40:351–401

Freudenburg WR, Muselli V (2010) Global warming estimates, media expectations, and the asymmetry of scientific challenge. Global Environ Chang 20:483–491

Hardin G (1968) The tragedy of the commons. Science 162:1243–1248

Hasson R, Löfgren A, Visser M (2012) Treatment effects of climate change risk on mitigation and adaptation behaviour in an experimental setting. S Afr J Econ 80:415–430

Jacquet J, Hagel K, Hauert C, Marotzke J, Röhl T, Milinski M (2013) Intra- and intergenerational discounting in the climate game. Nat Clim Chang 3:1025–1028

Lenton TM, Held H, Kriegler E, Hall JW, Lucht W, Rahmstorf S, Schellnhuber HJ (2008) Tipping elements in the earth's climate system. P Natl Acad Sci USA 105:1786–1793

Lewandowsky S, Oreskes N, Risbey JS, Newell BR (2015) Seepage: Climate change denial and its effect on the scientific community. Global Environ Chang 33:1–13

Lo AY (2015) National income and environmental concern: Observations from 35 countries. Public Understanding of Science: 0963662515581302

Lo AY, Chow AT (2015) The relationship between climate change concern and national wealth. Clim Chang 131:335–348

Loewenstein GF, Read D, Baumeister R (2003) Time and decision: Economic and psychological perspectives on intertemporal choice. Russell Sage Foundation, New York

Lorenzoni I, Pidgeon NF (2006) Public views on climate change: European and USA perspectives. Clim Chang 77:73–95

Meinshausen M, Meinshausen N, Hare W, Raper SCB, Frieler K, Knutti R, Frame DJ, Allen MR (2009) Greenhouse-gas emission targets for limiting global warming to 2 °C. Nature 458:1158–1163

Milinski M, Röhl T, Marotzke J (2011) Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. Clim Chang 109:807–814

Milinski M, Semmann D, Krambeck HJ, Marotzke J (2006) Stabilizing the Earth's climate is not a losing game: Supporting evidence from public goods experiments. P Natl Acad Sci USA 103:3994–3998

Milinski M, Sommerfeld RD, Krambeck HJ, Reed FA, Marotzke J (2008) The collective-risk social dilemma and the prevention of simulated dangerous climate change. P Natl Acad Sci USA 105:2291–2294

Moore FC, Diaz DB (2015) Temperature impacts on economic growth warrant stringent mitigation policy. Nat Clim Chang 5:127–131

Nordhaus W (2015) Climate clubs: Overcoming free-riding in international climate policy. Am Econ Rev 105:1339–1370

Pacheco JM, Vasconcelos VV, Santos FC (2014) Climate change governance, cooperation and self-organization. Phy Life Rev 11:573–586

Peters GP, Andrew RM, Boden T, Canadell JG, Ciais P, Le Quéré C, Marland G, Raupach MR, Wilson C (2013) The challenge to keep global warming below 2 °C. Nat Clim Chang 3:4–6

Roeckner E, Giorgetta MA, Crueger ME, Pongratz JP (2011) Historical and future anthropogenic emission pathways derived from coupled climate-carbon cycle simulations. Clim Chang 105:91–108

Santos FC, Pacheco JM (2011) Risk of collective failure provides an escape from the tragedy of the commons. P Natl Acad Sci USA 108:10421–10425

Schelling TC (1960) The strategy of conflict. Harvard University Press, Cambridge, MA

Schellnhuber HJ, Cramer W, Nakicenovic N, Wigley T, Yohe G (2006) Avoiding dangerous climate change. Cambridge University Press, Cambridge, UK
Spence A, Poortinga W, Pidgeon N (2012) The psychological distance of climate change. Risk Anal 32:957–972
Stewart R, Oppenheimer M, Rudyk B (2013) A new strategy for global climate protection. Clim Chang 120:1–12
Tavoni A, Dannenberg A, Kallis G, Löschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. P Natl Acad Sci USA 108:11825–11829
Vasconcelos VV, Santos FC, Pacheco JM, Levin SA (2014) Climate policies under wealth inequality. P Natl Acad Sci USA 111:2212–2216
Velders GJM, Andersen SO, Daniel JS, Fahey DW, McFarland M (2007) The importance of the Montreal Protocol in protecting climate. P Natl Acad Sci USA 104:4814–4819
Victor D (2015) The case for clubs. International Centre for Trade and Sustainable Development, Geneva
Weber EU (2010) What shapes perceptions of climate change? Wiley Interdisc Rev: Clim Chang 1:332–342