

[Click here to view linked References](#)

# Cooperation Studies of Catastrophe Avoidance: Implications for Climate Negotiations\*

Mark J. Hurlstone,<sup>1,2</sup> Annabel Price,<sup>1</sup> Susie Wang,<sup>1,2</sup> Zoe Leviston,<sup>1,2</sup>  
and Iain Walker<sup>1,2</sup>

<sup>1</sup>*The University of Western Australia*

<sup>2</sup>*Commonwealth Scientific and Industrial Research Organisation*

---

## Abstract

Despite more than two decades of international negotiations, the world still lacks a credible approach for addressing the threat of dangerous climate change. It is becoming increasingly apparent that climate negotiators need to be equipped with better strategies if a tragedy of the climate commons 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 avoid crossing a dangerous climate threshold. This literature has uncovered five key variables that influence the likelihood of averting disaster: (1) the risk of collective failure, (2) inequalities in responsibility, wealth, or risk, (3) uncertainty surrounding the threshold, (4) intergenerational discounting, and (5) the prospect of reward or punishment based on reputation. 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.

---

## 1. Introduction

Human greenhouse gas emissions are continuing to rise at unprecedented rates posing the risk that one day average global temperature will exceed a critical threshold (Alley et al., 2003; Lenton et al., 2008; Schellnhuber et al., 2006)—defined in the Copenhagen Accord as a 2°C increase in temperature relative to preindustrial levels. Catastrophe avoidance requires the collective action of many nations to reduce their emissions of greenhouse gases by 50% by 2050 so that atmospheric concentrations can

---

\*Correspondence regarding this article may be addressed to Mark Hurlstone, School of Psychology, University of Western Australia, Crawley 6009 WA, Australia. E-mail: [mark.hurlstone@uwa.edu.au](mailto:mark.hurlstone@uwa.edu.au) URL: [mark.hurlstone.github.io](http://mark.hurlstone.github.io)

be stabilised (Meinshausen et al., 2009; Peters et al., 2013; Roeckner et al., 2011). The protection of the global climate is a global public good that will benefit everyone, whether or not they help to deliver it. 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 a group of individuals acting in their self interest ruthlessly overexploit a common resource to the detriment of their collective long-term best interests. The tragedy of the commons tell us that when individuals, groups, or countries are free to overexploit a resource, they typically do, leaving the public good at risk of collapsing due to the susceptibility of cooperation to exploitation by self-interested parties (Milinski, 2006).

As the protection of the global climate depends upon the aggregate effort of many countries, international climate negotiations are crucial so that emission reduction targets can be set and countries can establish that they are not acting in isolation. However, despite more than twenty years of international climate negotiations following the establishment of the United Nations Framework Convention on Climate Change in 1992, the world still lacks a robust climate agreement. It is becoming increasingly apparent that climate negotiators need to be equipped with better strategies to make treaties easier to agree on. In this article, we consider the lessons for climate negotiations that can be learned from an emerging literature in which the problem of avoiding dangerous climate change has been simulated using laboratory cooperation experiments involving public good games.

## 2. A Climate Catastrophe Avoidance Game

Simulating the problem of avoiding dangerous climate change requires an experimental paradigm that mimics the fundamental characteristics of the real climate game. This game has at least three key attributes: (1) To avert catastrophe, humankind must reduce greenhouse gas emissions to prevent global mean temperature from exceeding a dangerous threshold. (2) Climate protection cannot be provided by the unilateral action of a single country, but instead depends upon the multilateral efforts of many nations to collectively reduce their emissions. (3) The provision of the global good entails a social dilemma—countries can choose to reduce their emissions, which comes at a short-term economic cost, or they can choose not to do so, but run the risk of dangerous climate change further down the line, with more serious economic, environmental, and health costs.

Milinski and colleagues (Milinski et al., 2008) have devised a public goods game known as the “collective-risk social dilemma” that satisfies these constraints. The game involves groups of six players. Each player is given an operating fund of €40 and must decide whether to invest €0, €2, or €4 in each of 10 rounds to a climate account without communicating. If at least €120 has been invested in the climate account

by the end of the game then dangerous 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 dangerous climate change occurs with a 90% probability 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 limit on warming identified in the Copenhagen Accord—whilst the player contributions are a metaphor for the level of investment of countries in emission reductions.

### 3. Factors Affecting Cooperation

We review five findings observed using the collective-risk game that cast light on the factors that spur and inhibit collective efforts to avoid dangerous climate change. Specifically, (1) we present evidence that the risk of collective failure is a major variable underpinning cooperation—the higher the perception of risk, the easier it becomes for groups to coordinate efforts to avert disaster. (2) We then show how inequality—either in terms of historic responsibility, wealth, or risk—acts as an impediment to cooperation, but this handicap can be attenuated through communication or by setting intermediate climate targets. (3) Next, we show how uncertainty about the location of the dangerous climate threshold breaks down cooperation, whereas certainty surrounding the threshold provides a powerful incentive for collective action. (4) Subsequently, we underscore how the tendency for individuals to discount the effects of climate change on future others extends to group cooperation settings and constitutes a formidable barrier to cooperation, but this barrier can be mitigated by building a sense of affinity between current actors and future generations. (5) Finally, we highlight the importance of reputation as a powerful tool for achieving stable group cooperation.

#### 3.1. Perception of risk

To avoid crossing a dangerous threshold, countries must be convinced that they will be seriously adversely affected by climate change. This was demonstrated in the original experiment using the collective-risk game by Milinski et al. (2008). Their experiment involved three conditions, each containing 10 groups of players. In the low-risk condition, the probability of dangerous climate change occurring 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, a tragedy of the commons was virtually assured—in the low-risk condition, all groups failed to reach the threshold, whilst in the moderate-risk condition only 1 group was successful. By contrast, in the high-risk condition 5 groups succeeded in reaching the threshold and the remaining groups came close. This experiment suggests that the perception that there is a high probability of being adversely affected by climate change is a necessary—but not sufficient—condition to avert catastrophe. These results mesh well with game theoretic analyses of the collective-risk game showing that coordination for a global good is best realised in small to medium sized groups where the perception of

risk is high (Pacheco et al., 2014; Santos & Pacheco, 2011; Vasconcelos et al., 2014).

Further evidence for the link between perceived climate risks and cooperation comes from an earlier experiment by Milinski et al. (2006). The experiment employed a public good game analogous to the collective-risk game, except that groups did not have to reach a threshold by the end of the game to avoid catastrophe. Investments in the climate account were instead used to finance an advertisement in a large German newspaper encouraging people to reduce their fossil fuel usage to mitigate climate change. In one condition, groups were given expert information about immediate, intermediate, and distant climate risks before the public good game in the form of a brief verbal passage (risk-informed), whereas in a second condition groups received no such information (risk-uninformed). The authors found that the mean level of investments in the climate account in the risk-informed condition were 33% higher than in the risk-uninformed condition (€59 vs. €44). This study demonstrates that people are willing to make altruistic investments in climate protection, and this willingness is enhanced when people are given expert information about climate risks.

These experiments suggest that the higher the perception of risk, the easier it is to coordinate efforts to escape a tragedy of the commons. Climate negotiators therefore need to be convinced of the high risks posed by dangerous climate change. If they have a misperception that the risk is low, then this will imperil efforts to coordinate for the global good. Accordingly, it is important that the expert information that feeds into climate negotiations—notably the assessments of the Intergovernmental Panel on Climate Change (IPCC)—present an accurate overview of the severity of global climate disruption. However, several authors have expressed concerns that the IPCC assessments have been conservative in their projections of the impacts of climate change and that this may be attributable to an effort by climate scientists to “err on the side of least drama” to avoid the charge by skeptic groups of being “alarmist” (Brysse et al., 2013; Freudenburg & Muselli, 2010; Lewandowsky et al., 2015). The experiment of Milinski et al. (2008) suggests it is important not to understate the severity of dangerous climate change, since a reduced perception of risk could hamper collective efforts to avert catastrophe. Finally, we note that in the laboratory even under high risk settings with a small group of players, the risk of collective failure is high. This suggests the risk of collective failure will be much larger in the real climate game where—under the United Nations (UN) negotiating framework—almost 200 countries are vying to reach an agreement. Since the free rider problem is exacerbated as group size becomes larger (Pacheco et al., 2014; Santos & Pacheco, 2011; Vasconcelos et al., 2014), the highest levels of cooperation should be realised when the task of avoiding dangerous climate change is broken down into smaller negotiating groups of countries.

### 3.2. Player inequality

The experiments considered so far have assumed that all players are equal. However, this is not an accurate reflection of the real climate game. In the real game, inequalities exist in terms of historic responsibility for mitigating climate change (wealthy countries have historically emitted more greenhouse gasses than poor countries), re-

source capacity (wealthy countries have more financial resources to combat climate change than poor countries), and risk (poor countries are more vulnerable to the effects of climate change than wealthy countries). Several studies have added greater ecological validity to the collective-risk game, by incorporating such inequalities.

Tavoni et al. (2011) examined the impact of inherited inequality using an augmented collective-risk game divided into “passive” and “active” phase components. In the passive phase (rounds 1-3), the computer determined the investments made by each player. In one condition (equal condition), all group members were forced to allocate €2 per round to the climate account, whereas in a second condition (unequal condition), half the players were forced to contribute €4 per round, whereas 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 invest on each round.

The authors also incorporated a communication manipulation. In one condition, at the end of rounds 3 and 7, players were able to announce how much they intended to invest in the climate account over the next three rounds, via the submission of non-binding pledges (with-pledge condition). These pledges were then made public to the other players so they could ascertain whether the threshold would be reached if players adhered to their commitments. This coordination mechanism was chosen based upon its similarity to the “pledge-and-review” instrument enshrined in the Copenhagen Accord. Contributions in this condition were contrasted with those in a further condition where communication between players was prohibited (no-pledge condition).

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 neutralised 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 their lesser resource capacity, and the poor players, in turn, were willing to trust that the rich players would fulfil their pledges. This experiment shows that if players are able to communicate their intentions so that they may coordinate their efforts then inequality need not be an impediment to cooperation.

The handicap of wealth inequality can also be attenuated through the introduction of an intermediate climate target, as demonstrated in an experiment 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 if they did not reach the 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. This condition was designed to emulate the two time horizons of climate negotiations—the long-term target of avoiding crossing the dangerous threshold of 2 °C of warming by 2050 and the short-term target of reducing emissions by 2020 embedded in the Copenhagen Accord. Cooperation in this condition was contrasted with that in a condition without an intermediate threshold.

Without an intermediate threshold, investments were highest amongst rich groups, followed by mixed groups, followed by poor groups (Figure 1: left panel). Although investments over the rounds were relatively stable in rich and mixed groups, they decreased monotonically in poor groups. All the rich groups; 60% of mixed groups; and none of the poor groups reached the final threshold. However, when an intermediate threshold was provided (Figure 1: right panel), the dynamics of investments changed, with all group types exhibiting a punctuated peak in investments in round 5. Investments were again highest amongst rich groups, lowest amongst poor groups, with mixed groups falling in between. Once again, rich groups 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 investments of poor players.

In addition to inequalities in responsibility and wealth, combined inequalities in wealth and risk can influence collective success. In an experiment by Burton-Chellew et al. (2013) 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 risks if they failed to reach the threshold. In the unequal-wealth condition, group members also faced the same risks, but differed in their resource capacity, 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, heterogeneity in the resource capacity of group members was induced in the same way as in the unequal-wealth condition, but the risks for rich and poor players if climate change occurred were also varied. In the rich-suffer condition, the risk was higher for rich than for poor players, whereas the reverse was true in the poor-suffer condition.

Cooperation was high 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 resource heterogeneity 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 the climate account to circumvent a catastrophe. The question is whether this assistance would be provided when poor players were more at risk. The answer is no—in the poor-suffer condition, cooperation

collapsed, with only 13% of groups managing to reach the threshold. Thus, when wealth and risk heterogeneity are combined, rich players are only willing to cooperate if it is in their self-interest.

What implications can be drawn from these experiments on inequalities? The results of Tavoni et al. (2011) and Milinski et al. (2011) suggest that two fundamental pillars of the Copenhagen Accord—its bottom-up pledge-and-review mechanism and its emphasis on short-term intermediate climate targets—may help foster an equitable and fair distribution of commitments to reduce emissions amongst rich and poor countries. This is noteworthy, because conflicts between rich and poor dominated the negotiations in Copenhagen and the resulting agreement was not expected to alleviate these equity and fairness concerns. Ironically, the results of the experiments suggest these features of the Copenhagen Accord may actually help encourage rich nations to take the lead in efforts to tackle climate change. Nevertheless, the results of Burton-Chellew et al. (2013) suggest that asymmetries in climate risks between rich and poor countries may undermine such cooperative interactions. However, a shortcoming of their poor-suffer condition is its assumption that poor countries are more likely to suffer from dangerous climate change than rich countries. Whilst it is true that poor countries will be affected more by ongoing climatic changes (IPCC, 2014), dangerous climate change will be so severe that in a globalised world all countries will be affected by it (Alley et al., 2003; Schellnhuber et al., 2006). Thus, a more ecologically valid scenario would be one in which the risks from dangerous climate change are the same for rich and poor, but the costs from those risks are larger for poor than for rich. We note that under such a scenario the imperative for rich countries to cooperate is much higher than under the conditions examined by Burton-Chellew et al. This matter aside, what lessons can we learn from their experiment? The key implication of their study is that climate negotiations need to make transparent that dangerous climate change will result in a tragedy of the commons that will bring ruin to all, not just a select few. If rich countries believe that they will be immune to the effects of dangerous climate change then this will reduce the likelihood of them taking the lead in reducing emissions.

### 3.3. Uncertainty surrounding the dangerous threshold

Another difference between the real climate game and the experiments examined so far is that the latter assume that the location of the dangerous threshold; the effort that needs to be undertaken to avoid it; and the costs of crossing it are known with certainty. However, in reality, much uncertainty pervades these and other aspects of the climate change problem. Although the Copenhagen Accord identifies a dangerous temperature threshold of 2°C, uncertainties nevertheless exist surrounding the specific concentration target that must be reached to avoid crossing it. Similarly, estimates of the expected damages resulting from catastrophic climate change differ widely. It therefore follows that any adequate laboratory analogue of the problem of avoiding dangerous climate change must incorporate such uncertainties. This has been done in a sequence of studies by Barrett and Dannenberg (2012, 2013, 2014) who have shown experimentally (Barrett & Dannenberg, 2014) and theoretically (Barrett, 2013) that

the existence of a dangerous climate threshold spurs cooperation, whereas uncertainty surrounding this threshold—but not the impact of crossing it—causes cooperation to collapse (Barrett & Dannenberg, 2012, 2013).

Barrett and Dannenberg (2012) examined the influence of uncertainty about the location of the threshold for avoiding catastrophe and the consequences of crossing it on cooperation. In their collective-risk game, groups consist of 10 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 goal is to reach an investment threshold  $T$  by the end of 10 rounds, otherwise a cost  $C$  is deducted from each player’s endowment.

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

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 + threshold uncertainty conditions, respectively, averted catastrophe. These results demonstrate that uncertainty surrounding the location of 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 (2013) 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 and 155, (2) 140 and 160, (3) 135 and 165, or (4) 100 and 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 circumvented catastrophe. These results show that even a small degree of uncertainty about the threshold is a major impediment to collective action. Thus, uncertainty about the threshold must be eliminated to maximise the chances of a successful outcome.

With regards the prospects of escaping a real tragedy of the climate commons, these experiments suggest that uncertainty about climate damages is relatively incon-



sequential, whereas uncertainty about the location of the threshold is crucial. Unless the threshold is known with certainty, it is unlikely that countries will be able to take collective action to avert catastrophe. Although the Copenhagen Accord identifies a dangerous threshold of a 2°C change in temperature, what it does not identify is the atmospheric concentration level required to avoid crossing it. Since it is the concentration level that must crucially be known with certainty to determine the degree of emission abatement required to avert dangerous climate change, this means that the real climate game is being played under a condition of threshold uncertainty, like in the experiments of Barrett and Dannenberg. Thus, a key priority for the climate negotiations is to identify a precise concentration level to ensure that countries avoid the temperature threshold.

### 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 horizon). Thus, a defining feature of climate change is its intergenerational nature, which involves trade-offs between the self and future others—the immediate self interest of the current generation must be balanced against the ethical and moral imperative to protect future generations from harm. 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 (Schelling, 1995, 2000). In individual decision making, both forms of discounting are known impediments to action on climate change (Lorenzoni & Pidgeon, 2006; Spence et al., 2012; Weber, 2010). Do intra- and inter-generational discounting also manifest in a group decision making setting resembling climate negotiations?

This question was addressed using the collective-risk game 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 the endowment was only awarded if catastrophe was circumvented. 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<sub>2</sub>, 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 inter-generational discounting, but intergenerational discounting was the stronger impediment to cooperation of the two. This experiment suggests that the intergenerational nature of climate change is likely a major obstacle to cooperation in climate negotiations.

How can this obstacle be surmounted? Hurlstone et al. (2015) conducted an experiment that incorporated the short-delay and intergenerational conditions of Jacquet et al. in conjunction with a third “affinity” condition. The affinity condition was the same as the intergenerational condition, except on rounds 1, 4, and 7, group members read a brief verbal passage before making their investment decisions. The passages were constructed to invoke a sense of affinity—a combination of empathy, perspective taking, and perceived connectedness—with future generations. One passage highlighted how past generations made many sacrifices to make the world we inherited a better place. It highlighted how we have a responsibility to reciprocate the kind acts performed by past generations for us by taking action to prevent dangerous climate change. A second passage highlighted the helpless nature of future generations by noting the power asymmetries between current and future generations—everything that we do will affect future generations, but nothing that they do will affect us. It underscored how we have an ethical and moral imperative to use our power responsibly by not imposing the burden of dangerous climate change on our descendants. A third passage highlighted the finiteness of human life, but underscored how efforts to tackle climate change represent an opportunity to leave a positive lasting legacy. The passages were motivated by research showing that building affinity with future generations by increasing intergenerational identification, highlighting power asymmetries, and appealing to people’s desire to leave positive legacies can increase intergenerational beneficence (Wade-Benzoni, 2008; Wade-Benzoni, Plunkett Tost, 2009; Wade-Benzoni et al., 2012).

Intergenerational discounting once again had a strong negative impact on cooperation—only 20% of groups in the intergenerational condition averted catastrophe, compared to 90% in the short-delay condition. However, the key result was that the verbal passages in the affinity condition mitigated the effect of intergenerational discounting—70% of groups in this condition managed to reach the threshold, a success rate which did not differ reliably from that in the short-delay condition.

These experiments suggest that intergenerational discounting is a major impediment to cooperation, but its effects can be lessened dramatically—and intergenerational beneficence enhanced—by crafting persuasive messages that build an affinity between current actors and future generations. The implication of these results for climate negotiations is that persuasive appeals for collective action to avoid dangerous climate change should be framed around intergenerational concerns, by highlighting

power asymmetries, the opportunity for legacy creation, and our responsibility to “pay forward” the debts we owe to past generations.

### 3.5. Reputation

In the experiments considered so far, the reputation of players cannot be used as a tool for aiding cooperation. However, if a positive reputation can be used as a currency by which to obtain rewards and avoid punishments, stable cooperation can be achieved. This was demonstrated in the experiment of Milinski et al. (2006; see also Milinski et al., 2002; Semmann et al., 2004) described earlier, in which some groups were given expert information about climate risks, whereas others were not. However, this manipulation was ancillary to the primary objective of the experiment, which was to establish whether reputation could be used as a tool to sustain cooperation.

In their experiment, rounds of the climate public good game were alternated with rounds of an indirect reciprocity game, where players could reward other players based on their reputation. In each round of the latter game, players adopted the role of potential “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. Thus, direct reciprocity—reciprocating the kind or unkind act of another—was not possible; interactions were instead based on indirect reciprocity—reciprocating the kind or unkind act of another to a third party.

The key manipulation was that on “odd” rounds of the public good game, investments in the climate account were made public, whereas on “even” rounds the investments were made anonymous. Thus, on rounds in which investments were made public, 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—viz. donors should offer support to players that cooperate in the public good game—whereas failure to cooperate affords a negative reputation, which may incur punishment—viz. donors should withhold support to players that do not cooperate in the public good game. By contrast, on the anonymous rounds, the incentive to cooperate and the disincentive to free-ride is removed because the reputation of players is unknown and cannot be used as a basis for deciding to donate or not to another player in the indirect reciprocity game.

The results are shown in Figure 2, which plots the average group investments on each round of the public good game in the risk-informed and risk-uninformed conditions. The data for both conditions exhibit a markedly scalloped appearance, with peaks on odd rounds in which investments in the public good game were made public and troughs on even rounds where investments were anonymous. It can also be seen that public investments were not just markedly higher, they were also more stable, whereas anonymous investments decreased monotonically over rounds. Investments were higher on public than anonymous rounds because cooperation was subsequently

rewarded in the indirect reciprocity game and defection was punished.

These results show that if altruistic investments in climate protection can be socially reinforced and selfish investments socially punished then reputation is a powerful tool for achieving stable cooperation. This suggests that the reputations of countries should be made salient during climate negotiations. How might this be accomplished? Countries could be ranked according to their emissions and investments in climate protection and this information could be used to create an “image score” that acts as a signal of their reputation (Dreber & Nowak, 2009; Pfeiffer & Nowak, 2006). To facilitate the transmission of the signal, countries could be further differentiated on the basis of these image scores into categories, such as “heavy lifters”, “fair contributors”, and “free-riders”. The “heavy lifters” and “fair contributors” would be socially rewarded with praise, whereas the “free-riders” would be socially punished through shaming. Is this proposal far-fetched? We think not. Indeed, the executive secretary of the UN, Christiana Figueres, recently lamented Australia’s lack of climate ambition stating that in terms of climate diplomacy, it has transitioned from “leadership to free-rider status”. Faced with such social disapproval, some free-riding countries may choose to cooperate simply out of a desire to preserve their good reputation. For other countries, shaming alone will not suffice. However, because those countries are engaged in multiple cooperative relationships, it will only be a matter of time before their poor reputations spill over into other situations requiring cooperation. The threat of other countries withholding their support during these interactions should provide a powerful incentive for defecting countries to invest in their reputations by joining collective efforts to tackle climate change.

#### 4. Summary

Avoiding dangerous climate change is perhaps the greatest collective action problem ever and is “the game we cannot afford to lose” (Dreber & Nowak, 2009). More than two decades of climate negotiations have failed to identify a cooperative solution to this problem. Laboratory cooperation experiments mimic the collective-risk social dilemma faced by countries playing the real climate game, and provide insights into the factors that might prevent a real tragedy of the climate commons. These experiments suggest that for cooperation to be achieved, (1) countries must be convinced of the very high risks of dangerous climate change, and (2) that failure to protect the global climate will bring ruin to all, not merely a select few. (3) They further suggest that knowledge of the temperature threshold for dangerous climate change will not prevent catastrophe, unless uncertainty surrounding the concentration level required to avoid crossing it can be reduced. (4) The intergenerational nature of climate change is a significant barrier to cooperation, but climate negotiators can attenuate its effects using affinity building messages that promote intergenerational identification, highlight power asymmetries between current and future actors, and appeal to people’s preference to leave positive legacies. Finally, (5) people care about their reputations, which provide a signal of their trustworthiness to others. Accordingly, the reputations

of countries should be made salient so that the threat of social punishment in other cooperative interactions provides an incentive for them to invest in climate protection.

## Figure Captions

**Figure 1:** Money invested in the climate account in each of 10 rounds of the climate game as a function of group composition in the experiment of Milinski et al. (2011). Left panel: investments without an intermediate threshold; Right panel: investments with an intermediate threshold. The broken horizontal line in each panel represents the average investment per round required to reach the €120 threshold.

**Figure 2:** Money invested in the climate account on public (odd) and anonymous (even) rounds of the public good game as a function of risk condition in the experiment of Milinski et al. (2006).

## References

- 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 (2013) Climate treaties and approaching catastrophes. *J Environ Econ Manag* 66: 235-250.
- Barrett S, Dannenberg A (2011) Climate negotiations under scientific certainty. *P Natl Acad Sci USA* 109: 17372-17376.
- Barrett S, Dannenberg A (2013) Sensitivity of collective action to uncertainty about climate tipping points. *Nature Clim Change* 4: 36-39.
- Barrett S, Dannenberg A Negotiating to avoid ‘gradual’ versus ‘dangerous’ climate change: an experimental test of two prisoners dilemmas. In: Cherry T, Hovi J, McEvoy DM, editors. *Towards a new climate agreement: conflict, resolution, and governance*. London: Routledge; 2014.
- 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.
- Burton-Chellaw MN, May RM, West SA (2013) Combined inequality in wealth and risk leads to disaster in the climate change game. *Climatic Change* 120: 815-830.
- Dreber A, Nowak MA (2009) Gambling for goods. *P Natl Acad Sci USA* 105: 2261-2262.
- Frederick S, Loewenstein G, O’Donoghue T (2002) Time discounting and time preference: A critical review. *Journal of Economic Literature* 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.
- Hurlstone MJ, Price A, Wang S, Leviston Z, Walker I (2015, under review) Building affinity with future generations mitigates intergenerational discounting in the climate game.
- Intergovernmental Panel on Climate Change (2014) *Climate change 2014: Impacts, adaptation, and vulnerability. Summary for policymakers*.
- Jacquet J, Hagel K, Hauert C, Marotzke J, Röhl T, Milinski M. (2013) Intra- and intergenerational discounting in the climate game. *Nature Clim Change* 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.

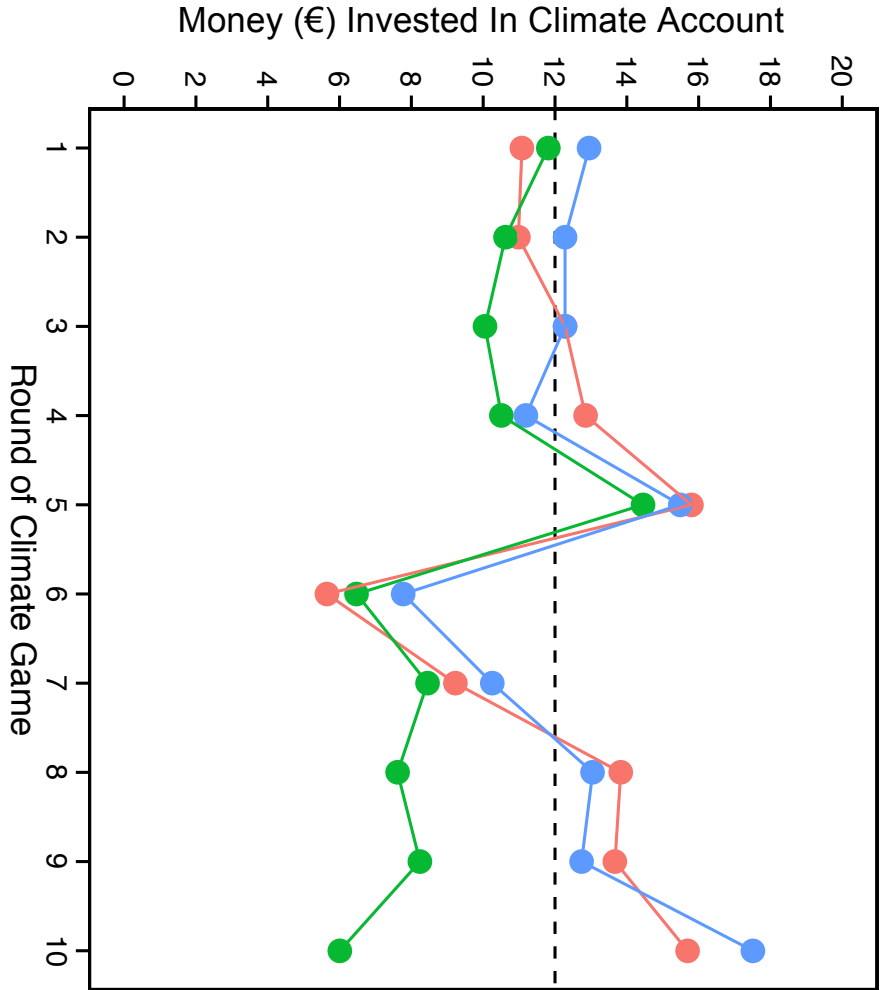
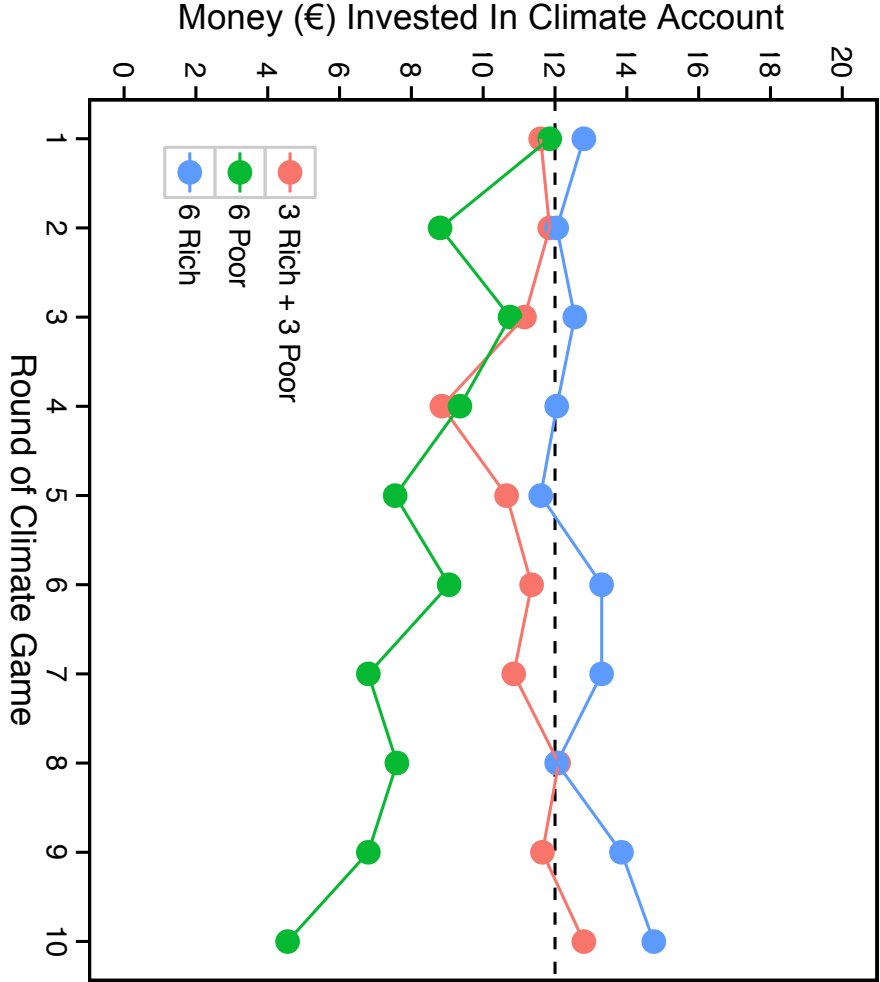
- Lewandowsky S, Oreskes N, Risbey JS, Newell BR (2015) Seepage: Climate change denial and its effect on the scientific community. *Global Environ Chang* 33: 1-13.
- Loewenstein GF, Read D, Baumeister R Time and decision: Economic and psychological perspectives on intertemporal choice. New York, NY: Russell Sage Foundation; 2003.
- Lorenzoni I, Pidgeon NF (2006) Public views on climate change: European and USA perspectives. *Climatic Change* 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. Reputation, personal identity and cooperation in a social dilemma. In: Kappeler PM, van Schaik CP, editors. *Cooperation in primates and humans*. Berlin Heidelberg: Springer-Verlag; 2006. pp. 265-278.
- Milinski M, Röhl T, Marotzke J (2011) Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic Change* 109: 807-814.
- Milinski M, Semmann D, Krambeck HJ (2002) Reputation helps solve the ‘tragedy of the commons’. *Nature* 415: 424-426.
- 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.
- 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. *Nature Clim Change* 3: 4-6.
- Pfeiffer, T., & Nowak, M. A. (2006). All in the game. *Nature*, 441, 583-584.
- Roeckner E, Giorgetta MA, Crueger ME, Pongratz JP (2011) Historical and future anthropogenic emission pathways derived from coupled climate-carbon cycle simulations. *Climatic Change* 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 (1995) Intergenerational discounting. *Energ policy* 23: 395-401.
- Schelling TC (2000) Intergenerational and international discounting. *Risk Anal* 20:



833-837.

- Schellnhuber HJ, Cramer W, Nakicenovic N, Wigley T, Yohe G Avoiding Dangerous Climate Change. Cambridge, UK: Cambridge Univ Press; 2006.
- Semmann D, Krambeck HJ, Milinski M (2004) Strategic investment in reputation. *Behav Ecol Sociobiol* 56: 248-252.
- Spence A, Poortinga W, Pidgeon N (2012) The psychological distance of climate change. *Risk Anal* 32: 957-972.
- 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.
- Vaconcelos VV, Santos FC, Pacheco JM, Levin SA (2014) Climate policies under wealth inequality. *P Natl Acad Sci USA* 111: 2212-2216.
- Wade-Benzoni KA (2008) Maple trees and weeping willows: The role of time, uncertainty, and affinity in intergenerational decisions. *Negotiation and Conflict Management Research* 1: 220-245.
- Wade-Benzoni KA, Tost LP (2009) The egoism and altruism of intergenerational behavior. *Pers Soc Psychol Review* 13: 165-193.
- Wade-Benzoni KA, Tost LP, Hernandez M, Larrick RP (2009) It's only a matter of time: Death, legacies, and intergenerational decisions. *Psychol Sci* 23: 704-709.
- Weber EU (2010) What shapes perceptions of climate change? *Wiley Interdisciplinary Reviews: Climate Change* 1: 332-342.

Figure



Figure

