b
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

University of Bern

Master Thesis

Insights from Behavioral Economics on Climate Negotiations

Author:
Remo Bebié

Cand. MSc. in Climate
Sciences with special
Qualifications in Economics

Supervisors:
Prof. Dr. Gunter Stephan
Prof. Dr. Ralph Winkler

Department of Economics
University of Bern

May 24, 2016

Contents

List of Figures

List of Tables

1 Introduction

Carbon based greenhouse gases are the most important drivers of global warming and the associated changes in climate. The earth's atmosphere is the main reservoir for human carbon emissions over the coming decades. In the long term, the world's oceans will take up most of that carbon. Both the ocean and the atmosphere are turbulent, dynamic systems that circulate around the globe, mixing over time. It therefore does not matter where carbon based greenhouse gases are emitted into the atmosphere. The global climate is affected regardless of where carbon is emitted through human activity, although the effects may differ regionally.

This property of the climate system renders it a public good. *It is shared by the entire human population, so it apparently does not pay for a single individual to invest in protecting it: the direct benefit that an individual gains from his or her investment is much smaller than the costs. The Earth's climate is therefore vulnerable to overexploitation — it faces a 'tragedy of the commons'* (Pfeiffer and Nowak, 2006, p. 583).

Taking into account the incentives at odds with a pareto-optimal outcome, traditional economic theory suggests that public goods are best provided by the public sector. In the context of climate change, this would require a supranational authority that could enforce mitigation globally. Since this type of authority does not exist, other solutions have to be found to overcome the climate change crisis.

The formation of the United Nations Framework Convention on Climate Change (UNFCCC) marked the starting point for an ongoing international effort to overcome this "tragedy" through negotiation and between country agreements. However, these negotiations have so far failed to avoid the tragedy of the commons described above. Although agreements have been reached in the past, important parties to the negotiations have either not ratified the agreements or failed to meet their reduction targets. The most recent achievements at COP21 were hailed as an international success. At least on the negotiation front, there finally seems to be substantial progress. The question remains, how this agreement will be turned into action and if the necessary steps are taken in time to avoid the worst.

The aim of this thesis is to investigate "make or break" factors for coordination to mitigate climate change based on experiments developed in behavioural economic game theory. One key challenge in climate negotiations is agreeing on how to share the economic costs of climate mitigation. Assuming that the total cost of keeping climate change below a certain level is known a priori, the situation appears not so different from a very complex game with a large number of players. In this case, the players have to agree on the share of the total cost, the country they represent is willing to bear. For this reason, a number of public good games have been framed to the context of climate negotiations in order to capture what factors, besides traditional economic rationality, may affect the outcomes.

This thesis is structured as follows: The following sections of the introduction provide an overview of the context of the thesis. They focus specifically on challenges that emerged in past efforts to mitigate climate change. For this purpose a brief summary of the UNFCCC history and past negitiation outcomes is provided. It will serve as a point of orientation in this thesis to remind us of the measures and mechanisms applied in negotiations.

The second section makes a general argument why behavioural economics matters in the context of climate change. It then investigates a number of study designs that used experimental game theory to identify key success factors for collaboration in a public good social dilemma situation. Section three places a focus in the aspect of vulnerability in the context of climate change.

This is followed by a description of the experiment performed in this thesis. Section four describes the methods applied as well as the experimental setup and parametrization. Section five then shows the results obtained from the experiment both for the entire sample, as well as for different sub-samples. Section six contains a discussion of the key findings and tries to explain the observations that were made.

The final section contains the conclusions drawn from this research effort. Additionally this thesis contains two appendices. Appendix I holds additional information regarding relevant literature that did not directly affect the experimental design. Appendix II contains screenshots from the experiment to give the reader an idea about what the user interface looked like for the participants in the lab.

1.1 The Kyoto Protocol

The Kyoto Protocol was the first binding climate agreement on a global scale. Negotiating parties were divided into Annex I and Non-Annex I members. Article 3 of the protocol (UNFCCC, 1998) states an emission reduction target of 5% below 1990 levels over the commitment period of 2008 to 2012. In the protocol's Annex B, a selection of Annex I countries were assigned an individual GHG target ranging from -8% to +10% for this period, depending on the state of development of the country belonging to the Annex B group. Countries were awarded some flexibility in the way they may achieve those targets and could make use of the Kyoto mechanisms to complement their direct national reduction efforts. The UNFCCC (2011, p.1) states in a 2011 factsheet on the Kyoto Protocol, that *industrialized countries must first and foremost take domestic action against climate change, but the Protocol allows them a certain degree of flexibility in meeting their emission reduction commitments through three innovative market-based mechanisms.*

The Kyoto Protocol specified three separate mechanisms for this purpose:

- Emissions Trading through carbon markets
- The Clean Development Mechanisms (CDM)
- Joint Implementation (JI)

The United States, at the time (and until rather recently) the biggest emitter of greenhouse gases, never ratified the protocol. The nations that committed to the first commitment period therefore covered roughly 50% of global emissions. This is of course also due to the fact that Non-Annex I countries, to which countries like China and India belong, were not assigned a reduction target within the Kyoto Protocol, based on the principle of common but differentiated responsibility. For the second commitment period this percentage declined further as Japan, the Russian Federation, Canada and New Zealand opted out, with the United States still not ratifying the protocol.

Although its success was limited, the Kyoto Protocol was the only successful internationally binding agreement on climate change mitigation before COP21. The nations that did ratify the protocol showed mixed success in achieving their specified targets but many have undertaken efforts of various strength to reduce their emissions as a result of the protocol.

1.2 COP15 in Copenhagen

The first attempt to reach a new global goal to mitigate climate change showed limited success in 2009 at COP15 in Copenhagen, with a complete breakdown of negotiations only narrowly avoided at the last minute. The Copenhagen Accord finally resulted from a secluded meeting between the leaders of the United States, China, India, Brazil, South Africa. The Conference of the Parties then took note of the accord. Measured against the goal of a reaching a new global climate deal negotiated by all parties of the UNFCCC, it was thus a failure. However, it may well have been a first step towards the new agreement now reached in Paris.

As Schneck (2009) notes, the Copenhagen Accord was equally significant for what it achieved as it was significant for what it lacked. Differential responsabilities of parties were affirmed and it was agreed that global warming should be kept below 2 ° C. Developed countries were given the target to raise $100 billion per year in financing by 2020, and $30 billion commitment for 2010-2012 period. The parties recognized crucial need to reduce emissions from forests and agreed to establish immediate mechanisms (including REDD+) to provide financial resources from developed countries. Schneck (2009) furthermore indicated that the Annex classifications, as they were applied during the negotiations leading up to the Kyoto Protocol, must be questioned as the weights in terms of economic development have shifted since then.

(Schneck, 2009, p. 4) noted that the Copenhagen Accord lacked specifically:

1. Legally binding requirements to uphold the Accord.
2. A date for completion of a binding agreement, or process to achieve one.
3. Global emission targets for 2020 or 2050 (although nations were encouraged to make individual pledges by 2010)
4. Specific rules and procedures for monitoring, reporting and verification (MRV) of emissions reductions

A number of nations did submit pledges before the introduction of the new INDC concept now applied at COP21. A total of 73 pre-2020 pledges, covering 83.1% of global GHG emissios, were submitted according to the World Resource Institute (WRI) WRI (2015), following the Copenhagen Accord and subsequent negotiations such as the Cancun Accord of 2010. The type of pre-2020 pledges by country are shown in Figure 1. Meetings of the parties subsequent to COP15 focused on quantifying emission reduction pledges for the period up to 2020 and clearing the path for a new global climate agreement at COP21 in 2015.

1.3 COP21: A new approach with INDCs

The 21^{st} conference of the parties to the UNFCCC (COP21) succeeded in reaching a global climate deal. Its setup was different to previous negotiations in a number of ways. First, the framing of the negotiations was complemented by a bottom-up approach. The new approach used pledges, called Intended Nationally Determined Contributions (INDCs), to drive countries

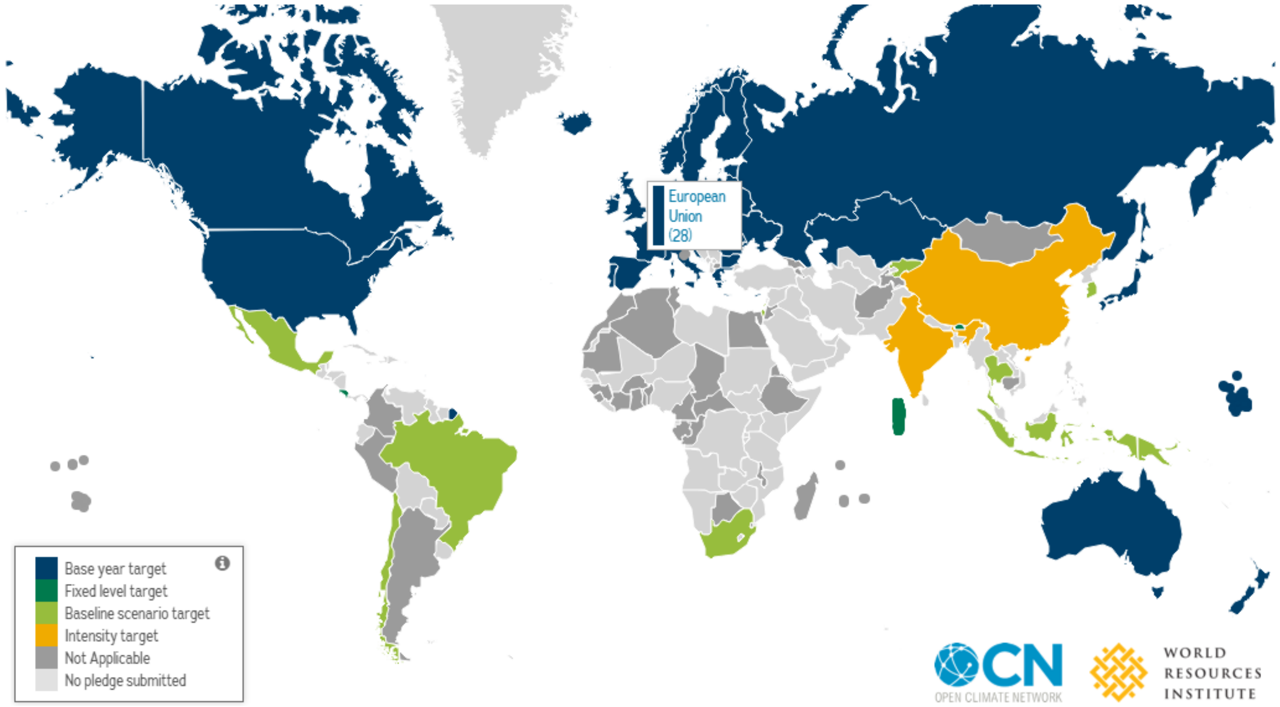


Figure 1: Greenhouse Gas Emissions Target Type of pre-2020 Pledges by Country as made available by the WRI (2015). Not applicable (dark grey shading) indicates that countries have submitted a pre-2020 pledge, but not in the form of a GHG target. Pre-2020 pledges refer to reduction targets concerning time-periods up to the year 2020.

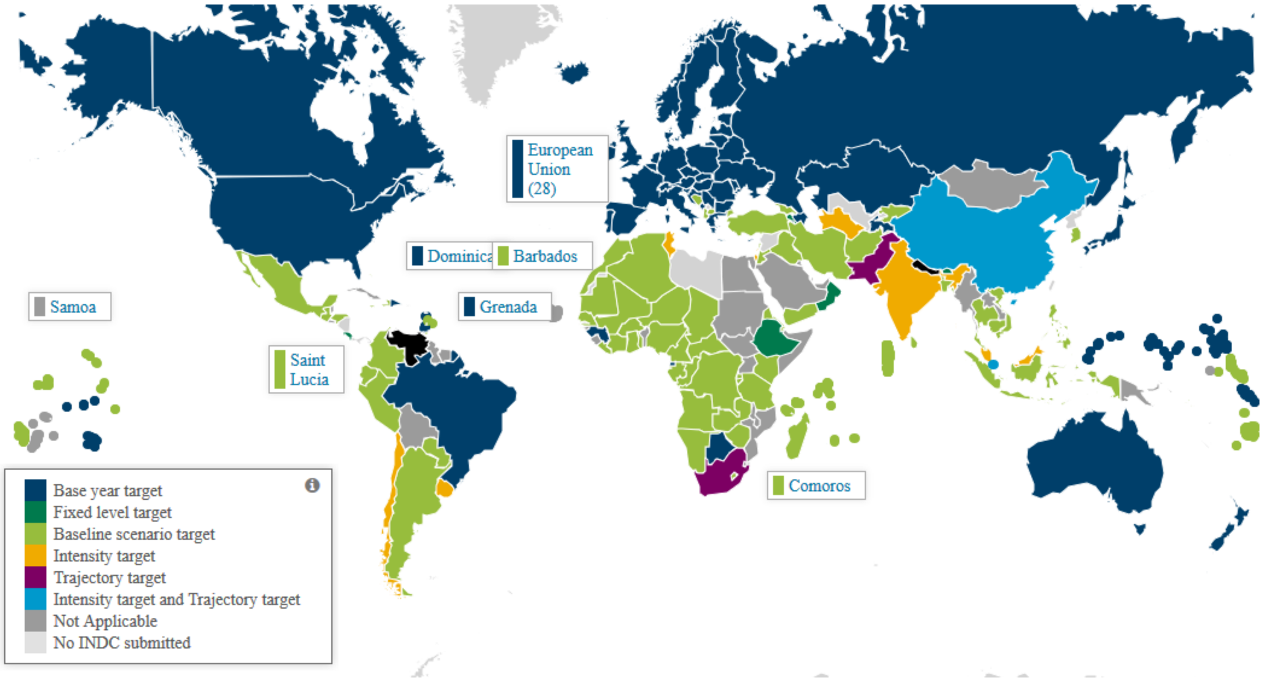


Figure 2: Greenhouse Gas Emissions Target Type in submitted INDCs, as of March 2016, by Country. Data shown as made available by the WRI (2015). Not applicable (dark grey shading) indicates that countries have submitted an INDC, which does not contain a GHG target.

to commit to mitigation and made no distinction between Annex I and Non-Annex I countries. INDCs are not legally binding, but represent the individual commitments countries are prepared to make before negotiations start. Secondly, the goal was to reach a new agreement on reduction of GHG emissions to enter into force by 2020, which is why INDCs contain targets beyond the year 2020 (typically targets for the year 2030). Pledges within INDCs can therefore not be directly compared to previous pledges, as the latter usually cover a different timeframe.

This new approach is fundamentally different to the top-down approach applied in the Kyoto Protocol, where reduction targets were negotiated between all parties, and industrialized nations were then assigned targets under the protocol which they did or did not ratify. Because COP21 deviated from the traditional Annex classifications, the setup was also fundamentally different to the negotiation approach taken at COP15.

By submitting INDCs, countries were encouraged to come forward and submit their pledges, and subject them to public scrutiny. The exact contents that should be included in INDCs was not officially defined, but a set of guidelines existed. According to the World Resource Institute (Levin et al., 2015, p. 21), *"the Lima Call for Climate Action references developing fair and ambitious INDCs that contribute towards achieving the objective of the Convention. The fairness and ambition of an INDC will be a value judgement; each Party will need to reflect on how it perceives fairness and ambition for itself and others, and how it will measure fairness and ambition. Information on the future level of emissions if the INDC is achieved, as well as emissions reductions that would result from implementing the INDC, can be helpful for evaluating the INDC against these criteria."*

The data provided by the WRI lets us observe how the situation changed after the INDC concept was introduced. As of March 6^{th} 2016, 161 INDCs were submitted, representing 188 countries. Collectively they cover 98,7% of global GHG emissions, according to the WRI WRI (2015).

Comparing Figures 1 and 2 shows how the types of pledges have evolved between the pre-2020 pledges and the pledges contained in the newly formed INDCs. Of course, the type of pledge per se does not indicate the strength of the pledge. Yet there is a rather striking difference between the percentage of actual GHG-targets submitted through the INDCs in 2015 compared to earlier pledges represented by the "pre-2020 pledge" category. Figure 3 shows how these percentages have changed:

1.4 Outcomes and Implications of INDC's

A major success of the COP21 negotiations was the formulation of a climate deal. Article 2 of the Paris Agreement contains the following objectives that were agreed upon by all members of the UNFCCC (United Nations, 2015):

- Holding the increase in the global average temperature to well below 2 ° C above pre-industrial levels and to pursue efforts to limit the temperature increase to 1.5 ° C above pre-industrial levels, recognizing that this would significantly reduce the risks and impacts of climate change
- Increasing the ability to adapt to the adverse impacts of climate change and foster climate resilience and low greenhouse gas emissions development, in a manner that does not

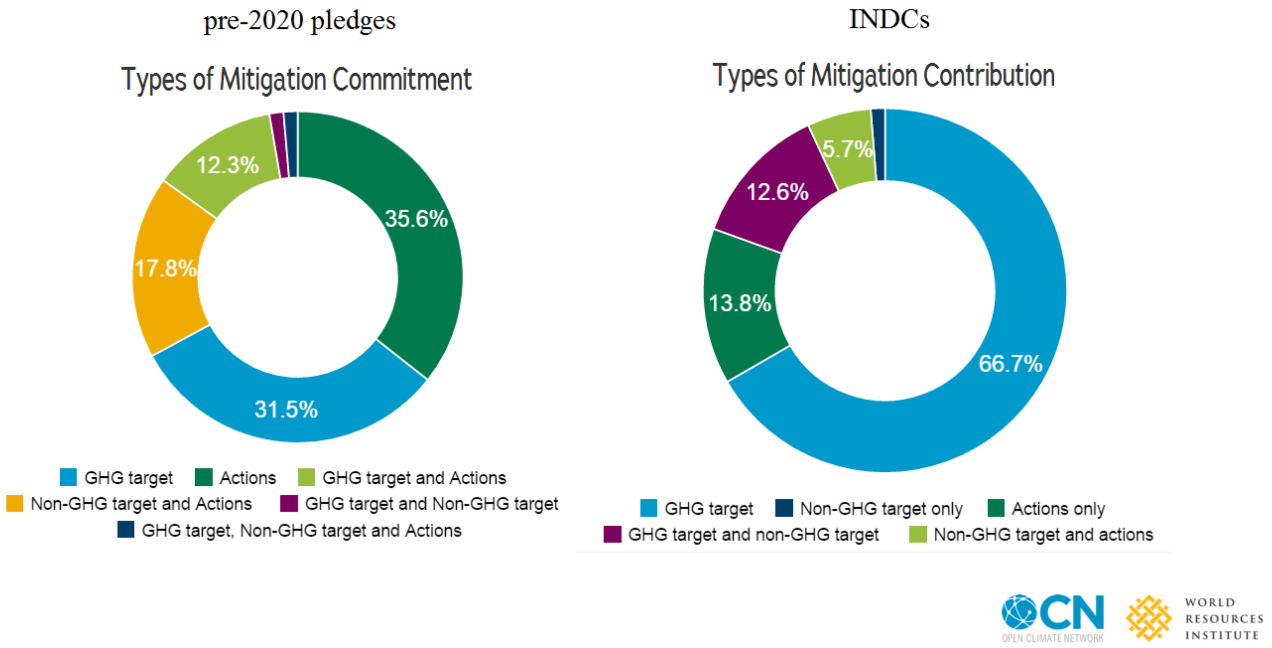


Figure 3: Types of pledges made in the pre-2020 setting compared to the pledges contained in INDCs as of March 6th 2016. The graphs are based on the data made available in the CAIT tool by the WRI (2015).

threaten food production

- Making finance flows consistent with a pathway towards low greenhouse gas emissions and climate- resilient development.

Comparing the commitments made by different countries over time or between countries is often challenging. Reasons include different emission base years to which reduction targets apply. Examples include targets relative to different scenarios instead of base years, intensity targets (e.g. GDP carbon intensity). Additionally, targets are often conditional on the actions by other parties, which again makes comparison difficult. A strong reduction target that is conditional on commitments by another party can be hard to compare to a weaker but unconditional reduction target. Such comparisons require a comprehensive assessment of individual pledges and their implications for emission pathways.

Shortly after India submitted its INDC on October 1st, the Climate Action Tracker (CAT) Hare et al. (2015) released a first assessment of what the pledges submitted within the deadline would mean for global emissions of greenhouse gases and the warming trajectory that can be derived. Underlying this assessment is the application of the data in the INDCs to climate models of which the median model prediction is shown. They find that the unconditional commitments made, including the individual unconditional INDCs submitted to the UNFCCC Secretariat as of March 2016, do not suffice to limit expected global warming to 2 ° C. But compared to policy projections in place in October 2015 before COP21, they suggest a noticeable improvement.

However, these results should be taken with a grain of salt. At ETH's Klimarunde in November of 2015, renowned Climate Scientist Reto Knutti pointed out that most INDCs actually only contain a 2030 target and that an assessment of warming implications requires information about long-term emission pathways. According to his assessment, the 2030 targets are in line with the RCP 4.5 emission range for that specific year only. The INDC's typically do not contain end of century emission goals.

Although many concerns remain about the salience of the agreement, COP21 facilitated a climate agreement between all members of the UNFCCC. With the INDC's most of these countries now also have suggested measures they will take to achieve those new goals of the UNFCCC. The conference was successful in bringing many emerging economies to the table and make them recognize the need for action by all.

The talk has been talked, now the walking needs to start. The international agreement to limit global warming to well below 2 ° C will require strong and immediate actions. The pledges made through the INDC's will only be effective if they are met with actions. So far, no specific binding actions have been negotiated. In this thesis I will try to examine some of the factors that may be necessary to turn talks into action. I will focus on insights from experimental behavioural economics to investigate the settings in which sufficient collaboration is most likely and to what extent such settings apply to climate negotiations.

2 Behavioral Economics of Cooperation

This section describes how behavioural economics matters in the context of climate negotiations. It also contains a number of game designs that investigate bounded rationality as an alternative to rational choice theory in the context of climate change mitigation. The literature criticising traditional economic concepts such as homo oeconomicus is very broad. Instead of trying to summarize this literature to the context of climate change myself, previous efforts by scholars much more versed in the field are used. Renowned economist Elinor Ostrom, along with others, provided a strong line of arguments why conventional economic theory may not be the best option for analysing how to reduce the threats of massive climate change (Ostrom, 2014). Her line of arguments are used to give an introduction as to why behavioural economics matter in the context of climate change mitigation.

2.1 The Conventional and Updated Theory of Collective Action

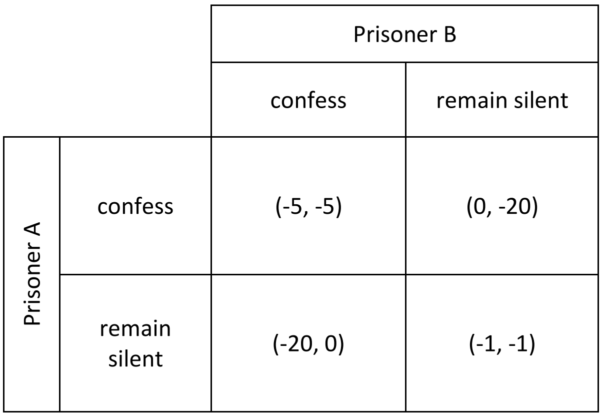
According to Ostrom, the term ”social dilemma” refers to situations where uncoordinated decisions based on short-term self interest leads to suboptimal outcomes for others and for self. Economists speak of dilemmas in this context because there is usually an alternative outcome that would be socially optimal. Socially optimal means that all players would achieve higher payoffs if they cooperated. Elinor Ostrom consistently speaks of socially optimal outcomes, which is a normative notion. To avoid making a normative statement one could instead call this type of outcome pareto-optimal. Not achieving what Ostrom calls the socially optimal outcome is also not pareto-optimal, if there is a way to increase at least one actor's benefit without creating a loss for another actor (Sen, 1993).

However, if actors pursue short-term maximization of material benefits they are predicted to fail in achieving the pareto-optimal outcome. The predicted outcome is characterized by a Nash equilibrium which rests upon the assumptions of rational choice theory. In a stable Nash equilibrium, no actor has an incentive to deviate from his or her strategy, because the individual material benefits would always be lower compared to the current strategy (Ostrom, 2014).

A classic example that illustrates such a situation in a simple way is the prisoner's dilemma. In this setup there are two players who have been caught whilst committing a crime. They are interrogated separately and each have to options: they can either remain silent or confess. These options lead to payoff matrix in table 1.

Given this specific parametrization, a dominant strategy exists for both players. Prisoner A will always be able to improve his payoff by choosing to confess no matter what Prisoner B chooses to do. The same is true vice versa for prisoner B. Therefore the Nash-Equilibrium in this case is for both prisoners to confess. However, the pareto-optimal choice would be for both prisoners to remain silent as this would decrease the disutility for both actors. Since this outcome is at odds with each prisoners isolated self interest, the situation is described as a dilemma.

Section 8.1 in the Appendix I contains another simple example of a setting where individuals' self-interested actions lead to an outcome that is not pareto-optimal. This type of behaviour is often described as consistent with rational choice theory or ”traditional assumptions”. Along with rational choice theory, a number of other assumptions are often applied in what Ostrom



		Prisoner B	
		confess	remain silent
Prisoner A	confess	(-5, -5)	(0, -20)
	remain silent	(-20, 0)	(-1, -1)

Table 1: Payoff-matrix for the classic prisoner's dilemma. Values in brackets are the payoffs for both players for a given set of choices. For example, if both players confess (-5, -5) describes the payoff for both prisoners, where the value on the left side in the brackets is prisoner A's payoff and the value on the right side is Player B's payoff.

calls the "conventional theory of collective action". She claims that in most game theoretic models of social dilemmas, the following assumptions are made:

1. All participants have complete and common knowledge of the exogenously fixed structure of the situation and of the payoffs to be received by all individuals under all combinations of strategies.

2. Decisions about actions are made independently and simultaneously.

3. Participants do not communicate with one another.

4. No central authority is present to enforce agreements among participants about their choices.

Ostrom argues that when rational choice theory is combined with the above assumptions, *the theoretical prediction derived from non-cooperative game theory is unambiguous zero cooperation* (Ostrom, 2014, p. 101). However, a review of the empirical support for this conventional theory of collective action reveals that these predictions have little empirical support in situations characterized as social dilemmas. While free-riding is often observed, many individuals facing collective action problems also cooperate (Poteete et al., 2010).

Rational choice theory works well in the context of provision and production of private goods in a highly competitive environment. But when used as a base for the theory of collective action, it does not do well in predicting outcomes of social dilemmas (Ostrom, 2014). Amartya Sen provides an insightful philosophical and empirical review of the adequacy of rational choice theory in general. Sen points out that human beings acting only according to rational choice theory would act foolishly in social contexts. A favourite passage from his essay is: *"The purely economic man is indeed close to being a social moron. Economic theory has been much* *preoccupied with this rational fool decked in the glory of his one all-purpose preference ordering. To make room for the different concepts related to his behaviour we need a more elaborate structure"* (Sen, 1977, p. 336). Another example is the concept of inequality aversion as described in Fehr et al. (2006), where individuals were empirically shown to have a preference for other's material well-being. Many instances have been found, even outside the realm of social dilemmas, where individuals seek both benefits for themselves and others (Poteete et al., 2010).

According to Ostrom (2014), any policy that aims to improve collective action to avoid social dilemmas must enhance the level of trust by participants that others will comply with the policy. Without this trust, compliance will be avoided by many. This particular argument was the inspiration to include a trust measurement in the following experiment.

The revision of the conventional theory of collective actions ends on an optimistic note. Rather than assuming that cooperation in social dilemmas is impossible due to the characteristics of the setting and the participants involved, it is presumed to occur under certain conditions. Among those conditions are the following:

1. Many of those affected have agreed on the need for changes in behaviour and see themselves as jointly sharing responsibility for future outcomes.
2. The reliability and frequency of information about the phenomena of concern are relatively high.
3. Participants know who else has agreed to change behaviour and that their conformance is being monitored.
4. Communication occurs among at least subsets of participants.

The following sections treat numerous studies that have examined a range of potential drivers and inhibitors of cooperation in social dilemmas. The purpose of these sections is (1) to investigate factors that have been studied in the past and (2) to aid the process of selecting a suitable game design and parametrization for our own experiment. At the same time the following sections summarize insights for policy or the setup of climate negotiations that are derived from the individual studies.

The focus lies on the game design that was also applied in the actual experiment. Commonly referred to as the "collective-risk social dilemma", this game showed the greatest promise to test our research question. First, a series of studies exists that applies a rather consistent game design to investigate numerous different factors that affect coordination and success in social dilemmas. Focusing on one particular game design has the advantage of making the results from our experiment directly comparable to previous work. Additionally, the collective-risk social dilemma captures the important characteristic of repeated and consistent action to mitigate a public bad.

The literature analysed identified the following key factors that affect cooperation and coordination:

2.1.1 Factors shown to increase Cooperation and Success

The following factors have been shown to increase the likelihood that groups in a collective-risk social dilemma will succeed in providing a public good:

1. Increasing loss probabilities (treated in section 2.2.1)
2. Pledge-Communication (treated in section 2.2.3)
3. Intermediate targets (treated in section 2.2.5

2.1.2 Factors shown to decrease Cooperation and Success

Among the key factors that have been reported to make coordination more difficult in the collective-risk social dilemma are the following:

1. Intergenerational Discounting (treated in section 2.2.7)
2. Inequality in wealth (treated in sections 2.2.3 and 2.2.5)
3. Threshold uncertainty (treated in section 2.3
4. Inequality in wealth and risk (treated in section 3.1

Before getting into the different setups of these experiment it is necessary to explain how the collective-risk social dilemma works.

2.2 The Collective-Risk Social Dilemma

Milinski, along with various other researchers, devised a series of games to investigate drivers and inhibitors to successful collective action. One factor that makes this series quite interesting, is that the game designs are quite similar in principle but vary in treatments to investigate different drivers and inhibitors of cooperation between multiple players in settings related to climate change mitigation. They framed the climate mitigation problem in a game design they named the collective-risk social dilemma.

Each game consists of 10 rounds and is played with a group of 6 players. In each round, the players can contribute a maximum of 10% of their total game endowments. In some cases choices are limited to 0, 50 or 100% of the endowment in each round. For example, if players had 40 € initial endowment, they could invest 0, 2, or 4 € in each round. After each round players are informed of the actions the other players in their group have taken. Typically participants are shown a table that identifies each player by a pseudonym and lists their contributions in current and past rounds. Most of the games are framed in the context of a collective climate change mitigation effort.

2.2.1 Game I: Varying loss probabilities

In a first game Milinski et al. (2008) varied the loss probabilities if climate change occurred for different groups. Any money that went into the pool was not paid back to the players but invested in a newspaper ad, as in Milinski et al. (2006). The authors applied three treatments: (*i*) 90% probability of total loss (*ii*) 50% probability of total loss and (*iii*) 10% probability of total loss. Total loss means that players lose their entire savings in the game.

All players received an endowment of 40 €. The threshold any group of 6 had to reach in order to avoid climate change was set at 120 €. As a result, if each player contributed 2 € in each round, the threshold would be met with each player contributing 20 €. The authors define three pure strategies (i.e. all players play the same strategy for the entire game) which lead to the following payoffs in each treatment:

Loss probability (%)	Free rider (€)	Fair sharer (€)	Altruist (€)
90	4	20	0
50	20	20	0
10	36	20	0

Table 2: The expected account values at the end of the game under three pure strategies (all players share the same strategy for the entire game). Free riders contribute 0 € each round, fair sharers 2 €, and altruists 4 €. At a 90% probability of account loss, the optimal strategy is to contribute 2 € each round to the collective. At a 10% probability of loss, the Free Rider strategy is rational, and at 50%, both of these strategies provide identical expected earnings (Milinski et al., 2008, p. 2292)

The expectation is that most groups in the 90% treatment would reach the threshold and avoid climate change, because the consequences of not reaching the threshold are dire, even for free riders. It can be shown that in this treatment, *"each course of the game that leads to exactly reaching the target sum of 120 €, irrespective of who contributes how much as long as each player invests less than 36 €, is a Nash equilibrium: No single player can gain by deviating from his or her strategy"* (Milinski et al., 2008, p. 2292). Conversely, in the 10% treatment, there is no reason to contribute to the pool under traditional rationality assumptions.

Figure 4 shows the average cumulative contributions made in the different treatments. On first sight, it appears that players exposed to a 90% loss probability performed fairly well, increasing their contributions towards the end and approaching the threshold. However, out of 10 groups in this treatment, only 5 reached the threshold, whilst the other 5 narrowly failed.

Milinski et al. (2008) report that groups exposed to a 90% loss probability collected an average 118.2 € ± 1.9 (mean ± SE). The 5 groups that failed in this treatment reached an average of 112.8 € ± 1.2. This means that the average player contributed 18.8 € of the 40 € endowment which leads to an expected payoff of a 2.12 €. The only pure strategy that leads to a lower payoff is pure altruism. From a total welfare perspective, narrowly failing to reach the threshold is the worst possible outcome of the game, with individual players losing close to the maximum amount and a failure to provide a public good.

Additionally, *"failure to achieve the target sum sometimes occurred in an extremely irrational way. Occasionally in the last round, it became clear that all of the contributions to the cumulative sum would have been in vain unless a large proportion of the players made a maximal contribution. Nevertheless, an insufficient number of players made this contribution, and the group just failed to reach the target sum"* (Milinski et al., 2008, p. 2293).

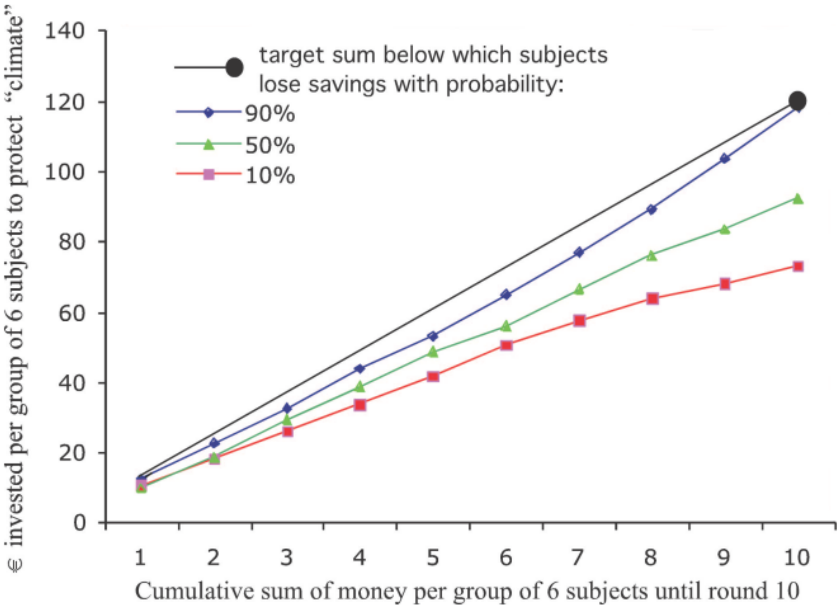


Figure 4: Cumulative sum of money per group and round provided for the climate account. The target sum to be achieved after 10 rounds was 120€; the treatments differed in the probability, i.e., 90%, 50%, and 10%, with which all subjects in a group lost their individual savings when the group did not supply the target sum for the climate account. (Milinski et al., 2008, p. 2293)

The same is true for the behaviour of individuals in the the 50% and 10% loss probability treatments. One of 10 groups in the 50% treatment did reach the threshold, while none of the groups in the 10% treatment succeeded. By itself this observation is not surprising and consistent with the predicted outcomes under traditional assumptions (see table 2). However, all groups raised a considerable amount of money to mitigate climate change but clearly failed to reach the threshold. With a 10% loss probability, the 10 groups raised 73.0€ ± 4.4, even though free-riding was clearly a dominant pure strategy under traditional assumptions.

2.2.2 Game I: Insights regarding Policy and Climate Negotiations

Milinski et al. (2008) acknowledge that the game design with groups of six individuals hardly represents realistic negotiation settings. Based on previous insights from game theory, they expect cooperation to become even more difficult in larger groups. Additionally, they point towards the fact that real climate change will not have uniform effects for all players. Some regions are much more vulnerable, while other regions may even expect to profit (at least in the short run).

They summarize their findings with three conclusions (Milinski et al., 2008):

1. People need to be convinced that there is a very high probability of negative effects from dangerous climate change, if they are to show an effective level of voluntary individual cooperation.
2. Individuals will not always behave according to traditional rationality. Fairness preferences matter and a climate protection program that is perceived as fair (i.e. all players

contribute a "fair share"), are more likely to avoid irrational behaviour, which is at odds with self-interest.

3. Large amounts of players make agreements more difficult. Climate and G8 summits may well be more likely to succeed because of their smaller numbers of participants.

2.2.3 Game II: Inequality and Pledge Systems

Tavoni et al. (2011a) ran the experiment described in section 2.2 with a fixed 50% probability of total endowment loss in case the threshold was not met. Instead of varying probabilities they introduced inherited endowment inequality and a continuous pledge system in different treatments.

If the threshold was met, the raised money was used to buy CO_2 emission certificates for the whole amount. If it was not met, half of the account would be used to buy emission certificates and the rest would be kept by the experimenters Tavoni et al. (2011b).

In order to simulate inheritance of past wealth and debt, the game commenced with three "inactive rounds". In these rounds the decision to contribute 0, 2 or 4€ was predetermined by the treatment and could not be influenced by players. Participants were informed of the other players' contribution choices in every round, including the inactive rounds were the choice was forced on the players. Thus by observing contributions in inactive rounds, subjects knew which players in their group were rich and poor.

Starting in round 4, players were again free to contribute 0, 2 or 4€. In the "Base" treatment all players were forced to contribute 2€ in the first three rounds. In the "Base-Unequal" treatment, 3 players were forced to contribute 4€ and the other 3 players 0€ in the first three rounds. This treatment resulted in rich and poor players, who had different remaining endowments when entering round 4. The endowments in round 4 in all treatments are shown in table 3 below.

Treatment	0€ per round	2€ per round	4€ per round
Symmetrical	(36)	(120)	(204)
w_{all} = 36€	17*	20	6
Asymmetrical	(36)	(120)	(204)
w_{rich} = 40€	20*	26	12
w_{poor} = 28€	14*	20	0

*Table 3: End payoffs (and corresponding climate account values for the group in parenthesis) arising if the three pure strategies were adopted by all players for the seven active rounds. In the symmetrical treatments (Base and Pledge), all group members begin active play having contributed 6€ in the previous three rounds, leaving them with a disposable endowment, w_{all} = 34€; in the asymmetrical treatments (Base-Unequal and Pledge-Unequal), three rich players have no prior contributions and the three poor players have prior contributions of 12€, leaving them with w_{rich} = 40€ and w_{poor} = 28€, respectively. *Expected values based on the 50% probability of account loss when the target sum of 120€ is not reached. (Tavoni et al., 2011a, p. 11826)*

In addition to the inequality treatment, they added an unbinding pledge mechanism to half of the groups. These treatments are called ”Pledge” and ”Pledge-Unequal”. Players could choose to announce what they planned to invest during the game, once after the 3 inactive rounds and once at the end of round 7.

The ”Base” treatment by Tavoni et al. (2011a) differs from to the 50% loss probability treatment in Milinski et al. (2008) in one important aspect. The forced ”fair-sharer” contributions in the three inactive rounds creates a lock-in. Under traditional assumptions, the rational pure strategy (for risk neutral or risk averse individuals) is to continue contributions to avoid foregoing past contributions. Contributions $\geq$ 22 € by all players are Nash equilibria, because they ensure a payoff of 18 € compared to an expected payoff of 17 € in the pure strategy to contribute 0 €(see table 3). The same is true for the ”Pledge” treatment.

In the asymmetrical treatments, the situation is different. Rich players stand to gain most when the threshold is reached. Risk-neutral poor players should be indifferent between pure strategies of contributing 0 € and 2 € per round under traditional assumptions. They will only prefer to avoid disastrous climate change if rich players contribute more and redistribute the burden.

As described in section 2.2.1, only 10% of the groups in the 50% loss probability treatment by Milinski et al. (2008) managed to reach the threshold and avoid potential disaster. Success rates and mean cumulative contributions (±) SE for the treatments by Tavoni et al. (2011a) (who apply a uniform 50% loss probability) are shown in Figure 5 below. Interestingly, success rates were substantially higher than in the 50% loss probability treatment by Milinski et al. (2008). In all treatments, except ”Base-unequal”, success rates were at least as high as in the 90% loss probability treatment in Milinski et al. (2008). The high success rate in the ”Base” treatment suggests that the lock-in effect from the three inactive rounds may be as potent in increasing collaboration as an increase in loss probability from 50% to 90%. At the same time, the highly irrational narrow failure by many groups observed by Milinski et al. (2008) is not reproduced in the ”Base” treatment (i.e. mean cumulative contributions in unsuccessful groups are 70 € compared to the 112.8 € in the 90% loss treatment by Milinski et al. (2008)).

Tavoni et al. (2011a) find that inequality substantially lowers the probability of success (Figure 5) but that this effect is dominated by a dramatic positive impact of the pledge mechanism. In the ”Base-Unequal” treatment, failing groups made the highest investments. From a total welfare perspective, high investments without reaching the threshold are the worst possible outcome of the game. The conclusion is that inequality not only lowers the probability of successful cooperation but also has a negative effect on coordination that are shown in drastic over-investment and narrow failure.

With respect to the pledge mechanism, the strongest effect was observed in the unequal treatment, where the success rate tripled with a pledge mechanism. Additionally the success rate (6 of 10) in the ”Pledge-Unequal” treatment is *is not significantly different from the 7 of 10 achieved by participants of the symmetrical Pledge treatment (P = 0.500)* (Tavoni et al., 2011a, p. 11827). The pledge mechanism appears very effective in mitigating the adverse effects of inequality in a setting, where the economic incentives to coordinate towards reaching the threshold are relatively weak, at least for risk-neutral players (see Table 3).

Tavoni et al. (2011a) identified two key factors which lead to success in treatments with a pledge mechanism: (*i*) the closer contributions were to the nonbinding pledges, the higher the

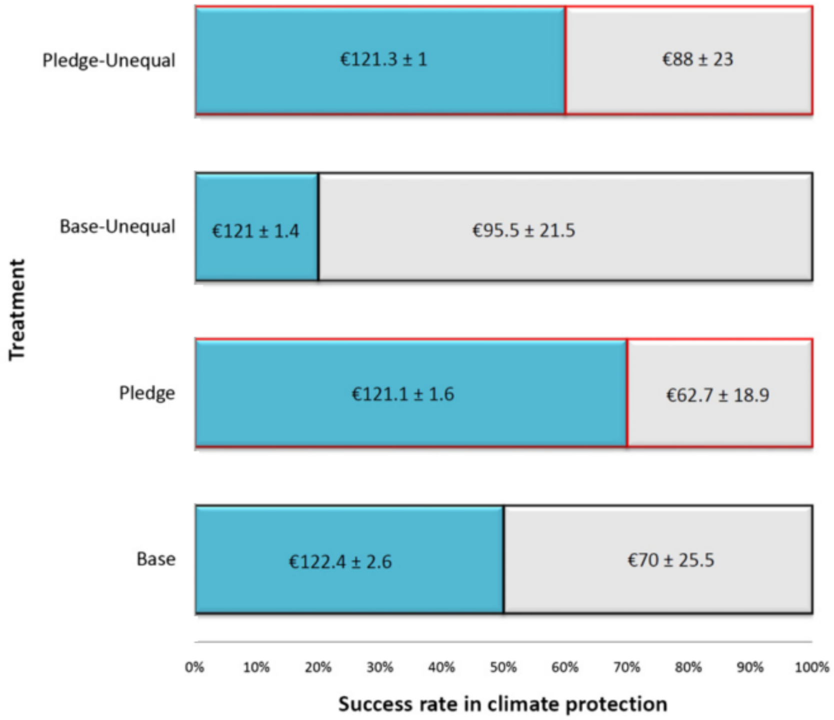


Figure 5: Success rate in avoiding dangerous climate change. The lower two treatments are symmetrical, and the upper two are asymmetrical. The blue sections of the bars indicate the % of successful groups, whereas the gray sections indicate the corresponding failures (with red contours for the treatments with communication). (Insets) For both group classes, the average investments (inclusive of the 36€ collected in the first three rounds) and SDs are shown. Only in Pledge was the outcome somewhat close to the rational prediction of all groups reaching the target (n =10; P = 0.082, binomial test). (Tavoni et al., 2011a, p. 11826)

probability of group success. Deviations between pledges and actual contributions lead to a significant decline in group success. (*ii*) successful groups in unequal treatments, on average, almost completely eliminated inequality over the 7 rounds following the inactive rounds in the beginning. This was true even for the two successful groups in the "Base-unequal" treatment. Furthermore, they found that *early signals by the rich of willingness to redistribute were decisive in the asymmetrical games.* (Tavoni et al., 2011a, p. 11828).

2.2.4 Game II: Insights regarding Policy and Climate Negotiations

Tavoni et al. (2011a) conclude that:

1. Unambiguous evidence exists that the poor are not willing to compensate for inaction of rich players
2. Early leadership by rich nations is critical
3. An appropriate coordination mechanism is instrumental to success
4. Communication and consent are not enough to tackle climate change

At the same time, the Tavoni et al. (2011a) remind the reader that extrapolation of these results to real world climate negotiations should be handled cautiously. Quite obviously, coordinating climate change mitigation is far more complex than coordinating on a known threshold. Apart from the much higher complexity of measures necessary, climate change is also associated with uncertainties in terms of expected damages. Additionally, asymmetries exist not only in wealth and carbon debt but also in adaptive capacity and risk exposure (Tavoni et al., 2011a, p. 11'828).

Nevertheless, the effect of the unbinding pledge mechanism is quite remarkable. Success rates increased both in symmetrical and asymmetrical treatments if unbinding pledges were available. At the same time, they reduced average cumulative contributions in failing groups by 10.4 % in symmetrical and 7.8% in asymmetrical treatments (see Figure 5). In a setting where the willingness by players to collaborate was weak, pledges thus lead to lower investment in mitigation, possibly because players revealed their unwillingness to mitigate through their pledges. In the game setting described above, this is positive from a total welfare perspective (i.e. cumulative wasted investments are reduced). But in a setting where effects are gradual instead of black and white, less mitigation means higher expected damage. Thus pledges could also have a dampening effect on mitigation if uniform willingness to reach a common target is not given.

In the asymmetrical treatments, fairness preferences likely mattered. Tavoni et al. (2011a) found evidence for strategic fairness in the participants, with poor players significantly more in favour of redistribution. It is also questionable if the inactive rounds lead to a credible rich and poor divide. At the beginning of the game all players had the same 40 € endowment. In the inactive rounds "poor" players were forced to contribute a high amount to the public account, while "rich" players could not contribute. The authors justify their approach with the goal of creating the perception that the inequality is a result of past action. But rather than just entering the game with a lower endowment, poor players had "lost" a substantial part of their endowment to a mitigation effort that would benefit all.

In reality, the divide comes from different economic growth over time. This left some countries richer as a result of industrialization, which in turn caused the problem at hand. Poor countries did not mitigate the problem, they simply had a smaller role in creating it. With respect to framing effects, this distinction may be important.

2.2.5 Game III: Inequality and Intermediate Targets

In this game Milinski et al. (2011) alter the endowments of the players. They assign operating funds, which can be invested in climate change mitigation and an additional endowment that cannot be invested. They divide the sample into poor and rich groups. Rich groups manage 40 € in operating funds and 60 € in endowment. Poor players only receive 50% of these funds (i.e. 20 € in operating funds and 30 € in endowment). This differentiation aimed to replicate assets with different grades of liquidity (e.g. as hard cash versus real estate). A climate mitigation threshold was again set at 120 € for groups of six individuals. In each round individuals could contribute 0, 2 or 4 €. Groups either consisted of 6 rich individuals, 6 poor individuals or an even mix of 3 rich and 3 poor individuals.

If the necessary funds were not collected, players would keep their remaining operating funds but lose their endowment with a fixed 90% probability. If the threshold was met, individuals kept both remaining operating funds and their endowment. Additionally, half the groups were exposed to an additional climate event: if they failed to raise an intermediate target amount of 60 € by the end of round 5, a climate event would occur during rounds 6 through 10, with a 20% probability in each round. The consequence was a loss of 10% of both operating funds and endowment, every time the event occurred. Such near-term adverse events aimed to mimic damages from heat waves or floods that would cause considerable, but not catastrophic damage.

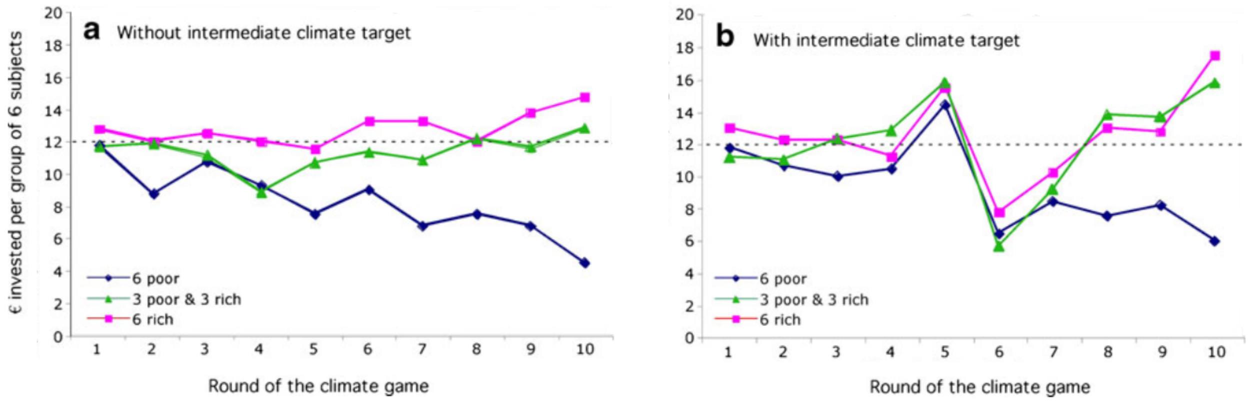


Figure 6: Money (€) invested per group against the climate target. Mean investment per group of 6 poor subjects (blue), 6 rich subjects (red) and mixed groups consisting of 3 poor and 3 rich subjects (green) in each of 10 rounds of the climate game; hatched line depicts the investment per round that is needed to reach the final climate target; ***a*** *without intermediate climate target announced;* ***b*** *with intermediate climate target announced. Figure adapted from Milinski et al. (2011, p. 810).*

The rich groups, 60% of mixed groups but not a single poor group reached the 120 € target. The authors found that *although the rich invested more than the poor both when among themselves and combined with poor subjects, they did not invest differently in mixed groups than when among themselves nor did the poor* (Milinski et al., 2011, p. 809)

The 100% success rate of rich players in this game (compared to the 50% success rate in the 90% loss probability treatment in Milinski et al. (2008)) may be explained by the higher absolute stakes in this version of the game (100 € versus 40 €). At the same time, in the 2008 version of the game, players faced complete loss, whereas in the present study only the endowment 60 € was at stake in groups without an intermediate target. The expected payoff for pure strategy fair-share (2 €) contributions was 80 € in this game and 20 € in the 2008 version. For pure strategy free-ride 0 € contributions, expected payoffs were 46 € in the present study and 4 € in the former. At least in relative terms, stakes were higher for players in the 2008 version of the game. Yet they failed far more frequently to avoid the loss.

Interestingly, the introduction of intermediate targets had a substantial effect on the investment behaviour of all groups. Figure 6 shows how the dynamics changed in groups with an intermediate target. All rich groups, most mixed groups and 60% of poor groups, met the intermediate target. In all groups contributions then dropped dramatically in round 6. In rich and mixed groups they then recovered, while in many poor groups investments stayed low. Success rates in reaching the final target remained almost unchanged for rich and mixed groups. But now

3 of 9 poor groups met the final target ((Milinski et al., 2011)). At the same time, mixed groups that failed to reach the final target now invested more and thus came closer to avoiding "disastrous climate change".

2.2.6 Game III: Insights regarding Policy and Climate Negotiations

Milinski et al. (2011) find that with intermediate targets, rich individuals in mixed groups were more likely to compensate the lower investment capability of poor players. This effect was not observed in treatments without intermediate targets. The mixed treatment comes closest to reality, where both liquid operating funds and illiquid endowments are more valuable in developed nations. In this treatment, Milinski et al. (2011) record a significant increase in cooperation when an intermediate target is introduced (the same is true for poor groups). While the success rate in mixed rates is not substantially higher with or without an intermediate target, the contributions of failing groups increased significantly with rich players more willing to compensate inequality. They call for a stronger focus on near-term damages and corresponding climate targets to increase chances of successful mitigation.

The authors conclude that *if one accepts round 5 as corresponding to year 2020 and round 10 to year 2050, we demonstrate that in principle long-term rational behaviour requiring cooperation is much harder to obtain than short-term rational cooperative behaviour* (Milinski et al., 2011, p. 812). It may not be convincing, that the rounds induced a perceived 2020 and 2050 situation. Nevertheless, they provide evidence for an apparently increased salience of the damage threat, if it occurs at a time within the game rather than in the end.

However, it seems unlikely that this observation is related to time preference. The long-term versus short-term problem addressed by the authors would likely be dominated by discounting. It seems unlikely that this concept played a decisive role in the present study, because subjects received payment immediately after the study. One wonders then how the effect of intermediate targets could be explained instead.

Conventional rationality cannot explain the significant difference observed in contributions after the introduction of intermediate threat and target. With a 20% occurrence probability in each of the 5 final rounds, the ex-ante probability of not experiencing a "near-term" climate event was 32.8% (0.8^5). With a 20% occurrence probability in each round, we would expect the "near-term" event to occur one time over all 5 rounds. The expected loss of operating funds and endowment was only 10% (one-time expected occurrence × 10% loss) of the sum of endowment and operating funds. Compared to the 90% loss probability of the entire endowment (if the 120 € threshold was not met) the stakes of not reaching the intermediate target were lower. Avoiding the intermediate damages is only rational if the disastrous final damages are also avoided. If the intermediate target is met but not the final target, the group invests 60 € to avoid a collective expected damage of 27.6 € for 6 rich players or 13.8 € for 6 poor players that play pure strategies of contributing 0 €. This calculation considers the 90% loss probability of the endowment if the final threshold is not met.

Of the treatments applied, the mixed treatment is certainly the most realistic when considering real world climate negotiations. At the same time, the uniform damage probability for all nations could be challenged. Although illiquid assets in poorer nations are less valuable in monetary terms, the loss probability is higher in many cases due to regional differences in vulnerability and exposure. This is not reflected in any of the treatments.

Finally, Burton-Chellew et al. (2013) point out that different total endowments but equal thresholds in different treatments makes comparison of treatments difficult. Because of this asymmetry it is much easier for rich groups than for poor groups to reach the threshold. The same is true for mixed groups, they have 25% less total endowment than rich-only groups.

2.2.7 Game IV: Intra- and Intergenerational Discounting

Jacquet et al. (2013) investigated the potential effects of intra- and intergenerational discounting on collective risk situations. The game design is very similar to Milinski et al. (2011). They make a distinction between operating funds 40 € and endowments 45 € that subjects received in the game. The operating funds that went into the climate account to prevent dangerous climate change, were used to finance an advertisement on climate protection in a daily German newspaper as in Milinski et al. (2006). Again, failure to meet a threshold of 120 € in the climate account lead to a 90% chance of losing the endowment, but not the remaining operating funds. In paying out the endowments the authors applied two different time horizons: Subjects were paid out their endowments in cash either one day or seven weeks after the experiment. In a third treatment, the endowments were invested in planting oak trees to mimic an intergenerational effect of successful mitigation. In all cases, the remaining operating funds not invested in the climate account were paid out directly after the experiment. Figure 7 illustrates the effects of the three treatments on investment behaviour.

Even under conventional assumptions, it is not surprising that the authors observed the lowest success rates in the treatment where endowments were invested in oak trees. None of the groups in this treatment reached the threshold. Time preference in favour of the present has long been known to influence individuals decisions. However, as shown in Panel B of Figure 7, investment in the first round are almost uniform in all treatments, followed by a steady decline in contributions that is strongest in the treatment with intergenerational payouts (oak tree planting).

Here, 7 out of 10 groups that received their endowments one day after the experiment succeeded in providing cumulative funds of 120 € to their climate accounts. In Milinski et al. (2011) the success rate in the treatment ”rich players and no intermediate target” was 100%. The present treatments are almost equivalent, with the rich subjects in Milinski et al. (2011) defending an endowment of 60 € instead of 45 € in the present study. Both groups' operating funds were 40 €. The concept of hyperbolic discounting as described by Kirby (1997) may have played a role here. It is conceivable that a payout as little as a day later from the present, could have had an impact on decisions. Since students were given invoices they had to redeem in person, it could be argued that the additional time necessary to pick up the cash may have also discouraged the students. It would be interesting to see, if the success rate is higher if students were automatically transferred the money to avoid additional transaction costs.

2.2.8 Game IV: Insights regarding Policy and Climate Negotiations

Jacquet et al. (2013) argue that the results show the power of intergenerational discounting to undermine cooperation. When we extrapolate these results to the reality of climate change mitigation, the picture is bleak. If short-term gains can be generated only from defecting,

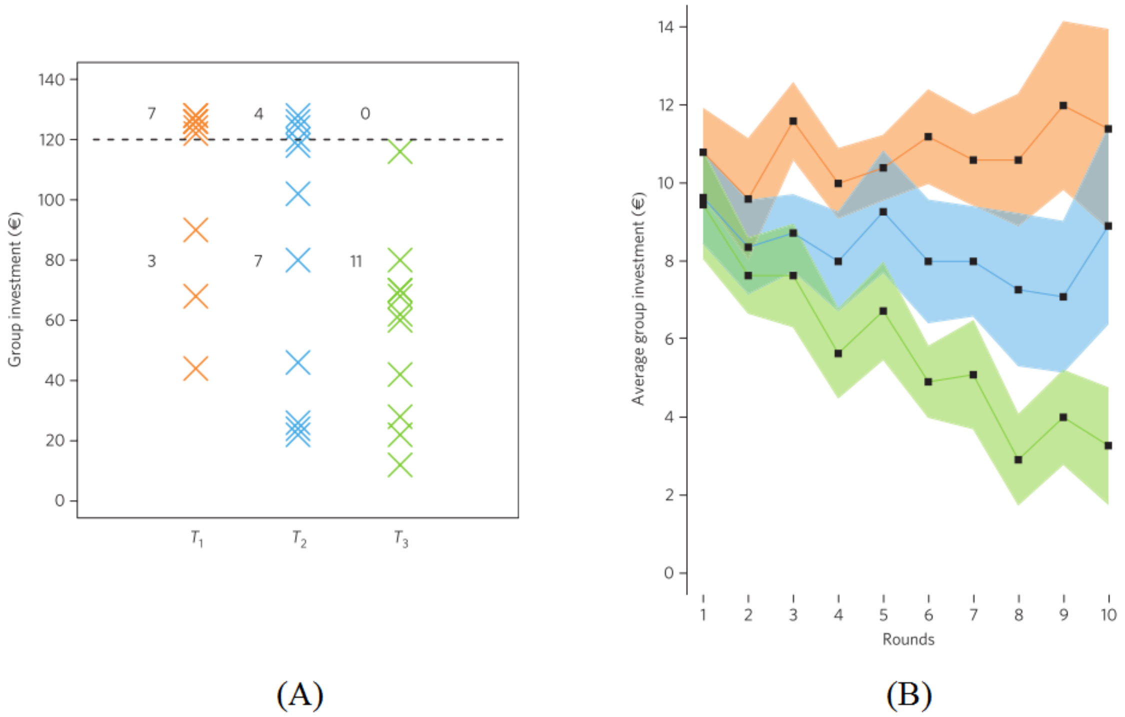


Figure 7: ***Panel A:*** *Group investments by treatment. T_1 (red) received their endowment the following day, T_2 (blue) received their endowment 7 weeks later, and T_3 (green) invested their endowment into planting trees. The dashed line represents the 120€ threshold that groups had to achieve to receive the endowment. In T_1 (red), 7 groups succeeded in reaching the target, 4 groups succeeded in reaching the target in T_2 (blue), and 0 groups succeeded in T_3, (green). The number of groups reaching the target differs significantly among the treatments.* ***Panel B:*** *Average group investment and standard error of the mean (coloured range) by treatment over the 10 rounds. Group investments across the three treatments were significantly different showing a trend of greater cooperation the closer in time the benefits were received. Investments were highest in T_1 (red), when the endowment was received the next day, compared with T_2 (blue), where participants received their endowment in 7 weeks, and much higher than those in T_3 (green) (Jacquet et al., 2013, pp. 1026-1027).*

cooperation will remain very difficult. The authors propose the introduction of short-term incentives to cooperate, which include punishment, reward and reputation.

Intergenerational discounting is likely one of the most important challenges to successful climate change mitigation. Punishment and reward could be potent instruments, but are difficult to implement for lack of a international legal body that could enforce action and punish defection. At the same time such measures do not change time preferences of individuals in favour of the future. Therefore defectors and free-riders can still realize short-term gains if they manage to avoid getting caught. Thus, the effectiveness of such measures may be limited even if fully implemented.

But one can question if it is true that short-term gains can only be generated from defecting. It appears that the most elegant way of solving this challenge would be to remove the existing trade-off between short-term gains and long-term damages. Recent developments in renewable energy technologies show promise in this respect. When these alternatives become economically superior to carbon emitting substitutes over relatively short investment horizons, there may be a silver lining to the bleak outlook suggested by the problem of intergenerational discounting.

2.3 Threshold and Damage Uncertainty

In addition to the context of the collective-risk social dilemma,Barrett and Dannenberg (2012, 2014a) investigated the effects uncertainties have on the success of collaboration in a one-shot public good game with a threshold. While the researchers aimed to get insights on processes relevant to climate change mitigation, the game was not framed in the context of climate change in contrast to other studies presented in this thesis. Participants were simply told that a public bad had to be avoided. This makes the results applicable to situations with a similar public bad social dilemmas. The authors show that increasing uncertainty about the (emission) threshold value that should not be exceeded, makes collaboration far less likely and changes the nature of their game design.

The design by Barrett and Dannenberg (2012, 2014a) contains N symmetrical players, which each own two operating funds (A and B) and an inactive endowment C that cannot be invested but can be lost to a public bad. Players can reduce hypothetical emissions by making use of two technologies. The contents of operating fund A can be invested in technology A. Operating funds B can be used to finance technology B. Per unit costs of reduction are constant in both technologies but different, with $c^A < c^B$. Reducing one unit of emissions with either technology creates a constant marginal benefit b for each player. The social marginal benefit to the whole group is thus also constant at bN. The emission units that can be reduced with both technologies are limited by the contents of operating funds A and B to q^A_{max} and q^B_{max} for each player i.

The authors chose the parameters such that

$$c^B > bN > c^A > b \tag{1}$$

is always satisfied. This means that without additional damages from a public bad, and considering purely the investment returns, the game is a prisoner's dilemma. The social benefit of reduction exceeds the private costs of any individual to contribute, but the private costs again exceed the private marginal benefit of reduction. As in equation 10, this leads to a Nash equilibrium of zero contribution from all players. Technology B would never be used in the absence of an additional threat as its marginal costs exceed the marginal private benefit and even the marginal social benefit.

In the certain threshold treatment, a minimum reduction above the threshold $\bar{Q}$ is needed to avoid a loss. The necessary reductions are feasible if the players collaborate, but require investments in technology B. This property is captured by defining

$$N(q^A_{max} + q^B_{max}) > \bar{Q} > Nq^A_{max} \tag{2}$$

Reductions below the applicable thresholds lead to a uniform damage X for all players. The total emission reduction made by all players is denoted by Q, with

$$Q = \sum_{i=1}^{N} q_i^A + q_i^B \tag{3}$$

If $Q < \bar{Q}$, the damage probability is 1. When $Q > \bar{Q}$ it is 0.

In the treatment with threshold uncertainty, $\bar{Q}$ is a random variable uniformly distributed between $\bar{Q}_{min}$ and $\bar{Q}_{max}$. As before, sufficient reductions are always feasible. The relation in equation 2, under threshold uncertainty becomes:

$$N(q_{max}^A + q_{max}^B) \geq \bar{Q}_{max} > \bar{Q}_{min} \geq N q_{max}^A \tag{4}$$

The damage probabilities are then defined as follows:

$$p_d = \begin{cases} 1 & for \quad Q < \bar{Q}_{min} \\ \frac{Q - \bar{Q}_{min}}{\bar{Q}_{max} - \bar{Q}_{min}} & for \quad Q \in (\bar{Q}_{min}, \bar{Q}_{max}) \\ 0 & for \quad Q > \bar{Q}_{max} \end{cases} \tag{5}$$

From the equations above it follows, that avoiding the catastrophic event requires an average investment into the costly technology B of $\bar{Q}/N$ under threshold certainty, and $\bar{Q}_{max}/N$ in the threshold uncertainty treatment. In both cases the maximum feasible amount q_{max}^A must be invested in technology A. The problem is a coordination game when the threshold is certain, as long as the private net cost of abating the damage is smaller than the damage itself for an individual player (see Figure 10, Panel A). This can be expressed as (Barrett and Dannenberg, 2012, p. 17373)

$$X \geq \underbrace{(c^B - b)\,\frac{\bar{Q}}{N} - (c^B - c^A)\, q_{max}^A}_{net\,private\,cost\,of\,abatement} \tag{6}$$

Here, $(c^B - b)\bar{Q}/N$ denotes the net cost of abatement if only technology B was used. $(c^B - c^A)\, q_{max}^A$ are the savings achieved by making full use of the cheaper technology A before turning to technology B. This condition is satisfied in all treatments with certain thresholds.

Thus, by adding a threshold and associated damage X as in equation 6, the underlying prisoner's dilemma (in the absence of X) can be transformed into a coordination game. This type of game is defined by the existence of two Nash equilibria, where all players choose the same strategy. No player has an incentive to change their decision by anticipating other players' decisions. If other players signal willingness to collaborate and one deems this information trustworthy, the most profitable strategy is always to collaborate as well. To defect as a single player in a group of collaborating players reduces the payoff of the defecting player compared to collaboration. Free riding does not pay in this case. In addition to abating the damage, zero contribution by all players is also a Nash equilibrium, albeit socially suboptimal (Barrett and Dannenberg, 2012). If all players signal unwillingness to collaborate, no single player can realize a private gain through abatement efforts.

2.3.1 Impact and Threshold Uncertainty

Barrett and Dannenberg (2012) investigate the effects of uncertainty about the damages that occur (impact uncertainty) and uncertainty about the threshold (threshold uncertainty). They find that impact uncertainty does not have a significant effect on participants' behaviour.

The game was played with N=10 players over one round. Parameter settings were as follows: For all players, operating funds were set at 11 € and split between accounts A (1 €) and B (10. Contributions to the public good were made by purchasing poker chips. Chips purchased from account A cost $c^A = 0.10$ € each and those from account B cost $c^B = 1$ €. Thus, $q^A_{max} = q^B_{max} =$ 10 chips and $N(q^A_{max} + q^B_{max}) = 200$ chips. Each chip invested in the public good generated a return of b = 0.05 € for each player in the game. The total welfare return for all players was thus bN = 0.5 €. Furthermore, each subject received an additional endowment of 20 € in fund C, that could not be used to purchase chips.

The players knew that they had to collectively contribute an amount $\bar{Q}$ of chips to avoid a damage X. In the certainty treatment, the authors defined $X = 15$ € and $\bar{Q} = 150$ chips. Under impact uncertainty, X was uniformly distributed between 10 and 20 €. Threshold uncertainty meant that $\bar{Q}$ was uniformly distributed between 100 ($\bar{Q} - min$) and 200 ($\bar{Q} - max$) chips. In a last treatment, impact and threshold uncertainty were combined.

In a first stage, all players had the opportunity to make a proposal on how many chips the group should invest collectively and a pledge on how much the player intended to contribute him- or herself. Proposals and pledges were not binding as in Tavoni et al. (2011a). At the end of the game, a computerized spinning wheel was used to determine the damage or threshold in the uncertainty treatments.

Barrett and Dannenberg (2012) observed that impact uncertainty had little effect on the success rate of groups but that threshold uncertainty was a literal game changer. Figure 8 shows how collaboration breaks down completely with threshold uncertainty introduced.

In contrast to impact uncertainty, threshold uncertainty transforms the game from a coordination game to a prisoner's dilemma. This transformation is explained in detail in section 2.3.2. In the treatments with certain thresholds, mean proposals were very close to the necessary 150 chips and mean pledges were close to the individual "fair share" of 15. Interestingly, actual contributions were often higher than pledges in treatments with certain thresholds (see Figure 9).

Barrett and Dannenberg (2012) find that under threshold uncertainty, communication tools were used strategically. In order to avoid the catastrophe with certainty, the maximum amount of 200 chips have to collected. Mean proposals and pledges are well below the necessary levels to achieve this. In this case, the actual contributions are often drastically lower than what players pledged (see Figure 9). In a post-game questionnaire, participants stated that they tried to motivate other players to contribute more by proposing lower amounts. *They thought that a proposal below 200 was more credible and so was more likely to stimulate contributions by others* (Barrett and Dannenberg, 2012, p. 17374).

However, it is important to keep in mind that participants played a one-shot game. If numerous subsequent rounds were added, such strategic behaviour would likely not prevail. Participant's pledges would soon lose their credibility if they refused to contribute in early rounds. In international climate negotiations, future negotiations and consequences from not keeping pledges

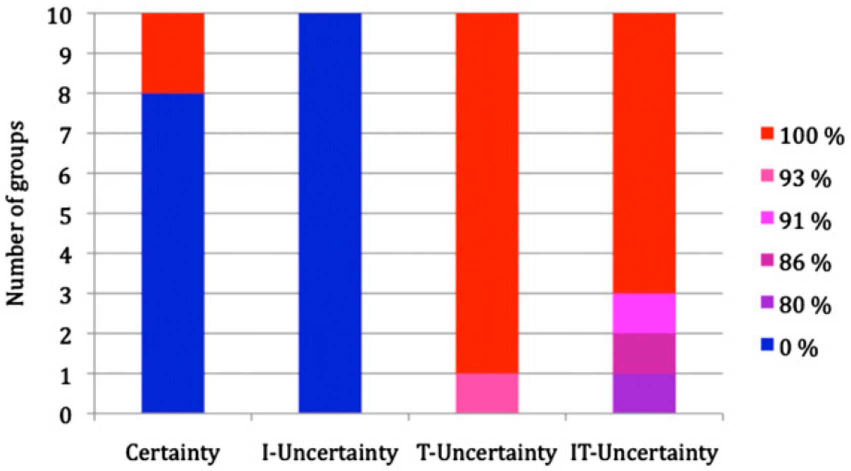


Figure 8: Probability of catastrophe by treatment. Catastrophe was avoided 8 of 10 times in the Certainty treatment and 10 of 10 times under Impact Uncertainty (I-Uncertainty). In contrast, the probability of catastrophe was reduced below 100% (to 93%) by only 1 of 10 groups under Threshold Uncertainty (T-Uncertainty) and by only 3 of 10 groups (to 91, 86, and 80%, respectively) under Impact-and-Threshold Uncertainty (IT-Uncertainty). In the four cases where the probability of catastrophe was reduced below 100%, the spinning wheel determined that the threshold was crossed every time. (Barrett and Dannenberg, 2012, pp. 17373).

are a reality. For example, many of the INDC's submitted to COP21 contain commitments that are conditional on other countries' actions.

2.3.2 Sensitivity to Threshold Uncertainty

In a second study, Barrett and Dannenberg (2014a) investigate the sensitivity of players to threshold uncertainty. In treatments with uncertain thresholds the situation becomes more complex. As hinted above, threshold uncertainty in the present game design can change the nature of the game from a coordination game to a prisoner's dilemma. The damage probabilities are defined as in equation 5. In this section I include the most essential pieces to understanding the game transformation from a prisoner's dilemma to a coordination game. Full formal proof for these relationships is provided in the supporting information to the study in Barrett and Dannenberg (2014b). The authors show that under uncertainty regarding the threshold's value, all players want to collectively abate $\bar{Q}_{max}$ if

$$XN \geq (c^B - bN)(\bar{Q}_{max} - Nq^A_{max}) \tag{7}$$

They then show, that if every other player abates $\bar{Q}_{max}/N$, each player will want to abate $\bar{Q}_{max}/N$. Thus, avoiding the damage becomes a Nash equilibrium, if

$$X \geq (c^B - b)(\bar{Q}_{max} - \bar{Q}_{min}). \tag{8}$$

If the condition in 8 does not hold, the game becomes a prisoner's dilemma. Rearranging

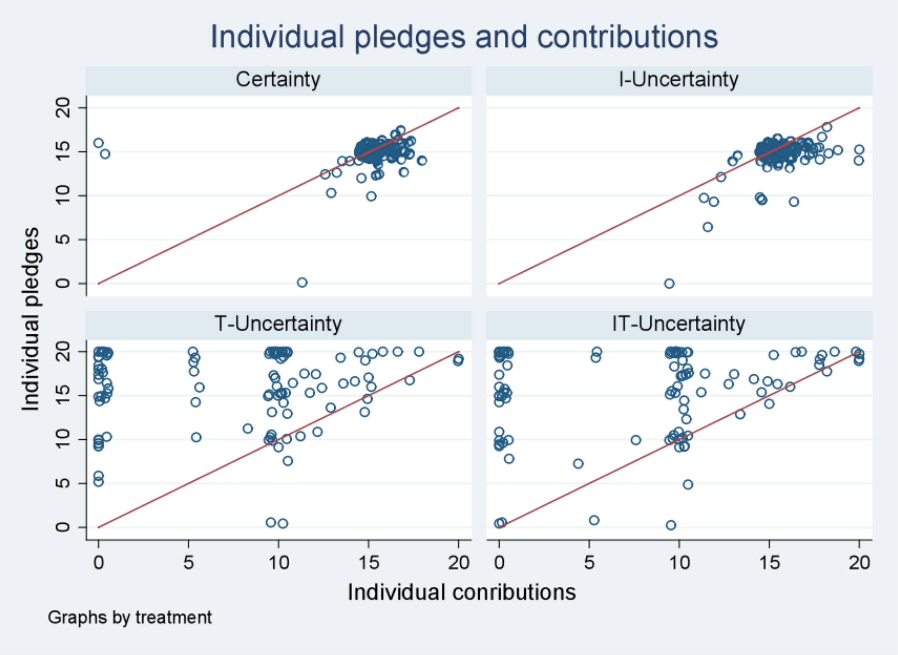


Figure 9: Pledges and actual contributions by treatment. In the Certainty and Impact Uncertainty treatments, pledges and contributions are tightly bunched, with contributions usually exceeding pledges. In the Threshold and Impact-and-Threshold Uncertainty treatments, values vary widely, with contributions usually falling far short of pledges. A small noise (3%) has been inserted to make all data points visible. (Barrett and Dannenberg, 2012, pp. 17374).

equation 8, the authors define

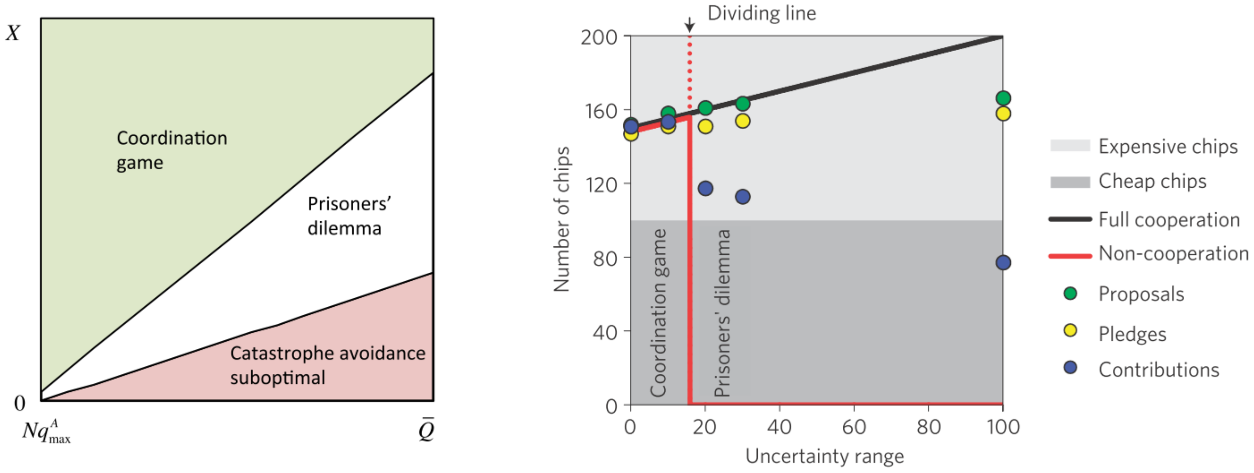
$$\phi = \frac{X}{c^B - b}. \tag{9}$$

If $\bar{Q}_{max} - \bar{Q}_{min} \leq \phi$, players are in a coordination game. Assuming that all players behave equally, each is expected to reduce emissions by $q_i^A = q_{max}^A$ and $q_i^B = \bar{Q}_{max}/N - q_{max}^A$. For $\bar{Q}_{max} - \bar{Q}_{min} \geq \phi$, the game becomes a prisoner's dilemma. In those scenarios players would choose $q_i^A = q_i^B = 0$ (Barrett and Dannenberg, 2014a).

Figure 10 illustrates the differences between the games with threshold certainty and uncertainty.

The black line, named full cooperation in Panel B of Figure 10, denotes the level $\bar{Q}_{max}$ necessary to avoid catastrophe with certainty. It's maximum lies at 200 chips, where $\bar{Q} \in [100, 200]$. Note that a certain threshold of $\bar{Q} = 200$ chips would result in a coordination game. Thus even though, the same amount of chips is necessary to avoid catastrophe with certainty in both cases, the case where $\bar{Q} \in [100, 200]$ is a prisoner's dilemma, because there is a probability greater than zero of avoiding the damage if aggregate contributions are below 200 chips.

The group behaviour in the different treatments was quite consistent with the authors expectations. With X fixed at 15 €, $c^B = 1$ and $b = 0.05$, ϕ from equation 9 was 15.8. This critical value is indicated by the red line in Panel B of figure 10. Groups in treatments, where $\bar{Q}_{max} - \bar{Q}_{min} \geq \phi$, were thus predicted to fail to abate catastrophe. Figure 11 shows the results for the different treatments.



A: Threshold certain

B: Threshold uncertain

*Figure 10: **A**: Certainty model. Red area shows values for X and $\bar{Q}$ for which players are collectively better off not avoiding catastrophe; here, $X < (c^B - b)\bar{Q}/N - q^A_{max}$. In the green area, catastrophe avoidance is a coordination game; here, $X \geq (c^B - b)\bar{Q}/N - (c^B - c^A)q^A_{max}$. In the white area, avoiding catastrophe is a prisoner's dilemma; here, if all other participants play $\bar{Q}/N$, each player prefers to abate 0. With certainty, a prisoner's dilemma arises only if $b > 0$. (Barrett and Dannenberg, 2012, pp. 17373). **B**: Treatment means versus predicted values. Mean contributions are consistent with the predicted values to the left of the dividing line (ϕ). To the right of the dividing line, mean contributions lie between the full cooperative and the predicted (non-cooperative) values. Mean proposals and mean pledges match the full cooperative values to the left of the dividing line; to the right, a wedge opens up between these values as the uncertainty range widens (Barrett and Dannenberg, 2014a, p. 37).*

2.3.3 Insights for Policy and Climate Negotiations

Barrett and Dannenberg (2012) conclude that their research is consistent with behaviour observed in past climate negotiations. Their experimental results, under threshold uncertainty, suggest that countries would strategically contribute less than their fair share of the proposed total amount. The actual contributions are then again expected to be even lower than what was pledged. Taking the Copenhagen accord as a reference, countries indeed did not make pledges that are sufficient to limit global warming to 2 ° C. The Kyoto Protocol offers numerous examples of countries not keeping the goals that they ratified. According to the authors, the climate change game is a prisoner's dilemma not just because of the need for collective action, but because the relevant thresholds that avoid ”dangerous climate change” are uncertain.

Based on these results, Barrett and Dannenberg (2012, p. 17375) argue that if a threshold for ”abrupt and catastrophic” climate change could be identified with certainty, free riding would be disciplined, countries would pledge their fair share to collective action and would act upon their pledges. Barrett and Dannenberg (2014a) add evidence that a sufficiently small range of threshold uncertainty may also do the trick.

However, the authors acknowledge that exact thresholds that would completely avoid damages, are not a reality in the climate system. As a result they point to alternative mechanisms, such as trade restrictions in the Montreal Protocol. Such measures would effectively transform the prisoner's dilemma at hand back into a coordination game. Another proposition is the use of

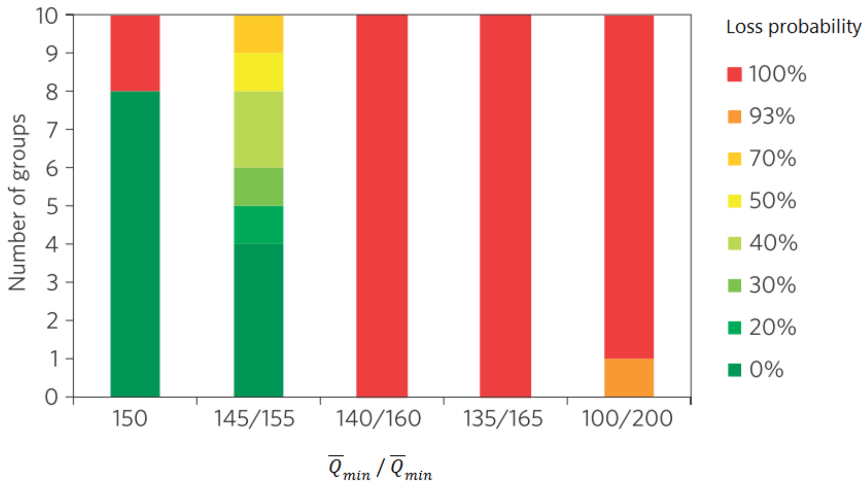


Figure 11: Probability of catastrophe by treatment. In 150, catastrophe is avoided eight out of ten times. In 145/155, catastrophe is avoided four out of ten times with probability 100% and in the other six cases with probability between 30 and 80%. In 140/160 and 135/165, catastrophe is never avoided. In 100/200, catastrophe occurs nine out of ten times with probability 100% and once with probability 93% (Barrett and Dannenberg, 2014a, p. 37).

clean technology standards that exhibit network externalities. This means that the returns to all participating countries increase with the number of countries participating.

The effects of uncertainty about the necessary abatement demonstrated by Barrett and Dannenberg (2012, 2014a) are impressive. They have identified a single consistent ”make or break” factor for success in this game design. It is especially notable that uncertainty about sufficient abatement levels clearly dominates uncertainty about damages to be expected (see Figure 8).

At the same time, the game design by Barrett and Dannenberg (2012, 2014a) is a one shot game. Players make one single pledge and one single contribution that determines the outcome. This is a characteristic that may take away some of the applicability of the results from this study to the real situation of climate negotiations. The authors emphasize the importance of using additional measures to transform the prisoner's dilemma caused by threshold uncertainty into a coordination game. Theoretically, it should not matter if a prisoner's dilemma is repeated over a finite amount of rounds. Rational agents would solve the game through backwards induction, considering first the last round of the game and then going back to the first round. If defection is a dominant strategy in the last round of a sequence of identical game rounds, it is a dominant strategy in the second to last round as well, and thus in all prior rounds as well. Rational agents would anticipate this and defect in the first round already.

However, past experiments have shown, that even finite repetition of games framed as prisoner's dilemmas will often lead to enhanced cooperation in reality (Andreoni and Miller, 1993; Cooper et al., 1996). It has been shown that cooperation can be substantial up to the last couple of rounds in finite repetition of games. Among the explanatory factors are altruism and reputation building in situations with incomplete information about player's agendas and type. The latter factor seems to be particularly important in the context of climate change mitigation, as success is only possible through sustained collaboration over an extensive time frame. This problem cannot be solved through a one-time commitment and corresponding action. It is this characteristic that is not reflected in the game design by Barrett and Dannenberg (2012, 2014a). It could be argued that targets in past negotiations were generally lower than the necessary amounts, because countries worry about credibility issues. Committing to a goal that is very difficult to attain would lead to reputation damage and international pressure that affects the country's position in future negotiations.

Simple repetition of the game as in Andreoni and Miller (1993) or Cooper et al. (1996) is also not applicable to the climate change context. In abating climate change, outcomes in prior rounds have an effect on the attainability of success in following rounds. The game design by Milinski et al. (2008) captures this important characteristic.

Perhaps for this reason, Dannenberg et al. (2014) also applied threshold uncertainty to a collective-risk social dilemma game. The parametrization was equivalent to Milinski et al. (2008) with all players managing an endowment of 40 € and the option to invest 0, 2, or 4 € in each of 10 rounds. The game setup also included a pledge mechanism but in contrast to Tavoni et al. (2011a) the players could only make pledges as to the group investment they viewed as adequate but not to their personal intention of contribution. In the multi-round collective-risk social dilemma, the above derivation of Nash equilibria and theoretical game transformation is no longer feasible (see section 4).

Treatments differed in terms of the threshold that had to be avoided, where in one case the threshold was certain at 120 € (certainty treatment). In the other two cases the threshold was uncertain. In the ”risk” treatment participants were informed that the threshold was equally likely to take any 20 € incremental between 0 and 240. In this case the expected value of the threshold was still 120 € In the ”ambiguity” treatment the probability distribution of the threshold value was unknown to the participants. In the ambiguity treatment the expected threshold value was also unknown. Similar to the one-shot game above, contributions were substantially lower in treatments with increasing uncertainty about the threshold (Dannenberg et al., 2014). The probability of providing the public good decreased dramatically with threshold uncertainty. This suggests that the findings for the one-shot game also apply to a setting with repeated contribution rounds such as in the collective-risk social dilemma.

2.4 Strategic Fairness

The notion that nations apply fairness norms strategically in climate negotiations is likely not surprising for anyone familiar with the history of the UNFCCC. A study by Brick and Visser (2014) shows that the strength of the agreements increases if players' incentives to act strategically are removed.

Brick and Visser (2014, p. 94) advocate the *use of burden-sharing mechanisms which are not subject to subjective interpretations of what is fair – for example, deriving the abatement burden from the anticipated costs and benefits of climate policy across countries.* The INDC approach introduced at COP21 may be a first step in this direction. With INDCs, the focus is shifted away from a fair division of a global burden towards an action plan on a country level. This regional focus may put stronger emphasis on local costs and benefits from climate action which leaves strategic fairness considerations less important. For a detailed summary of their work please refer to section 8.2 in Appendix I.

3 A Focus on Vulnerability

This section provides insights from several studies that investigate the aspect of vulnerability in the context of climate negotiations. The role vulnerability plays in climate policy design and negotiations is unclear. A common argument is that the most vulnerable countries are not major contributors to climate change. Steves and Teytelboym (2013) argue that those countries focus more on adaptation rather than mitigation and that this is why high vulnerability is not strongly correlated with strong climate policy. Instead, they find a rather strong correlation between policy strength and country income, with a marked leadership of select Northern European countries.

The countries that are most vulnerable to climate change (e.g. AOSIS member states) are not the ones that are causing the problem. Therefore their importance in terms of collective mitigation is limited. At the same time, there are marked differences in vulnerability between major players in the climate change game. Buys et al. Buys et al. (2009) have constructed a dataset from various sources to classify different countries according to two dimensions of vulnerability. On the one hand, they determine countries' "impact vulnerability", the extent to which a country is exposed to expected negative effects of climate change. In this category, weather events and sea level rise are considered. On the other hand, they investigate "source vulnerability", which indicates the extent to which countries have access to fossil fuels and renewable energy sources and the potential income and employment shocks that may result from stricter regulation of carbon emissions. They use these results to anticipate the degree to which countries are likely to pursue an agreement in international climate negotiations. Figure 12 shows the classification of countries in the categories "source vulnerability" and "impact vulnerability".

Major emitters such as China, India or Indonesia are exposed to higher impact vulnerability than their western counterparts but the source vulnerability is high or at least medium for most major players in the negotiations. Any country exposed to high levels of both impact and source vulnerability faces a strong tradeoff. While stronger regulations would limit expected damages in the future, they also imply negative economic effects. Buys et al. (2009) draw four main conclusions from their analysis: *First, with six dimensions of vulnerability affecting country stakes, even neighboring states in the same region can have very different orientations toward a global protocol. Policy analysis and dialogue should therefore be tailored to specific conditions in each country. Second, despite significant country-level variation in each region, our analysis does indicate sufficient regional clustering to warrant some attention to regional strategies. Third, even with good information and programs tailored to country conditions, our results suggest that many countries will resist a global protocol unless they are compensated for disadvantages associated with source vulnerability. Many countries have persistently unfavourable stakes in emissions reduction, no matter how we index their relative vulnerability. As a group, furthermore, they account for almost half of all CO_2 emissions from the World Bank's partner countries.*

The "impact vulnerability" measure in Buys et al. (2009) is quite broad. By considering mainly the risks from sea level rise and extreme weather events, climate scientist may argue that they are missing some important points. One of those points could be food security. Crop losses leading to higher food prices would have stronger effects in poorer countries (Ivanic and Martin, 2008; Headey, 2013).

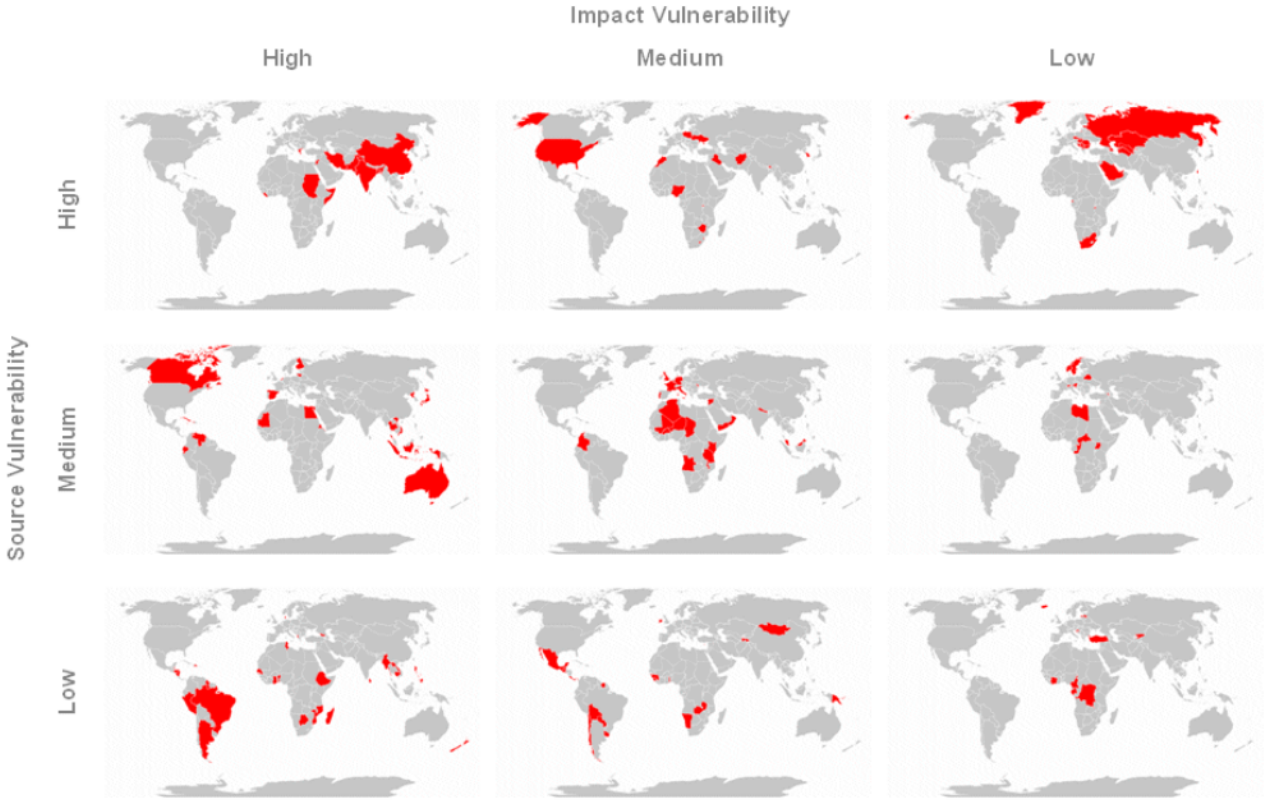


Figure 12: Classification of countries according to "source vulnerability" and "impact vulnerability". Countries marked red in the map in the top left corner show both high impact and source vulnerability. (Buys et al., 2009, p. 50).

Other initiatives have attempted to create a more complete picture of vulnerability to the effects of climate change. One of those initiatives is the ND-GAIN Index Based on the latest available data, there is a weak positive correlation between countries GHG emissions and their vulnerability to climate change. Figure 13 shows the vulnerability assessments by the ND-GAIN Index for the year 2014.

Thus even though climate mitigation may be at odds with economic incentives as shown by Buys et al. (2009), vulnerability may still be an important factor in negotiations regarding climate change. Given the dynamics of worldwide economic development, I expect the role of vulnerability to become stronger in the future. This is because the effects of climate change will become more extreme in the decades to come, which will make vulnerability more salient.

The experimental literature suggests that when different vulnerability levels come into play, co-operation becomes difficult. Risk asymmetry leads to increased free-riding (Blanco et al., 2014) by the parties less affected, mitigation of fairness criteria applied by players (Gampfer, 2014). It has also been shown to lead to a breakdown of collaboration in the collective risk social dilemma game (Burton-Chellew et al., 2013). In the following subsections , I introduce two additional studies with an experimental game design that focused specifically on vulnerability.

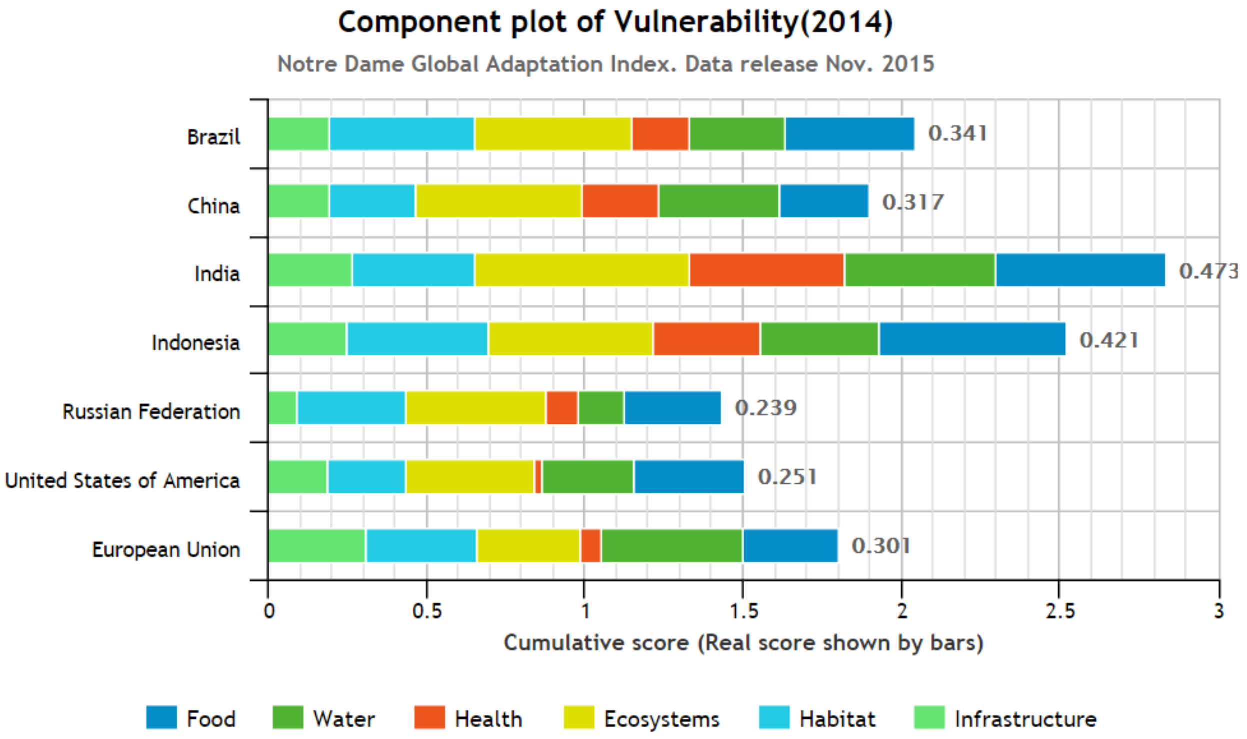


Figure 13: Vulnerability measures a country's exposure, sensitivity and ability to adapt to the negative impact of climate change. ND-GAIN measures the overall vulnerability by considering vulnerability in six life-supporting sectors: food, water, health, ecosystem service, human habitat and infrastructure. Overall scores shown by bar. A negative score indicates that the score was not computed because it lacked enough data; bars still show available data. (Regan et al., 2015)

3.1 Collective-Risk Social Dilemma with Inequality in Wealth and Risk

This study by Burton-Chellew et al. (2013) again applied a collective.risk social dilemma setting. Groups of 6 players had to spend half of their collective endowment to mitigate climate change or face total endowment loss with a certain probability. Endowments were measured in monetary units (MU's). Total group endowments were 240 MU and avoiding climate change required a collective investment of 120 MU. Four treatments were applied that are illustrated in Panel B of Figure 14 (Burton-Chellew et al., 2013)

- Egalitarian: All group members had the same endowment (40 MU) and were exposed to the same risk of total endowment loss from climate change (A: P = 0.8 or B: P = 0.7).
- Unequal-wealth: Two players received 80 MU and four players 20 MU each. Loss probabilities were uniform as in the egalitarian treatment
- Rich-suffer: Same unequal wealth distribution but rich were at greater risk of loss (A: P = 0.95 or B: P = 0.90) than poor (A: P = 0.65 and B: P = 0.50)
- Poor-suffer: Same unequal wealth distribution but poor were at greater risk of loss (A: P = 0.95 or B: P = 0.90) than rich (A: P = 0.65 and B: P = 0.50)

The authors applied two probability blocks (A and B) which is why there were always two possible loss probabilities in each treatment. In Block A, the "mean" level of risk over both groups is P = 0.8 (in unequal case 0.95 and 0.65) if the threshold to mitigate climate change is not met. In Block B the mean level of risk is P = 0.7 (in unequal case 0.9 and 0.5).

The threshold was reached in 1 out of 8 groups in the poor-suffer treatment (12.5% success rate) and in 18 out of 24 groups (75% success rate) that were exposed to the other treatments. The authors found no significant differences between the latter three treatments and also not between the two probability blocks A and B (Burton-Chellew et al., 2013, p. 823). In contrast to Tavoni et al. (2011a) and Milinski et al. (2011) they also did not find an isolated effect of inequality on the success rate of groups. They did identify a significant reduction in success rates only in the poor-suffer treatment compared to the other treatments.

However, it is interesting to note that in block A (with $P_{mean} = 0.8$) 0 of 4 poor-suffer groups succeeded and 8 out of 12 of the other groups did. In block B (with $P_{mean} = 0.7$), 1 of 4 poor-suffer groups succeeded and 10 out of 12 of the other groups. With only 4 poor-suffer groups in each block, meaningful statistical analysis is difficult. Higher success rates in treatments with lower loss probability stand in contradiction to the findings of Milinski et al. (2008). One wonders if a larger sample size would have resulted in a significant difference between treatments.

Figure 14 shows the success rates of groups in different treatments. Over both probability blocks, each treatment was applied to 8 groups of 6 players.

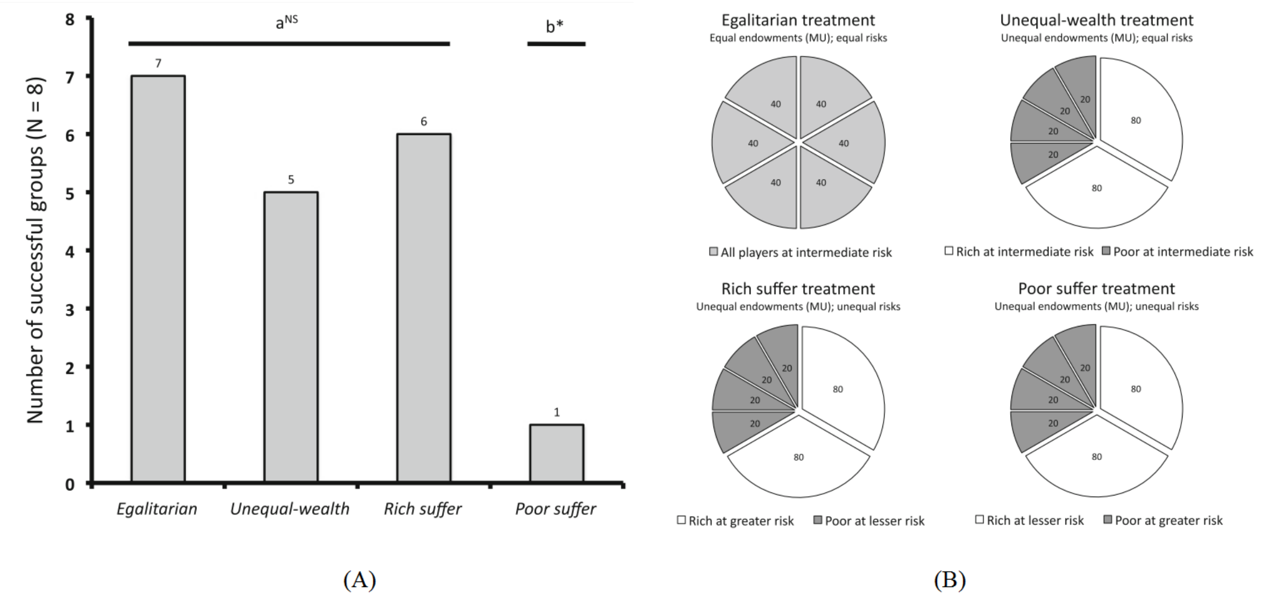


Figure 14: ***Panel A:*** *Cooperation collapses when the poor suffer. The proportion of groups which were successful in preventing severe climate change was significantly lower in the poor suffer treatment (b*, P=0.002), compared with the other three, which were not significantly different (a^{NS}, P=0.626).* ***Panel B:*** *Experimental design: resources and risk. (Burton-Chellew et al., 2013, pp. 818-824).*

Note that P_{mean} is not the the mean risk weighted by group assets, because in the rich suffer treatment 160 MUs are lost with P = 0.95 and 80 MU with P = 0.65. Expected MU distributed in pure strategy of non contribution: $0.05x160+0.35x80 = 36MU$. In the poor suffer treatment this is: $0.05x80 + 0.35x160 = 60MU$. Thus the expected payoff on a group level when climate change is not prevented is almost twice as high in the poor suffer compared to the rich suffer treatment. From a total welfare perspective avoiding climate change is more attractive in the rich suffer than in the poor suffer treatment. Adjusting the endowment levels to 60 MU for rich and 30 MU for poor would lead to the same expected group payout in these treatments and eliminate this divergence by splitting group endowment equally among rich and poor.

The authors find that inequality in both wealth and risk in favour of rich players, leads to a collapse of coordination. In this setting they observe a significant reduction in contributions from rich players to reach the threshold. In the unequal-wealth treatment, rich players contributed 56% of their funds on average, while in the poor suffer treatment, rich players reduced this share to 36% (Burton-Chellew et al., 2013). Such a reduction is rational under traditional assumptions as the expected payoff without successful mitigation is higher for rich players when they are less and the poor more vulnerable. Nevertheless, many groups in the poor suffer treatment came very close to reaching the threshold. Of 9 groups, one succeeded and 8 failed. But 4 of the failing groups collected over 100 MUs. As described in Milinski et al. (2008) they created the worst possible outcome by narrowly failing to gather the required funds.

Burton-Chellew et al. (2013) provide yet another example of this highly irrational behaviour. Out of 32 groups, 4 gathered less than 100 MUs and thus clearly failed (three of them were in poor suffer treatments). 9 groups across all treatments failed but gathered over 100 MUs throughout the 10 rounds. Almost 30% of groups showed a high level of collaboration but still failed in this most irrational way of just missing the threshold.

The authors also found a correlation of contribution rates in the games with participant's beliefs about real climate change. More sceptical individuals contributed less to the common pool. Under traditional assumptions, this should not have mattered as one's conviction about climate change had no impact on the individual payoffs. Since the public good fund was also not used to counteract climate change, there was no rational reason for participants to behave in this way.

3.2 Vulnerability and Fairness Preferences

While researching fairness preferences of individuals in the climate change context, Gampfer (2014) also applied a difference in vulnerability. Apart from a base treatment he introduced a simple inequality treatment and a treatment where both inequality and different levels of vulnerability are applied. Instead of the collective action social dilemma he used a one-shot ultimatum game that is played repeatedly.

Gampfer (2014) found that the introduction of vulnerability mitigated fairness considerations in the game. In the simple inequality treatment, richer subjects offered to pay more to eliminate the income asymmetries. When different levels of vulnerability were introduced, this effect became much weaker. At the same time, the frequency of successful agreements declined from the base to the unequal treatment, and again from the unequal to the treatment with combined inequality in wealth and vulnerability. This result is consistent with the previously covered literature.

In the base treatment, players on average propose that their counterparts take over about 53% of the total burden. The mean accepted burden share by respondents was just shy of 52% of the burden. This implies that the proposer on average gets away with offering to take over a little less than half of the burden.

The other four treatments are asymmetrical: In the LowCap treatment, a poor (LowCap) player makes a proposal to a rich (HiCap) player. The poor individuals offered to carry an average of 26% of the total burden (leaving 74% for the rich player). The mean share their rich counterparts accept to take over was a little lower, at 71% on average. In the HiCap treatment, the situation is reversed with the rich player making the poor an offer. On average, the rich offered to take over roughly 63% of the burden when vulnerability is equal for both players. In this case the poor on average accepted any offer higher than 67%.

When vulnerability was introduced, both the rich player's generosity and the poor players boldness in making low offers declined. The burden shares were much closer to the baseline. Gampfer (2014) concludes that higher vulnerability of poor players mitigates the fairness norms applied in the game.

One additional finding is particularly interesting. Across most treatments, the mean offered shares by proposers and the mean accepted shares by respondents were not significantly different. The only exception was the treatment, where poor, vulnerable players were responding to offers from rich, less vulnerable players (Gampfer, 2014, p. 71). When in the respondent's role, the vulnerable poor players on average rejected offers that were higher than what they themselves would have proposed in the proposer role. Their average proposals were to take over 42% of the burden. But in the responder role they refused, on average, to accept a burden of more than 36%. In this case the mean accepted offer by poor, vulnerable players, was significantly higher than the mean offer by rich players with low vulnerability.

This particular finding inspired the hypothesis that: *needs-based fairness is more important to disadvantaged actors if they have less influence on the burden-sharing outcome. Responders' agency opportunities are lower than proposers', since they are second movers who can only accept or reject a given cost distribution. Disadvantaged proposers might be prepared to make a higher payment offer to ensure acceptance and avoid catastrophe, since this offer is at least their own decision. But disadvantaged responders simply confronted with a low, "unfair" offer might tend to reject such unfair treatment, regardless of the potential material consequences. (Gampfer, 2014, p. 74)*

4 Methods

Our experiment took place at ETH's Descision Science Laboratory (DeSciL) from February 8^{th} to February 10^{th} in four separate sessions lasting one hour each. The zTree Software Package developed by Fischbacher (2007) was used to program the design and perform the experiment in the lab. The participants were part of the English speaking subject pool of the University of Zurich and ETH Zurich. The total subject pool of these two Universities contains roughly 10'000 individuals. The English speaking pool contains numerous international students from India, China and other Asian countries. This is why the English speaking pool was chosen as the international participants create a more realistic setting for simulating climate negotiations. Since the participants of the study were guaranteed anonymity (as is policy in experiments run at DeSciL) their countries of origins are unknown and it was thus not possible to test for such factors.

120 students participated in the experiment, 60 in each of the two treatments applied. With 6 students per group, a total of 20 Groups were observed. Treatments differed in one aspect: In the so called pledge treatment, participants had three opportunities to state their intended investments and communicate those to the other group members. Apart from this, the treatments were identical. Both treatments had 60 subjects, equivalent to 10 groups.

4.1 Experimental Design and Procedure

The aim of this experiment was to test, whether a pledge-communication tool as described in Tavoni et al. (2011a) would mitigate the collapse in collaboration in a setting with combined inequality in wealth and risk, described by Burton-Chellew et al. (2013).

For this purpose two versions of the "collective risk social dilemma" were programmed, keeping close to the design by Burton-Chellew et al. (2013). The two versions were (1) Base-Vulnerability and (2) Pledge-Vulnerability. Both versions had a non-egalitarian distribution of endowments and unequal risk levels. Two "rich" players received 80 CHF and the remaining four "poor" players received 40 CHF each. The poor were at a greater relative risk ($P = 0.95$) than the rich ($P = 0.65$) of losing their endowments if they failed to raise sufficient funds. The treatments differed in only one aspect. In Pledge-Vulnerability the players were given the chance to make declarations of intent regarding their own planned investments. The first chance was at the beginning of the experiment, before investments started. Subsequently, there were two more opportunities to revise the declarations (after investment rounds 3 and 6).

As in Burton-Chellew et al. (2013), every game lasted for 10 rounds and participants could only invest a maximum of 10% of their endowment per round. After each round a table informed the participants of the initial endowments, the investments of the other five group members and the cumulative group spending so far. In the Pledge-Vulnerability version, the table also contained the latest planned catastrophe account that contained the declarations of intent of all players in the group. Players were also reminded of the target amount, the history of their own spending and how much of their original endowment remained. In modelling the Pledge-Mechanism the design by Tavoni et al. (2011a) served as a basis.

A particular interest was trust building in the collective risk social dilemma. Thus participants were asked in each round to state how much they trusted the other group members to contribute enough to avoid the catastrophe. Figure 15 shows the different stages of the game that were programmed. The game loop describes the different game stages. A single game loop describes one round of the game. The loop was repeated ten times.

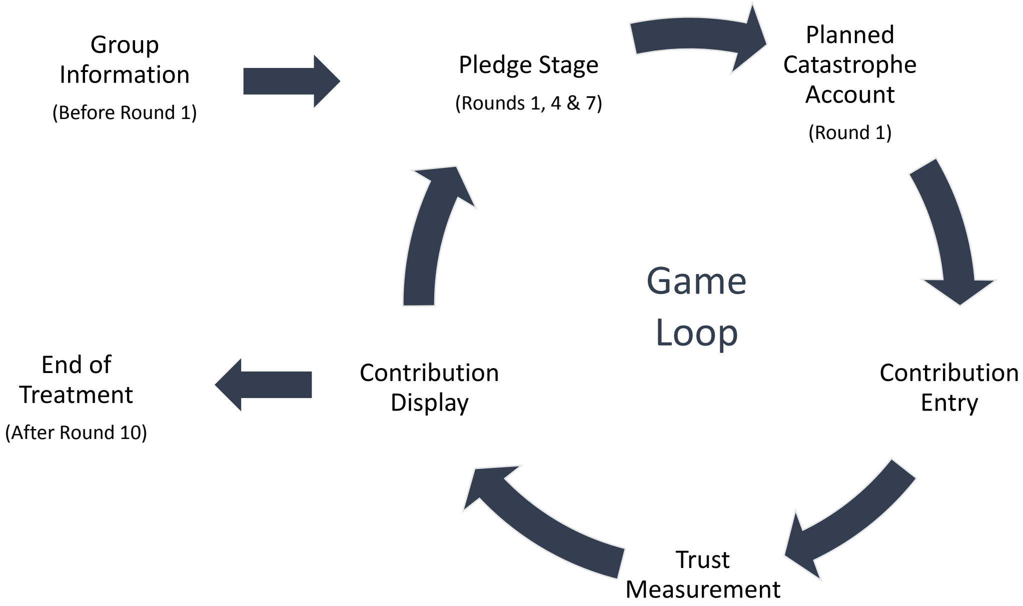


Figure 15: Different stages modelled using zTree. Before the experiment started, subjects entered an information stage where they learned their player type and the pseudonyms and player types of the other players in their group. In the base treatment, the pledge stage and the planned catastrophe account stage were simply deleted. Otherwise the design was equivalent in both treatments.

The collective-risk social dilemma was chosen for a number of reasons. First of all, many of the studies applying this concept investigated factors that were of particular interest. These factors include pledge-communication, income inequality, inequality in risk, differing overall risk levels, discounting and intermediate target definition. Additionally, the necessity of repeated and consistent contributions over several rounds come closest to the reality of climate change mitigation. While the design may not be a good model for repeated investments over the course of decades with changing political and economic concerns, it does model an important characteristic of coordinated international efforts of climate change mitigation: the ability to adjust one's actions based on the observation of the actions of one's peers. According to Ostrom (2014), this is one of the central factors that applies to the climate change social dilemma. This aspect is not modelled in most of the other designs that were considered.

As noted by Tavoni et al. (2011a), the multiplicity of equilibria in this type of game makes classification very difficult. This was an n-person public good game with a stochastic threshold and played over 10 rounds, of which all 10 allowed freedom of choice over several different investment decisions.

In principle any trajectory leading to a group investment of 160 CHF, irrespective of individual contributions (as long as they are below the rational limit for the player type), is a Nash equilibrium. Unilateral deviations are non-profitable in such cases. At the same time the Nash equilibria leading prevention of the catastrophe are unstable: While group investments of 160 CHF are payoff-dominant compared to the free-riding equilibrium, a single downward deviation by one player may leave defecting as the best response for all remaining players (Dannenberg et al., 2014).

Unstable in this context means that the equilibrium that leads to catastrophe avoidance is sensitive to the deviation of a single-player. Given a downward deviation by just one player, the best-response may change for the remaining players in the group. In the free-ride equilibrium this is not the case. When a majority of players apply the pure-strategy of contributing nothing and playing the odds, their best response is not affected if a single player starts to make investments. Keeping investments at zero will still maximize the expected payoff in this case.

Some Nash-Equilibria can easily be identified, assuming that players are rational agents and follow pure strategies over the course of the entire game. Table 4 shows the expected payoffs by player type for a number of selected pure strategies. The following pure strategies are Nash equilibria: (1) Free Ride (stable), (2) Equal Proportion (unstable) and (3) Weak Transfer (unstable). In these strategies, no risk neutral player has an incentive to deviate from their strategy as this would lead to a lower expected payoff for the defecting player.

In Strong Transfer, risk neutral rich players can gain from defecting because this leads to a higher expected payoff than cooperating in Strong Transfer. This is because rich players' contributions in this pure strategy is beyond their rational limit for contributions (see table 4). If all players chose a pure strategy according to the rational limit or pure altruism, there would be a substantial over-investment and thus incentives to reduce one's own contribution.

4.2 Deviations from previous studies

The game was designed mainly on the basis of Burton-Chellew et al. (2013)'s example. The same loss probabilities and a very similar game interface and wording was applied. Burton-Chellew et al. (2013) provided their zTree files, which were used as a reference design. The literature on collective risk social dilemma games contains numerous previous studies which were used to calibrate the design. In many of these studies, participants could only choose between three investments options (0%, 50% or 100% of endowment) in each round, which correspond to pure strategies if applied consistently. In this experiment, participants could invest any whole number of CHF in each round. Burton-Chellew et al. (2013) also used this approach and it was decided that this would lead to a more realistic setting, since the choice between fixed amounts is neither a reality in government nor in private climate finance.

Keeping with the design by Burton-Chellew et al. (2013), the money invested into preserving the public good was not donated to a charity. In Milinski et al. (2008); Tavoni et al. (2011a) the money invested by participants into preventing a climate crisis was invested into a form of climate protection. Burton-Chellew et al. (2013) reported that the participants' beliefs about the reality of climate change influenced their decisions. Participants that thought climate change was a hoax had invested less. Such effects were not of interest and an effort was made to prevent such factors from influencing participants' behaviour. That is why, instead of a climate change disaster, participants were told that the game was simulating a setting where a catastrophe had to be avoided, without providing any further detail.

Player Type	Loss Risk	Free Ride	Equal Proportion	Weak Transfer [a]	Strong Transfer [b]	Altruism	Rational Limit [c]
-	in %	in CHF	in CHF	in CHF	in CHF	in CHF	in CHF
rich	65	0 (28)	40 (40)	50 (30)	60 (20)	80 (0)	52
poor	95	0 (2)	20 (20)	15 (25)	10 (30)	40 (0)	38
GROUP	80	0 (64)	160 (160)	160 (160)	160 (160)	320 (0)	256

Table 4: Pure strategy contributions (and corresponding expected payouts in parenthesis) arising if the above pure strategies were adopted by all players over all 10 rounds. Note that Weak Transfer and Strong Transfer contain two separate pure strategies for rich and poor players.

[a] *This "pure" strategy implies that rich players consistently invest 5 CHF and poor players 1.5 CHF in each round. To achieve this, poor players would have to invest 1 CHF in 5 rounds and 2 CHF in 5 rounds.*

[b] *This "pure" strategy implies that rich players would always invest 6 CHF and poor players 1 CHF in each round.*

[c] *This is the limit any rational, self-interested and risk-neutral player should be willing to spend given their resources and personal risk, as otherwise they will be better off fully defecting.*

Finally, the difference in endowments (80 CHF and 40 CHF) was less extreme than in Burton-Chellew et al. (2013) (80 MU and 20 MU), because the group's total capital at stake had to be equal to an egalitarian treatment where all participants had the same endowments and risk. In Burton-Chellew et al. (2013) rich players owned two thirds of the group's endowment. When loss probabilities are shifted, with the average loss probability remaining the same, this leads to lower stakes on a group level when the rich are at a lower risk of losing.

4.3 Hypotheses

The aim of the experiment was to test whether a pledge-communication mechanism would replicate the increased group success as reported by Tavoni et al. (2011a) in a setting with inequality in wealth and risk. Based on the results by Burton-Chellew et al. (2013) the expectation was for groups to fail very frequently as Burton-Chellew et al. (2013) had reported a failure rate of close to 90% in a design that was very similar to our base-treatment. The expectation was that the communication mechanism in the form of simple pledges would increase trust within the groups and that this would lead to better coordination and thus higher success rates.

The main hypotheses can be summarized as follows:

H_1 : A strong difference in vulnerability leads to mitigation of fairness. This results in low success rates in treatments where communication between subjects is not allowed. Rich, less vulnerable players will contribute less than 50% of their assets towards the provision of the public good. This hypothesis is based on the findings of Burton-Chellew et al. (2013); Gampfer (2014), where it had been shown that rich players cooperated less frequently when combined inequality in wealth and risk was modelled.

H_2 : Pledge-communication increases the groups' ability to coordinate their investments. This hypothesis is based on the work by Tavoni et al. (2011a) that showed an increase in success rates when pledge-communication was introduced in a setting with unequal endowments.

H_3 : Pledge-communication provides additional information and builds trust among players that other players will invest enough to provide the public good. Therefore, groups that have access to such a mechanism will be more successful in providing the public good. It was assumed that trust-building is the mechanism that leads to higher success rates.

H_4 : When poor, vulnerable players are empowered, they will contribute more to the provision of a public good. The option to voice their opinion provides such an empowerment, that may lead to such behaviour. Poor, vulnerable players in the pledge-treatment will therefore contribute more than those in the base treatment. This hypothesis is based on the results reported by Gampfer (2014), who showed that poor, highly vulnerable players in the proposer role of an ultimatum game were willing to contribute more than when they were in the responder role.

5 Results

The treatments did not have significant effects on success rates or contribution levels. Out of the 20 groups, 17 succeeded in preventing the catastrophe and 3 failed. In the base-vulnerability treatment 9 of 10 groups succeeded and in the pledge-vulnerability treatment the success rate was 8 out of 10. Table 5.1.1 summarizes the investment behaviour in both treatments:

	Pledge-Treatment	Base-Treatment
Success Rate	80 %	90 %
Average Rich Player Contribution	39.45 CHF	43.45 CHF
Average Poor Player Contribution	18.18 CHF	19.13 CHF

Table 5: Observed success rates by treatment and average total contributions (over all 10 rounds of the experiment) by player type.

These observations are not in line with our initial hypotheses as far more participants were expected to fail. Burton-Chellew et al. (2013) reported a failure rate of almost 90% in a treatment that is very similar to our base treatment in terms of risk and endowment parametrization (see section 3.1).

Apart from treatment effects on success rates, a focus was placed on trust and fairness perception in the experiment. Trust levels were recorded in each round for all players. Trust was expected to increase in treatments that had a pledge-communication mechanism, compared to treatments that had no communication.

The results section is structured as follows: A first subsection compares the combined sample results to report where differences between treatments were found. The second subsection considers the groups exposed to the pledge treatments separately as some variables (e.g. pledged contributions and pledge credibility) were only recorded for subjects in the pledge treatment.

5.1 Results for the combined sample

5.1.1 Contributions

As mentioned above, no significant differences in the chosen total contributions between treatments were found. However, the distribution of individual contributions is more spread out in the treatment with communication. Figure 16 contains a number of histograms that illustrate these differences.

When subjects had the chance to communicate, the distribution of chosen contribution levels is generally broader and less concentrated around peaks. The histograms showing data for both player types, separated by treatment, contain two ranges: The contributions in the higher range (30 - 60 CHF) typically come from rich players, while those below 25 CHF are typically contributions by poor players.

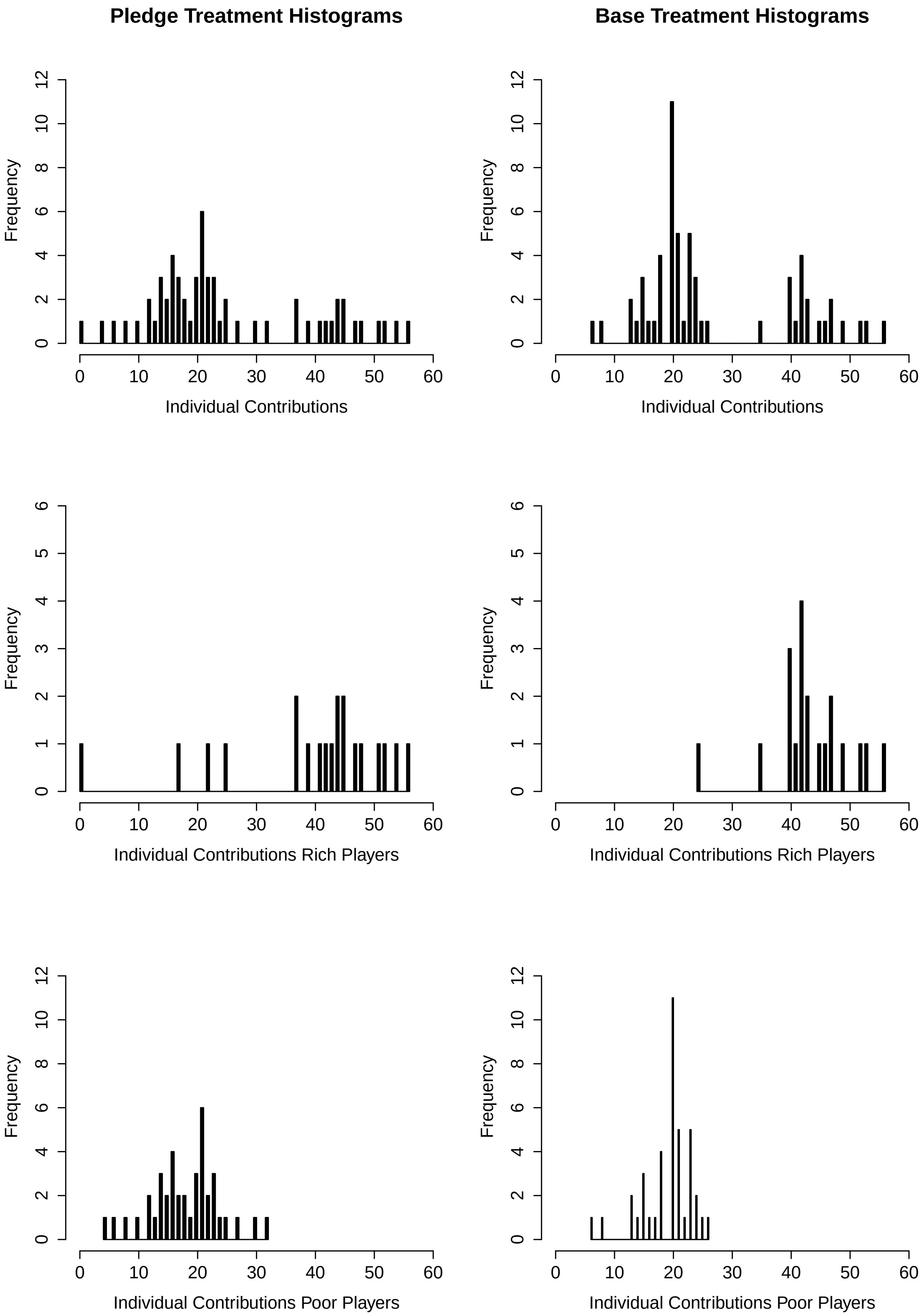


Figure 16: Histograms of individual contributions seperated by treatments. The left column contains the data from groups in the pledge-treatment, the right column that of groups in the base-treatment.

However, in the pledge-treatment this range separation is much less clear than in the base treatment. The histograms that separate rich and poor players show that in the pledge-treatment there is a stronger overlap of chosen total contribution levels between rich and poor players. In the base-treatment this separation is more clear. The distribution of contribution choices in the pledge-treatment were generally broader.

If communication and better coordination leads to more homogeneous contributions, the observations would imply that pledge-communication did not lead to better coordination. Quite the opposite would have been the case.

In fact, considering the average contributions in individual rounds, an interesting picture emerges. Poor players were, on average, much more consistent in their contribution choices. The choices also do not differ notably between treatments. When rich players are isolated, the situation is quite different. In the first five rounds of the game, rich players in the pledge-treatment contributed substantially less than those in the base treatment. When individual contributions by these players are summed up over the first five rounds, the sum of contributions by rich players in the pledge-treatment is significantly lower (Mann-Whitney Test, W = 125.5, p-value = 0.04438). Figure 17 shows how average contributions developed throughout the experiment.

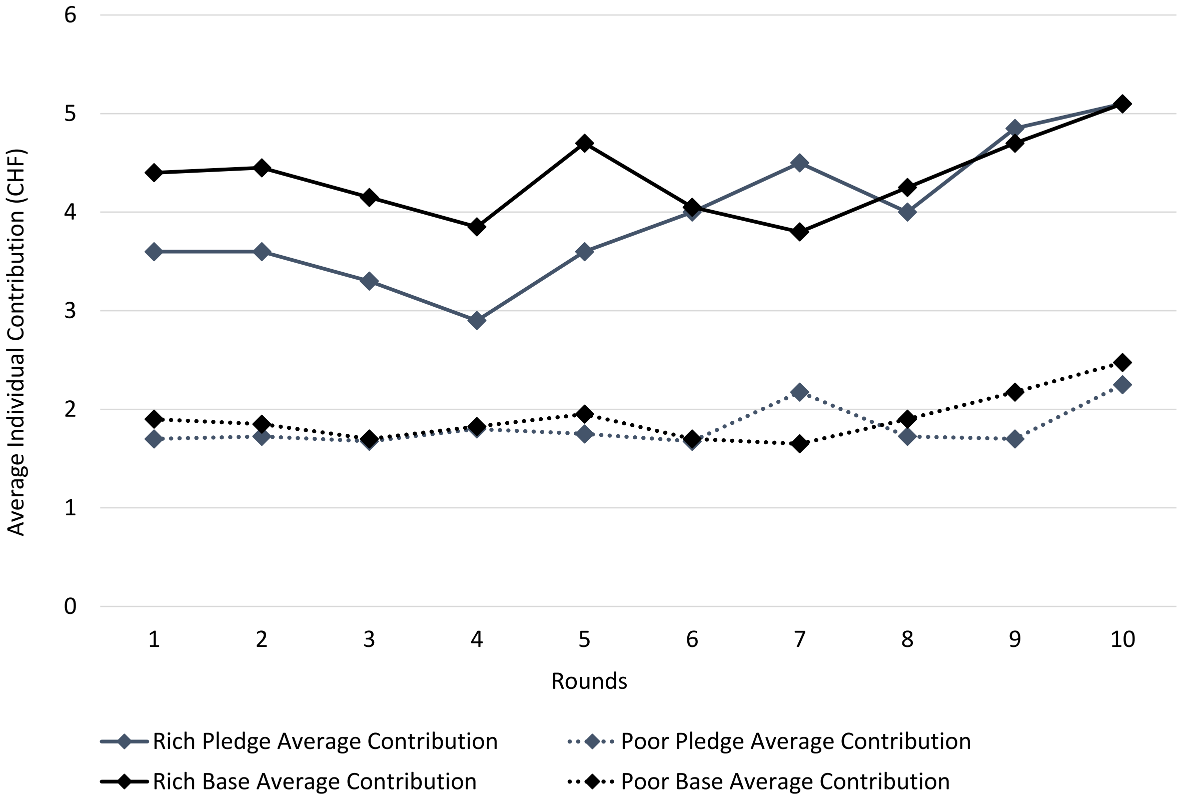


Figure 17: Average contributions throughout the entire experiment, separated by treatment and player type. Rich players in the pledge-treatment invested substantially less than their counterparts in the base-treatment.

5.1.2 Trust Building

Tavoni et al. (2011a) described increasing group success when communication in the form of repeated pledges was allowed in the groups. In groups with inequality in wealth, the positive effect was statistically significant. It was assumed that this form of communication built trust within groups, as players had the opportunity to reveal their intended actions to other group members before making their investments. As a result, success would mainly depend on whether players actually stick to their pledges. Actions in accordance with pledges should then lead to higher trust within the group.

But the trust levels in groups with pledge-communication were consistently lower compared to the groups without a pledge mechanism. The average trust levels over all 10 rounds, are significantly lower in groups with the pledge mechanism. (Mann-Whitney-Test, $W = 948.5, p-value = 7.833e^{-06}$). Figure 18 shows how the mean trust levels in each treatment (N=60 for each treatment) evolved over the 10 rounds of the experiment.

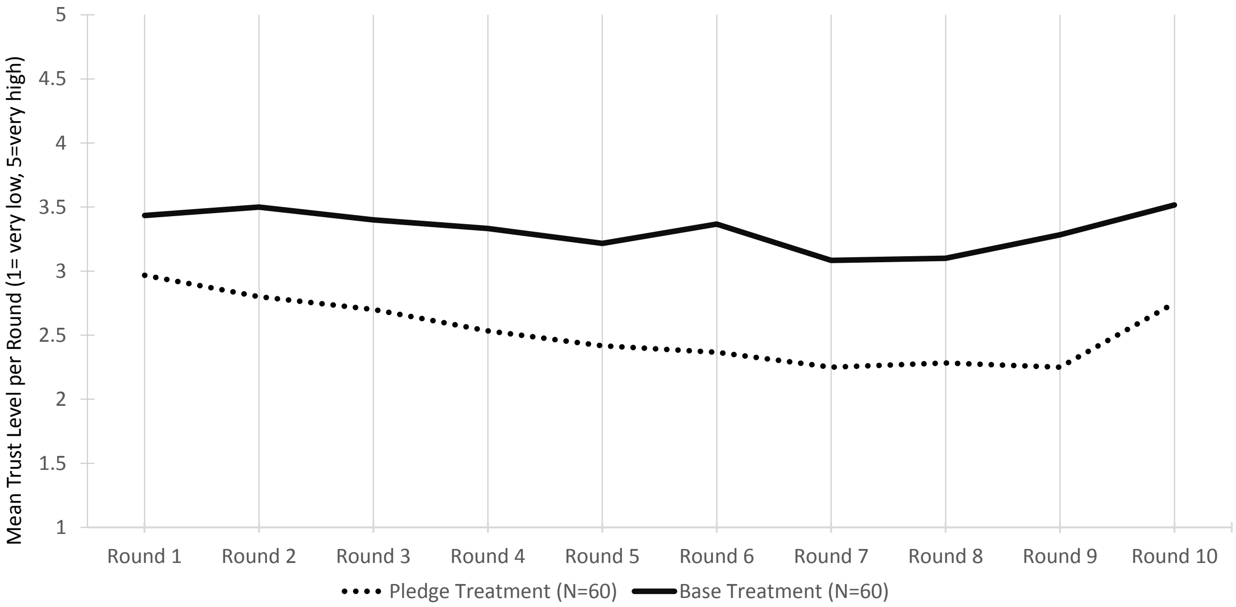
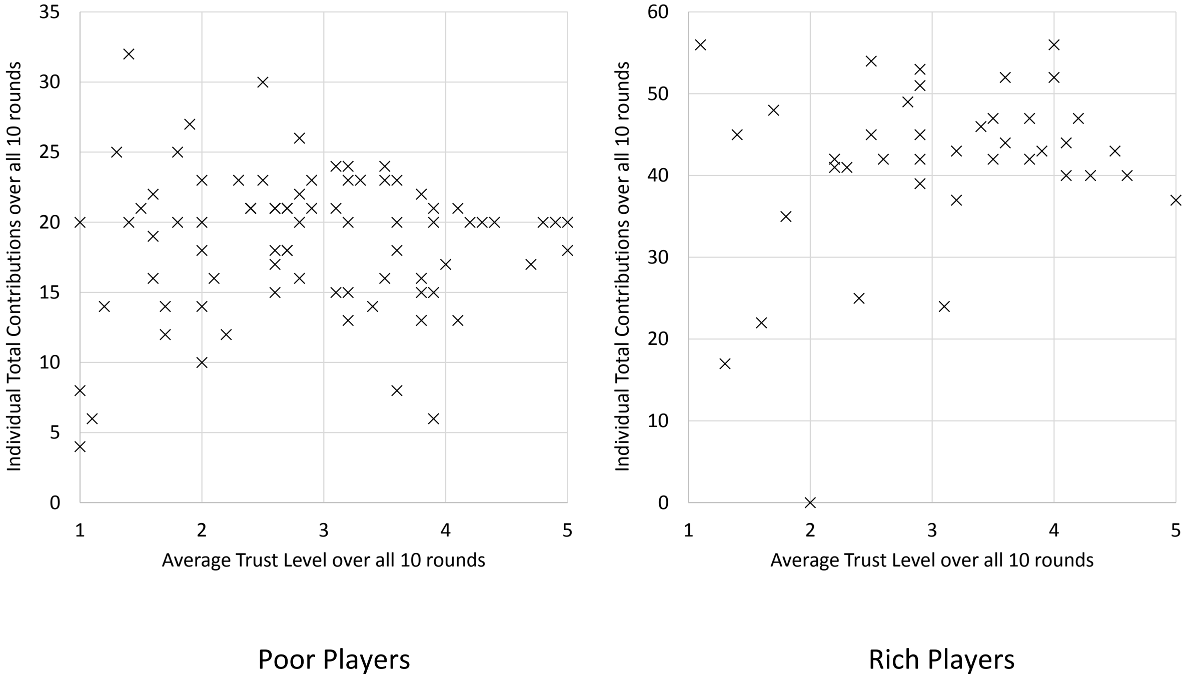


Figure 18: Over all 10 rounds, the mean trust levels expressed by participants in the pledge treatment were consistently lower than in the base treatment. N=120, with 60 participants in each treatment.

Across both treatments a positive correlation (r = 0.449) between the sum of group investments in the first round of the experiment and the trust levels expressed in round two was observed. Trust levels in round two were chosen for this analysos because, at the time of stating their trust levels in the first round, the participants had not yet observed what their fellow group members invested. In order to take other's actions into account when making up one's mind about how much to trust other group members, a first observation of other's actions is necessary. An almost identical correlation (r = 0.447) was found between first round total group investments and the average trust levels over all 10 rounds. When giving the first trust assessment, players only had the pledged amounts displayed in the planned catastrophe account to make their assessment. Interestingly, average trust levels went down from the first to the second round in groups with a pledge-mechanism, while they went slightly up in treatments without communication (see figure 18.

Although group total contributions in the first round correlated with average trust levels during the game, trust is a poor predictor for individual contributions. No meaningful relationship between an individuals' expressed trust level and their actual contributions could be identified. Since trust level statements were not made public to the other group members there was no incentive to make strategic trust statements. Across the board, correlations between individuals' total contributions and their average trust levels over all 10 rounds were quite low. This was tested for the entire sample as well as various sub-samples: (1) pledge-treatment only, (2) base-treatment only, (3) rich players, (4) poor players, (5) rich players isolated by treatment, (6) poor players isolated by treatments. No indication was found that the extent to which players trusted their group to contribute enough to avoid the catastrophe affected their investment decisions. Whether players thought it was hopeless or whether they were convinced the group would avoid the catastrophe, the contributions differed only slightly. This again is surprising as the intuitive expectation was to find a positive relationship between trust and contribution levels on an individual subject level. Figure 19 shows how trust levels and contributions were distributed in the experiment.



Poor Players

Rich Players

Figure 19: Average trust level stated by subjects over all 10 rounds of the game and their total contribution (N = 120)

In the post-game questionnaire participants were asked to assess the trust levels in their groups in the first round and the last round of the game. While there was no significant difference between the means of the assessments, the distribution of the perceived group trust level in the last round was much broader, with more participants stating they perceived either high or low levels of trust. Thus throughout the game participants perceived trust levels to have moved away from neutral towards both sides of the spectrum. Figure 20 shows how participants thought the trust levels in their group evolved from the first to the last round.

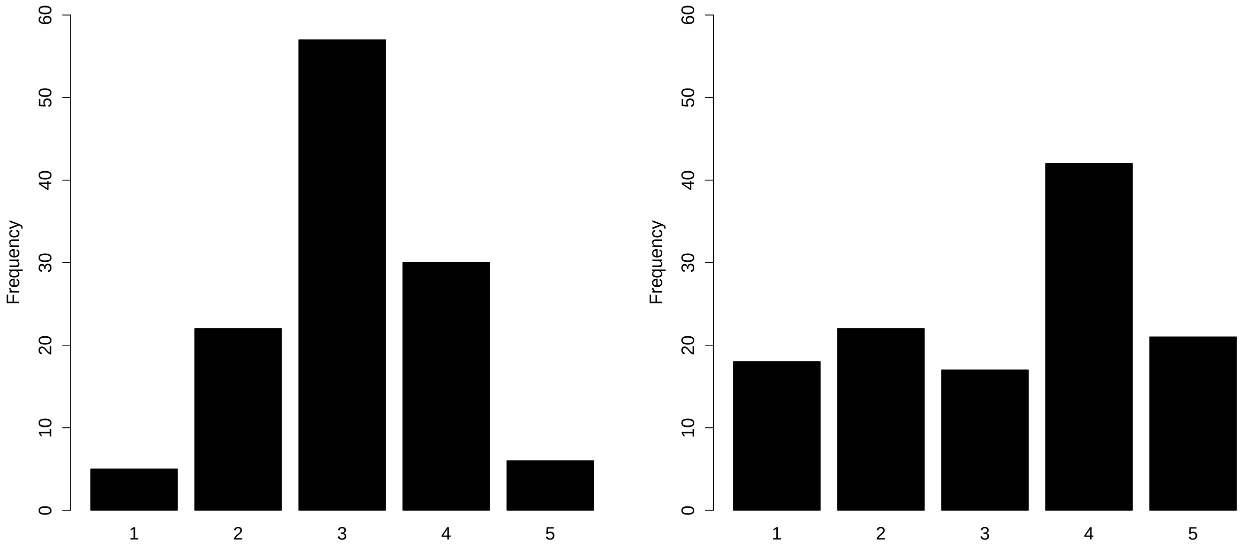


*Figure 20: Participants were asked after the experiment: **A:** What do you think was the the most frequently stated Trust Level in your Group in the **first** round? **B:** What do you think was the the most frequently stated Trust Level in your Group in the **last** round? Frequency of chosen answers. (1 = very low, 2 = low, 3 = medium, 4 = high, 5 = very high).*

5.1.3 Fairness

After the experiment participants were asked to state their idea of a fair total contribution for rich and poor players. Additionally subjects were asked to assess the fairness of the actions by the other players and the fairness of their own actions.

On average, both rich and poor players stated their own actual contribution when assessing a fair contribution for their own player type. This suggests that the majority of players thought that they acted in a fair way, which, although not compatible with standard economic assumptions, is not particularly surprising given the literature on the role of fairness in public goods games.

What is more surprising is that a strong consensus between poor and rich players exists about what fair contributions would have been for both player types. No significant differences in rich and poor players' stated amounts of fair contributions for both player types were found.

On average both rich an poor players agreed that a contribution of slightly more than 40 CHF (just above 50% of the endowment) would have been fair for rich players. For poor players this figure was just below 20 CHF, which would have been 50% of the endowment. It appears that the subjects in the aggregate applied a fairness norm consistent with equal proportion contributions, although this may also be influenced by the framing of the experiment to introduce a target investment of 50% of the group's total endowment. Figure 21 shows the mean assessments of fair contributions and actual contributions separated by player type (rich and poor).

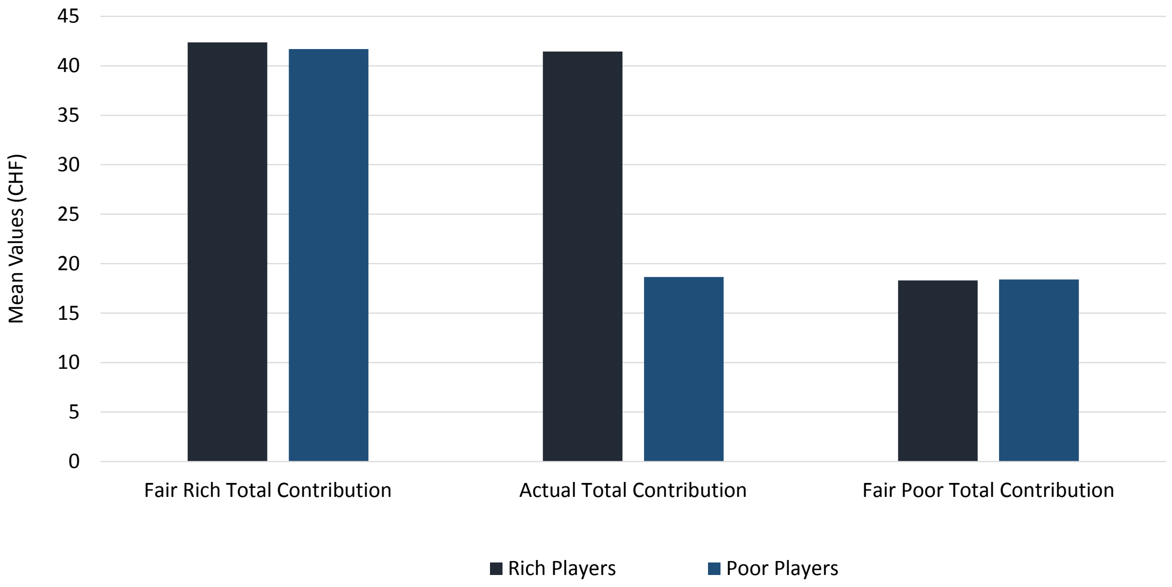


Figure 21: In the aggregate, there was a very strong agreement on what fair contributions would be for both fair and poor players. The assessments by poor players did not differ significantly from those of rich players. Additionally, both player types acted in accordance with the consensus on fair contributions on average.

At least in the aggregate, players acted according to the experiment fairness consensus. Thus one would expect that the majority of players would assess the behaviour of their fellow players as fair. However, this was not the case. On average, players assessed the fairness of their group members just above "Neutral" (mean: 3.2). When assessing their own fairness the average value was close to "Fair" (mean: 3.8). The assessment of own fairness is significantly higher than that of others' fairness (Mann-Whitney-Test, $W = 5212.5, p - value = 0.0001322$). A possible reason for this may be single players in groups that clearly deviated from the average assessment of fair contributions. This may have overshadowed the fair behaviour of the majority of the other group members and may have led to these lower assessments.

5.1.4 Gender Effects

A final focus were differences in behaviour between genders in the experiment. To avoid possible framing effects, participants stated their gender only at the very end of the experiment. In the total sample, there were 61 males and 59 females. The investment behaviour and fair contribution assessments did not differ significantly between males and females. The stated trust levels were also not different between women and men. By chance, only 15 females acted as rich players compared to 25 males. Interestingly, when asked to assess their own actions in terms of fairness, women in the rich player role stated a substantially lower level than their male counterparts. In poor players such a difference was not found. While rich female players did not act differently from their counterparts in terms of investments, it appears they did not feel as good about their choices in terms of fairness. Figure 22 shows the assessments of own fairness versus other players' fairness for different player types.

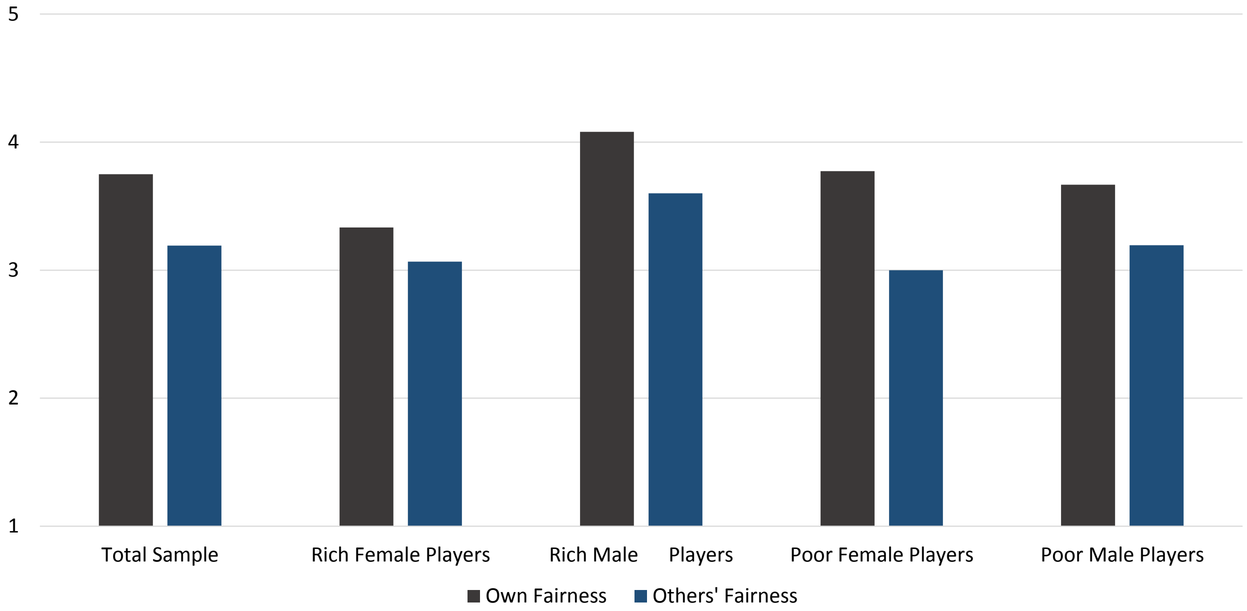


Figure 22: Comparison of assessment of own fairness versus others' fairness for the total sample and different player types (1 = Very Unfair, 2 = Unfair, 3 = Neutral, 4 = Fair, 5 = Very Fair). A significant difference was observed between rich females (N=15) and males (N=25) in how they assessed their own fairness. Rich women assessed their own fairness as lower than their male counterparts. Rich men also assessed others' contributions higher although the difference is not statistically significant compared to rich women's assessment of others' fairness.

5.2 Results for the Pledge Treatment Groups

Participants in the pledge treatment had the chance to make declarations of intent on three separate occasions during the game. Groups were expected to generally pledge to contribute sufficient amounts to avoid the catastrophe.

The opposite was the case. Of the 10 groups in this treatment, none consistently pledged more than the necessary amount of 160 CHF over all three opportunities to state pledges. Only two groups reached aggregate pledges of more than 160 CHF at some point in the game. At the same time, 8 out of 10 groups invested enough to avoid the catastrophe. Figure 24 compares the pledges made by participants to their actual contributions in the experiment.

Interestingly, even when given the chance to revise pledges in later rounds, group total pledges exceeded the threshold of 160 CHF in only one instance. The group pledges stated at the last opportunity (after round 6) were all below the necessary threshold of 160 CHF. On average subjects did not substantially alter their pledges when given the chance. From first to second pledge, players raised their pledges by 4.9% on average. From the second to the final pledge, they raised pledges by 0.8%. The evolution of the sum of in-group pledges and the modifications individuals made to their pledges when given the chance are shown in figure 23.

However, Panel B in figure 23 shows that few player kept their pledges constant. Most either increased or decreased their pledges throughout the game. Any points to the right of the blue dividing line indicate an increase in pledges compared to the first round pledge, while points to the left of the divide show individuals that have reduced their pledges in the course of the game.

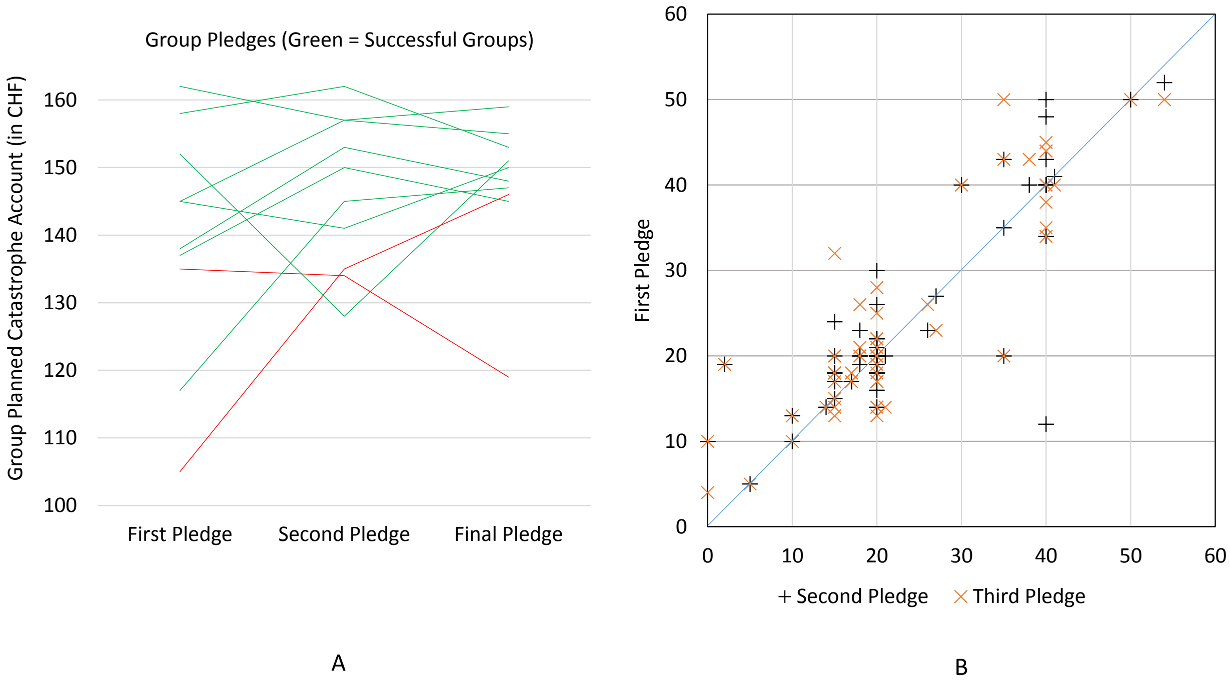


Figure 23: ***A:*** *Development of sum of in-group pledges (planned catastrophe account) over all three pledge rounds.* ***B:*** *First round individual pledges plotted against corresponding individual pledges in later rounds.*

While a clear upward or downward tendency in terms of pledge evolution throughout the game cannot be identified, almost all individuals ended up investing more than what they promised in their first pledges. Towards the end of the experiment, the actual contributions are more in line with the revised pledges. But still the majority of players invested more than what they pledged. Figure 24 shows how pledges and actual total contributions compare over the course of the game.

Panel D in 24 compares first round pledges with total contributions on a group level. Given the data on individuals it is not surprising that most groups invested substantially more than the sum the pledges in their planned catastrophe account.

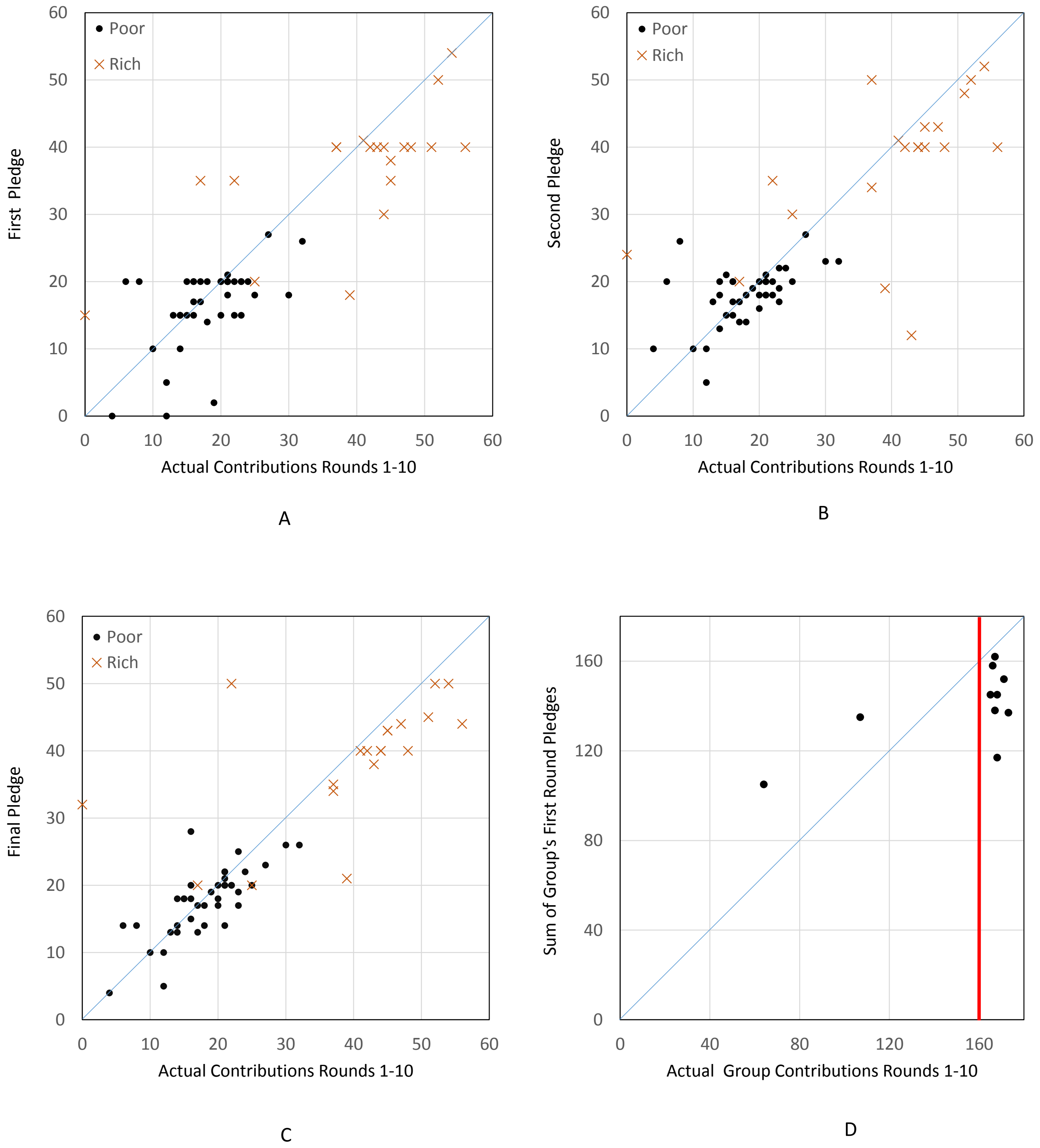


Figure 24: ***A:*** *First pledges and actual individual total contributions. A distinction is made between rich and poor players by choosing different colors.* ***B:*** *Second pledges and actual individual total contributions.* ***C:*** *Final pledges and actual individual total contributions.* ***D:*** *Sum of in-group first pledges and actual group total investments. The red border shows the necessary threshold investment to avoid the catastrophe. Pledges always pertained to intended total contributions over all 10 rounds.*

5.2.1 Willingness to Discipline Strategic Inaction

At the end of the experiment participants stated whether they believed their fellow group members would act according to their pledges. This was asked for each pledge round. 57% of the players believed that other players would act according to their first pledge. The credibility of the following pledges went down to 50% for the pledge in round 4 and to 33% for the final pledge respectively. Participants were also asked about how they would react if another player in their group invested a lot less than what he or she declared. Very few participants stated that they would discipline such actions by investing less themselves. Figure 25 shows participants responses. Out of 60 subjects, five individuals stated that they would react by also investing a lot less. Nine individuals said they would invest less. All other players said they would either not react or actually invest more.

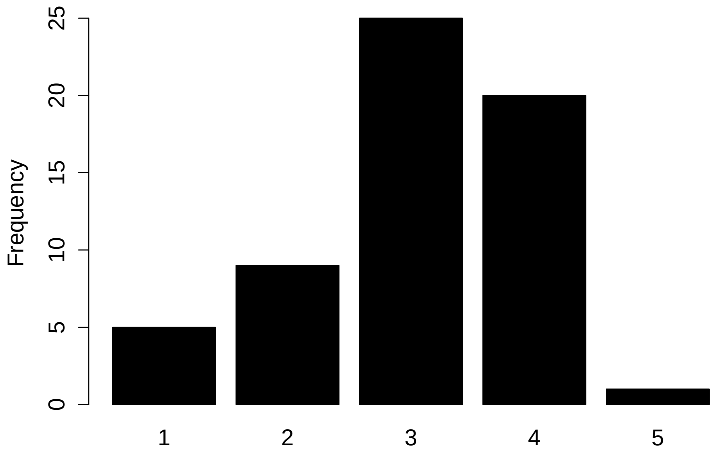


Figure 25: Frequency of chosen answers in the post-game questionnaire for subjects in the pledge treatment. Participants were answering the question: How would you react if another player in your group invested a lot less than they declared they would? Possible answers were: I would invest 1 = a lot less, 2 = less, 3 = the same amount, 4 = more, 5 = a lot more.

6 Discussion

This section provides possible explanations for the observations made in this experiment. By and large, the initial hypotheses were not confirmed. But the way in which deviations occurred are quite interesting and offer room for interpretation and a deeper discussion. Please recall our main hypotheses:

H_1 : A strong difference in vulnerability leads to mitigation of fairness. This results in low success rates in treatments where communication between subjects is not allowed. Rich, less vulnerable players will contribute less than 50% of their assets towards the provision of the public good. This hypothesis is based on the findings of Burton-Chellew et al. (2013); Gampfer (2014).

H_2 : Pledge-communication increases the groups' ability to coordinate their investments. This hypothesis is based on the work by Tavoni et al. (2011a).

H_3 : Pledge-communication provides additional information and builds trust among players that other players will invest enough to provide the public good. Therefore, groups that have access to such a mechanism will be more successful in providing the public good. Trust-building is the main factor that leads to higher success rates.

H_4 : When poor, vulnerable players are empowered, they will contribute more to the provision of a public good. The option to voice their opinion provides such an empowerment, that may lead to such behaviour. Poor, vulnerable players in the pledge-treatment will therefore contribute more than those in the base treatment. This hypothesis is based on the results reported by Gampfer (2014).

H_1 was not confirmed. Success rates were extraordinarily high in both treatments. In Burton-Chellew et al. (2013) wealth inequalities were more extreme, with rich players owning two thirds of the group's capital (i.e. 80 CHF) and poor players owning only one third (i.e. 20 CHF). This experiment deviated from these values because compared to a base treatment with equal loss probabilities, the group capital at stake in Burton-Chellew et al. (2013) is not equal in treatments with risk inequality.

For this to be fulfilled, both rich and poor players must collectively own 50% of the group's total endowment (or the risk inequality must be parametrized in a different way). Thus this deviation and equalization of group endowments may be one of the reasons for the observations made. If this were true it would follow that in Burton-Chellew et al. (2013) the breakdown of collaboration was primarily caused by the lower stakes on a group level, and not because of different levels of vulnerability. But the group stakes alone do not provide a satisfying answer as the literature hold numerous examples where success rates were lower in games with very similar parametrization (e.g. Milinski et al. (2008).

Another possible explanation is a difference in the salience of the threat that subjects were mitigating. Previous studies either used the terms "climate disaster" (Milinski et al., 2006, 2008; Tavoni et al., 2011a; Milinski et al., 2011; Jacquet et al., 2013; Burton-Chellew et al., 2013) or "project to avoid a loss" (Barrett and Dannenberg, 2012, 2014a). Participants were told that they were trying to avoid a "catastrophe". Compared to a climate disaster or a project to avoid a loss, the term catastrophe may be more salient. When imagining a catastrophe, participants may think of different events depending on their background. But the term catastrophe itself likely implies a very bad and unfavourable event to all of them. In addition, compared to a climate disaster, the threat of a catastrophe may be perceived as more imminent (e.g. avalanche, tsunami, earthquake, meteor hitting earth).

Of course, the expected payoffs in the game are not affected by such factors. But Burton-Chellew et al. (2013) report that participants' real world beliefs affected their decisions even though they had no effect on expected payoffs. For example, participants that were more sceptical of climate change and humanity's ability and obligation to prevent it, invested a lower proportion of their endowments in the games. The game was framed in a neutral way to avoid such effects but still asked participants in a post-game questionnaire, whether they thought climate change was a serious threat. Out of 120 subjects, 118 believed that climate change is a serious threat.

H_2 was not confirmed. In contrast to Tavoni et al. (2011a), success rates did not differ significantly between treatments in our study. For this reason it is impossible to say, whether pledge-communication increased coordination. Given the high success rates in both treatments, any effects from communication may have been dominated by another factor. A possible explanation is the different loss probability which made our experiment a higher-stake game. In Tavoni et al. (2011a) the loss probability was 50% for all players, whereas in our case it was 65% for rich players and 95% for poor players. With a lower loss probability the incentive to provide the public good is also lower. In our study it was very advantageous from a social benefit perspective to provide the public good. On a group level, investing 160 CHF lead to a 100 CHF premium in expected group payoff compared to investing nothing at all. In Tavoni et al. (2011a) this premium was 18 € with an active group investment of 84 €.

Perhaps the most interesting observation is that H_3 was disproved. The results indicate the exact opposite. The fact that players communicated intentions to invest less than what was necessary to avoid the catastrophe, could explain the higher distrust observed. One indeed wonders why pledged contributions that do not suffice to avoid a catastrophe should increase the subjects trust in their fellow group members to contribute enough.

Intrigued by this observation we wondered if this type of behaviour had occurred in previous studies. Tavoni et al. (2011a) provided the group pledge data from their study. It turns out that their subjects also generally pledged less than what was necessary. Figure 26 shows the aggregate pledges on group levels before the first active investment decision for a total of 30 groups, ranked in ascending order. 8 out of 10 groups invested enough to provide the public good in our study, while in Tavoni et al. (2011a) it was 13 out of 20 groups. Interestingly, 2 out of the 5 groups that had group pledges higher than the necessary 84 € failed to provide the public good.

For group success it apparently did not matter if pledges were sufficiently high on aggregate. In Tavoni et al. (2011a) 5 groups pledged to contribute at least the necessary amount of 84 €. Of those, three were successful (60%). That leaves 15 groups that pledged to contribute less than 84 €. Of these, 10 succeeded in providing the public good (67%). In our study, the single group that pledged to contribute at least 160 CHF succeeded in avoiding the catastrophe. 9 groups pledged to contribute less, of which 7 succeeded (78%).

One possible explanation for this behaviour is that some individuals pursued a strategy of scaring their fellow players into higher investments by stating low pledges and signalling their own unwillingness to contribute enough to provide the public good. The observed behaviour

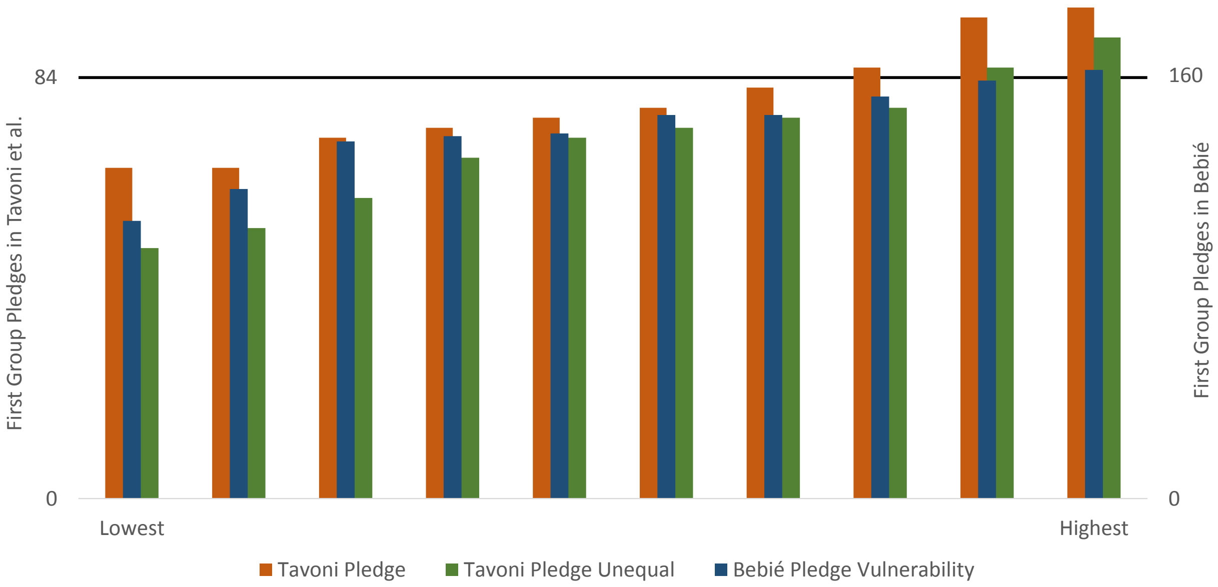


Figure 26: Group aggregate pledges stated at the first opportunity to declare intentions. The black line shows the necessary investments to provide the public good. In the Tavoni groups this amount was 84€ (120€ minus the inactive round investments of 36€), whereas in our study it was 160 CHF. Tavoni et al. (2011a) provided the group pledges from their two pledge treatments, which were here combined with our own results. Out of 30 groups, only 6 pledged to invest enough and of those two failed to invest enough. In 24 groups, the individual pledges would have been insufficient to provide the public good. 18 of these groups still suceeded.

reminds one of a chicken game situation, where some players finally give in and invest more to avoid the worst and some players continue to free-ride, betting that their counterparts will not dare to let the catastrophe happen. In the comment section of the questionnaire, various participants voiced their frustration with single players in the group that were not pulling their weight and relied on others to save them.

Evidence for strategic behaviour is provided in figure 17 in section 5.1.1 at least for rich players. Rich players in the pledge-treatment invested significantly less than their counterparts in the base treatment in the first 5 rounds. Towards the end of the game they adjusted their contributions upwards to avoid the catastrophe. The histograms shown in figure 16 also point towards more diverse investment choices in rich players in the pledge-treatment.

Support for strategic communication also exists in Barrett and Dannenberg (2012) where impact and threshold uncertainty was introduced. In a post-game questionnaire, participants in treatments with threshold uncertainty stated that they tried to motivate other players to contribute more by proposing lower amounts than what was necessary to provide the public good. *They thought that a proposal below 200 was more credible and so was more likely to stimulate contributions by others* (Barrett and Dannenberg, 2012, p. 17374). But even in the treatment with certain thresholds, where the game was a coordination game, the actual contributions in scenarios without threshold uncertainty were typically higher than what the players pledged, although the difference is much less substantial.

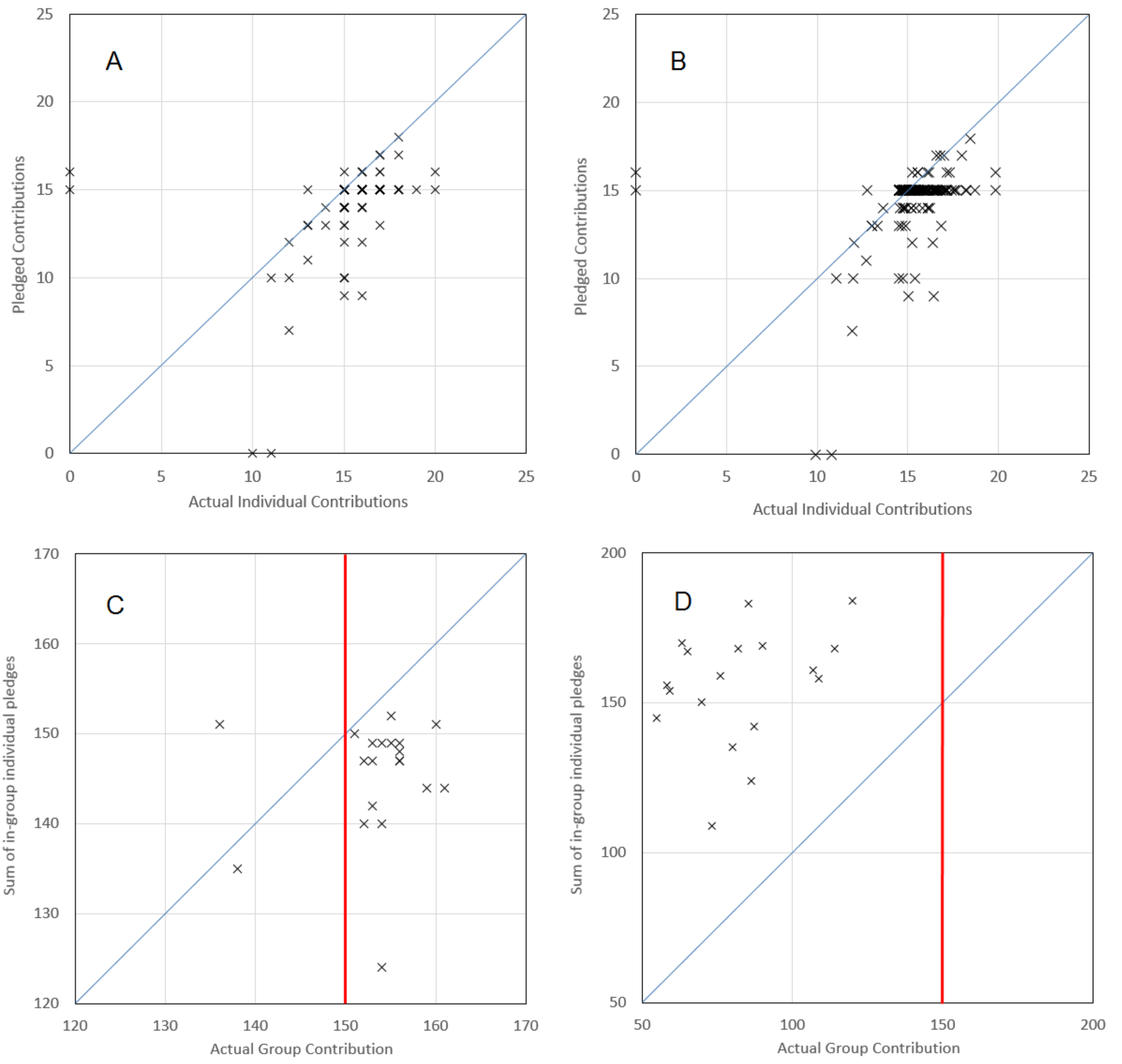


Figure 27: Visual representation of data provided by Barrett and Dannenberg (2012): ***A****: Pledges and actual contributions from individuals that were exposed to threshold certainty (N=200).* ***B****: Pledges and jittered (random noise added to avoid overlapping points) actual contributions from individuals that were exposed to threshold certainty (N=200).* ***C****: Sum of in-group individual pledges and group total contributions for groups that were exposed to threshold certainty (N=20).* ***D****: Sum of in-group individual pledges and group total contributions for groups that were exposed to* ***threshold uncertainty*** *(N=20).*

Barrett and Dannenberg (2012) kindly agreed to share their data for a detailed analysis. Their data adds further evidence for strategic behaviour. Although, they applied a one-shot game to test threshold uncertainty, the behaviour of subjects resembles our observations quite closely. In treatments without threshold uncertainty (in which the one-shot game was a coordination game) most groups made pledges that were below the necessary threshold investment of 150 units. Figure 27 shows the individual and group level investment behaviour in their study.

When comparing these results to figure 24 in the results section, the similarities in contribution behaviour on an individual, as well as on a group level are striking.

Many subjects appear to have tried to nudge their fellow group members into higher contributions by making pledges that were insufficient if followed by all individuals in a group. But by and large they did not dare to act on their threat of low contributions as most players ended up investing more than they pledged to contribute.

A comparison with the results by Barrett and Dannenberg (2012) may also show evidence that our collective-risk social dilemma was a coordination game rather than a prisoner's dilemma. When comparing panels C and D in figure 27 to panel D in figure 24, it seems that our own observations are much more in line with the observations Barrett and Dannenberg (2012) made for the case where the game was a coordination game. They showed that the introduction of threshold uncertainty lead to a prisoner's dilemma in a one-shot game. In the case of the multi-round collective risk social dilemma, a clear classification is difficult as mentioned by Tavoni et al. (2011a).

Finally it is also surprising that the majority of subjects stated that they would either keep their investments constant or even increase their own contributions if a player in their group invested substantially less than what that player pledged to contribute (see Figure 25). It was expected that in such scenarios players would state an intention to penalize such behaviour by reducing their own contributions as well.

H_4 was not confirmed. Contributions by poor players did not differ significantly between treatments. On average poor players invested 18.7 CHF (46,8% of their endowments) towards the provision of the public good. It appears that adding a pledge-mechanism which allowed poor players to voice their opinion did not affect their willingness to invest more. Of all the results, this was the least surprising. The switch from the responder to the proposer role described by Gampfer (2014) is a much stronger power shift than the provision of a communication tool. Thus it was to be expected, that any effect in our experiment would be weaker.

Fairness Assessments

In the aggregate a rather strong consensus on what rich and poor players deemed fair contributions for both player types emerged. And on average at least, players acted according to what they thought to be a fair contribution for their own player type. Thus it may appear puzzling at first that players saw their own actions as substantially fairer compared to those of the other members in their group. But this phenomenon has long been documented in the literature and is often described as an "egocentric fairness bias" (Messick et al., 1985; Thompson and Loewenstein, 1992). The phenomenon describes situations where individuals view their own actions as fairer than others or - when asked to make fair allocations of any good - allocate to themselves a larger share than they would to a third party. The evidence from our experiment thus shows the presence of an egocentric fairness bias.

An interesting debate is whether strategic fairness, as described for example by Brick and Visser (2014) is rooted in the same egocentric bias or whether the individuals in their study (and others) knowingly applied fairness norms strategically. The distinction is important as the egocentric fairness bias observed in our experiment is likely not applied knowingly by the subjects. Therefore there is a potential to mitigate its effect on the outcomes of the collective-risk social dilemma. For example, one could inform the participants of the bias' existence in the instructions to the experiment and advocate subjects to take this information into account when making their decisions. Perhaps this would result in a change in negotiation behaviour. In contrast, when negotiating parties knowingly apply fairness norms that work to their advantage, improving outcomes by taking countermeasures that mitigate this behaviour are likely to fail.

A final notable point concerns a possible gender bias: While women and men in the experiment did not act differently or feel different levels of trust compared to men, women in the rich player role apparently felt their actions were less fair compared to their male counterparts (see Figure 22. In poor players such a difference was not found. Since the number of observations is rather low, one should be careful to interpret too much into this divergence.

6.1 External Validity of Results

The game design with a threshold to avoid disastrous consequences does reflect certain aspects of climate negotiations but clearly fails to reflect others. At COP21 a 2 ° C global warming limit was set as a threshold to avoid "dangerous climate change". In this respect an effort to reach a pre-defined threshold is realistic. However, the term "dangerous climate change" is much more vague compared to the consequences modelled in the games described above.

Additionally, there is no scientific consensus that reaching this goal would actually avoid dangerous climate change (threshold uncertainty). In fact there are considerable changes to be expected even if global warming were limited to below 2 ° C (IPCC, 2012). Delegates involved in negotiations are likely aware of this scientific consensus.

Furthermore, while the uncertainty associated with adverse effects from climate change remain large, a scenario with total loss of endowment is not plausible at least on a national level. Conversely, players in the game did not face loss of their entire wealth (i.e. only their game-endowment was at stake). The actual expected consequences of climate change include such perilous examples as loss of lives, property and land or famine.

As a result, the insights from the studies summarized above should be taken with a grain of salt. None of the authors cited in this thesis claim to have created conditions that are equivalent to actual climate negotiations. Instead they all point to the limitation of applicability of results from these studies to real world situations.

7 Conclusion

This experiment set out to investigate the importance of communication for trust building and enhancing cooperation within groups. We used a social dilemma situation that is particularly challenging. In this situation, as in the climate change challenge, actors have to overcome inequality in wealth and risk and coordinate countermeasures to a common threat. As in the real world example, some actors are more vulnerable to the consequences of inaction and they are often the ones that do not control the lion's share of resources to mitigate the problem. The existing literature suggests that combined inequality in wealth and risk leads to disaster (Burton-Chellew et al., 2013), but that communication in the form of pledges results in enhanced cooperation and greater group success in settings with inequality (Tavoni et al., 2011a).

Our results point in a different direction. First of all, groups generally did well in coordinating to avoid catastrophic loss in a setting with inequality in wealth and risk. Compared to the breakdown of cooperation reported by Burton-Chellew et al. (2013) our design must have had at least one element that sparked collaboration even in the absence of communication. As argued above, the slight difference in parametrization is hardly a satisfying explanation. Framing the game as a "neutral" catastrophe scenario rather than a climate change mitigation task may have changed the mentality of players in the game. Although changing the name of the threat doesn't change the structure of expected payoffs, a psychological bias may be at work.

Most previous studies that applied a collective-risk social dilemma were framed in the context of climate change. This is a rather specific threat. Thus participants likely differed in how they viewed the severity of this threat. Climate change will materialize over a medium to long-term time horizon. One can only imagine the perilous consequences that may occur in the future. And even those will not be concentrated in a single catastrophic event but be spread out over time. Finally it is at least debatable if we have truly observed actual catastrophic events that were clearly caused by climate change. All of these factors may make climate change a somewhat obscure and less tangible threat than a "neutral" catastrophe.

Regarding the impact of communication, we found strong evidence that this tool, when used strategically, mitigates trust among actors. In our experiment, numerous subjects exploited the pledge-mechanism to make pledges that likely aimed to force other players in their group to contribute more. Few appear to have stated their true contribution intentions. The fact that most players finally invested more than what they pledged provides strong evidence for the presence of strategic communication.

Perhaps the most worrying observation in the present experiment is the strategic behaviour by rich players, when communication was available (see figure 17). Rich players' contributions in early rounds were significantly lower when pledge-communication was available. This suggests that the communication tool encouraged some privileged actors to take advantage of their situation and delay their actions towards a later stage in the game. In the context of climate change mitigation, where immediate action is necessary, such behaviour must be avoided. This result in mind, the effectiveness of communication in overcoming difficult social dilemmas may need to be further investigated at least when applied to the context of climate change, where timing of actions is critical.

A closer look at additional data from studies that inspired this experiment revealed that the strategic behaviour we observed was not an isolated incident within our own experiment. In numerous previous studies subjects pledged to contribute amounts that were (1) insufficient to provide the public good if adopted by all and (2) lower than what subjects actually contributed in the experiment. Pledge-communication was shown to increase group success in Tavoni et al. (2011a). But the mechanism at work may have been quite different to what one might intuitively assume. Tavoni et al. (2011a, p. 11827) state that, *"as the difference between cumulative contributions and pledged amounts increases, the probability of a player being in a successful group decreases significantly"*. This implies that staying true to one's pledge increases the chances of success. Intuitively one may assume that the reason for this is trust building within the group. As players back up their pledges with corresponding action, the pledge credibility increases, which leads to higher trust among players.

But this observation fails to account for the fact that, had all players acted exactly according to their pledges, 15 out of 20 groups would have failed in Tavoni et al. (2011a) (see figure 26). In reality 13 out of 20 groups succeeded, because most players contributed more than what they pledged. Readiness to deviate from pledges was actually instrumental for group success both in Tavoni et al. (2011a) and our own experiment.

This suggests that pledge-communication made defecting players more successful in scaring their fellow group members into action. Pledges transformed the experiment into a game with characteristics of a chicken game. Imagine a situation where two drivers are driving towards each other. One eventually has to deviate from the chosen line to avoid a crash. Whoever goes out of line first loses the game. If both players refuse to drive out of the way, the cars crash and both players lose. This is a classic chicken game situation and it does resemble some of the behaviour observed in this experiment.

A key question is whether the same was true in Tavoni et al. (2011a). With lower stakes, provision of the public good is less attractive. Therefore such a chicken game may be less strong because the effect of letting the group fail is less severe. Considering the car analogy above, one can imagine that the drivers are driving at each other at lower speed and that a crash would be less deadly. However, the fact that many subjects invested more than what they pledged points towards strategic communication in Tavoni et al. (2011a) in the same way as in the present experiment.

The role trust played in this experiment is not entirely clear: Trust did not seem to play a decisive role in how players chose their contributions, at least in the present study. Figure 19 shows the lack of an obvious link between the trust that subjects had in the other members of their groups and their investment choices. For rich players, there is a slight positive correlation between trust and contributions. Among poor players a relationship is non-existent. However, the contribution bandwidth decreases in poor players with increasing trust levels. Perhaps this indicates a higher sense of confidence in one's own choices when trust in one's group members was high.

For policy-making and strategies in climate negotiations the present study provides a number of insights: Strategic communication has clearly been shown to mitigate trust among players. Ostrom (2014) emphasised the importance of trust for conformance with policies of any kind in the context of social dilemmas. Thus, applying strategic communication in policy-making or negotiations may have a similar trust-mitigating effect as it did in our experiment. Actors aiming to increase multilateral trust in such processes would therefore do well to avoid such strategies.

Additionally the benefit of strategic communication comes at the cost of having to reinforce one's strategy with inaction until one's counterparts are sufficiently scared to cave in. This is a gamble that carries a risk. Applied to climate change mitigation, this type of gamble will lead to higher costs in the future and lower chances of mitigating severe damages. Negotiators striving for immediate international action should keep this in mind when devising pledges in the future.

Finally, assuming that threat salience was among the factors that sparked higher readiness to avoid the disaster, it would follow that action will become stronger as people become more scared of climate change. Threat salience will increase with time as the effects of climate change become more visible. Unfortunately this is no cause for optimism as it implies that necessary action may be delayed until threat salience becomes great enough. Given what we know about the climate system, by that time it will be too late to fully reverse the path already taken. Policy-makers will have to find alternatives to drive action before widespread fear of the effects of climate change becomes sufficient to put the necessary pressure on governments. Ostrom (2014) discusses polycentric approaches as possible solutions. Initiatives like the C40 Cities Climate Leadership Group are promising examples where substantial impacts can be generated without the need for global agreements.

7.1 Suggestions for Future Research

Future research should focus on mechanisms that discourage strategic communication in such settings. An obvious possibility is to penalize any kind of deviation from stated pledges (positive or negative). For example, a fixed amount could be deducted from players' private accounts if they deviate from pledges. Of course this is not a solution that should be applied to the real-life situation. Penalizing countries for taking action that is stronger than their pledges would clearly not help climate change mitigation.

A softer approach would be to inform subjects before the experiment about the effects of strategic communication on trust within groups. Perhaps knowing the impact of strategic communication can keep actors from applying it in an experimental setting.

An additional idea would be to increase the costs of avoiding the disaster towards the end of the experiment. Early action would then be more cost-effective than efforts at a later point in the game. Such a parametrization would make strategic communication less attractive, or at least make it more costly to delay one's own actions to scare others. This would also come closer to modelling the real situation of coordinating climate change mitigation over time.

A final factor that is not modelled in the games described in this thesis is reliance on future generations to fix the problems caused by the current generation. A way to take this into account would be to split the game into two sections: In the first section groups play rounds one to five of a mitigation game. Another group would then enter the game at the beginning of round six and have to finish the game until round ten. The first group would be informed that another group of subjects will take over their task when they are finished. Throughout the game, any investment gap to a fixed target amount would result in ever increasing damages. Both groups would be paid out the amounts remaining in their private accounts when their part of the experiment finishes. This type of intergenerational discounting and subjects' sense of responsibility for the well-being of actors in the future, whom they do not know, would be very interesting to investigate as it applies to the reality of climate change mitigation.

Acknowledgements

A number of individuals have helped me in various ways in writing this thesis. I would like to thank my supervisor, Prof. Dr. Gunter Stephan for his advice and support, which made the experiment presented in this thesis possible. I also thank Prof. Dr. Ralph Winkler for helpful insights and valuable feedback. Additionally, Dr. Elisabeth Gsottbauer's advice and encouragements were instrumental for the successful implementation of the game in zTree and the subsequent experiment at ETH. The same is true for Dr. Robert Gampfer who took the time to explain his own experiment and factors to consider when programming a game. Dr. Maxwell Burton-Chellew of Oxford University provided his design for a similar game which served as a basis for our own design. Without his aid, conducting this experiment in such a short timeframe would have been much more difficult. Finally, Dr. Stefan Wehrli showed great readiness to help me test the experiment at the Decision Science Laboratory of ETH.

In the analysis of the results I was supported by Dr. Alessandro Tavoni of LSE, Prof. Dr. Scott Barrett of Columbia University and Prof. Dr. Astrid Dannenberg of University of Kassel. These authors have shared data from their own experiments that were instrumental to solidify the evidence for strategic communication.

Finally I was able to count on the support of family and friends, who helped me in many different ways along the way. I would like to extend a special thanks to Pascale Gut, Rosmarie, Bruno and Stefan Bebié, Daniela Seiler, Fabian Baumgartner, Gian Egeler, Fabian Langensteiner and Luciano Raveane.

References

James Andreoni and John H. Miller. Rational cooperation in the finitely repeated prisoners' dilemma: Experimental Evidence. *The Economic Journal*, 10327(418):570–585, 1993. ISSN 00220531. doi: 10.1016/0022-0531(82)90029-1.

Scott Barrett and Astrid Dannenberg. Climate negotiations under scientific uncertainty. *Pnas*, 109(43):17372–17376, 2012. doi: 10.1073/pnas.1208417109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1208417109.

Scott Barrett and Astrid Dannenberg. Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change*, 4(1):36–39, 2014a. ISSN 1758-678X. doi: 10.1038/nclimate2059. URL `http://www.nature.com/doifinder/10.1038/nclimate2059`.

Scott Barrett and Astrid Dannenberg. Supporting Information: Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change*, 4(1):1–20, 2014b. ISSN 1758-678X. doi: 10.1038/nclimate2059. URL `http://www.nature.com/doifinder/10.1038/nclimate2059`.

Esther Blanco, E Glenn Dutcher, and Tobias Haller. *To mitigate or to adapt? Collective action under asymmetries in vulnerability to losses.* 2014. ISBN 5125072788.

Kerri Brick and Martine Visser. What is fair? An experimental guide to climate negotiations. *European Economic Review*, 74(June):1–57, 2014. ISSN 00142921. doi: 10.1016/j.euroecorev.2014.11.010. URL `http://dx.doi.org/10.1016/j.euroecorev.2014.11.010`.

Maxwell N. Burton-Chellew, Robert M. May, and Stuart a. West. Combined inequality in wealth and risk leads to disaster in the climate change game. *Climatic Change*, 120(4):815–830, 2013. ISSN 0165-0009. doi: 10.1007/s10584-013-0856-7. URL `http://link.springer.com/10.1007/s10584-013-0856-7`.

Piet Buys, Uwe Deichmann, Craig Meisner, Thao Ton That, and David Wheeler. Country stakes in climate change negotiations: two dimensions of vulnerability. *Climate Policy*, 9(3):288–305, 2009. ISSN 14693062. doi: 10.3763/cpol.2007.0466.

Fredrik Carlsson, Mitesh Kataria, Alan Krupnick, Elina Lampi, Å sa Löfgren, Ping Qin, and Thomas Sterner. A fair share: Burden-sharing preferences in the United States and China. *Resource and Energy Economics*, 35(1):1–17, 2013. ISSN 09287655. doi: 10.1016/j.reseneeco.2012.11.001. URL `http://dx.doi.org/10.1016/j.reseneeco.2012.11.001`.

Russell Cooper, Douglas V. DeJong, Robert Forsythe, and Thomas W. Ross. Cooperation without Reputation : Experimental Evidence from Prisoner ' s Dilemma Games. *Games and Economic Behavior*, 12(2):187–218, 1996. ISSN 08998256. doi: 10.1006/game.1996.0013.

Astrid Dannenberg, Bodo Sturm, and Carsten Vogt. Do equity preferences matter for climate negotiators? An experimental investigation. *Environmental and Resource Economics*, 47(1):91–109, 2010. ISSN 09246460. doi: 10.1007/s10640-010-9366-5.

Astrid Dannenberg, Andreas Löschel, Gabriele Paolacci, Christiane Reif, and Alessandro Tavoni. On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds. *Environmental and Resource Economics*, pages 365–383, 2014. ISSN 09246460. doi: 10.1007/s10640-014-9796-6. URL `http://dx.doi.org/10.1007/s10640-014-9796-6`.

Ernst Fehr, Michael Naef, and Klaus M. Schmidt. in and Maximin Preferences Aversion , Inequality Efficiency , Comment Simple Distribution Experiments :. *American Economic Review*, 96(5):1912–1917, 2006. ISSN 0002-8282. doi: 10.1257/aer.96.5.1912.

Urs Fischbacher. Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007. ISSN 13864157. doi: 10.1007/s10683-006-9159-4.

Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397–404, 2001. ISSN 01651765. doi: 10.1016/S0165-1765(01)00394-9.

Robert Gampfer. Do individuals care about fairness in burden sharing for climate change mitigation? Evidence from a lab experiment. *Climatic Change*, 124(1-2):65–77, 2014. ISSN 01650009. doi: 10.1007/s10584-014-1091-6.

Bill Hare, Niklas Höhne, Kornelius Blok, Lousie Jettery, and Johannes Gütschow. Climate Action Tracker (CAT), 2015. URL `http://climateactiontracker.org/`.

Derek D. Headey. The impact of the global food crisis on self-assessed food security. *World Bank Economic Review*, 27(1):1–27, 2013. ISSN 02586770. doi: 10.1093/wber/lhs033.

IPCC. Summary for policymakers. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor,. Technical report, Cambridge University Press, Cambridge, UK, and New York, NY, USA, 2012. URL `http://ebooks.cambridge.org/ref/id/CBO9781139177245`.

Maros Ivanic and Will Martin. Implications of higher global food prices for poverty in low-income countries. *Agricultural Economics*, 39(SUPPL. 1):405–416, 2008. ISSN 01695150. doi: 10.1111/j.1574-0862.2008.00347.x.

Jennifer Jacquet, Kristin Hagel, Christoph Hauert, Jochem Marotzke, Torsten Röhl, and Manfred Milinski. Intra- and intergenerational discounting in the climate game. *Nature Climate Change*, 3(12):1025–1028, 2013. ISSN 1758-678X. doi: 10.1038/nclimate2024. URL `http://www.nature.com/doifinder/10.1038/nclimate2024`.

Martin Kesternich, Andreas Löschel, and Andreas Ziegler. Negotiating Weights for Burden Sharing Rules among Heterogeneous Parties : Empirical Evidence from a Survey among Delegates in International Climate Negotiations. *ZEW DIis*, (14), 2014.

Kris N Kirby. Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, 126(1):54–70, 1997. ISSN 1939-2222. doi: 10.1037/0096-3445.126.1.54. URL `http://www.researchgate.net/publication/232437900_Bidding_on_the_future_Evidence_against_normative_discounting_of_delayed_rewards`.

Kelly Levin, David Rich, Yamil Bonduki, Michael Comstock, and Dennis Tirpak. Designing and Preparing Intended Nationally Determined Contributions (INDCs). Technical report, World Resource Institute, 2015.

David M Messick, Suzanne Bloom, Janet P Boldizar, and Charles D Samuelson. Why we are fairer than others. *Journal of Experimental Social Psychology*, 21(5):480–500, 1985. doi: 10.1016/0022-1031(85)90031-9.

Manfred Milinski, Dirk Semmann, Hans-Jürgen Krambeck, and Jochem Marotzke. Stabilizing the earth's climate is not a losing game: supporting evidence from public goods experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11): 3994–3998, 2006. ISSN 0027-8424. doi: 10.1073/pnas.0504902103.

Manfred Milinski, Ralf D Sommerfeld, Hans-Jürgen Krambeck, Floyd a Reed, and Jochem Marotzke. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7):2291–2294, 2008. ISSN 0027-8424. doi: 10.1073/pnas.0709546105. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2268129&tool=pmcentrez&rendertype=abstract`.

Manfred Milinski, Torsten Röhl, and Jochem Marotzke. Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic Change*, 109(3-4):807–814, 2011. ISSN 0165-0009. doi: 10.1007/s10584-011-0319-y. URL `http://link.springer.com/10.1007/s10584-011-0319-y`.

Elinor Ostrom. A polycentric approach for coping with climate change. *Annals of Economics and Finance*, 15(1):97–134, 2014. ISSN 15297373. doi: doi:10.1596/1813-9450-5095.

Thomas Pfeiffer and Martin a Nowak. All in the game. *Nature*, 441(June):583–584, 2006. ISSN 0028-0836.

Amy R Poteete, Marco A Janssen, and Elinor Ostrom. *Working together: collective action, the commons, and multiple methods in practice.* Princeton University Press, 2010.

Patrick Regan, Joyce Coffee, Nitesh Chawla, Chen Chen, Martin Murillo, Meghan Doherty, Jessica Hellmann, and Ian Noble. Notre Dame Global Adaptation Index, 2015. URL `http://index.nd-gain.org:8080/index_main.py?tool_type=basic`.

Joshua Schneck. NI Summary of COP 15 Outcomes. 2009.

Amartya Sen. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs*, 6(4):317–344, 1977. ISSN 00483915. doi: 10.2307/2264946.

Amartya Sen. Markets and freedoms: achievements and limitations of the market mechanism in promoting individual freedoms. *Oxford Economic Papers*, pages 519–541, 1993.

Franklin Steves and Alexander Teytelboym. Political Economy of Climate Change Policy. *Working Paper*, (October), 2013. URL `http://www.smithschool.ox.ac.uk/research/library/Steves_Teytelboym_WorkingPaper.pdf`.

Stockholm Environment Institute and EcoEquity. Carbon Equity Reference Calculator, 2015. URL `http://calculator.climateequityreference.org/`.

A. Tavoni, A. Dannenberg, G. Kallis, and A. Loschel. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, 108(29):11825–11829, 2011a. ISSN 0027-8424. doi: 10.1073/pnas.1102493108. URL `http://www.pnas.org/cgi/doi/10.1073/pnas.1102493108`.

A. Tavoni, A. Dannenberg, G. Kallis, and A. Loschel. Supporting Information to Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, 108(29):1–10, 2011b. ISSN 16136829. doi: 10.1073/pnas.1201800109.

Leigh Thompson and George Loewenstein. Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, 51(2):176–197, 1992. ISSN 07495978. doi: 10.1016/0749-5978(92)90010-5.

UNFCCC. Kyoto Protocol To the United Nations Framework. Technical report, UNFCCC, 1998. URL `http://unfccc.int/resource/docs/convkp/kpeng.pdf`.

UNFCCC. Fact sheet the Kyoto Protocol. Technical Report December 1997, 2011. URL `https://unfccc.int/files/press/backgrounders/application/pdf/fact_sheet_the_kyoto_protocol.pdf`.

United Nations. Adoption of the Paris Agreement. *Conference of the Parties on its twenty-first session*, 21932(December):32, 2015. URL `http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf`.

WRI. World Resources Institute: CAIT Climate Data Explorer, 2015. URL `http://cait.wri.org/`.

8 Appendix I

This section contains summaries of two studies that are important in the context of this thesis but have not directly influenced the experimental design. They are included for reasons of completeness and to provide additional information on factors not directly investigated in this thesis.

8.1 Altruism and Reputation

In 2006, Milinski et al. Milinski et al. (2006) published the results of a study investigating altruism and reputation in the setting of climate change mitigation. At the time, they stated that evidence from public goods games, aimed to find cooperative solutions to a tragedy of the commons, usually showed that the collective benefit was not produced in past experiments, Milinski et al. (2006, p. 3994). In many public good games, individuals are asked to invest in a pool that is then distributed between all players at the end of the game, yielding a return. They describe a standard example of a game with four players, each endowed with 10\$. Players are then asked to invest any percentage of their endowment into a public fund. The experimenter will double the total funds invested and divide the sum between all four players equally, regardless of their own contribution. The payoff function for individuals can be written as follows (adopted from a similar design by Fischbacher et al. (2001)):

$$\pi_i = 10 - g_i + 0.5 \times \sum_{j=1}^{4} g_j \tag{10}$$

π_i is the payoff, g_i the investment into the Pool, and $\sum_{j=1}^{4} g_j$ is the sum of the investments made by all players. The marginal benefit from investing 1\$ is therefore 0.5\$, which under standard economic rationality would lead to no investments because the marginal cost of investing (1\$) is greater than the marginal benefit (0.5\$).

At the same time, the payoff for each player could be doubled, if all players collaborated and invested $g_i = 10\$$. This would bring the pool to 40\$, which is then doubled and each player is assigned 20\$. However, a free-riding player in a cooperation scenario would end up with 25\$, if the three other players invested and he or she did not. This renders the game a prisoner's dilemma, where self-interest is at odds with a group's collective interest. Milinski et al. (2006) point to a study by Fischbacher et al. (2001) that shows consistent cooperation is not achieved in such a setting. In subsequent work, Milinski et al. (2008) also observed obstacles to cooperation in a setting that simulates climate negotiations.

The game designed by Milinski et al. (2006) asked individuals to contribute to a climate pool. In half of the rounds these contributions were nonanonymous and in the other half they were anonymous. Between anonymous and nonanonymous rounds, so called indirect reciprocity rounds were played to allow reputation building. The authors hypothesised that *because players would risk their reputation if they did not cooperate in a public goods game, that was alternated with the indirect reciprocity game; alternating rounds of these two games may induce cooperation in the public goods game* Milinski et al. (2006, p. 3995). The pool financed an ad in a newspaper that aimed to educate the broader public on the subject of climate change and contained calls to action for mitigation (e.g. turning down one's thermostat). In this case, the investments made would not result in any kind of financial return to the player. Half of the participants received information about climate change prior to the experiment while the other half was not. They are referred to in the study as "well-informed" and "little-informed".

8.1.1 A Reputation Game Design

Participants were divided into groups of six and each was assigned a pseudonym for the duration of the game. The game started with an indirect reciprocity round, with the aim to introduce reputation building into the game. Each reciprocity round had six subrounds of "donor-recipient interaction". In these subrounds, each of the six players once acted as a donor and once as a recipient. A donor could choose to give 1.50 € to the recipient. The experimenter matched the donation, resulting in a 3.00 € donation for the recipient and a 1.50 € loss for the donor. "Indirect reciprocity", in this context, meant that players could never directly punish "their" donor for not donating. If A was a donor to B, B would never be a donor to A. But all players were informed of whether an individual player, marked by a pseudonym, did or did not donate in each round.

The second round was a nonanonymous contribution round, where all players chose to donate 0, 1 or 2 € to the public climate pool. The value of their contributions were then made public to all players in their group. This was followed by another indirect reciprocity round, identical to the first reciprocity round, where all players knew of each player, whether he or she had donated in the first round and contributed to the climate pool in the second round. The fourth round was an anonymous contribution round, where players again chose to contribute 0, 1 or 2 € towards the climate pool. However, this time the other players were not informed of the decisions each player made. These four rounds constitute one cycle of the game, which was then repeated five times with a total of 20 rounds played.

8.1.2 Rational Reputation Effects and Irrational Altruism?

Figure 28 shows the average cooperation rate in each round of the game, distinguishing between little-informed and well-informed groups. Generally, well-informed groups contributed more often to the climate pool both in anonymous and nonanonymous rounds. The reciprocity rounds show no consistent differences in cooperation between the two groups. Well-informed groups showed a higher contribution rate in the nonanonymous climate pool rounds than in the reciprocity rounds. In anonymous rounds many players chose to contribute to the climate pool, even though they knew that their behaviour would not be communicated to the other players. The contribution rate in anonymous rounds shows a decreasing trend in both well-informed and little-informed groups, while the contribution rate in nonanonymous rounds is consistently high.

The authors then tested whether defectors in nonanonymous climate contribution rounds were punished in subsequent indirect reciprocity rounds. They found evidence for such punishment and conclude that *investments in sustaining the global climate are socially rewarded, and the refusal to do so is socially punished*(Milinski et al., 2006, p.3995).

Since individuals could choose to invest 0, 1 or 2 € to the climate pool, the contribution rate

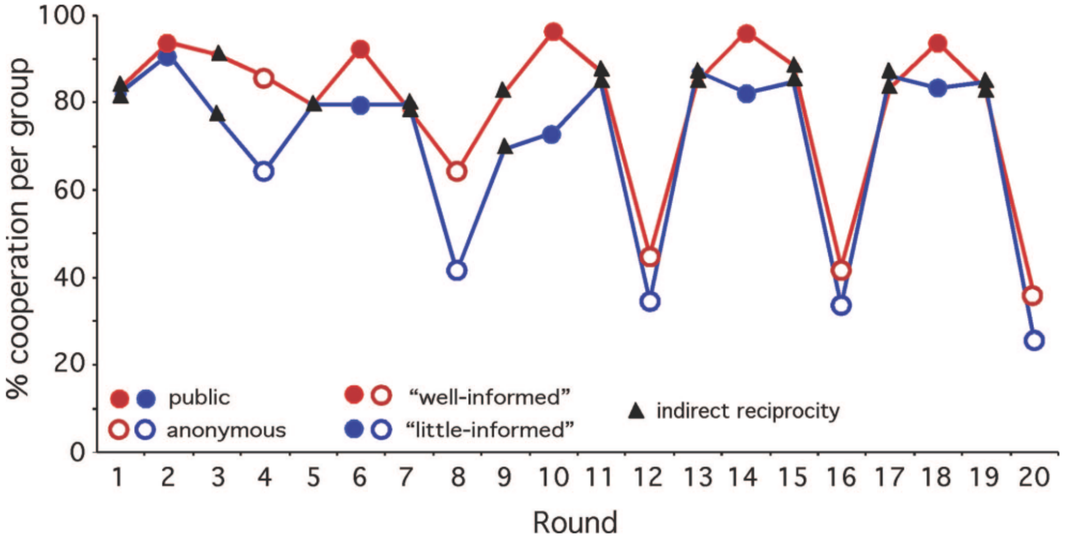


Figure 28: Percentage of cooperation (yes) per group of six subjects in each round of the climate public goods game (circles) and each round of the indirect reciprocity game (triangles). Rounds of the climate public goods game were either anonymous (open circle) or nonanonymous (filled circles). In one treatment (well-informed), the groups received additional expert information about the state of the global climate (red symbols), in the other treatment (little-informed); the groups received no additional information (blue symbols). Figure as in Milinski et al. Milinski et al. (2006, p.3995).

shown in Figure 28 does not show the strength of the commitment to contribute to a climate mitigation effort. Figure 29 allows for his analysis, showing the average group contribution to the climate pool in each round. With six players in each group, the maximum contribution in each round was 12 €.

While contribution rates in nonanonymous rounds remained constant, the value of the contributions decreased towards the end of the game in well-informed groups. This is not the case for little-informed groups as their contributions stabilize after an initial dip. In anonymous rounds there is a decreasing trend in both types of groups with the contributions converging towards the end of the game. At no point in the game did contributions reach zero.

Contributions were consistently higher in nonanonymous contribution, showing that reputation matters. Players that failed to build positive reputation were punished within the group when possible as described above. However, if a contribution is made to avoid punishment, it is no longer altruistic. It may even be rational in the traditional economic sense if the contribution leads to a reward in the following reciprocity rounds that exceeds the contribution.

Milinski et al. (2006) report that little-informed groups contributed an average of 14.55 € (yielding 29.1 € to the pool when doubled by the experimenter) in the five anonymous rounds of the public goods game. This corresponds to an average of 0.485 € per player per round, or 0.97 € if one considers the doubling by the experimenter. The authors regard this sum as "basically altruistic". The individuals that chose to contribute to the climate pool in anonymous rounds could not have expected to be rewarded for their behaviour in subsequent indirect reciprocity rounds and their investment would serve to educate the general public. Such behaviour is clearly inconsistent with traditional self-interested economic rationality.

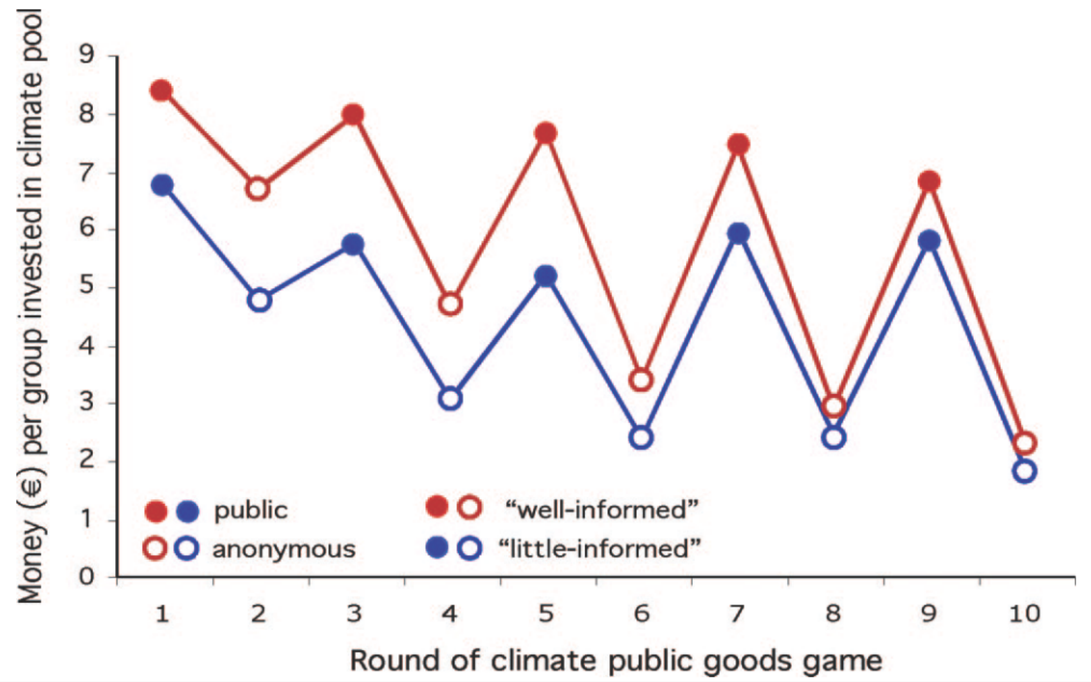


Figure 29: Money (€) per group of six subjects invested in climate pool in each round of the nonanonymous (filled) and anonymous (open) climate public goods game. In one treatment (well-informed), the groups received additional expert information about the state of the global climate (red); in the other treatment (little-informed), the groups received no additional information (blue). Milinski et al. (2006, p.3995).

8.2 Strategic Fairness

Equity is another factor in negotiating an international climate agreement that is often discussed. There is a broad literature covering equity in the context of climate change mitigation, discussing different approaches for fair burden sharing amongst nations. The Stockholm Environment Institute and EcoEquity (2015) even developed a climate equity reference calculator , where mitigation obligations for different nations can be explored under different fairness concepts. A number of studies have investigated so called strategic fairness (Carlsson et al., 2013; Dannenberg et al., 2010; Kesternich et al., 2014). The term implies that nations may choose the equity principle that is most advantageous for themselves in burden sharing negotiations, instead of the principle that they would objectively view as fair.

Brick and Visser (2014) conducted an interesting study that investigates strategic fairness with participants from different countries. Their sample includes participants from the US, China, India, the EU and South Africa. From each country, 51 students and 14 practitioners took part in the study. The participants were divided into groups of 5 individuals. Each group consisted of one representative of each of the five countries.

The experiment uses a threshold public good game, with the following payoff setup: (Brick and Visser, 2014, . 84)

$$\pi_i = \beta_i(y_i - c_i) + \alpha_i(g_j) \quad \text{if} \quad g_i \geq T \tag{11}$$

Each player had an endowment y_i in experimental currency units (ECU) and chose a contribution c_i to to the common account, denoted as g_i. If the pool exceeded a certain threshold T, each group member receives a return of α_i. If $g_j \geq$ T, g_j was multiplied by 1.5 and distributed equally between all 5 players. This translates to $\alpha_i = 0.3$. Any endowment not invested in the common account remained in the individuals private account, yielding a return of $\beta_i = 1$. Then Brick & Visser Brick and Visser (2014, p. 84) added a consequence for not reaching the threshold:

$$\pi_i = \beta_i(y_i - c_i)\lambda_i \quad \text{if} \quad g_i < t \tag{12}$$

where λ_i is < 1. λ_i reflects each nation's vulnerability to climate change and takes a value of 0.75 for the EU and United States and 0.5 for China, India and South Africa. This meant that the latter group of countries would lose 50% of their remaining endowment, while the EU and United States would only lose 25% if the threshold T is not met. Additionally, players did not receive the same initial endowments. Their endowments depend on the country they represent in the study, with the endowments reflecting the different income levels in those countries.

Brick & Visser frame a reduction goal by informing the participants that the US, EU, China, India and South Africa had agreed at a climate summit, to reduce their collective emissions to 62 MtCO2 by the year 2050, which corresponds to half of their collective emissions in 1990 (which is stated to be 124 MtCO_2e in Table 3 in Brick & Visser Brick and Visser (2014, p. 86)). It was stressed that no agreement had been made on how this emission reduction was to be achieved.

In reality, the actual collective pure CO_2 emissions of these countries in 1990 were around 12'400 MtCO_2 (according to WRI (2015) data available in 2015). Brick & Visser Brick and Visser (2014) appear to have used the pure CO_2 data and not total greenhouse gas emission data measured in CO_2e, provided in the WRI's CAIT tool. Perhaps the authors have used the smaller numbers for easier comparison by the study's participants, because the relative emissions of the countries are stated correctly, although the absolute values are off by a factor of 100 (i.e. US emissions are stated as 50 MtCO_2e when they were in fact 4'912 MtCO_2 and the US total greenhouse gas emissions were 5'744 MtCO_2e in 1990 according to CAIT WRI (2015)). Since the relative contributions are stated accurately, I assume that this deviation did not affect participant's behaviour as it is unlikely that they knew the actual absolute emission values by heart.

The participants were then shown a Table that showed the past emissions of countries and their projected emissions until 2030 as a point of reference.

8.2.1 Equity principles

The game deviates further from traditional game designs, because individuals could not allocate their endowments freely. Instead they had to choose their contribution according to one of four allocations that correspond to different equity principles common to the context of climate negotiations. The equity principles are defined as follows by Brick & Visser Brick and Visser (2014, p. 86):

1. **Equal per capita entitlement to emissions (egalitarian principle)** ***(EPC)***: If the population of your country represents x

2. **Equal percentage reduction of current emissions (sovereignty rule)** ***(EPR)***: If your country's current emissions amount to x% of global emissions, you should get x% of global emissions entitlements.

3. **Historical polluter-pays rule** ***(HPP)***: The abatement burden is allocated according to historical responsibility (1980– 2000). If your country's emissions between 1980 and 2000 amount to x% of global emissions in that time, you are responsible for x% of the reduction target.

4. **Future polluter-pays rule** ***(FPP)***: The abatement burden is allocated according to future responsibility (2010–2030). If your country's projected emissions between 2010 and 2030 will amount to x% of projected global emissions in that time, you are responsible for x% of the reduction target.

Each equity principle is linked to a respective contribution by players, depending on the country they represent. The game was framed such that the mitigation target (and thus the game's threshold) would always be achieved if all 5 players in a group chose the same equity principle (e.g. equal per capita emissions). The individual initial endowments of the players and their contributions under different equity principles are illustrated in Table 6 below. It is a simplified version of the information provided by Brick & Visser Brick and Visser (2014, p. 87, Table 4). In each Panel of the Table, the sum of ECU contributions is ≥ 77, ECUs which is defined as the minimum sum of investment to reach the threshold T. In the equal per capita (EPC) treatment, the sum of investments is 82 and thus substantially above the minimum threshold. This is because , in this treatment, India is entitled to increase its emissions until 2050, which must be compensated for by the other countries (see Brick & Visser Brick and Visser (2014, p.87).

8.2.2 Own Nationality vs. Random Nationality

Each player played two different rounds of the game, differentiated by the ON and RN treatments. The treatments are defined as follows (see Brick & Visser Brick and Visser (2014):

1. **ON-Treatment**: In the own nation (ON) treatment, players represent their own countries.
2. **RN-Treatment**: The random nation (RN) treatment randomly assigns a nation to each participant after they have chosen an equity principle.

Players had to choose an equity principle for both roles. In the RN treatment their payoff was that of the nation they were randomly assigned to. Brick & Voss Brick and Visser (2014) use these treatments to investigate the strategic use of equity principles. Since in the RN treatment, players did not know which nation they represented, any difference in behaviour from the same player between treatments points towards strategic fairness principle choice. If players apply different equity principles in the two treatments, they reveal that they do not

Panel A: equal per capita emissions (EPC)	ECU endowment	**ECU contribution**	% of endowment
US	70	**35**	50.0
EU	70	**21**	30.0
China	50	**25**	50.0
India	25	**0**	0.0
SA	15	**1**	6.7
Panel B: equal percentage of reduction of current emissions (EPR)	ECU endowment	**ECU contribution**	% of endowment
US	70	**25**	35.7
EU	70	**18**	25.7
China	50	**27**	54.0
India	25	**6**	24.0
SA	15	**1**	6.7
Panel C: historical polluter-pays (HPP)	ECU endowment	**ECU contribution**	% of endowment
US	70	**31**	44.3
EU	70	**25**	35.7
China	50	**15**	30.0
India	25	**4**	16.0
SA	15	**2**	13.3
Panel D: future polluter-pays (FPP)	ECU endowment	**ECU contribution**	% of endowment
US	70	**20**	28.6
EU	70	**14**	20.0
China	50	**34**	68.0
India	25	**8**	32.0
SA	15	**2**	13.3

Table 6: Contributions under application of different equity principles based on Brick and Visser (2014, p. 87)

chose the principles according to their pure fairness preference but that there is a strategic component to the selection. If the principles selected do not differ, the results would point towards a non-strategic (if not pure) fairness preference.

8.2.3 Evidence of strategic behaviour is not uniform

Figure 30 shows the burden sharing principles participants selected, grouped by nations. For each nation the ON and RN treatment are shown for comparison. Considering individual nations there are considerable differences between the two treatments. At the same time, all principles were selected by at least a few individuals from each country.

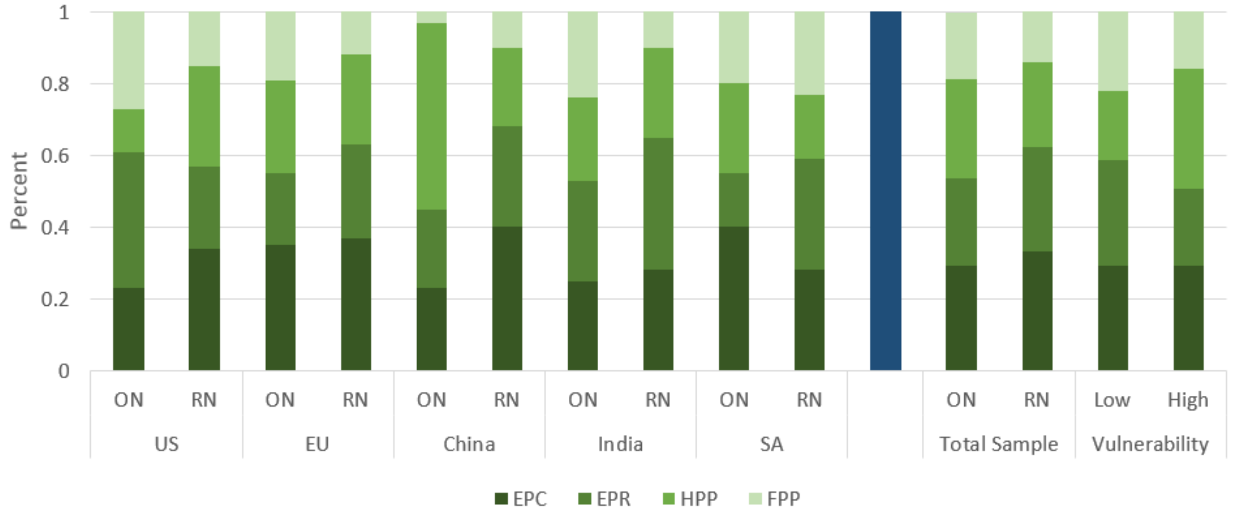


Figure 30: Burden sharing principles, by nationality, for ON and RN. Notes: ON: own nationality; RN: random nationality based on data provided in Brick & Visser Brick and Visser (2014, p. 90) and extended to include low and high vulnerability. For the vulnerability comparison, only data from ON treatments are included (where participant represent their own nationality). Low vulnerability contains US and EU (N=130), high vulnerability includes China, India and South Africa (N=195)

For each country an order of costliness of different equity principles can be made. This order only considers the ECU contributions and not the expected payoffs for the countries. The costs are ranked from highest to lowest by Brick & Visser Brick and Visser (2014, p. 87) as follows for each country:

1. US: EPC > HPP > EPR > FPP
2. EU: HPP > EPC > EPR > FPP
3. China: FPP > EPR > EPC > HPP
4. India: FPP > EPR > HPP > EPC
5. SA: HPP = FPP > EPC = EPR

There was a significant difference in the success rate of groups to reach the threshold between the ON and RN treatments. *In own nationality (ON), approximately 35% of groups met the target. This proportion increased significantly to 57% in RN when participants were to be allocated a* *nationality at random (McNemar Chi-square test: p = 0.022)* Brick & Visser Brick and Visser (2014, p. 92). They find that the greater success rate was primarily due to higher contribution by American and Chinese players in the RN treatment.

Chinese and American participants show clear differences in burden sharing principle choice between the treatments. Their choices of equity principle become much more similar in the RN treatment while they are quite different in the ON treatment (see Figure 30). Participants from China and the United States chose equity principles in line with material self-interest. Specifically Brick & Visser Brick and Visser (2014) state that participants from these nations opted for equity principles with higher costs to their own nation in the RN treatment, where they are randomly assigned to a nation for payoff. It can thus be argued that these players did not state their true fairness preferences in at least one of the treatments in the game and may have strategically deviated from their true preferences.

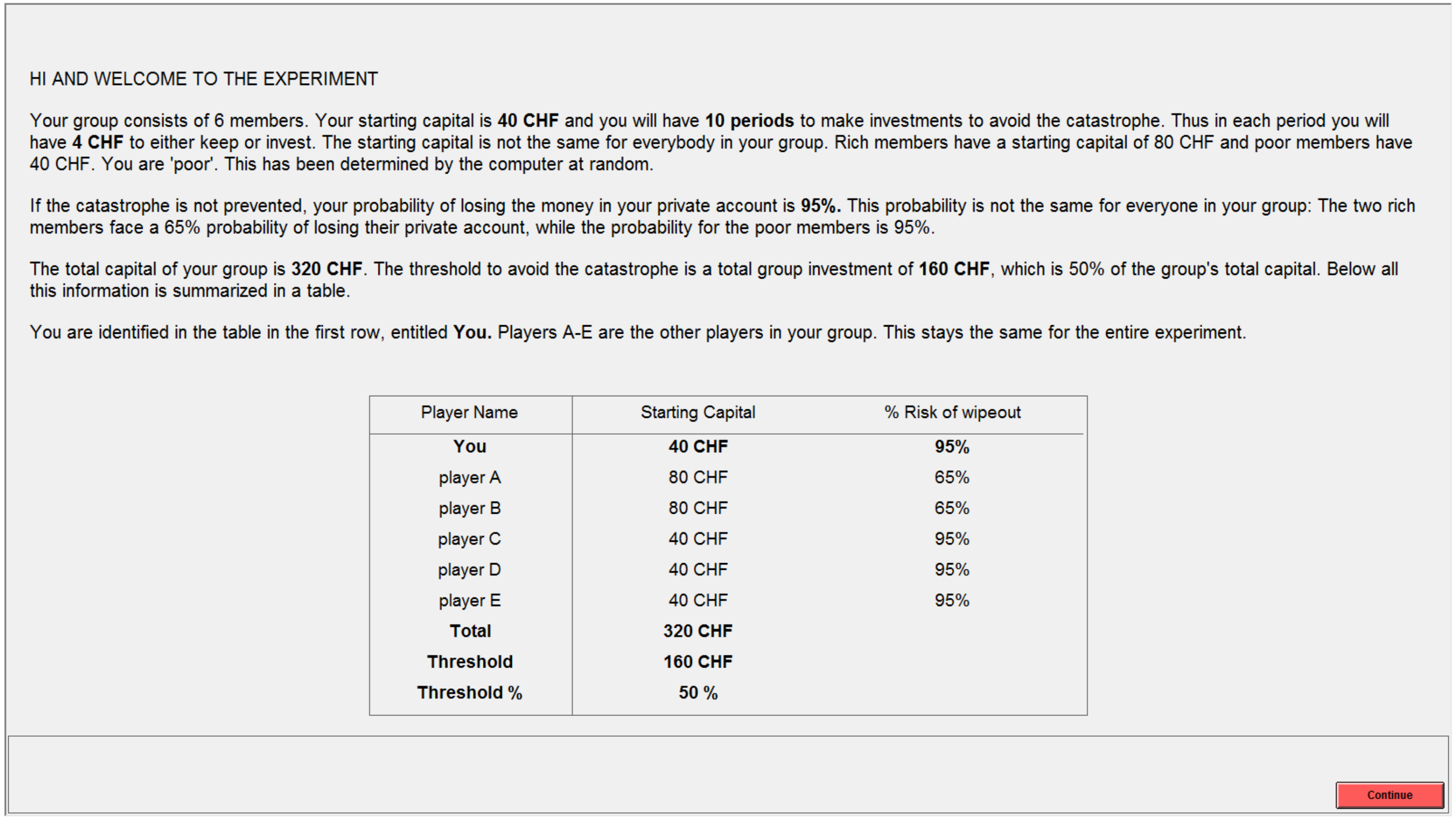
In the ON treatment American players most commonly selected the equal percentage reduction (EPR) or future polluter-pays principles (64% of participants selected one of those two rules). Looking at the cost ranking above, we see that those are the least costly options for US players and thus in line with material self interest. Brick & Visser Brick and Visser (2014)see the selection of the equal percentage reduction rule as consistent with the refusal of the US to ratify the Kyoto Protocol without a tit-for-tat contribution from emerging economies. In the random nationality (RN) treatment, the situation reversed: Now the majority of US players (62%) chose either equal per capita emissions (EPC) or historical polluter-pays (HPP)rule, which are most costly to players assigned to US-payoffs. Brick & Visser Brick and Visser (2014, p.94) argue that *American players are following a maximin strategy in this treatment, assuring themselves of the highest possible payoff in the advent of being allocated to either China or India.*

Chinese participants also acted according to material self interest in the own nationality (ON) treatment. Over half of them selected the historical polluter-pays (HPP) rule, while the second most popular rule was equal per capita emissions (EPC). Those are the least costly options for Chinese players. Brick & Visser Brick and Visser (2014) found this to be in line with BASIC countries' emphasis on equity and historical responsibility. In the random nationality treatment, Brick & Visser Brick and Visser (2014) observed a significant decrease in choice of the historical polluter-pays rule and a significant increase in the choice of the equal per capita emissions rule amongst Chinese players. One could argue that, in contrast to US players, we cannot speak of a preference reversal among Chinese players, but rather a shift from the most favourable to the second most favourable equity principle.

The same cannot be said about participants from the European Union. Statistical analysis by Brick & Visser Brick and Visser (2014) revealed that for players from this region, there was no significant difference in the choice of burden sharing principle. Additionally, the choice of burden sharing rules was also not in line with material self interest for European participants in the ON treatment. European players most commonly chose equal per capita reductions, followed by the historical polluter-pays rule. These are the most costly rules for European players in the ON treatment. Indian players did not show self-interested selection behaviour as they chose each principle with almost identical frequency in the ON treatment (see figure 30. For South African players the contributions do not differ greatly between the different burden sharing principles and their contributions matter little for the collective sum raised (see Table 6). It is therefore difficult to determine whether material self interest plays a role in their selection of burden sharing principles.

Vulnerability was not a focus of this study although differing vulnerability levels were introduced in the study design. I have therefore included a vulnerability distinction in Figure 30. The most obvious difference is that players exposed to higher vulnerability preferred the HPP over EPR and FPP, whereas low vulnerability players did the opposite. The frequencies with which EPC was chosen is practically equal for both types.

9 Appendix II



HI AND WELCOME TO THE EXPERIMENT

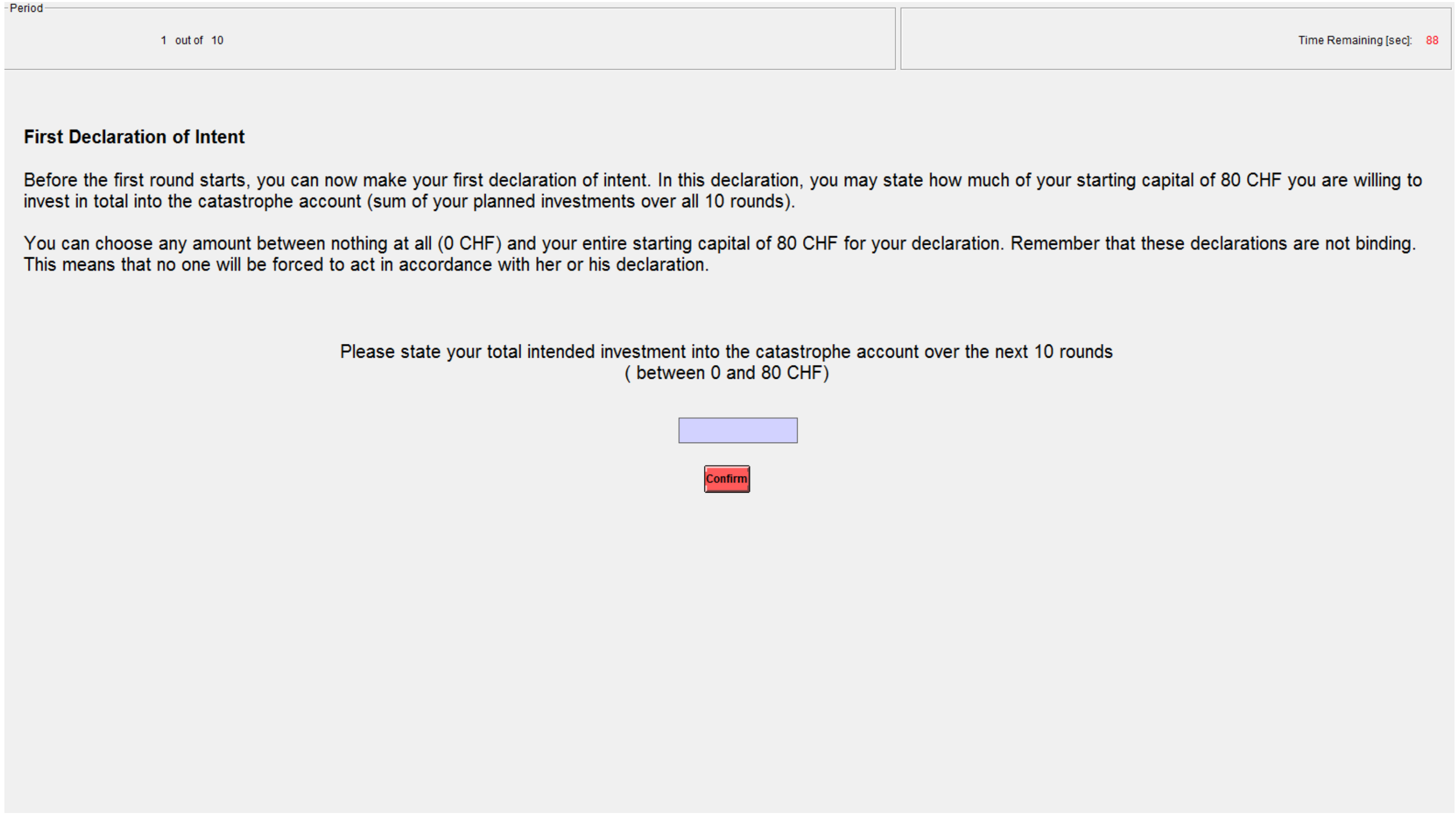
Your group consists of 6 members. Your starting capital is **40 CHF** and you will have **10 periods** to make investments to avoid the catastrophe. Thus in each period you will have **4 CHF** to either keep or invest. The starting capital is not the same for everybody in your group. Rich members have a starting capital of 80 CHF and poor members have 40 CHF. You are 'poor'. This has been determined by the computer at random.

If the catastrophe is not prevented, your probability of losing the money in your private account is **95%.** This probability is not the same for everyone in your group: The two rich members face a 65% probability of losing their private account, while the probability for the poor members is 95%.

The total capital of your group is **320 CHF**. The threshold to avoid the catastrophe is a total group investment of **160 CHF**, which is 50% of the group's total capital. Below all this information is summarized in a table.

You are identified in the table in the first row, entitled **You.** Players A-E are the other players in your group. This stays the same for the entire experiment.

Player Name	Starting Capital	% Risk of wipeout
You	**40 CHF**	**95%**
player A	80 CHF	65%
player B	80 CHF	65%
player C	40 CHF	95%
player D	40 CHF	95%
player E	40 CHF	95%
Total	**320 CHF**	
Threshold	**160 CHF**	
Threshold %	**50 %**	

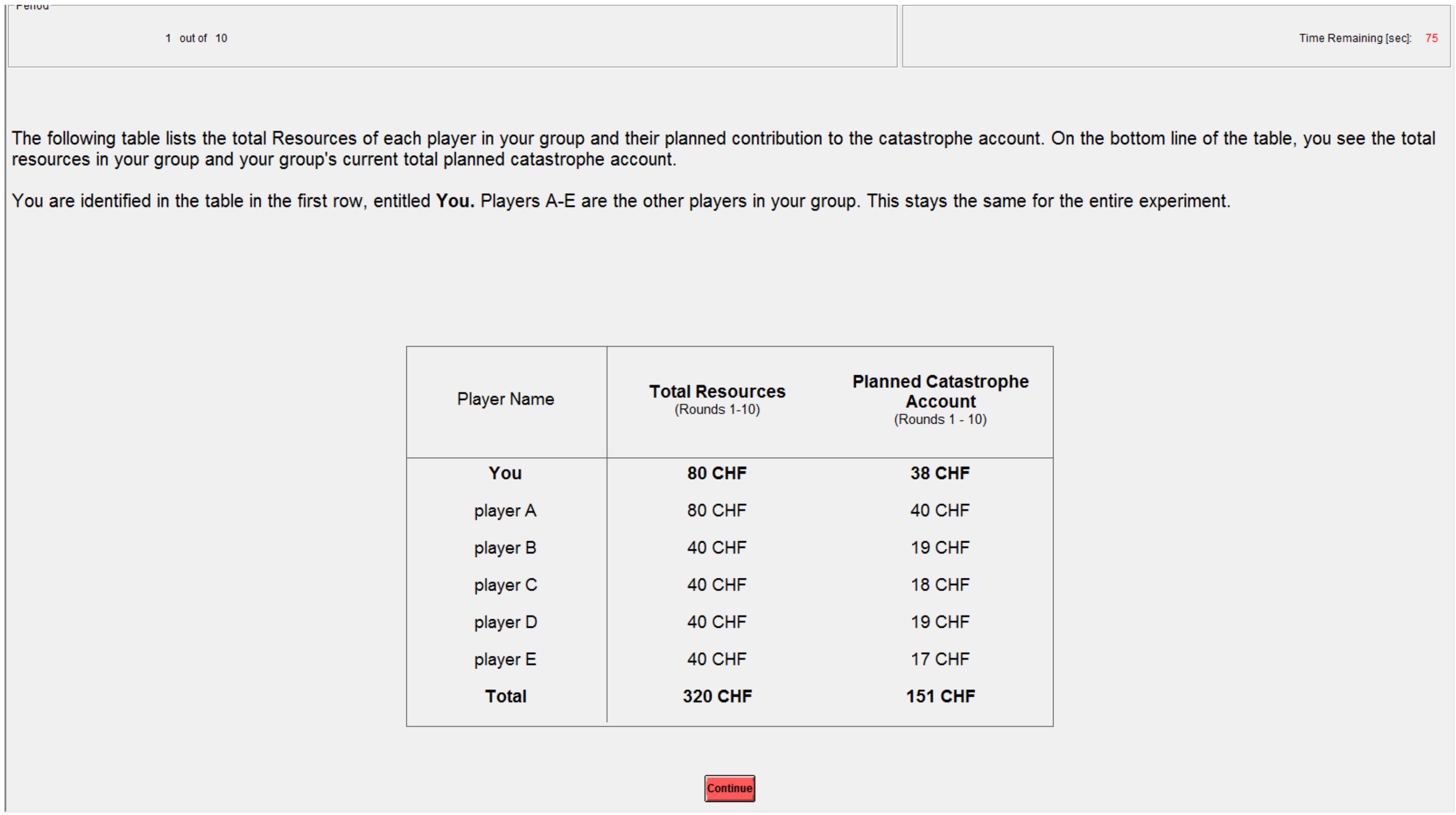


First Declaration of Intent

Before the first round starts, you can now make your first declaration of intent. In this declaration, you may state how much of your starting capital of 80 CHF you are willing to invest in total into the catastrophe account (sum of your planned investments over all 10 rounds).

You can choose any amount between nothing at all (0 CHF) and your entire starting capital of 80 CHF for your declaration. Remember that these declarations are not binding. This means that no one will be forced to act in accordance with her or his declaration.

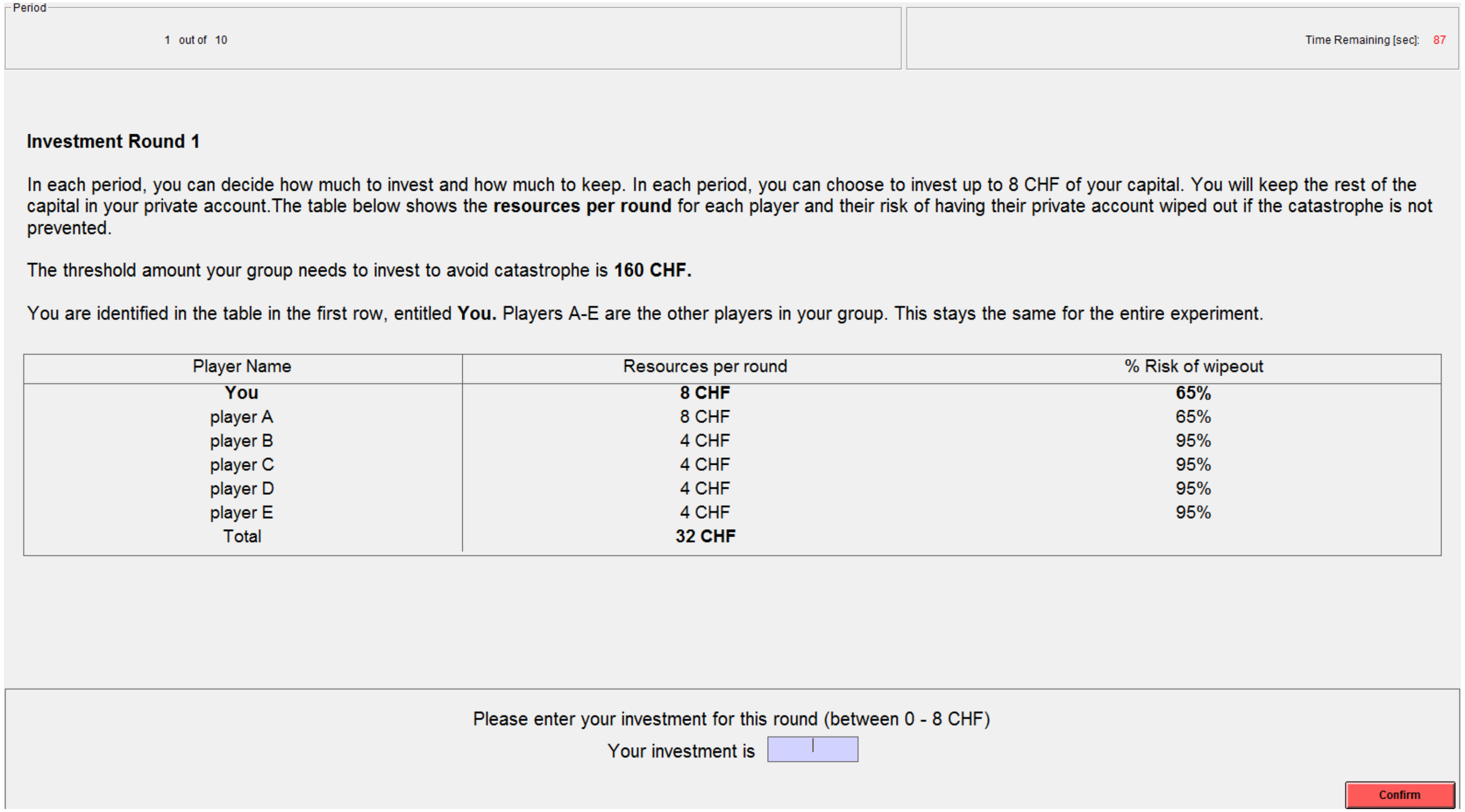
Please state your total intended investment into the catastrophe account over the next 10 rounds
(between 0 and 80 CHF)



The following table lists the total Resources of each player in your group and their planned contribution to the catastrophe account. On the bottom line of the table, you see the total resources in your group and your group's current total planned catastrophe account.

You are identified in the table in the first row, entitled **You.** Players A-E are the other players in your group. This stays the same for the entire experiment.

Player Name	**Total Resources** (Rounds 1-10)	**Planned Catastrophe Account** (Rounds 1 - 10)
You	**80 CHF**	**38 CHF**
player A	80 CHF	40 CHF
player B	40 CHF	19 CHF
player C	40 CHF	18 CHF
player D	40 CHF	19 CHF
player E	40 CHF	17 CHF
Total	**320 CHF**	**151 CHF**



Investment Round 1

In each period, you can decide how much to invest and how much to keep. In each period, you can choose to invest up to 8 CHF of your capital. You will keep the rest of the capital in your private account.The table below shows the **resources per round** for each player and their risk of having their private account wiped out if the catastrophe is not prevented.

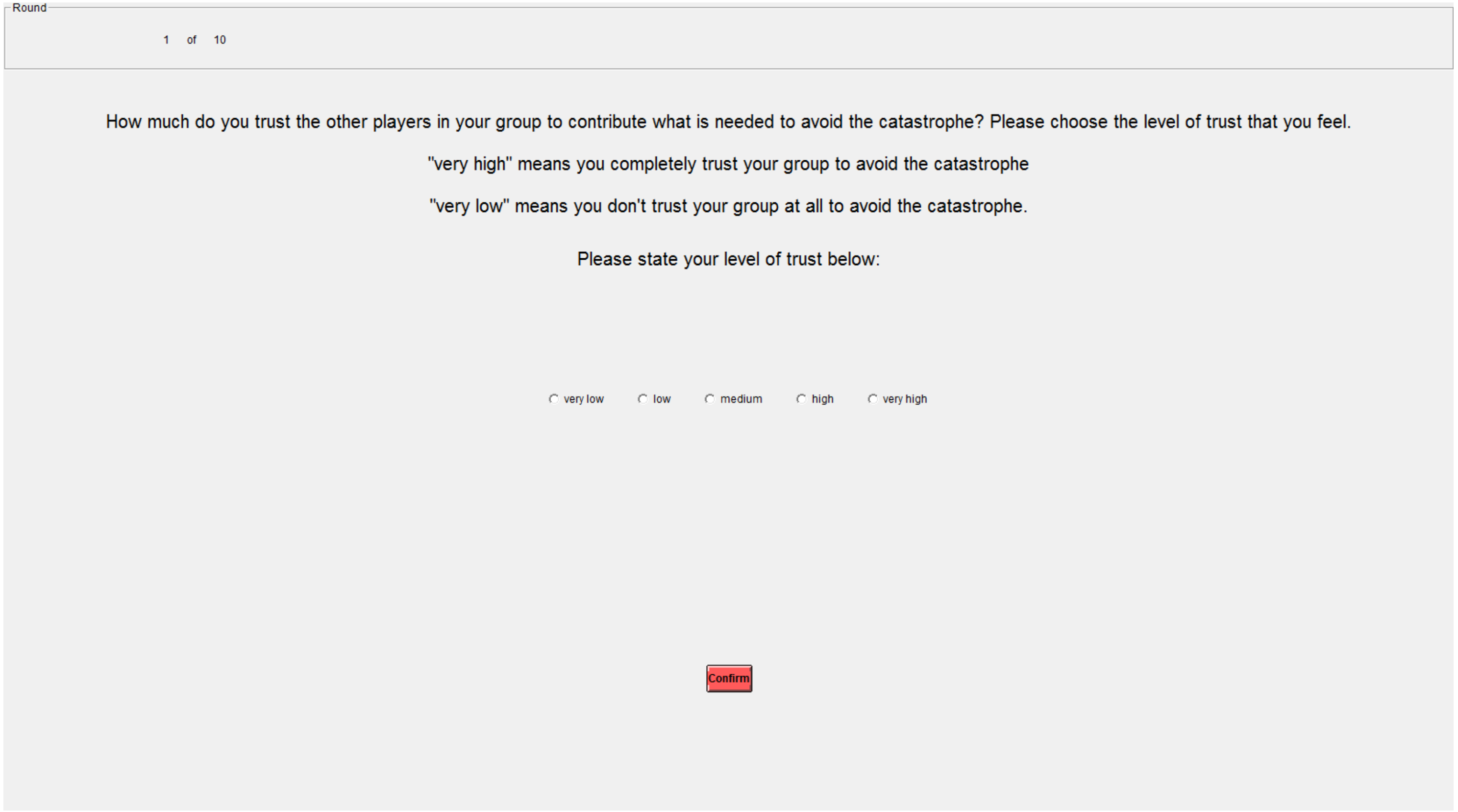
The threshold amount your group needs to invest to avoid catastrophe is **160 CHF.**

You are identified in the table in the first row, entitled **You.** Players A-E are the other players in your group. This stays the same for the entire experiment.

Player Name	Resources per round	% Risk of wipeout
You	**8 CHF**	**65%**
player A	8 CHF	65%
player B	4 CHF	95%
player C	4 CHF	95%
player D	4 CHF	95%
player E	4 CHF	95%
Total	**32 CHF**	

Please enter your investment for this round (between 0 - 8 CHF)

Your investment is



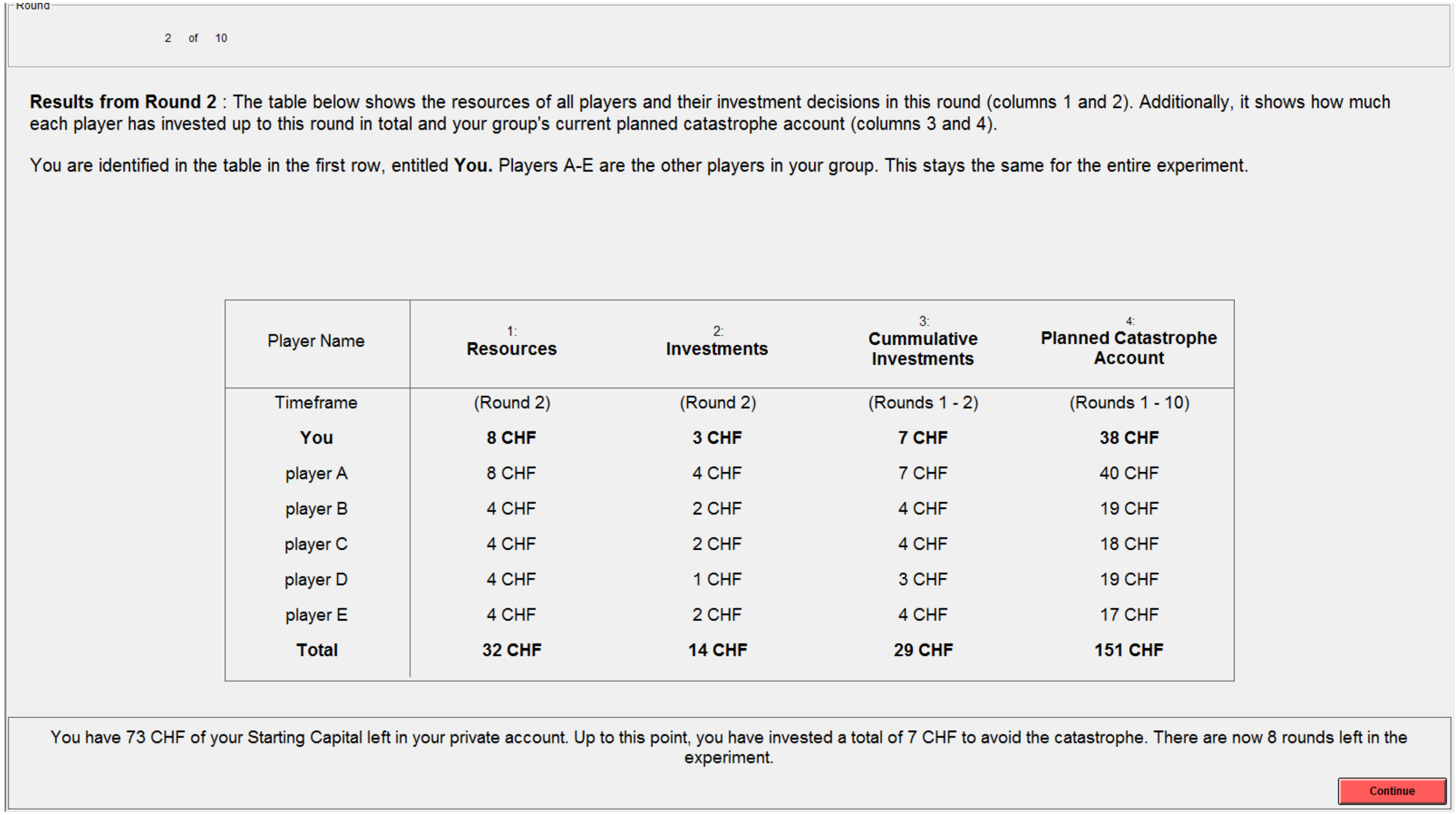
How much do you trust the other players in your group to contribute what is needed to avoid the catastrophe? Please choose the level of trust that you feel.

"very high" means you completely trust your group to avoid the catastrophe

"very low" means you don't trust your group at all to avoid the catastrophe.

Please state your level of trust below:

○ very low ○ low ○ medium ○ high ○ very high



Results from Round 2 : The table below shows the resources of all players and their investment decisions in this round (columns 1 and 2). Additionally, it shows how much each player has invested up to this round in total and your group's current planned catastrophe account (columns 3 and 4).

You are identified in the table in the first row, entitled **You.** Players A-E are the other players in your group. This stays the same for the entire experiment.

Player Name	1: **Resources**	2: **Investments**	3: **Cummulative Investments**	4: **Planned Catastrophe Account**
Timeframe	(Round 2)	(Round 2)	(Rounds 1 - 2)	(Rounds 1 - 10)
You	**8 CHF**	**3 CHF**	**7 CHF**	**38 CHF**
player A	8 CHF	4 CHF	7 CHF	40 CHF
player B	4 CHF	2 CHF	4 CHF	19 CHF
player C	4 CHF	2 CHF	4 CHF	18 CHF
player D	4 CHF	1 CHF	3 CHF	19 CHF
player E	4 CHF	2 CHF	4 CHF	17 CHF
Total	**32 CHF**	**14 CHF**	**29 CHF**	**151 CHF**

You have 73 CHF of your Starting Capital left in your private account. Up to this point, you have invested a total of 7 CHF to avoid the catastrophe. There are now 8 rounds left in the experiment.

Figure 31: Screenshots showing the user interface associated with each stage of the game in order of appearance: Group Information, Pledge Stage, Planned Catastrophe Account, Contribution Entry, Trust Measurement, Contribution Display. See figure 15 for a overview of the different stages.

Declaration

under Art. 28 Para. 2 RSL 05

Last, first name: Bebié, Remo Andreas

Matriculation number: 09-709-122

Programme: MSc in Climate Sciences with special qualification in Economics
Bachelor □ Master ☒ Dissertation □

Thesis title: Insights from Behavioral Economics on Climate Negotiations

Thesis supervisor: Prof. Dr. Gunter Stephan
Prof. Dr. Ralph Winkler

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgement has been made in the text. In accordance with academic rules and ethical conduct, I have fully cited and referenced all material and results that are not original to this work. I am well aware of the fact that, on the basis of Article 36 Paragraph 1 Letter o of the University Law of 5 September 1996, the Senate is entitled to deny the title awarded on the basis of this work if proven otherwise. I grant inspection of my thesis.

Zürich, 20.3.2016

R. Bebié
....
Signature