University of Technology Sydney

Economics Discpline Group

Cooperation among equal-minded subjects in a dynamic collective social dilemma

Experimental Design Proposal

by Anna-Mateja Delas

Contact details:	annamateja.delas@gmail.com
	+41 632 206 865
Submission date:	April 20, 2017
Course:	Experimental Economics
Lecturer:	Dr. Jingjing Zhang
Semester:	Autumn 2017

Contents

1 Motivation

Climate change is often stated to be the most complicated environmental problem society has ever faced. The associated complexity of the problem is due to the large asymmetry between countries in terms of economic development and historical emissions. Moreover, future damages of a warmer planet can not well be predicted since the global carbon cycle, future emissions, and technology improvement are highly uncertain.

Climate negotiations, organised by the United Nations Framework Convention on Climate Change (UNFCCC), have so far not realised what they were meant to have and total annual emissions continue to rise exponentially indicating global warming to be well around 4-6° Celsius (by overshooting the atmospheric carbon stock at 550 ppm being agreed upon) [IPCC, 2014].[1] The latest agreement, adopted in Paris in December 2015, shows no overall progress and is based on the assumption that high polluters (such as China, the US, India, Brazil, or Canada) will curb their carbon emissions without any binding mechanism for enforcement for measuring and controlling CO_2 (neither firm- nor country-wide), and without any substantial sanction system or monetary pressure (e.g. carbon tax or permit) in place.[2] As countries exhibit different levels of willingness to abate global warming, which in turn is part of their position, there has recently been progress by opening up new negotiation tables, which involve countries with a higher similarity in their positions.[3] Such smaller negotiation coalitions consisting of more equally-minded players are recently referred to as *climate clubs* (e.g. Victor [2015]).[4] In contrast to e.g. the Annex and Non-Annex parties of the Kyoto Protocol, climate clubs are seen to contain more „economic glue“ to hold a cooperative coalition together [Nordhaus, 2015]. Biermann et al. [2009] name in this regard four types of benefits resulting from negotiations conducted in clubs; these involve speed (faster negotiations), higher ambition („narrow but deep“) versus („broad but shallow“), easier participation (fewer barriers), and equity advantages („regional“ solutions).[5]

Leaning on such observations the idea for the following paper is to test effectiveness in achieving a critical level of social investments in order to avoid a public bad for *willing* and *non-willling* subjects in an economic laboratory experiment.[6] My research question is thus stated as follows: *Will a separation of agents into willing and non-willing sets make it more likely to achieve a certain threshold of a public good under a collective social dilemma*

[1] Stiglitz [2010] in this regard famously noted that pretty speeches can only bring you so far".

[2] < *http://unfccc.int/paris_agreement/items/9485.php* >

[3] These new, smaller negotiation strands can be imagined as subsets of the annual meetings of all parties under the Kyoto Protocol. E.g. < *http://unfccc.int/key_documents/bali_road_map/items/6447.php* >.

[4] If thinking of one single sub-coalition only, Dessler and Parson [2010] suggest another term; *the coalition of the willing.*

[5] Such advantages might in the aggregate thus overweigh the universality advantage of the Kyoto Protocol (i.e. sub-set versus entire set of parties).

[6] The adjective "willing" (or "determined") in this regard means whether an individual is willing to achieve a public threshold beforehand.

setting? Thus, the novelty and aim of this question is to find whether the dimension of intrinsic individual willingness to avoid a public bad can act as a cooperation incentive, and moreover recognise the delicate but highly powerful feature of intrinsic motivation if matched with the right group environment. In particular, it is predicted that if players are split into sub-sets consisting of like-minded (equally-minded) subjects, they are more likely to achieve the threshold.[7]

2 Related literature

Climate change in game-theoretical terms, often known as global emissions game, is recognised to be a prisoner's dilemma game with additional coordination features of the type of a chicken game [Barrett, 2003]. However, translating the prisoner's dilemma in this context into a public goods game is misleading, since social marginal returns are not truly present.[8] Moreover, it is not implicitly about realising a certain level of a public good but rather about avoiding a public bad if a necessary level of social investments is achieved. In particular, thresholds represent dangerous tipping points, such as the collapse of the thermohaline circulation or an accelerated melt-down of polar ice caps, which would trigger discontinuous and irreversible acceleration in climate change [Lenton et al., 2008]. To capture such a payoff structure the game analysed in this paper is very similar to the one used in Milinski et al. [2008] who set up a model to test a *collective-risk social dilemma*. In their game, investments into the public good (i.e. avoidance of disastrous climate change) are lost, meaning the social marginal per capita return on collective investments (in the following denoted by a) is set zero.[9] Above that the entire money is at risk if a critical

[7] Willing subjects would in this sense lose their motivation to contribute if they are placed in a set where free-riders dominate. Similar to the hawk-dove game, which is a game analysed in evolutionary dynamics in game theory, it is believed that the prevailing dynamics would lead to an extinction of doves and thus resulting in a population consisting of hawks (free-riders) only. Translated to my research question this could explain why intrinsic motivation might decline if subjects play with many free-riders.

[8] One could argue at this point that the case of no-refunding returns is too strict. Imagine the Netherlands building an embankment to protect itself from climate-induced flooding. If it fails to provide the necessary level of embankment, investments so far do not get lost (the dam is only halfly built). However, I do not investigate the possibility to insure oneself against climate change (i.e. climate adaptation). In here the idea is to achieve a necessary level of atmospheric carbon stock such that a dangerous tipping point - a climate catastrophe - does not occur [Lenton et al., 2008]. One can also think that the costs of not managing to provide the threshold are infinitely large such that payoffs get zero for everyone. The global emissions game can be thought of a function such as $\pi_i = B(e_i) - e(a_i) - D(E)$ where the individual profit of a an agent i (e.g. country, firm) can be thought as of benefits resulting of individual emissions less abated individual emissions (which come at an individual abatement cost (a_i)) less the damage induced by global emissions E.

[9] Their dilemma game is special in that all investments into the public good (a „climate / group account") are lost, and differs from the commonly used payoff function in the prisoner's dilemma where the payoff P_i (being dependent on i's contribution to the public good $x_i \in [0, e_i]$ and the sum in the group account determined by $\sum_{j=1}^{N} x_j = x_1 + ... + x_i + ... + x_N$) is stated as $P_i(x_i, x_j) = e_i - x_i + a \sum_{j=1}^{N} x_j$ subject

threshold of total social investments is not achieved.[10] This is different to pure *threshold public goods games*, which have been investigated with $a > 0$ and no risks attributed to the private endowments in case of failing to meet the threshold (e.g. Cadsby and Maynes [1999], Croson and Marks [2001]).[11] Very few groups in Milinski et al. [2008] achieve the threshold (in fact only 50% of all treatment groups achieve the threshold under 90% chance of losing everything). Milinski et al. [2008] played the game in groups of 6 people over 10 periods and with initial individual endowments of € 40. Every player could contribute € 0 (free-riding strategy), 2 (fair strategy), or 4 (altruistic strategy) to the group account. If at round 10 the threshold of € $120 = \bar{X}$ was not achieved (i.e. if everybody played the fair strategy on average), then $P_i = 0$ with a certain probability. The tipping point was assigned with different probabilities varying from 0.1, 0.5, to 0.9, that account for the uncertainty in the occurrence of the environmental disaster. Investments into the climate account happened anonymously, and after each round the strategies of all subjects were communicated. Expected payoffs were thus given as follows (1):

$$E(P_i(x_{it})) = \begin{cases} (1-\pi)(40 - \sum_{t=1}^{10} x_{it}) + \pi * 0 & if\ \bar{X} < 120 \\ 40 - \sum_{t=1}^{10} x_{it} & if\ \bar{X} \geq 120 \end{cases}$$

$$s.t.\ x_{it} \in \{0, 2, 4\},\ \pi \in \{0.1, 0.5, 0.9\},\ \bar{X} = \sum_{t=1}^{10} \sum_{j=1}^{6} x_{jt} \tag{1}$$

Milinski et al. [2008] conclude that the higher the probability π, the more likely the threshold $\bar{X}$ is achieved.[12]

Variations of the Milinski et al. [2008] experiment include the analysis of the effect on achieving a sufficient amount in the climate account if not only the probability distribution π is known, but $\bar{X}$ either [Danneberg et al., 2011], adding very high ambiguity to the climate change game. In fact, since the earth's tipping points are associated with a great amount of uncertainty (Lontzek et al. [2015]), introducing stochastic rather than

to $a > 0$, where e_i indicates i's initial endowment and a is the refunding factor with $a > 1 > \frac{1}{N}$, and thereby suggesting that all individuals would be better off if they contributed to the public good (the sum of all contributions of N players), i.e. played by the non-dominant strategy, and thus the socially better outcome be realised. Milinski et al. [2008] assume $a = 0$.

[10] The uncertainty component in Milinski et al. [2008] is defined in such that the value on the private account is at risk if a critical threshold $\bar{X}$ (a tipping point) is not achieved, thus $P_i(x_i, \bar{X})$.

[11] Note here that while *public goods games* emphasise more on the provision of a public good with possibly positive marginal social returns, *collective social dilemma games* are about the avoidance of a public bad with no returns. At this point it is worth noting that the emissions game can also find applications in labour economics; e.g. if team members fail to reach a target (finishing a prototype, deliver work on time, etc.), all accumulated investments become lost (getting sacked, etc.).

[12] For $\pi = 0.9$ only 50% of all groups achieved $\bar{X}$, whereas for π=0.5 only 10% groups, and for π=0.1 none. Milinski et al. [2008] conclude that if individuals are sufficiently informed about π and its impacts then $\bar{X}$ is more likely to be achieved (suggesting that provision of information matters highly (i.e. "informing people" [Milinski et al., 2011: 2294])) (i). Further, there is also a positive effect if these impacts are sufficiently known (if π is known) (ii).

deterministic thresholds captures another dimension of the emissions game. Dannenberg et al. [2011] find that the uncertainty about $\bar{X}$ has a significant negative effect on preventing a collective damage, whilst unknown probability distributions yield mixed results.[13] Tavoni et al. [2011] analyse inequality and communication with π=0.5 and conclude that inequality in initial individual endowments e_i has a negative impact on reaching $\bar{X}$. Communication on the other hand yields very desirable results. Burton-Chellew and May [2013] model both inequality in e_i and different risks of climate vulnerability. They conclude that rich subjects invest less in the avoidance of a climate catastrophe since they are less exposed to climate change (i.e. the richer, the safer [Wildavsky [1989]).

My study is also closely related to studies on group composition effects such as Gächter and Thöni [2005] and Burlando and Guala [2005] who test cooperation amongst like-minded people in public goods experiments. Gächter and Thöni [2005] for instance conclude that higher cooperation is maintainable within groups of similar cooperative attitudes compared to randomly composed groups.

3 Experimental design

The experiment consists of three stages. All stages are conceptualised as collective social dilemma games (i.e. including a threshold). The first stage aims to identify a subject's type and whether she is willing to contribute to prevent a public bad in a one-shot game. Second and third stages are constructed as multiple-period games, either being played by four equally-sized subgroups or by all players of a group (i.e. no subgroups). In all stages the amount of subjects playing simultaneously together is twelve. The collective social dilemma captures lost investments (i.e. the marginal social return is set zero in a threshold public goods game), the sure outcome that everyone loses everything if the threshold is not achieved (contrary to Milinski et al. [2008]), as well as an inherent coordination problem. The experimental game is framed neutrally (see instructions) without using any words associated to climate change or destruction of the environment. Even though the motivation for this study originates from observing recent climate change politics, allowing a neutral frame is believed to gain more general insights on cooperation and human behaviour.

First stage (one-shot game)

Similar to Gächter and Thöni [2005], I include a one-shot game in order to determine the individual contribution types (i.e. to „extract“ the degree of ex ante willingness inherent to each subject). The information on individual contribution willingness is used for the latter stages to assign subjects into subgroups. In this stage, each player has the

[13] Lenton [2008] however state that tipping points are unknown, implying their probability distributions to be even more unknown. Weitzman [2011] proposes tipping points to have grave impacts if they happen, but imply small probability to occur ("fat tails - low probability").

possibility to contribute a number $x_i \in \{0,\ 0.5,\ ...,\ 3.5,\ 4\}$ of her initial endowment being four monetary units to a group account. Thus the decision is how much of the initial endowment the subject wants to leave on her private account and how much she wants to invest in avoiding to lose everything. The subjects thus face the following payoff:

$$P_i(x_i) = \begin{cases} 4 - x_i & if\ \bar{X} \geq 30 \\ 0 & if\ \bar{X} < 30 \end{cases}$$

$$\text{subject to } x_i \in \{0, 0.5, ..., 3.5, 4\},\ \bar{X} = \sum_{j=1}^{12} x_j,\ j \in 1, ..., i, ..., 12 \tag{2}$$

Equation 2 reflects that the one-shot payoff depends on own contribution levels as well as those of eleven other members. On average that would mean that every player contributes 2.5 tokens to achieve the threshold being set at 30 tokens. Moreover, to reflect a coordination dimension, if at least three people coordinate at the contribution levels 3.5 or 4, then their invested contributions are doubled towards $\bar{X}$ creating an incentive to cooperate with highest contribution levels. This not only reflects the climate change game to exhibit additional features of a coordination game, but also compensates for dropping the variable of a positive marginal social return. Deception is excluded by letting the subjects know that their behaviour in this stage will affect the subgroup arrangement in the next stages. Moreover, I seek to avoid possible experimenter demand effects, in case subjects would understand the logic behind subgroup assignment.

Second/third stage (ten-period game with/without subgroups)

In the second and third stage subjects play a ten-period collective social dilemma game. The difference between both stages is the group composition being either four subgroups (of three people each) or the entire group (twelve people). At the beginning of each stage each individual is endowed with 40 monetary units on her private account. In every period each player makes an allocation decision between the individual and the group account. The individual contribution possibilities in each period are the same as in stage one. If subjects do not reach the predefined threshold level after the last period, they get a payoff of zero. The threshold is set at 300 tokens, which means that every period an average contribution of 2.5 tokens is needed to reach the threshold. Equation 3 gives the payoff after the last round.

$$P_i(x_{it}) = \begin{cases} 40 - \sum_{t=1}^{10} x_{it} & if\ \bar{X} \geq 300 \\ 0 & if\ \bar{X} < 300 \end{cases}$$

$$\text{subject to } x_{it} \in \{0, 0.5, ..., 3.5, 4\},\ \bar{X} = \sum_{t=1}^{10} \sum_{j=1}^{12} x_{jt},\ j \in 1, ..., i, ..., 12 \tag{3}$$

Analogously to stage one, there exists the possibility of doubling individual contributions in the group account if at least three people manage to contribute 3.5 or 4 tokens.

Table 0: Sessions to be conducted

Sessions	1	2	3	4	5	6
Second stage	Baseline	Type	Baseline	Random	Random	Type
Third stage	Type	Baseline	Random	Baseline	Type	Random

In one stage all twelve players of the group play together *(BASELINE-treatment)*. In the other stage players are divided into four equally-sized new subgroups either randomly *(RANDOM-treatment)* or according to their cooperation type from stage one *(TYPE-treatment)*. While the *BASELINE-treatment* acts as a control treatment for whether subgroups in general matter, the *RANDOM-treatment* acts as a control for whether homogeneous subgroups of similar cooperative attitutudes matter. In four out of six sessions one of the two stages consists of the *BASELINE-treatment*. In half of all sessions the other stage consists of the *RANDOM-treatment*, in the other half it consists of the *TYPE-treatment*.

The experiment consists of six experimental sessions (table 0) involving twelve individuals participating in one session at a time (within subjects design). One drawback of the within subjects design is that individuals may experience learning effects across the two stages. In order to prevent this undesirable effect, the re-shuffling of treatments allows to minimise this effect by statistically measuring whether learning plays a role.

The reason of using a within subjects design is to examine individual behaviour in different environments (built by others' contribution types). Since contribution behaviour is individual - though highly influenceable by others - by having the same subject participating in two out of three treatments the effect of the others' influence on the same subject can be isolated while holding the individual contribution willingness constant provided that stage one manages to successfully extract this inherent feature.

Subjects are matched into four subgroups in the *TYPE-treatment*, i.e. one *VHIGH-s* (highest quarter of contributors from stage one), one *HIGH-s* (second highest quarter of contributors from stage one), one *LOW-s* (second lowest quarter of contributors from stage one), and one *VLOW-s* (lowest quarter of contributors from stage one). In case of ties (i.e. in case that the amount of people in one sub-treatment is exhausted for individuals with same contribution levels from stage one, a die determines to which closest subgroup they are matched to).

In the *BASELINE-treatment* subjects are given feedback on the average contribution and cumulative total contributions towards $\bar{X}$ after every period. In the *TYPE-* and *RAND-treatements* they see not only the average contribution and cumulative total contributions but also the the individual contributions in their subgroup. In the *TYPE-treatment* and *RAND-treatment* subjects are moreover informed about the efforts of the other three members in their subgroup from stage one (a similar approach was used by Gächter and

Thöni [2005]). The underlying argument is that individual behaviour should be influenced only by her subgroup but at the same time is not explicitly made aware of which subgroup she is assigned to. However, subjects are informed about the contributions of the other members of the subgroup, so they can form a priori expectations of their contribution type.[14]

Details

72 undergratuate students (12 subjects * 6 sessions) from the University of Technology Sydney are tested.[15] This allows for 1'152 observations (72 students* 21 periods).

All three stages are paid in case the threshold is achieved. The first stage is paid according to what is left on the private account. Since this reflects a one-shot game the payoff will be quintupled (by stating that every token left on the private account is worth 5 AUD). To simplify the payment procedure, I define a ranking function for the second and third stage. Subjects are ranked according to the amount on their individual account after the final round. This leads to a discrete ranking function $R_i(P_i)$ (with $R_i'(P_i) > 0$). To each rank a certain payoff is assigned. The highest rank (1) receives 24 AUD, the second-highest rank (2) paying 2 AUD less yielding 22 AUD, and so on, with the lowest rank (12) getting 2 AUD (see instructions). In case of a tie, affected subjecs are paid an average payoff. If the threshold is not achieved, subjects are only given a show-up fee amounting to 10 AUD.

Since the experimental game is structurally very simple, no practice test is required beforehand. At the end of the game subjects are given a questionnaire to fill out including questions about their pro-social attitutes and socio-economic characteristics.

The experiment is conducted in z-tree [Fischbacher, 2007] and *Google forms*.[16]

4 Hypotheses

Five hypotheses (formulated as predictions) are tested.

> *Hypothesis 1:* Subjects in *VHIGH-s* exhibit similar behaviour in the start of the game in stages two and three as in their elicitation of types in stage one.

I predict that even though I use a one-shot game to elicit individual contribution preferences subjects will exhibit similar behaviour in the multiple-period games later on, at least in the beginning of the treatments. Implicitly this prediction also holds that subjects are correctly matched to their „well-deserved" subgroup in the *TYPE-treatment*.

[14] Having said that, *group identity* should not be the driver of individual contribution behaviour.

[15] Even though most of them are economics students, they are expected to be unfamiliar with the concepts of free-riding in public goods games, etc.

[16] $< https://www.google.com/intl/de/forms/about/ >$

Hypothesis 2: Subjects are more likely to individually contribute more on average and achieve the threshold in the *TYPE-treatment* than in the *BASELINE-treatment*.

The prediction here is that if subjects are placed into groups, they contribute more than if they are not placed into environments of equally-minded subjects. Besides analysising contribution averages, this hypothesis also allows to analyse individual behaviour of a cooperative individual in both treatments and how dispersed individual contributions become over time. By implictly assuming that behaviour depends partly on the behaviours of others, altruistic subjects will have an even higher motivation to contribute when they are in the same group with other high contributors. This effect is supposed to be strong enough to compensate for the low contribution levels of subjects in the lower sub-treatments.

Hypothesis 3: Subjects are more likely to individually contribute more on average and achieve the threshold in the *TYPE-treatment* than in the *RANDOM-treatment*.

This hypothesis is similar to the former, but emphasises that subgroup assignment matters. In fact, I expect similar contribution levels between *RANDOM-treatment* and *BASELINE-treatment*, since randomly composed subgroups should behave not any different than the entire group. The beneficial effect of building subgroups would hence only apply to the *TYPE-treatment*.

Hypothesis 4: Coordination to double contributions at levels 3.5 and/or 4 is more likely to happen in *VHIGH-s* than in other sub-treatments.

Because subjects in *VHIGH-s* are (if correctly matched) by definition very high contributors, they should be more likely to coordinate on higher contribution levels.

Hypothesis 5: Inequality and fairness preferences for subjects in *VHIGH-s* play no significant role.

The predicition is that subjects in *VHIGH-s* contribute regardless of low contributions by members in other subgroups and hence are motivated by achieving the threshold and not by lowering their inequality with respect to free-riders in other subgroups.

5 Results

Possible modelling:

In the following a model is presented which investigates free-riding, cooperative, and altruistic preferences.[17]

$$x_{it} = \alpha + \beta \quad \text{subject to} \quad \alpha \ \in (0, 2.5], \quad \beta \in (0, 1.5] \tag{4}$$

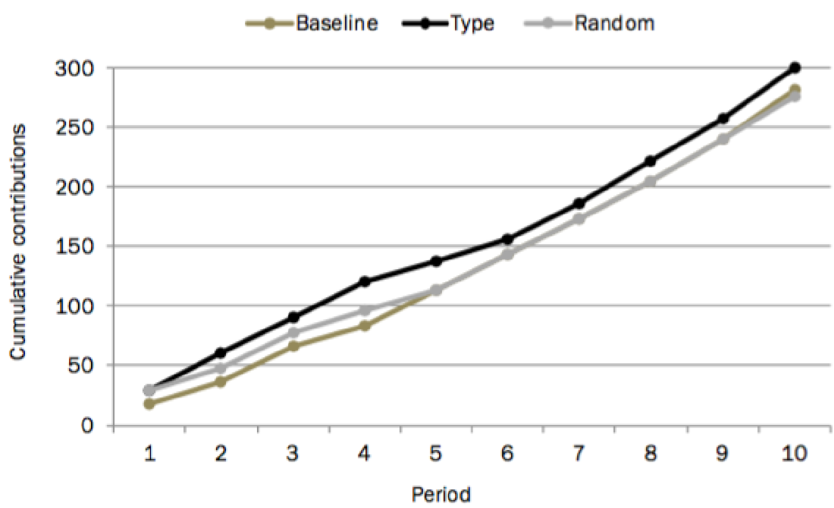
In (4) the variable α denotes cooperative and β altruistic preferences of individual i at period t ($x_{it} = 0$ denotes free-riding). I could specify every observation x_{it} by one of those three preference types over time. Moreover, by making x_{it} dependent on the sum of all contributions (5) I could analyse to what degree i is a conditional cooperator [Fischbacher et al., 2001]. This could be accompanied by an eliciation of beliefs similar to what Fischbacher and Gächter [2010] have done, to map beliefs to actions based upon conditional behaviour.

$$x_{i,t+1}(\sum_{j=1}^{N} x_{j,t}) = \gamma_i(\sum_{j=1}^{N} x_{j,t}) * x_{i,t} \tag{5}$$

Possible graphing:

One possible graph would be as shown below (figure 1).

Figure 1: Cumulative contributions after every period of all sessions per treatment



[17] Altruistic preferences is the variable I am most interested in since it captures the notion of intrinstic movitation (inherent contribution willingness).

Power test:

I would like the positive predictive value (PPV $= \frac{(1-\beta)R}{(1-\beta)R+\alpha}$) [Ionnidis, 2005] to be rather high, with a preferred statistical power $(1-\beta)$ of 0.8 and a statistical significance (α) of 0.05. However, I cannot calculate the pre-study odds (R) since my study is novel and thus have no indicator showing how many times an effect is expected to be truly non-null. I believe that by including 72 statistically indepenent observations under the obstacle of having no comparable studies I can still conclude findings with sufficient care.

6 Discussion

Evidence suggests that a set of players does not perform sufficiently well in achieving a predefined threshold in collective (risk-) social dilemma or public goods games. Especially observations from how climate change negotiations have recently been conducted indicate that socially better solutions (stricter CO_2 targets) can be achieved by smaller groups consisting of equally-minded countries. The key idea of this paper is that this is due to the heterogeneity of players implying different levels of cooperation behaviour. As a consequence, cooperative individuals might resign in playing fair or altruistic strategies because of the free-riding strategy of non-cooperative players. Splitting the players into groups according to their willingness to contribute and emphasising information on contribution behaviour of their group should reduce the problem. I design a laboratory experiment in order to test this. I predict that there is a positive effect of letting people play in a group of equally-minded people.

There are several specifications of the experimental design that should further be reconsidered. First, group size of twelve subjects might be too large to let subjects play in the *BASELINE-treatment*, and higher contribution levels (and/or achieving the threshold) in the *TYPE-treatments* might arise because of easier coordination and not because of playing with equally-minded group members. I.e., larger group could possibly exacerbate coordination among players and could lead to more variation in the data. However, having twelve people play in every treatment allows for separation into more subtreatments allowing for more observations of behaviour being dependent on contribution levels of others. Second, a great obstacle in my experimental design is how to successfully uncover true contribution preferences. The one-shot game in stage one allows a very simple eliciation procedure (Gächter & Thöni [2005]) but possesses obvious limitations. To counteract this drawback one could introduce a treatment where subjects assign themselves into groups (with means of communication allowed) in contrast to being anonymously assigned by the experimenter based on stage one. Also, by not explicitly telling subjects that they will be assigned to groups depending on their behaviour from the one-shot game, they might regret their decision later on (which obviously wants to be avoided).

One further feature, which could be added to my design, would be a treatment where cooperative subjects receive full information about the behaviour of others (meaning all individual contribution levels after every period). On the other hand, less cooperative subjects would only be informed about their group members' contribution levels. Altering the amount of information would certainly be interesting to investigate, but would not allow to answer my original research question without endogeneity concerns.

Further research should consider including communication (non-anonymous players). This modification is assumed to increase efficiency [e.g. Tavoni et al., 2011] and would closer reflect actual negotiations. Another part worth investigating is whether a smaller group (consisting of e.g. *VHIGH-s* + *HIGH-s* + *LOW-s*), by not letting free-riders (*VHLOW-s*) participate, managed to come up for all the predefined level of avoiding a public bad. Then on average everyone would need to contribute 3.33 (=300/(10*9)), which is more than the 2.5 tokens needed if all participated. Balancing out such group size effects, one could adjust the threshold to be a relative (per capita) threshold (=2.5*9*10=225) rather than a fixed absolute threshold (i.e. 300).

Since it remains unclear how to disentangle intrinsic motivation (i.e. altruistic preferences) from conditional cooperation as well as to better understand individual's motivation and rationality of decisions, it might be helpful to perform eliciation of beliefs (in all stages and before every subsequent period in all treatments) to check whether individual beliefs map to actions [Fischbacher and Gächter, 2010]. This would also help to establish a proper model for the results section.

At last, since the motivation of the research question for this paper originates from climate change negotiations, the neutral framing of the collective social dilemma game could be replaced by environmental language (e.g. Nordhaus [2015], Milinski et al. [2008]). It is in the essence of this study to extract individual contribution behaviour and how this might be affected by behaviour of other subjects, but contribution behaviour differs among social, risk, intertemporal, and eventually environmental preferences.[18] It therefore lies on future research - in particular that of *economic foundations of environmental preferences* - to investigate whether environmental preferences in specific differ from other types of preferences (mainly social, risk, and time preferences), and how those come into existence and change over time.[19]

[18] For example, while Milinski et al. [2008] investigate different probabilities associated to the threshold, they fail to analyse the difference between risk and environmental preferences and how those, independently of each other, drive their results.

[19] Moving away from individual level (and closing the loop from the motivational part in my proposal), this might help to explain why some countries such as Germany (having the highest unconditional CO_2 reduction target [UNEP, 2016] (making it a cooperator independent of the behaviour of other countries)) are extremely willing to abate greenhouse gases - besides political reasons (Christoff and Eckersley [2011], Wurzel and Connelly [2011]).

References

Barrett, S. [2003], Environment and statecraft: The strategy of environmental treaty-making. Oxford: Oxford University Press.

Biermann, F., and Pattberg, P. [2009], Global environmental governance reconsidered, *Global Environmental Change, 8,* 35-54.

Burlando, R. M., and Guala, F. [2005], Heterogeneous agents in public goods experiments, *Experimental Economics, 8,* 35-54.

Burton-Chellew, M. N., May, R. M., West, S. A. [2013], Combined inequality in wealth and risk leads to disaster in the climate change game, *Climatic Change, 120,* 815-830.

Cadsby, C. B., and Maynes, E. [1999], Voluntary provision of threshold public goods with continuous contributions: experimental evidence, *Journal of Public Economics, 71,* 53-73.

Cappelen, A. W., Nygaard, K., Sorenson, E. O., and Tungodden, B. [2011], Social preferences in the lab: A comparison of students and a representative population; *accessible via* < *http* : //*www.econstor.eu*
/*handle*/10419/49506 >.

Christoff, P., and Eckersley, R. [2011], Comparing state responses, In: J. S. Dryzek, R. B. Norgaard and D. Schlosberg (eds.), *The Oxford handbook of climate change and society, 431-448,* Oxford: Oxford University Press.

Croson, R. T. A., and Marks, M. B. R. [2000], Step returns in threshold public goods: A meta- and experimental analysis, *Experimental Economics, 2,* 239-259.

Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., and Tavoni, A. [2011], Coordination under threshold uncertainty in a public goods game, *Discussion Paper, No. 11-065, provided in cooperatiion with the Center of European Economic Research,* available at < *http* : //*hdl.handle.net*/10419/51354 >.

Dessler, A. E., and Parson, E. A. [2010], *The science and politics of global climate change: a guide to the debate.* Cambridge: Cambridge University Press.

Dryzek, J. S., Norgaard, R. B., and Schlosberg, D. (eds.) [2011], *The Oxford handbook of climate change and society, 144-160,* Oxford: Oxford University Press.

Fischbacher, U., and Gächter, S. [2010], Social preferences, beliefs, and the dynamics of free riding in public goods experiments, *American Economic Review, 100,* 541-556.

Fischbacher, U., Gächter, S., and Fehr, E. [2001], Are people conditionally cooperative? Evidence from a public goods experiment, *Economics Letters, 71,* 397-404.

Fischbacher, U. [2007], z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics, 10,* 171-178.

Gächter, S., and Thöni, C.[2005], Social learning and voluntary cooperation among like-minded people, *Journal of the European Economic Association, 3,* 303-314.

Ionnidis, J. P. A. [2005], Why most published research findings are false, *PLOS Medicine*, accessible via $< https://doi.org/10.1371/journal.pmed.0020124 >$.

IPCC [2013], *Assessment Report V*, accesssible via $< https://www.ipcc.ch/report/ar5 >$.

KP[2012], *Doha amendment to the Kyoto Protocol,* accessible via $< http://unfccc.int/files/kyoto\_protocol/application/pdf/kp\_doha\_amendment\_english.pdf >$

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J. [2008], Tipping elements in the earth's climate system, *PNAS, 105,* 1786-1793.

Lontzek, S. T., Cai, Y. C., Judd, K. L., and Lenton, T. M. [2015], Stochastic integrated assessment of climate tipping points indicates the need for stricter climate policy, *Nature Climate Change, 5,* 441-444.

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A., Marotzke, J. [2008], The collective-risk social dilemma and the prevention of simulated dangerous climate change, *PNAS, 105,* 2291-2294.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. J. [2011], Measuring social value orientation, *Judgment and Decision Making, 6,* 771-781.

Nordhaus, W. [2015], Climate clubs: Overcoming free-riding in international climate policy, *American Economic Review, 105,* 1339-13370.

Tavoni, A., Dannenberg, A., Kallis, G., and Löschel, A. [2011], Inequality, communication, and the avoidance of disastrous climate change in a public goods game, *PNAS, 108,* 11825-11829.

UNEP [2016], *The emission gap report: A UNEP synthesis report,* accessible via $< http://www.unep.org/pdf/2016gapreport.pdf >$.

Stiglitz, J. E. [2010], *Overcoming the Copenhagen failure*, accessible via $< http://host.madison.com/news/opinion/column/joseph-e-stiglitz-overcoming-the-copenhagen-failure/article\_a43e6c32-ef98-5c38-9a64-ae22b5732127.html >$.

Victor, D. G. [2015], The case for climate clubs, Working paper, accessible via $< http :$

$//e15initiative.org/publications/the-case-for-climate-clubs/>$.

Sturm, B., and Weimann, J. [2001], Experimente in der Umweltökonomik, FEMM working paper No. 7/2001.

Weitzman, M. W. [2011], Fat-tailed uncertainty in the economics of climate change, *Review of Environmental Economics and Policy, 5,* 275-292.

Wildavski, A. [1981], Richer is safer, *Financial Analysts Journal, 37,* 19-22.

Wurzel, R., and Connelly, J. (eds.) [2011], *The European Union as a leader in international climate change politics.* Oxfordshire: Routledge.

All weblinks cited in this paper were accessed on April 15, 2017.

Appendix

Attached to the appendix are the instructions.

Instructions

You are now taking part in an economic experiment. It is expected to last about 30 minutes. In order to understand the rules and to increase your chance of achieving a higher payoff from the experiment, it is important that you study the instructions very carefully.

During the experiment conversation is not allowed. If you have questions, please raise your hand and one of the experimenters will answer your question personally. Violation of the rules about communication and anonymity will lead to exclusion from the experiment and all payments.

At the end of the experiment you will receive a monetary payment according to your performance in this experiment. We will determine the payment as follows:

- The experiment consists of 3 stages (in the following noted by *stage 1, stage 2,* and *stage 3*).
- In all stages you are paid according to what is left on your private account.
- In *stage 1* every token is worth AUD 5.
- For each of the subsequent stages, *stage 2* and *stage 3*, you will be ranked according to your performance (to what is left on your private account) after each of both stages is over. Your ranking translates to a certain payment. The person that achieves the highest payoff left on her private account is ranked as 1, the second highest is ranked as 2, etc. The following table summarizes the payoffs in AUD according to the rank:

Rank	1	2	3	4	5	6	7	8	9	10	1	12
Payment	24	22	20	18	16	14	12	10	8	6	4	2

In case of a tie, the average amount will be paid. The following table shows the payment when the rank for the 2^{nd}-3^{rd} position as well as the 6^{th}-8^{th} position is the same:

Rank	1	2	3	4	5	6	7	8	9	10	1	12
Payment	24	21	21	18	16	12	12	12	8	6	4	2

In the following, the stages of the experiment are explained in detail.

Stage 1

In this stage you take simultaneously part with 11 other participants. There exist two types of accounts – individual accounts for each participant and one group account. The group account is an investment into avoiding that you lose everything. All participants are endowed with 4 monetary units on their individual accounts. The group account is empty. Each participant should decide how to split the endowment between her individual account and the group account. All participants submit their contributions simultaneously. You make your decision once, where you decide if you want to donate 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, or 4 tokens to the group account. Note that you cannot type any other amount into the box on your screen but only those.

In this stage you are faced with the following problem. Your contributions to the group account are investments that are used to prevent that you lose everything on your private account after the stage is over (and thus get 0).

If the sum of the contributions from all 12 participants (including you) is more or equal than 30, you will be paid according to what is left on your private account. If the amount on the group account is less than 30, all money on your private account is lost and you get 0.

There exists the possibility that your contribution towards the group account gets doubled if you and 2 other participants contribute 3.5 or 4 tokens. That is, if 3 participants contribute 3.5 or 4 tokens their contributions will be doubled towards their group account. *Example:* Imagine you contribute 3.5 tokens and 2 other participants contribute 4 tokens to the group account. Then the investments in the group account increase by an additional of 11.5 tokens.

Once you and the 11 other participants make a decision, you will see whether the threshold of 30 tokens in the group account was achieved.

Moreover, note that your behavior in this stage of the game will affect the arrangement of the game in the next stages.

If you have any questions, please ask them now. To start the experiment, please wait until the START bottom appears on your screen.

Please click the NEXT once this stage is over.

Stage 2

In this stage you again take simultaneously part with 11 other participants. However, you are placed in a group with 2 other participants. The groups are determined according to your performance in the 1st stage. On your screen, you will see the contributions from stage 1 of your group members.

There are again 2 types of accounts – individual accounts and a group account for all 12 participants (including you). You will play the game for 10 periods. Each participant is endowed with 40 tokens on the individual account at the beginning of the game. In each period you decide whether you want to donate 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, or 4 tokens to the group account. Note that you cannot type any other amount into the box on your screen but only those. Please wait for the computer's announcement that the period is over before you submit your contribution for the next period. You will get information about each individual contribution level of all group members of your group after each period, but you will not be informed about the contributions of the other groups. Furthermore, you will see how much has been contributed towards the group account by all 12 participants (including you), as well as total cumulative contributions towards the group account after every period. You will learn your payoff and your rank at the end of this stage.

In this stage you are faced with the following problem. Your contributions to the group account are investments that are used to prevent that you lose everything on your private account after the stage is over (and thus get 0).

If the sum of the contributions from all groups after 10 periods is more or equal than 300, you will be paid according to what is left on your private account. In this case your rank is determined according to the amount on your individual account. The person with the highest amount receives rank 1 and the person with the lowest – rank 12.

If the amount on the group account is less than 300, all money on your private account is lost.

There exists the possibility that your contribution towards the group account gets doubled if you and 2 other participants contribute 3.5 or 4 tokens. That is, if 3 participants contribute 3.5 or 4 tokens their contributions will be doubled towards their group account. *Example:* Imagine you contribute 3.5 tokens and 2 other participants contribute 4 tokens to the group account. Then the investments in the group account increase by an additional of 11.5 tokens.

If you have any questions, please ask them now. To start the experiment, please wait until the START bottom appears on your screen.

After every period you will be able to view the contributions of your group members as well as total cumulative distributions towards the group account in the right corner of your screen.

Please click the NEXT bottom to type your contribution in the next period, and repeat the above procedure until the stage is over.

Stage 3

The rules of this stage are the same as in the previous stage (stage 2), but you are not split into groups. That is, you again take simultaneously part with 11 other participants. All participants are endowed with 40 monetary units. The game is played for 10 periods. In each period you make a decision whether to place 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, or 4 tokens into the group account. At the end of the 10th period the sum of the contributions is compared to the threshold level of 300. If it is above or equal to the threshold, your rank will be determined according to the amount on your individual account. If the total contribution is below 300, you lose everything (you get 0).

There exists the possibility that your contribution towards the group account gets doubled if you and 2 other participants contribute 3.5 or 4 tokens. That is, if 3 participants contribute 3.5 or 4 tokens their contributions will be doubled towards their group account. *Example:* Imagine you contribute 3.5 tokens and 2 other participants contribute 4 tokens to the group account. Then the investments in the group account increase by an additional of 11.5 tokens.

You will be informed about the average individual contributions (including total and total cumulative contributions) after each period.

If you have any questions, please ask them now. To start the experiment, please wait until the START bottom appears. After every period you will see the contributions made in this period in the right corner of your screen.

Please click the NEXT bottom to type your contribution in the next period, and repeat the above procedure until the stage is over.

Thank you for your participation in the experiment.