Wealth and Vulnerability to Climate Change: An Experimental Study on Burden Sharing among Heterogeneous Agents

Ilona Reindl[1] ⓘ

Abstract

Agents in global climate negotiations differ with respect to their vulnerability to the negative consequences of climate change, but also their ability to contribute to its prevention. Due to this multidimensional heterogeneity, agents disagree about how the costs of emission reduction ought to be shared and, as a consequence, efficiency is low. This experiment varies the two dimensions separately in a controlled setting. The results show that in groups that succeed in reaching a predefined threshold, the rich and the more vulnerable, ceteris paribus, tend to carry a larger share of the burden. Surprisingly, groups are most likely to master the challenge when poverty coincides with high vulnerability. In this case the rich and less vulnerable abstain from interpreting fair burden sharing in a self-serving manner. Instead, they seem to acknowledge the double-disadvantaged position of the poor and more vulnerable and voluntarily carry a larger share of the burden.

✉ Ilona Reindl
  ilona.reindl@univie.ac.at

1 Faculty of Economics, Business and Statistics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

1 Introduction

Preventing climate change and global warming are among the biggest challenges of our time. In 2015, international climate negotiators agreed in Paris to keep the rise in global mean surface temperature "well below" 2 °C. Reaching this target is thought to prevent catastrophic climate change,[1] but requires a considerable reduction in greenhouse gas emissions. This, in turn, requires collective efforts by a substantial number of countries, which is notoriously difficult since emission reductions are costly for the country undertaking the effort but beneficial for all (Nordhaus 2018).

This "collective action problem at the global scale" (IPCC 2014) is aggravated by the heterogeneity of countries, in particular of developed and developing nations. The latter tend to be more vulnerable to the consequences of climate change than the former since climate change intensifies already existing risks of extreme weather events such as floods, storms, and droughts. These risks tend to be more pronounced in developing countries, also because their economies tend to rely on climate-sensitive sectors like agriculture or fishery for food security and employment and often lack the ability to cope with climatic shocks. In addition, developing countries also dispose of fewer (financial) resources to invest into vulnerability-reducing adaption measures and can contribute less to the prevention of climate change in the first place (Stern 2007; DARA 2012; IPCC 2014). In sum, developing countries are both more vulnerable and poorer than developed countries.

These asymmetries induce conflicting perceptions of how the burden of emission reduction ought to be distributed among nations. While an equal distribution of the burden could be considered a default option, international climate negotiators early on agreed on the principle of "common but differentiated responsibilities" (United Nations 1992; Ringius et al. 2002). Frequently discussed criteria for determining such a differentiated distribution are a country's vulnerability to the negative consequences of climate change and its ability to pay for emission reduction (Hayward 2012; Gampfer 2014). Standard burden-sharing rules yield, however, conflicting implications since wealth and vulnerability tend to come in conjunction. Burden sharing based on a country's wealth would require higher contributions by developed countries (*ability-to-pay rule*), i.e. by those who are less affected by the negative consequences of climate change. Using vulnerability as a distribution criterion would entail larger contributions by developing countries (*beneficiary-pays rule*), i.e. by those who are more vulnerable but usually dispose of fewer resources to contribute to climate change prevention. This availability of multiple legitimate behavioral rules creates normative conflict—a situation in which agents tend to endorse distributions that favor their private interests (Bicchieri 2005). And indeed, self-serving interpretations of fair

Table 1 Treatment variations

Wealth	Vulnerability	
	Homogeneous	Heterogeneous
Homogeneous	*BASE*	*VUL*
Heterogeneous	*WLT*	*WLT-VUL*

---

[1] The expression *catastrophic climate change* refers to the threat of "abrupt, unpredictable, and potentially irreversible changes that have massively disruptive and large-scale impacts" that are triggered if global temperatures pass a certain threshold (Molina et al. 2014, p 6).

burden sharing are an important reason for the recurrent deadlock of international climate negotiations (Lange et al. 2010; Carlsson et al. 2013; Brick and Visser 2015).

This study uses the climate change game (Milinski et al. 2008) to disentangle the effects of heterogeneity in wealth and vulnerability (i) on a group's ability to reach a predefined collective threshold, reflecting the prevention of catastrophic climate change, and (ii) on how the burden is shared conditional on the group's composition. In the experiment the participants were matched in groups. In each round group members could individually contribute part of their private funds to a public account in order to reach the collective threshold by the end of ten rounds. If the group missed the threshold each member experienced a predefined loss. Conditional on the treatment (see Table 1), group members differed with respect to their contribution capacities (wealth) or the loss they experienced (vulnerability), or both simultaneously. The experimental design thus allowed for an investigation of the effect of group heterogeneity in none (*BASE*), one (*VUL* and *WLT*) or both of these dimensions (*WLT-VUL*) in a systematic and controlled way. By matching agents who were both poor and more vulnerable with agents who were both rich and less vulnerable in *WLT-VUL*, the potentially most conflicting composition regarding preferences for burden sharing was brought to the lab. Individual burden sharing preferences were elicited behind a veil of ignorance at the beginning of the experiment to assess the effects of (i) ex-ante normative disagreement on the group level and (ii) self-serving deviations from ex-ante preferences on the individual level.

The results of the experiment show that if group members differ with respect to wealth (*WLT*), groups are more likely to reach the threshold if the rich carry a larger share of the burden. When members differ with respect to vulnerability (*VUL*), groups are more successful if the more vulnerable contribute more. Interestingly, the results show that groups are most successful and efficient in case of multidimensional heterogeneity (*WLT-VUL*), i.e. when low wealth coincides with high vulnerability and vice versa. In this case burden sharing mirrors the one observed in *WLT*, with the difference that groups are significantly more likely to reach the threshold. The additional vulnerability of the poor seems to facilitate coordination by emphasizing the responsibility of the rich to assist those in need and carry a larger share of the burden.

These findings are in line with burden sharing preferences elicited behind a veil of ignorance. Before being informed about their own type, participants state that it would be fair that the rich, despite being less vulnerable, bear more of the burden in *WLT-VUL*. Ex-ante normative disagreement about how the burden should be shared is relatively low in homogenous groups (*BASE*) and groups that are heterogeneous with respect to wealth (*WLT*), but it is significantly larger in groups that are heterogeneous with respect to vulnerability (*VUL*). As a consequence, the latter coordinate on a broader variety of different burden-sharing rules than other groups (who mostly coordinate on the same burden-sharing rule) and have more difficulties in reaching the threshold. In general, groups are more likely to succeed if individuals comply with the contribution they indicated as fair behind a veil of ignorance, while self-serving deviations from ex-ante preferences are detrimental to group success.

This study adds to a small body of literature that uses the climate change game to study climate change prevention behavior in an experimental setting. Multidimensional and correlated heterogeneity in wealth and vulnerability has been studied before (Burton-Chellew et al. 2013; Gampfer 2014), however, these studies were not able to systematically disentangle the effects of wealth and vulnerability. The results of this study show that groups are surprisingly successful in solving the coordination problem that evolves when the poor are also more vulnerable, and vice versa. The double-disadvantaged position of the poor

seems to facilitate coordination by making the responsibility of the rich more salient. In the experiment, the rich readily assume this responsibility, resulting in an impressive success rate of the respective groups.

The remainder of this paper is organized as follows: Sect. 2 gives an overview of the previous literature, Sect. 3 outlines the experimental design and Sect. 4 the theoretical expectations. The results are presented in Sect. 5 and discussed in Sect. 6.

2 Previous Research

This study builds on and complements a slim experimental literature using the so-called climate change game developed by Milinski et al. (2008) to study burden sharing in heterogeneous groups. It focuses on two of the main asymmetries and sources of conflict between climate negotiators, wealth and vulnerability. So far, these two dimensions have not been studied systematically in a 2×2 design. Such a design allows isolating how heterogeneity in wealth and/or vulnerability affect burden sharing and the likelihood that groups reach the threshold, conditional on the type of heterogeneity.

The climate change game is a variant of a threshold public good game (TPGG). It was developed by Milinski et al. (2008) as a means to study experimentally the social dilemma that comes with the prevention of climate change.[2] In the game, group members can contribute their private funds either to a public account or keep them in their private account. After ten rounds of play, the sum of individual contributions to the public account has to exceed a certain threshold, otherwise each group member loses her private funds with a predetermined probability. Individual contributions are considered to reflect emission reduction efforts and the loss of private funds the costs of catastrophic climate change. Coordinating on reaching the threshold is in the collective interest of the group. This collective effort can, however, be thwarted by group members choosing to free-ride. In Milinski et al. (2008) group members are homogenous with respect to wealth and vulnerability, which turns the game into a coordination game with equal burden sharing as the normatively appealing and hence focal burden-sharing rule. In reality, agents are, however, heterogeneous and find different burden-sharing rules normatively appealing, which makes coordination more difficult.

The climate change game has been used before to study the effect of *heterogeneity in wealth*. Both Tavoni et al. (2011) and Burton-Chellew et al. (2013) found that heterogeneous groups are less likely to reach the threshold and if they do, the rich tend to carry a larger share of the burden than the poor. These studies provide different explanations for the contribution behavior of the rich. While Tavoni et al. (2011) explain the larger contributions with a concern for fairness of the rich, Burton-Chellew et al. (2013) argue that they contribute more since they have relatively more at stake if the threshold is missed. These

---

[2] Note that Milinski et al.'s way to model climate change deviates from the conventional approach chosen by environmental economists to model climate change in the context of (self-enforcing) international environmental agreements, where parties first choose to be part of an agreement or not and then choose their actions as signatories or non-signatories, respectively (e.g. Barrett (1994), Carraro and Siniscalco (1998) or Finus (2008)). Consequently, Milinski et al. (2008) have been criticized for providing subjects with no possibility to negotiate a treaty or to communicate as well as for assuming a certain threshold in their climate change game (Barrett 2013). The present study responds to this critique by providing subjects with a possibility to communicate (pledge mechanism) and by commenting on the assumption of a certain threshold in Sects. 4.1 (Equilibria and Best Responses) and 6 (Discussion and Conclusion).

findings are not surprising given the previous evidence from TPGGs, in which wealth heterogeneity was repeatedly found to have a negative effect on the provision of the public good (e.g. Rapoport and Suleiman 1993; Bagnoli and McKee 1991; Bernard et al. 2014).

Contrary to heterogeneity in wealth *heterogeneity in vulnerability* has not been addressed systematically in the climate change game. To the best of my knowledge, only Burton-Chellew et al. (2013) varied the group members' vulnerability, albeit not in a ceteris paribus manner. As a consequence, an isolation of the effect was not possible. In comparable PGGs and TPGGs, in which subjects benefited differently from the provision of the public good, subjects were found to support conflicting fairness principles, which was detrimental to group success (e.g. Reuben and Riedl 2013; Nikiforakis et al. 2012; Bagnoli and McKee 1991; Bernard et al. 2014).

If heterogeneity in one dimension is already detrimental to group success, what are the consequences if agents are *heterogeneous in both wealth and vulnerability*? Again, to the best of my knowledge, only Burton-Chellew et al. (2013) have addressed this question before in a climate change game. They found that groups are more likely to reach the threshold if the rich face a higher risk than the poor. In this case the rich are willing to carry a larger share of the burden. On the other hand, if the poor face a higher risk, the rich reduce their contributions and groups are more likely to fail. Gampfer (2014) observed similar patterns in an ultimatum game setting. In both studies the weaker bargaining position of the poor seemed to mitigate fairness considerations of the rich and induce a more self-serving behavior which was detrimental to group success. In support of this interpretation, Brick and Visser (2015) provide evidence that the adverse effect of heterogeneity is due to a self-serving application of fairness principles. In their variant of a climate change game, groups were more likely to fail if subjects decided as representatives of their country, which were informed about their own and others' wealth and vulnerability, rather than behind a veil of ignorance which precluded self-serving behavior.

3 Experimental Design

Subjects played ten rounds of a climate change game in groups of four. At the beginning of each round, each subject received a certain number of tokens as operating funds o_i and could decide to contribute any integer amount of these funds to a public account. The remaining tokens were allocated to her private account. Subjects liquidity in each round was constrained to o_i, i.e. saving for future rounds and taking loans on funds to be received in future rounds were precluded by design. In any given round, y_i denotes an individual i's contribution and Y_j a group's collective contribution to the public account. Until the end of round 10, $\sum Y_j$ had to exceed the threshold T of 80 tokens, otherwise each group member lost a certain amount of tokens L_i from her private account. The size of o_i and L_i was a function of a subject's type and was common knowledge. To reach the threshold T, contributions of more than one subject were required since the total amount of operating funds an individual received across all rounds (O_i) was insufficient to reach the threshold (i.e. $O_i < T$ for all types). At the beginning of the experimental session, each subject received an additional endowment E of 40 tokens in her private account which was not available for investment but at stake if T was not reached. The level of E was set such that negative payoffs were precluded even if a subject spent its entire operating funds and experienced the maximum loss.

In successful groups, i.e. if a group's joint contribution to the public account $\sum Y_j$ exceeded T after ten rounds, an individual i's payoff π_i was

$$\pi_i = E + O_i - Y_i \quad \text{if } \sum Y_j \geq T \tag{1}$$

where Y_i denotes the sum of contributions of individual i across ten rounds.

In failing groups, i.e. if the group's contributions $\sum Y_j$ did not exceed T after ten rounds, group members experienced a loss and an individual i's payoff π_i was

$$\pi_i = E + O_i - Y_i - L_i \quad \text{if } \sum Y_j < T \tag{2}$$

Unlike in linear one-shot PGGs subjects did not receive a marginal per capita return (*mpcr*) from contributions to the public account. Instead, subjects benefited indirectly from contributing to the public account by experiencing no loss if the threshold was reached.

The primary interest of this study is to disclose differences in the behavior of individuals and groups when heterogeneities are multidimensional compared to when they are onedimensional or absent. Therefore, four different treatments were implemented (see Table 1) using a between-subjects design and partner matching. In the *baseline treatment (BASE)* subjects were homogenous with respect to wealth and vulnerability. All subjects received 4 tokens per round as operating funds ($o_i = 4$) which could be invested in the public account. If the group did not reach the threshold by round 10, each subject lost 30 tokens ($L_i = 30$). In the *vulnerability treatment (VUL)* subjects were still equally wealthy ($o_i = 4$) but differed in their degree of vulnerability. Two subjects were more vulnerable and experienced a loss of 40 tokens if the threshold was not reached ($L_i = 40$), the other two subjects were less vulnerable and lost only 20 tokens ($L_i = 20$). In the *wealth treatment (WLT)* subjects had an equal degree of vulnerability ($L_i = 30$) but differed with respect to their wealth. Two subjects were rich and received 5 tokens per round as operating funds ($o_i = 5$), while the other two poor subjects received only 3 tokens per round ($o_i = 3$). In the *wealth and vulnerability treatment (WLT-VUL)* subjects were heterogeneous with respect to both dimensions. More precisely, both poor subjects were more vulnerable ($o_i = 3, L_i = 40$), while both rich subjects were less vulnerable ($o_i = 5, L_i = 20$). Matching rich/less vulnerable and poor/more vulnerable subjects is of particular interest. It not only reflects reality better than the other treatments, but there is also good reason to expect the existence of normative conflict among group members since it is not a priori clear which criterion, wealth or vulnerability, should be used to determine burden sharing (see Sect. 4.3).[3] The total wealth of the group and the total potential loss for the group were kept constant across treatments ($\sum O_i = 160$; $\sum L_i = 120$) to rule out that efficiency concerns drive the results.[4]

---

[3] In the reverse case of rich/more vulnerable and poor/less vulnerable normative conflict is less likely since both criteria, wealth and vulnerability, suggest higher contributions of the first.

[4] Efficiency in this experiment is defined as the amount lost or wasted due to missing the threshold (see Sect. 5.6). By modeling vulnerability as a fixed number of tokens to be lost, the potential loss for each subject remained constant across rounds irrespective of her contributions to the public account in previous rounds. In contrast, modeling vulnerability in relative terms (as it is done in previous studies, e.g. by Milinski et al. (2008), Tavoni et al. (2011) or Burton-Chellew et al. (2013)) would have implied that the potential loss for each subject would have decreased with previous contributions since contributions would have reduced the number of tokens left in her private account (i.e. the number of tokens that would have been lost with a predefined probability). That is, the motivation of subjects with a preference for efficiency to contribute to the public account might have decreased conditional on their previous contributions. By modeling vulnerability in absolute terms and hence decoupling the amount to be lost from previous con-

Table 2 Expected payoffs by treatment and type conditional on equilibrium and burden-sharing rule

Treatment	N	Type	Parameters			Free-rider equilibrium		Threshold equilibrium					
								Equal b. sharing		Ability to pay		Beneficiary pays	
			E_i	O_i	L_i	Y_i	π_i	Y_i	π_i	Y_i	π_i	Y_i	π_i
BASE	40	Basic	40	40	30	0	50	20	60	n.a	n.a	n.a	n.a
VUL	44	Less vuln	40	40	20	0	60	20	60	n.a	n.a	10	70
		More vuln	40	40	40	0	40	20	60	n.a	n.a	30	50
WLT	44	Rich	40	50	30	0	60	20	70	30	60	n.a	n.a
		Poor	40	30	30	0	40	20	50	10	60	n.a	n.a
WLT-VUL	40	Rich/less vuln	40	50	20	0	70	20	70	30	60*	10	80
		Poor/more vul	40	30	40	0	30	20	50	10	60	30	40

E_i = endowment; O_i = total operating funds; L_i = loss if threshold missed; Y_i = total contribution; π_i = payoff

*no equlibrium

This implies that o_i and L_i were set at the average value whenever subjects were homogenous with respect to the respective dimension. For an overview of treatment parameters see Table 2.

Subjects had complete information during the entire session. The types and, consequently, wealth and vulnerability of each group member were common knowledge. At the end of each round a table was displayed showing the group's contribution as well as an individual breakdown of the group members' contributions in the current round and in total. Pseudonyms were used to share this information.

At the very beginning of the experiment subjects' burden sharing preferences were elicited behind a veil of ignorance. Subjects were asked to indicate for each group member the average contribution per round they would consider fair. For this purpose, they were informed about the type of each group member (but not their own member ID or type) and could specify a tokens contribution between 0 and o_i as fair for the respective subject. In addition, subjects were asked to indicate on a scale from 1 (extremely unlikely) to 7 (extremely likely) how likely they think it is that the threshold will be reached by round 10. This question was repeated at the beginning of each round, jointly with a question asking subjects to assess the number of tokens they expect the group as a whole to contribute in the current round. Following Gächter and Renner (2010), beliefs were not incentivized in order to prevent their elicitation influencing contribution behavior. Since beliefs were elicited in the same way across all treatments I have no reason to assume that this procedure explains any treatment effects.

To facilitate coordination, subjects were provided with a simple communication device. Following the procedure in Tavoni et al. (2011), subjects had to pledge simultaneously and publicly at the end of round 3 and round 7 which amount they intended to contribute in the

---

Footnote 4 (continued)

tributions, I tried to reduce the effect of such efficiency concerns to a minimum. A similar way to model vulnerability has been chosen by Barrett and Dannenberg (2014a, b).

remaining rounds of the game.[5] Combining the information about intended future contributions of the other group members with the information about cumulative contributions in the past allowed subjects to assess whether the intended contributions would be enough to reach the threshold or whether adjustments were needed. That is, although pledges were non-binding, they were nevertheless not only "cheap talk" (Bochet et al. 2006) since they facilitated coordination. Again, the procedure was the same across treatments. Further analyses with respect to the pledge mechanism are provided in Appendix 1.

3.1 Procedures

The experiment was programmed using z-tree (Fischbacher 2007) and subjects were recruited via ORSEE (Greiner 2015). After entering the laboratory subjects were assigned randomly to a treatment and a group. The treatment number was displayed on the first screen allowing the lab assistants to hand out the corresponding instructions to each participant (instructions are included in Appendix 2). Before the start of the experiment subjects were provided with a series of control questions that would help them to understand the game. After ten rounds of play, the sum of tokens in the possession of each participant was converted into Euro (4 tokens = €1) and paid to the participant in private before leaving the lab. After completing the experiment and before payment, participants were asked to complete a questionnaire including questions regarding their satisfaction with the outcome of the game, a cognitive reflection test, several socio-demographic questions as well as questions from the social preference module developed by Falk et al. (2016) to assess, for example, their risk taking and reciprocity. Participants received an additional payment of €5 for completing the post-experimental questionnaire. Neutral language was used throughout the experiment.

4 Equilibria and Expectations

4.1 Equilibria and Best Responses

The climate change game is a variant of a *n*-person cumulative TPGG played for 10 rounds. Contrary to standard TPGGs, as implemented for example by Bagnoli and McKee (1991), Rapoport and Suleiman (1993) or Croson and Marks (2000), the collective target is to avoid a collective bad rather than creating a public good. Contributions happen simultaneously in each round, but there is an element of sequentiality since subjects learn others' contributions after each round. Contributions to the public account are not returned if the threshold is missed (no refund), neither are contributions above the threshold (no rebate of over-contributions).[6]

---

[5] The pledge mechanism reflects the mechanism agreed on in the Lima Call for Climate Action in 2014 (UNFCCC 2014), which asks its signatories to submit national emission-reduction pledges (*INDC: Intended Nationally Determined Contributions*) that should be sufficient to keep global warming below 1.5 °C.

[6] For studies on the impact of different levels of refund and rebate on success rates see the meta-study by Croson and Marks (2000) or the experimental papers by Rauchdobler et al. (2010) and Marks and Croson (1999).

Contrary to standard PGGs, which have a unique Nash equilibrium at zero contributions, the existence of a threshold turns the game into a coordination game that is characterized by two sets of equilibria. One is the *free-rider equilibrium* where all subjects contribute zero in each round. Assuming rational and selfish agents, this equilibrium is an inefficient, pure-strategy Nash equilibrium since free-riding is the individually best response if subjects expect everybody else to free-ride. There is only one pattern of contributions leading to this equilibrium, namely zero total contributions by all. The second focal contribution level is the *threshold equilibrium* where subjects collectively contribute just enough for the threshold to be reached but not passed, i.e. the allocation of funds is sufficient and efficient. There is a continuum of contribution patterns leading to this equilibrium, which distributes the contribution burden in different ways among the group members. Each of these patterns is a Nash equilibrium if it satisfies the following two constraints: (1) the sum of contributions is exactly equal to the threshold (*efficiency constraint:* $\sum Y_j = T$), and (2) no group member contributes more than the amount she would lose otherwise (*individual rationality constraint:* $\sum Y_i \leq L_i \forall i$) (Croson and Marks 2000). Contributing to the public account is collectively rational since the efficient threshold equilibrium pareto-dominates the inefficient free-rider equilibrium with collective payoffs equal to 240 and 200 tokens, respectively.

A profound theoretical analysis of the game is beyond the scope of this experimental study, but a few general comments can be made regarding individual best-response behavior. A rational and selfish agent should only contribute to the public account if and as long as she perceives her contributions to be critical to reach the threshold. Once the threshold cannot be reached anymore due to insufficient collective contributions in previous rounds, she should stop contributing. Similarly, if total contributions are about to exceed the threshold, a subject's best response is to reduce her own contributions as long as she expects the threshold to still be reached. This includes some strategic uncertainty since miscoordination and thus the waste of all previous contributions is possible (Bagnoli and McKee 1991; Croson and Marks 2000; Dannenberg et al. 2015). Note that in the final rounds of the game it might be optimal for a subject to contribute more than her ex-ante willingness to pay, such that $Y_i > L_i$. The reason is that, rationally, a subject should treat her contributions in previous rounds as sunk cost. She should assess in each round anew whether her contribution is critical to reach the threshold and base this decision on the information she receives at the end of each round about the group's cumulative contribution so far and her beliefs about the future contributions of the other group members.

4.2 Equilibrium Selection: The Role of Fairness Preferences

Given the large number of contribution patterns leading to the efficient threshold equilibrium, other selection criteria are needed to predict more precisely which contribution patterns are likely to be observed. Fairness principles are useful indicators to identify contribution patterns that serve as focal points and facilitate coordination (Schelling 1960). Considering the notions of equality and equity, three salient burden-sharing rules can be identified in the experimental setting at hand. The *equal burden-sharing rule* requires equal

contributions by all players. It reflects the principle of equality with respect to contributions and, in case of *BASE*, also with respect to outcomes. The *ability-to-pay rule* requires individual contributions to be proportional to wealth and hence respects the principle of equity but also of equality in outcomes. The third salient rule is the *beneficiary-pays rule*,[7] which suggests that those who are expected to benefit more, i.e. lose more when the threshold is missed, contribute relatively more to the public account. In this case the principle of equity is respected, but there is neither equality in contributions nor outcomes (Ringius et al. 2002; Konow 2003; Bernard et al. 2014).

Table 2 gives an overview of individual total contributions Y_i and the expected payoffs π_i for all treatments and types conditional on the equilibrium (free-rider equilibrium or threshold equilibrium) and the burden-sharing rule (equal burden sharing, ability-to-pay or beneficiary-pays rule). Note that the loss L_i also reflects the maximum total amount subjects are ex-ante willing to contribute to the public account if they expect to be critical for the threshold being reached, taking into consideration the efficiency constraint and the individual rationality constraint. At this contribution level subjects are ex-ante indifferent between free-riding and contributing the respective number of tokens.

The identification of normatively appealing burden-sharing rules narrows down the set of contribution patterns that are likely to be chosen by subjects (Reuben and Riedl 2013). There exists, however, a variety of legitimate burden-sharing rules in all treatments except *BASE*. This provides subjects with the flexibility to choose the burden-sharing rule that requires them to contribute the lowest amount, while still being able to justify their relatively low contributions as being fair. Such a self-serving interpretation of fairness principles allows subjects to reduce the tension between their conflicting desires for utility maximization and fairness (Konow 2000). Unfortunately, though, coordination will fail if all subjects interpret the fairness argument in their own favor. Burden-sharing rules might, nevertheless, facilitate coordination if subjects are motivated by inequality aversion and willing to contribute more than their self-interested fair share.

4.3 Expectations

The following expectations regarding the outcomes of the experiment are based on the previous literature and the theoretical thoughts presented above.

4.3.1 Threshold Achievement

Homogenous groups dispose of a unique focal burden-sharing rule, while a set of legitimate burden-sharing rules is available to members of heterogeneous groups. I therefore expect heterogeneous groups *(VUL, WLT, WLT-VUL)* to be less likely to reach the threshold than homogenous groups *(BASE)*. Assuming self-regarding utility-maximization, agents' self-serving burden sharing preferences are the most conflicting in *WLT-VUL* (see

---

[7] Beneficiaries are defined as agents who benefit from contributions to the public account, i.e. emission reduction efforts and the prevention of climate change. Note that the literature sometimes considers those to be beneficiaries who benefit(ed) from emissions, e.g. through economic development (Hayward 2012).

below) and I therefore expect to observe the lowest share of groups to reach the threshold in this treatment. If the rich/less vulnerable are, however, motivated by inequality aversion, they might be willing to compensate for the lower wealth of the poor, resulting in a larger share of successful groups in *WLT-VUL*.

4.3.2 Burden Sharing Preferences

Burden sharing preferences are elicited behind a veil of ignorance at the beginning of the experiment. Since subjects are ignorant of their type at this point in time, ex-ante preferences should not be biased by self-interest. I therefore expect most subjects to reveal preferences in line with one of the salient burden-sharing rules identified above, conditional on their preferences for equality or equity: the equal burden-sharing rule in *BASE*; the ability-to-pay rule and the equal burden-sharing rule in *WLT*; the beneficiary-pays rule or the equal burden-sharing rule in *VUL*. In *WLT-VUL* I expect subjects to reveal preferences for a larger variety of different burden-sharing rules, since both wealth and vulnerability can serve as reference points for those with a preference for equity, and the equal burden-sharing rule is still appealing to those with a preference for equality.

4.3.3 Individual Contribution Behavior

Under the assumption of self-regarding utility-maximization, I expect to observe contributions in line with the following burden-sharing rules (see also Table 2): In *VUL*, less vulnerable subjects are expected to contribute according to the beneficiary-pays rule, more vulnerable subjects in line with the equal burden-sharing rule. If subjects differ with respect to wealth (*WLT*), poor subjects are expected to follow the ability-to-pay rule, while rich subjects are more likely to adhere to the equal burden-sharing rule. In *WLT-VUL,* the self-interested preferences of both types are even more conflicting: The rich/less vulnerable are expected to contribute in line with the beneficiary-pays rule, the poor/more vulnerable in line with the ability-to-pay rule. As a consequence, assuming self-regarding utility-maximization, I expect the lowest average individual contributions in *WLT-VUL.*

Contributions might deviate from these expectations if the assumption of self-regarding utility-maximization is relaxed. If subjects are motivated by inequality aversion, I expect the rich in *WLT* to contribute more and compensate for the lower wealth of the poor; in *VUL*, the less vulnerable might be willing to share the burden equally in order to avoid inequality in earnings; given the perfect correlation of wealth and vulnerability in *WLT-VUL*, I also expect the rich/less vulnerable to carry a larger share of the burden than predicted under the assumption of self-regarding utility-maximization.

4.3.4 Burden Sharing in Groups

While fairness rules can be a source of conflict if interpreted in a self-interested way, they can also be the basis of agreement if group members manage to coordinate on one of the rules. Based on the expectations regarding burden sharing preferences, I expect that groups are most likely to succeed in reaching the threshold if they coordinate on equal burden sharing in *BASE*, the ability-to-pay rule in *WLT* and the beneficiary-pays rule in *VUL*. Coordination is expected to be more difficult in *WLT-VUL*. Implementing the ability-to-pay

rule is no equilibrium in this treatment unless subjects are motivated by inequality aversion since it would require a level of contributions of the rich that exceeds their maximum willingness to pay (see Table 2). The question is thus whether the rich/less vulnerable crowd out or whether they are motivated by inequality aversion and feel obliged to assist the poor/more vulnerable with higher contributions.

5 Results

The experiment was implemented in September and October 2016 in the VCEE lab in Vienna. In total, 168 subjects participated in 42 groups: ten groups of four in both *BASE* and *WLT-VUL*, eleven groups of four in both *VUL* and *WLT*. Among the participants were slightly more women than men (57% and 43%, respectively) and subjects were on average 24 years old. Subjects were mostly students (99%) from a broad range of studies offered at the University of Vienna, one fifth (20%) of them studied economics or business. An overview of basic descriptive statistics is provided in Appendix 3. The experiment lasted approximately 75 min and subjects earned on average €18.8, with a minimum of €10 and a maximum of €24.

The presentation of the results starts with an analysis of threshold achievement in Sect. 5.1. In order to gain a better understanding of the observed treatment differences, ex-ante burden sharing preferences are analyzed in Sect. 5.2, actual burden sharing in Sect. 5.3, and potential self-serving biases in individual contribution behavior in Sect. 5.4. Section 5.5 looks at the dynamics across rounds and Sect. 5.6 finally examines efficiency and earnings.

5.1 Threshold Achievement

Groups were in general quite successful in avoiding a collective loss: 79% of all groups managed to reach the threshold of 80 tokens by round 10. Groups were most likely to miss the threshold in the treatments with one-dimensional heterogeneity: 36% in *VUL* and 27% in *WLT* did not reach the threshold. Success rates in these treatments were lower than in *BASE*, where only 20% of the groups failed, but the differences are not significant (Wilcoxon rank-sum tests, $p = 0.099$ and $p = 0.437$, respectively). These findings deviate at a first glance from those of previous studies, which found that heterogeneous groups in general perform worse than homogenous groups (e.g. Bernard et al. 2014; Burton-Chellew et al. 2013). The reason could be that these studies operationalized vulnerability in a different way (as losing everything with a certain probability, see Sect. 3) and did not provide subjects with the possibility to communicate. As has been shown by Tavoni et al. (2011), introducing a simple pledge mechanism, as used in this experiment, can eliminate the adverse effect of heterogeneity on group success.

Due to their multidimensional heterogeneity, groups in *WLT-VUL* were expected to have the most difficulties to coordinate. Surprisingly, *all* groups in *WLT-VUL* managed to reach the threshold and were thus significantly more likely to do so than in any other treatment (Wilcoxon rank-sum tests, $p = 0.003$ for *WLT-VUL* vs. *BASE*, $p = 0.000$ for both *WLT-VUL* vs. *VUL* and *WLT-VUL* vs. *WLT*). The following analyses shed light on the preferences and behavior underlying the observed treatment differences.
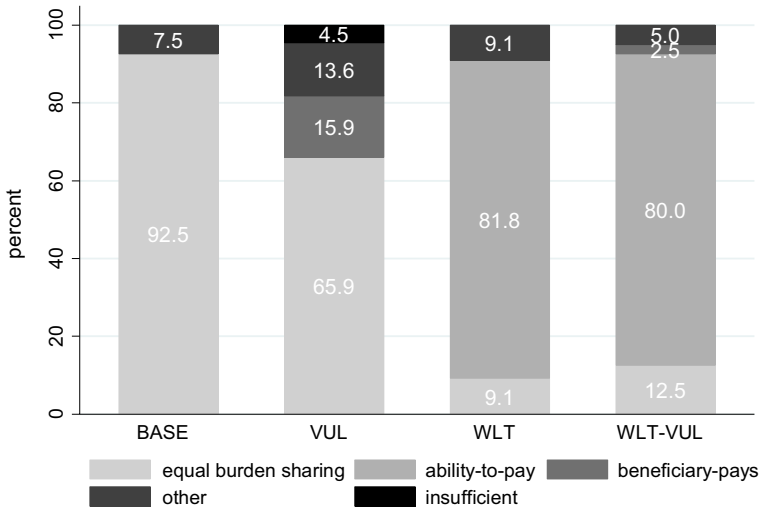
Fig. 1 Individual preferences for burden sharing behind a veil of ignorance

5.2 Burden Sharing Preferences

Individual burden sharing preferences were elicited behind a veil of ignorance by asking subjects to indicate the average contribution per round they would consider fair for each member ($\tilde{y}_j$ for all j). Since subjects knew the type of each group member but not their own type, these ex-ante preferences should not be biased by self-serving considerations. Subjects' answers were classified in four mutually exclusive categories according to the following procedure: if subjects allocated in total higher contributions to the rich types, i.e. $\sum \tilde{y}_{r1+r2} > \sum \tilde{y}_{p1+p2}$, and to each rich type at least as many tokens as to each poor type, i.e. $min\left(\tilde{y}_{r1}, \tilde{y}_{r2}\right) \geq max\left(\tilde{y}_{p1}, \tilde{y}_{p2}\right)$, they were classified as preferring the *ability-to-pay rule*; if they assigned in total more tokens to the more vulnerable types ($\sum \tilde{y}_{mv1+mv2} > \sum \tilde{y}_{lv1+lv2}$) and at least as many tokens to each more vulnerable subject than to each less vulnerable subject ($min\left(\tilde{y}_{mv1}, \tilde{y}_{mv2}\right) \geq max\left(\tilde{y}_{lv1}, \tilde{y}_{lv2}\right)$), they were classified as preferring the *beneficiary-pays rule*; subjects who considered equal contributions to be fair were classified as preferring the *equal burden-sharing rule*; if subjects' allocation did not correspond to any of these rules, they were classified as *other*.[8] Importantly, if subjects assigned a sum of contributions *insufficient* to reach the threshold, they were classified as such.

Figure 1 shows how much support the different burden-sharing rules received in each treatment. Overall, 89.9% of subjects supported one of the three burden-sharing rules specified in the theoretical part. There were, however, significant differences between treatments (χ^2-test, $p=0.000$). Subjects in *BASE* expressed a clear preference for the equal burden-sharing rule (92.5%). Equal burden sharing was also the most popular allocation in *VUL* (65.9%), but found only little support in *WLT* (9.1%) and *WLT-VUL* (12.5%). In both treatments with wealth heterogeneity, the ability-to-pay rule was the preferred allocation

---

[8] The allocation decisions of most subjects classified as *other* did not show a clear pattern. Their decisions could be explained by tremble, difficulties of understanding the task or irrationality.

with 81.8% in *WLT* and 80.0% in *WLT-VUL*. The beneficiary-pays rule found some support in *VUL* (15.9%), but hardly any in *WLT-VUL* (2.5%).

The clear preference for the ability-to-pay rule over the beneficiary-pays rule in *WLT-VUL* is remarkable. The results of *WLT* and *VUL* show that if heterogeneity existed in only one of the dimensions, subjects supported very different burden-sharing rules conditional on the type of heterogeneity. As a consequence one would expect to observe much more diversity in subjects' burden sharing preferences in *WLT-VUL* or, alternatively, that subjects choose equal burden sharing as a compromise. However, this is not the case and subjects clearly consider the wealth dimension to be the more valid criterion for burden sharing.

5.2.1 Normative Disagreement

Following Bernard et al. (2014), I assessed the level of normative disagreement by calculating the number of different burden-sharing rules that was supported by the members of a group. Normative disagreement was lowest in *BASE* with groups supporting on average 1.3 different burden-sharing rules. There was slightly more disagreement in *WLT* and *WLT-VUL*, with groups supporting on average 1.64 and 1.6 different rules, respectively. The significantly highest level of normative disagreement was observed in *VUL* with an average of 2.1 (Wilcoxon rank-sum test, $p = 0.000$ with respect to all other treatments). Since *VUL* is also the treatment with the lowest success rate, a plausible hypothesis would be that normative disagreement hindered groups from coordinating successfully. However, the correlation between normative disagreement and success rate is only weak and insignificant (Spearman's $r = -0.054$, $p = 0.488$) and does not support this hypothesis. This is in line with Bernard et al. (2014), who argue that subjects might not have very strong intrinsic preferences for a certain burden-sharing rule but rather consider them as useful focal points and are willing to compromise on their initially preferred allocation.

In the following section I analyze which burden-sharing rules groups finally coordinate on and whether the distribution mirrors the one of ex-ante preferences.

5.3 Burden Sharing in Groups

A similar procedure as in Sect. 5.2 is used to categorize actual burden sharing on the group level. Groups were classified as following the *ability-to-pay rule* if (i) the sum of contributions of both rich types was larger than the sum of contributions of both poor types, i.e. $\sum Y_{r1+r2} > \sum Y_{p1+p2}$, (ii) the smaller sum of contributions of a rich type was at least as high as the larger sum of contributions of a poor type, i.e. $min(Y_{r1}, Y_{r2}) \geq max(Y_{p1}, Y_{p2})$, and (iii) the range between the average contributions of the rich and the poor was larger than 10 tokens, i.e. $\left(\frac{\sum Y_{r1+r2}}{2} - \frac{\sum Y_{p1+p2}}{2} \right) > 10$. The last condition guarantees that the rich contributed substantially more than the poor and burden sharing is thus better captured by the ability-to-pay rule than the equal burden-sharing rule. Similarly, groups were classified as following the *beneficiary-pays rule* if (i) both more vulnerable types contributed in total more than both less vulnerable types, i.e. $\sum Y_{mv1+mv2} > \sum Y_{lv1+lv2}$, (ii) the smaller sum of contributions of a more vulnerable type was at least as high as the larger sum of contributions of a less vulnerable type, i.e. $min(Y_{mv1}, Y_{mv2}) \geq max(Y_{lv1}, Y_{lv2})$, and (iii) the range between the average contributions of the more vulnerable and the less vulnerable was larger than 10 tokens, i.e. $\left(\frac{\sum Y_{mv1+mv2}}{2} - \frac{\sum Y_{lv1+lv2}}{2} \right) > 10$. If the observed burden sharing pattern did not conform to any one of these rules and contributions of all members did not
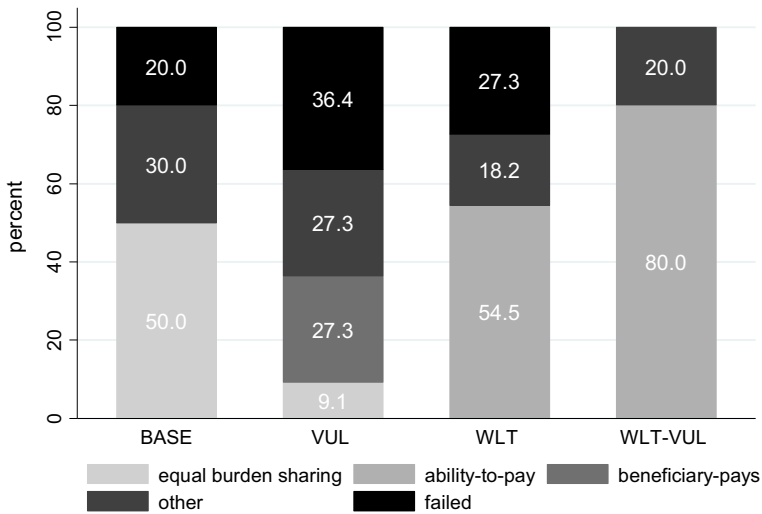
Fig. 2 Burden sharing in groups

deviate by more than 25% from the equal contribution level of 20 tokens, i.e. ranged between a minimum of 15 and a maximum of 25 tokens, groups were classified as following the *equal burden-sharing rule*. Groups that missed the threshold were classified as *failed* since no final statement can be made about burden sharing in these groups. All remaining groups were classified as *other*.[9]

Figure 2 shows for each treatment the share of groups following the different burden-sharing rules. Just as with burden sharing preferences, there are significant treatment differences with respect to the implemented burden-sharing rules (χ^2-test, $p=0.000$). In addition, several observations can be made when comparing the burden sharing preferences depicted in Fig. 1 and the actual burden sharing depicted in Fig. 2. Remember that subjects were not informed about the burden sharing preferences of the other group members and could hence not use this information to coordinate their contributions. First, the strong preference for the ability-to-pay rule in *WLT* and *WLT-VUL* directly translates into an impressive share of groups actually implementing this rule. In 74.8% of the successful groups in *WLT* (54.4% of all groups) and in 80% of the groups in *WLT-VUL* the rich carry a larger share of the burden than the poor. Second, in *VUL* the beneficiary-pays rule gained quite some popularity, with 42.9% of successful groups (27.3% of all groups) implementing this rule. Equally surprising is the fact that in this treatment only one group shared the burden equally, having in mind the strong preference for this rule subjects expressed ex-ante. One explanation for this outcome could be the self-serving behavior of the less vulnerable, which urges the more vulnerable to carry a larger share of the burden. This observation is also in line with Bicchieri (2005), who argues that a social norm is a behavioral rule that many people know exists but do not necessarily follow. Third, at a first glance there is a rather low share of groups following the equal-burden-sharing rule in *BASE*, namely

---

[9] The results presented in this section are robust to decreasing the range to 5 tokens. This reduces the number of equal sharing groups slightly in *BASE*. Increasing the range to 15 tokens, on the other hand, increases the number of groups that have to be classified as other across all treatments.
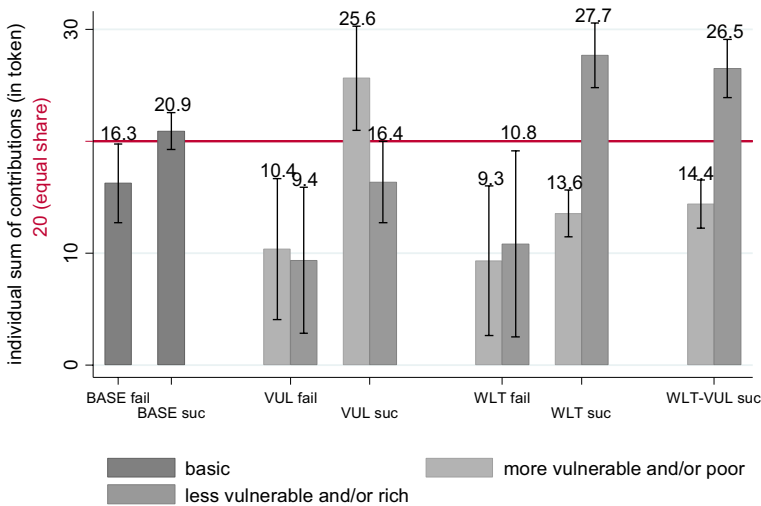
Fig. 3 Average sum of contributions by type, conditional on treatment and group success

62.5% of the successful groups (50% of all groups). This is due to several groups in which three subjects adhered to equal burden sharing and a fourth subject contributed an inefficiently high amount of tokens (>25 tokens). Including these inefficient groups, the share of equally sharing groups among successful groups increases to 87.5% (70% of all groups) in *BASE*.

Overall, when a certain burden-sharing rule was popular ex-ante, groups were also likely to coordinate on this rule. When, however, burden sharing preferences were more heterogeneous, as in *VUL*, groups also coordinated on a variety of different burden-sharing rules and had in general more difficulties coordinating successfully compared to the other treatments.

5.4 Individual Contributions

In *WLT*, we saw that a majority of successful groups implemented the ability-to-pay rule. Figure 3 shows that the rich subjects contributed significantly more than 20 tokens, compensating for the lower wealth of the poor who contributed significantly less than this (Wilcoxon signed rank tests, $p = 0.000$ for both types; the red line in Fig. 3 marks the equal share of 20 token). The difference between the contributions of both types is significant (Wilcoxon rank-sum test, $p = 0.000$). In *VUL* burden sharing was more diverse and not many groups strictly implemented the beneficiary-pays rule as it was defined in Sect. 5.3. On the individual level, however, we find that, on average, the more vulnerable contributed significantly more than the less vulnerable (Wilcoxon rank-sum test, $p = 0.002$). That is, the more vulnerable contributed significantly more and the less vulnerable less than the equal share of 20 tokens (Wilcoxon signed rank tests, $p = 0.017$ and $p = 0.084$, respectively).

The distribution of individual contributions in *WLT-VUL* closely mirrors the distribution in *WLT*: the rich/less vulnerable types contributed significantly more than 20 tokens, the poor/more vulnerable types significantly less (Wilcoxon signed rank tests, $p = 0.000$ for both types), and the difference between the contributions of both types is highly significant (Wilcoxon rank-sum test, $p = 0.000$). Treatment comparisons reveal that there is

no significant difference between the contributions of the rich/less vulnerable in *WLT-VUL* and the rich in *WLT*, but the difference is highly significant with respect to the less vulnerable in *VUL* (Wilcoxon rank-sum tests, $p=0.161$ and $p=0.000$, respectively). Likewise, the contributions of the poor/more vulnerable in *WLT-VUL* are not significantly different from those of the poor in *WLT*, but significantly different from those of the more vulnerable in *VUL* (Wilcoxon rank-sum tests, $p=0.220$ and $p=0.044$, respectively). This suggests that the contribution decisions of subjects in *WLT-VUL* are guided by the distribution of wealth in the group.

The picture looks quite different for failing groups, where all types in all treatments contributed (significantly) less than 20 tokens (Wilcoxon signed rank tests, $p=0.057$ for equal types, $p=0.020$ for more and less vulnerable types, $p=0.059$ for rich types, $p=0.035$ for poor types). In particular the rich in *WLT* and the more vulnerable in *VUL* reduced their contributions substantially compared to their peers in successful groups. In consequence, in failing groups there are no longer any significant differences in contribution behavior between the rich and the poor in *WLT* or the more and less vulnerable in *VUL* (Wilcoxon rank-sum tests, $p=0.872$ and $p=0.708$, respectively).

To sum up, these findings suggest that if groups are heterogeneous regarding their members' wealth, the willingness of the rich to carry a larger share of the burden is decisive for group success. Likewise, if groups are heterogeneous with respect to vulnerability, the more vulnerable subjects have to take the lead to ensure group success. In case of multidimensional and correlated heterogeneity the wealth dimension seems to be more salient and drive contribution decisions. Considering the significantly higher success rate of groups in *WLT-VUL* than in *WLT*, the role of vulnerability should, however, not be underestimated. In line with needs-based distributive justice norms, it is plausible to assume that if low wealth coincides with high vulnerability, the rich might feel a stronger responsibility to assist those in need and to take care that the collective bad, which would affect the already poor disproportionately, is prevented (Konow 2010; Gampfer 2014). The effect of heterogeneity in vulnerability on burden sharing is thus reversed when it coincides with heterogeneity in wealth.

5.4.1 Self-Serving Behavior

The results presented so far suggest that a self-serving interpretation of burden sharing rules is the reason why many groups fail. And indeed, column 1 in Table 3 confirms that differences in group success cannot be attributed to differences in ex-ante preferences (with respect to contributions of their type) since there are no significant differences in ex-ante preferences between members of successful and failing groups (Wilcoxon rank-sum tests, $p=0.239$ for equal types, $p=0.606$ for more vulnerable types, $p=0.383$ for less vulnerable types, $p=0.755$ for poor types, $p=0.256$ for rich types).

Notably, in successful groups, subjects' average contributions overall (column 2 in Table 3) and in round 1 (column 3) do not deviate significantly from their ex-ante preferences (column 1) (Wilcoxon signed rank tests, $p=0.069$ and $p=0.549$, respectively). We find, however, evidence for self-serving behavior in failing groups: Subjects who belonged to groups that ultimately failed contributed significantly less not only on average (column 4), but already in the first round (column 5) than what they had specified ex-ante as fair (Wilcoxon signed rank tests, $p=0.000$ and $p=0.021$, respectively).

These findings suggest that groups are more likely to succeed if subjects contribute what they consider a fair contribution for their type behind a veil of ignorance and

Table 3 Ex-ante preferences, average contributions and 1st round contributions by treatment and type, conditional on group success

Treatment	Type	Ex-ante preference	*Successful*		*Failing*	
			Average	Round 1	Average	Round 1
		(1)	(2)	(3)	(4)	(5)
BASE	Basic	2.17	2.13	2.09	2.13	1.63
		(0.54)	(0.66)	(0.46)	(0.83)	(0.42)
VUL	Less vulnerable	1.89	1.57	1.64	1.5	0.94
		(0.71)	(0.65)	(0.63)	(0.76)	(0.78)
	More vulnerable	2.25	2.43	2.56	1.38	1.04
		(0.75)	(0.76)	(0.81)	(0.74)	(0.76)
WLT	Rich	2.84	3.13	2.77	2.17	1.08
		(0.70)	(0.72)	(0.54)	(1.94)	(0.79)
	Poor	1.57	1.63	1.36	1.5	0.93
		(0.66)	(0.62)	(0.39)	(1.05)	(0.64)
WLT-VUL	Rich/less vuln	2.92	3.2	2.65		
		(0.99)	(1.01)	(0.55)		
	Poor/more vuln	1.5	1.5	1.44		
		(0.49)	(0.69)	(0.46)		

abstain from adjusting their contributions in a self-serving way. See Appendix 4 for an OLS regression backing these results.

5.5 Dynamics

This section takes a look at factors that might determine a change in individual contributions in a given round relative to the previous round: others' contributions, individual beliefs about others' contributions as well as beliefs about the group reaching the threshold.

Model (1) in Table 4 shows that subjects adjusted their contributions relative to others' contributions in the previous round: if the other group members contributed on average more (less) in the previous round, subjects increased (decreased) their contribution in the current round significantly (*contr others—contr ego, lag1*). They also increased (decreased) their contributions if they corrected their beliefs about others' contributions upwards (downwards) relative to the previous round (Δ *belief group contribution*). Both effects are evidence for a conditional willingness to contribute. Surprisingly, there is no evidence that subjects adjusted their contributions if their beliefs regarding the likelihood of reaching the threshold changed (Δ *threshold belief*). One reason might be that subjects mostly adjusted their beliefs not at all or only slightly from one round to the next. In 65% of all rounds subjects did not adjust their beliefs and in 27% only by one unit (on a scale from 1 to 7). Theoretically, subjects should, however, adjust their beliefs and hence contributions drastically as soon as the threshold is out of reach due to insufficient contributions in previous rounds. Including a dummy variable that takes the value one in the round in which the threshold moves definitely out of reach (*target out of reach*), I indeed find

Table 4 Panel model—determinants of change in individual contributions

DV: Δ individ. contribution	All	Successful	Failing
	(1)	(2)	(3)
Contr others—contr ego, lag1	0.301***	0.271***	0.414***
	(0.026)	(0.027)	(0.040)
Δ belief group contribution	0.125***	0.143***	0.065**
	(0.016)	(0.019)	(0.024)
Δ threshold belief	−0.016	−0.035	0.059
	(0.034)	(0.037)	(0.087)
Target out of reach	−0.422*		−0.447*
	(0.181)		(0.224)
Constant	−0.030*	−0.023	−0.043
	(0.014)	(0.015)	(0.049)
Observations	1,512	1,188	324
R-squared	0.247	0.239	0.306
N	168	132	36

Random effects regression; standard errors clustered on the group level; clustered standard errors in parentheses; $^{+}p<0.1$, $* p<0.05$, $** p<0.01$, $*** p<0.001$

a strong and significant negative effect on contributions. Subjects seemed to understand that further contributions are in vain and rationally decreased their contributions. 94% of the respective subjects reduced their contributions to the rational level of zero. All results presented in this section are robust to considering only successful groups in model (2) and only failing groups in model (3).

To sum up, the results provide evidence for subjects' conditional willingness to contribute: Subjects adjust their contributions to others' contributions in previous rounds and also to changes in their beliefs regarding others' contributions. Once the threshold cannot be reached any longer, subjects respond rationally by decreasing their contributions immediately and significantly. A more detailed analysis of the role of beliefs is provided in Appendix 5.

5.6 Efficiency and Earnings

Finally, two alternative performance concepts besides the success rate are discussed: Efficiency and the distribution of earnings among group members.

5.6.1 Efficiency

When groups missed the threshold, their members not only experienced a loss, but their entire contributions to the public account were wasted. As a consequence, successful groups were strikingly more efficient than failing groups. While successful groups wasted on average 2.9 tokens due to excess contributions, failing groups contributed—and thus wasted—on average 45.6 tokens.

Among successful groups, groups in *VUL* had the largest difficulties in coordinating. They wasted on average 4 tokens, significantly more than groups in *WLT-VUL* (1.8 tokens)

and *WLT* (2.5 tokens) (Wilcoxon rank-sum tests, $p = 0.000$ and $p = 0.006$, respectively). Among failing groups, groups in *BASE* did a particularly bad job by failing at a very high average contribution level of 65 tokens compared to the relatively lower levels in *VUL* (39.5 tokens) and *WLT* (40.3 tokens) (Wilcoxon rank-sum tests, $p = 0.003$ and $p = 0.000$, respectively).

5.6.2 Earnings

Coordinating on the threshold in general paid off for subjects, not least because of the high (low) level of efficiency in successful (failing) groups. Members of successful groups earned significantly more than members of failing groups across all treatments (Wilcoxon rank-sum tests, $p = 0.000$ for all treatments). Figure 4 shows that the less vulnerable in *VUL*, the rich in *WLT* and the rich/less vulnerable in *WLT-VUL* earned on average significantly more than the respective other types, both in successful (Wilcoxon rank-sum tests, $p = 0.002$ for *VUL* and *WLT*, $p = 0.000$ for *WLT-VUL*) and failing groups (Wilcoxon rank-sum tests, $p = 0.002$ for *VUL*, $p = 0.006$ for *WLT*).

Remember, however, the differences in the initial wealth situation in *VUL*, on the one hand, and *WLT* and *WLT-VUL*, on the other hand. In *VUL*, both types were provided with the same total amount of operating funds. This implies that if the less vulnerable earned more, this is because they contributed less to the public account than the more vulnerable. In *WLT* and *WLT-VUL*, on the other hand, rich subjects earned more despite contributing a larger share to the public account because they had higher operating funds than the poor in the first place. From this perspective, inequality decreased significantly from 20 to 5.88 tokens in *WLT* and from 20 to 7.9 tokens in *WLT-VUL*, while it increased significantly in *VUL* from zero to 9.3 tokens (Wilcoxon signed rank tests, $p = 0.000$ for all comparisons). This is also illustrated by deviations from the red line in Fig. 4 marking 60 tokens, which is the amount individual group members would earn if contributions were fully efficient and all group members earned the same.

The findings in *WLT* and *WLT-VUL* are in line with those of Tavoni et al. (2011) who found that successful groups are able to eliminate initial wealth heterogeneity. The opposite
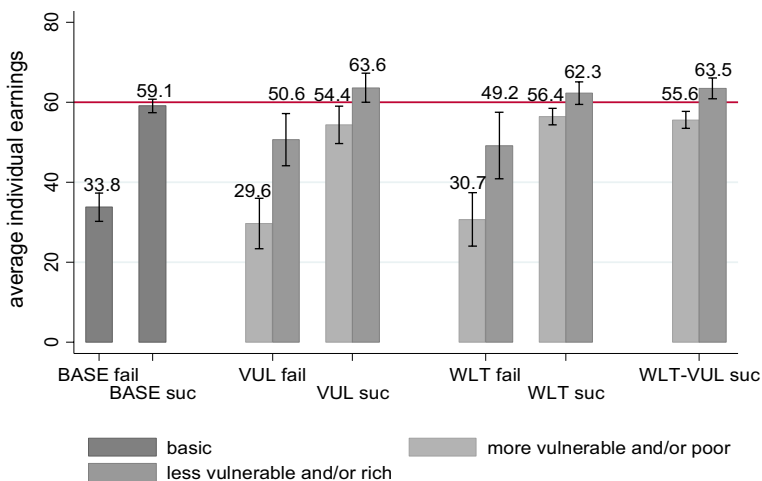

Fig. 4 Average individual earnings conditional on type, treatment and group success

happened, however, in *VUL* where initial equality in wealth was eliminated. In this treatment payoff equality seemed to be less of a concern for subjects. One explanation might be that the wealth dimension was less salient and subjects focused on the different benefits the various types would derive from reaching the threshold.

6 Discussion and Conclusion

The results of this experiment provide new insights into the role of vulnerability, wealth and fair burden sharing when heterogeneous groups seek to overcome a collective action problem. While heterogeneity of wealth has been studied in the experimental literature on climate change before, heterogeneity in vulnerability has hardly ever been addressed (see Sect. 2 for references). Importantly, the interaction between both dimensions has not been studied systematically before. This is problematic since the conflict of interest between countries involved in climate change negotiations is triggered to a large extent by the fact that their wealth and vulnerability tend to be correlated. As a consequence, climate negotiators tend to support conflicting burden-sharing rules.

The most surprising result of this study is that groups are very effective in solving the dilemma that emerges when low wealth coincides with high vulnerability, and vice versa. While heterogeneity in vulnerability by itself poses a challenging coordination problem for groups, it seems to have a facilitating effect once it coincides with wealth heterogeneity. Needs-based distributive justice norms provide a plausible explanation for this effect (Konow 2010; Gampfer 2014): The fact, that the already poor would in addition suffer disproportionately from the collective bad makes the responsibility of the rich to take a larger share of the burden—and thus assist those in need—more salient. In the experiment, the rich seem to readily assume this responsibility, resulting in an impressive success rate and a high efficiency of contributions in the respective groups. These findings challenge those of Burton-Chellew et al. (2013) and Gampfer (2014) who find that vulnerability has the opposite effect, i.e., that it significantly decreases the willingness to pay of the rich and less vulnerable. As a consequence, in their experiments these groups were significantly less successful in reaching the threshold. One reason for the diverging results might be the provision of a pledge mechanism as in Tavoni et al. (2011), which facilitated coordination in this experiment.

The results presented in this paper suggest that the focus on a needs-based perspective in the ongoing climate negotiations where countries that are both poor and more vulnerable negotiate with countries that are both rich and less vulnerable might be promising. Emphasizing the double-disadvantaged position of the first might trigger inequality aversion in the second and in turn facilitate coordination for reaching the collective target. However, the real-world situation is of course much more complex than the one modeled in the lab, where subjects were confronted with a relatively simple task. Other aspects than the heterogeneities studied here might explain the difficulties of international climate negotiators to reach a consensus on how to share the burden of emission reductions—not only compared to the groups in this experiment but also compared to other more successful international negotiations characterized by the heterogeneity of its agents, such as those on the Montreal Protocol on Substances that Deplete the Ozone Layer that involved side payments from developed to developing countries (Benedick 1998).

First, the focus of my study was on two spatial dimensions of climate change, wealth and vulnerability, ignoring a historical dimension. Economic wealth is often attributed

to large green-house gas emissions in the past, implying a historical responsibility of wealthier nations for the current state of the world's climate (Fuessel 2010). From a fairness perspective, this bolsters the claim of developing countries for larger contributions from developed countries. The implications of historical responsibility on burden sharing have been studied in lab experiments by Tavoni et al. (2011) and Gampfer (2014), but it would be an interesting avenue for further research to study the interaction of all three dimensions (wealth, vulnerability, historical responsibility) in the given experimental setting. Second, to keep things simple, I abstained from including uncertainty regarding the location of the threshold and the consequences of missing it. Remember that the introduction of a certain threshold turns the game into a coordination game with two Nash equilibria. Previous experimental studies show that while uncertainty about the consequences of missing the threshold does not impede coordination, uncertainty about its location is detrimental for success (Barrett and Dannenberg 2012, 2014b; Brown and Kroll 2017). The attempt of international climate negotiators to establish the 2 °C-target as a threshold to coordinate on falls short of providing the degree of certainty required for successful coordination (because countries only control their emissions directly, while the effect of the emissions on temperature is uncertain). Again, it would be worthwhile to study how introducing uncertainty about the threshold, which would turn the situation from a coordination game into a prisoner's dilemma, affects success rates in the current experimental setting. While I would expect success rates to drop across treatments, I would expect the main treatment effect to be robust to such a change in parameters. Third, decisions in my experiment are made by individuals, while governments are the decision makers in climate negotiations. What individuals perceive as fair matters, however, since governments—at least in democracies—consider the public opinion prevailing in their country when sitting at the negotiation table. Not least because the outcomes of the negotiations have to be translated into domestic policy measures, for example, on energy supply and use, which directly affect the public (Gampfer 2014). Fourth, one has to consider the—often considerable—power of special interest groups who would lose from a climate agreement (e.g. the oil industry) and lobby against a climate agreement although this might be in the better interest of the country in general (Stiglitz 2015). Fifth, participants in the experiment were students, who are not representatives of the general population, and care has to be taken when generalizing the results (Henrich et al. 2010). Further aspects not considered in the experiment are the potential gradual benefits from investing in climate change prevention, for example, savings due to a more efficient use of energy, or the potential gradual consequences of climate change as opposed to the catastrophic event modeled in this study. For experimental studies considering the first see Barrett and Dannenberg (2012); for experiments on the latter see Barrett and Dannenberg (2014a) or Freytag et al. (2014).

This study was motivated by conflicting interests in international climate negotiations between developing countries, on the one hand, and developed countries, on the other hand. The bold categorization of developing countries as poor and more vulnerable and developed countries as rich and less vulnerable, assuming a perfect correlation of wealth and vulnerability, admittedly disregards the large heterogeneity within these groups of countries. Note, however, that wealth and vulnerability are important criteria for the formation of coalitions of countries who work together in the climate negotiations. An example is the group of Small Island Developing States, a coalition of 40 developing island states that are particularly vulnerable to a rise in sea-levels (UNFCCC 2021). Note, also, that conflicting interests arising from heterogeneity in wealth and vulnerability exist not only

on the country level, but also on the individual level. This notion is also supported by the Intergovernmental Panel on Climate Change, which states that risks from climate change are "generally greater for disadvantaged people and communities in countries at all levels of development" (IPCC 2014).

Appendix 1: Pledge Mechanism

Subjects in the experiment were asked at the end of round 3 and round 7 to make non-binding pledges about their contribution intentions for the remaining rounds. Comparing subjects' pledges with their actual contributions in the respective rounds, I find that subjects in general respected their pledges. The two upper panels in Fig. 5 show that observations scatter closely around the diagonal line in successful groups. Wilcoxon signed rank tests find no significant differences between pledges and contributions in round 3 or round 7 ($p = 0.707$ and $p = 0.667$, respectively). In failing groups, depicted in the two lower panels of Fig. 5, contributions tended to be significantly lower than pledges in round 3 (Wilcoxon signed rank test, $p = 0.000$). While pledges in round 3 were not significantly different between successful and failing groups (Wilcoxon rank-sum test, $p = 0.380$), pledges in round 7 were significantly lower in failing groups (Wilcoxon rank-sum test, $p = 0.000$). There are no significant treatment differences.

These findings suggest that pledges—if respected—can serve as a useful coordination mechanism allowing groups to reach the threshold. This is in line with Tavoni et al. (2011) who found that groups that are able to communicate through a simple pledge mechanism
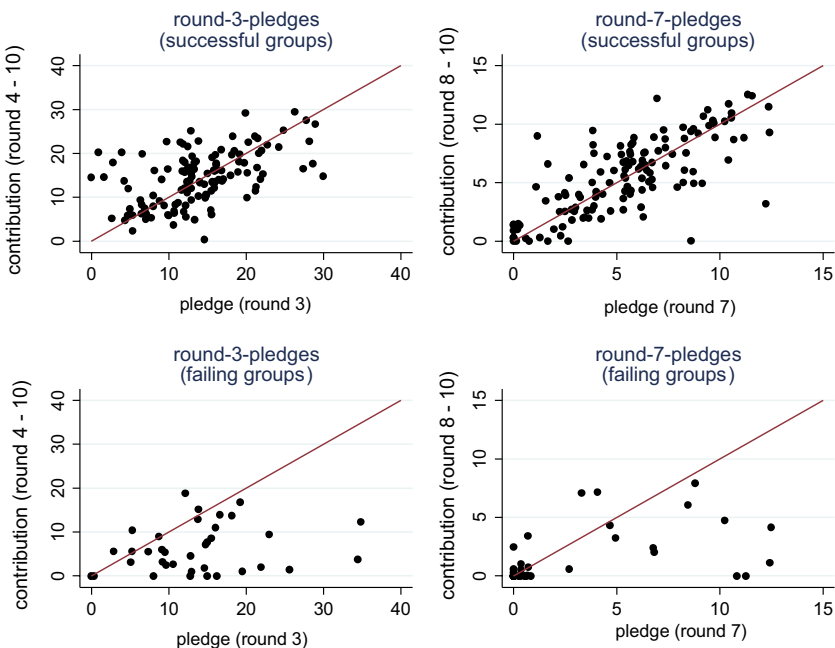

Fig. 5 Contribution-pledge comparison for successful (top) and failing groups (bottom)

are significantly more successful in reaching the threshold than groups that have no means to communicate.

Appendix 2: Experimental Instructions

Sample Instructions for WLT-VUL [Translated from German]

Welcome to the experiment and thank you for your participation. Please do not talk to the other participants of the experiment.

During the experiment, you and the other participants will be asked to make decisions. Your payout depends on your own decisions and the decisions of the other participants. You are paid individually, in private and in cash after the experiment. During the experiment, you will not be using Euro, but tokens (game coins). They are converted after the experiment at the following exchange rate: 1 token = 0.25 Euro.

Please take sufficient time to read the explanations and to make your decisions. The experiment is completely anonymous. Neither you nor the other participants will be told the identity of the players, either during the experiment or after it.

If you have questions, please raise your hand. One of the lab assistants will come to you and answer your question. The experiment will take about 60 min.

Proceedings of the Experiment

At the beginning of the experiment, all participants are randomly divided into groups of 4 players. They stay in the same group throughout the experiment and play with the same people. All members of a group receive the same information and face the same decisions. All decisions in the experiment are anonymous. For this purpose, each player is randomly assigned a player ID (a letter between A and D). The player IDs remain the same throughout the experiment. Your own player ID will be displayed in the top right-hand corner of your screen throughout the experiment (see figure on page 2).

In each group there are two purple players and two orange players. At the beginning of the experiment you will learn your own type of player as well as the type of player of the other members in your group. You can recognize the player type by the color in which the player ID is written (see figure on page 2).
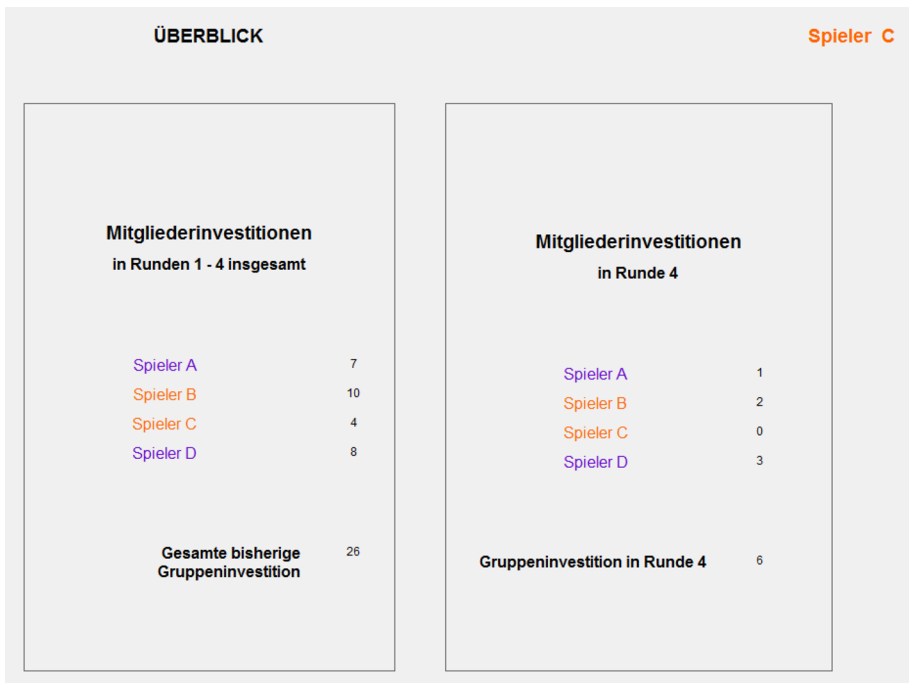
At the beginning of the experiment, all participants will receive an initial endowment of 40 tokens. Earnings during the experiment are added to this endowment while losses are deducted.

The experiment consists of 10 rounds. At the beginning of each round each player receives a budget (round budget). The exact amount of tokens a player receives as a round budget depends on the type of player. Purple players receive 5 tokens, orange players 3 tokens. From this round budget you can invest an amount of your choice in a joint project with the other group members. Purple players can invest between 0 and 5 tokens in the group project, orange players between 0 and 3 tokens. Any token that you do not invest in the group project will be added to your private wealth. In total, purple players possess round budgets of 50 tokens (10 round × 5 tokens per round = 50 tokens), while orange players possess a total of 30 tokens (10 round × 3 tokens per round = 30 tokens).

After all members of a group have made an investment decision, all members will receive the following information (see figure below):

(1) Amount of tokens each group member has invested in the group project in the current round (member investment in round x)
(2) Amount of tokens all group members together have invested in the group project in the current round (group investment in round x). This equals the sum of all investments in (1).
(3) Amount of tokens that each group member has invested in the group project in all rounds played so far (total member investment in rounds 1—x).
(4) Amount of tokens that all group members have jointly invested in the group project in all rounds played so far (total group investment so far). This equals the sum of all investments in (3).

ÜBERBLICK		Spieler C
Mitgliederinvestitionen in Runden 1 - 4 insgesamt		**Mitgliederinvestitionen** in Runde 4
Spieler A	7	Spieler A — 1
Spieler B	10	Spieler B — 2
Spieler C	4	Spieler C — 0
Spieler D	8	Spieler D — 3
Gesamte bisherige Gruppeninvestition	26	**Gruppeninvestition in Runde 4** — 6

All tokens from the round budget that you do not invest in the group project will be added to your initial endowment. During the experiment, you will be kept informed about the current amount of your private wealth.

Private wealth = initial endowment + tokens from the round budgets you did not invest in the group project

Payment

Your payoff at the end of the experiment will depend on how many tokens your group has invested in the group project in total in all 10 rounds. There is a target value of 80 tokens. If this target value is reached, all members of the group will be paid the private wealth they accumulated up to the 10th round.

If the total investment in the group project at the end of round 10 is less than the target value of 80 tokens, each group member loses a certain amount of tokens. The lost tokens are deducted from the private wealth and the remaining wealth is paid out. The exact amount of tokens a player loses depends on the player type: Purple players lose 20 tokens, orange players lose 40 tokens.

Examples

Example 1:   You are an orange type and you have invested a total of 23 tokens in the group project. Your group has invested in total more than 80 tokens in the group project, i.e. the target value has been reached. Your payout is therefore 40 tokens (initial endowment) + 7 tokens (uninvested round budget [30-23]) = 47 tokens.

Example 2:   You are an orange type and you have invested a total of 23 tokens in the group project. Your group has invested in total less than 80 tokens in the group project, i.e. the target value was not reached. Your payout is therefore 40 tokens (initial endowment) + 7 tokens (uninvested round budget [30-23]) − 40 tokens (loss for orange types) = 7 tokens.

Example 3:   You are a purple type and you have invested a total of 23 tokens in the group project. Your group has invested in total more than 80 tokens in the group project, i.e. the target value has been reached. Your payout is therefore 40 tokens (initial endowment) + 27 tokens (uninvested round budget [50-23]) = 67 tokens.

Example 4:   You are a purple type and you have invested a total of 23 tokens in the group project. Your group has invested in total less than 80 tokens in the group project, i.e. the target value was not reached. Your payout is therefore 40 tokens (initial endowment) + 27 tokens (uninvested round budget [50-23]) − 20 tokens (loss for purple types) = 47 tokens.

Payment (if target value reached) = private wealth
Payment (if target value not reached) = private wealth – loss

Announcements

Twice during the experiment, at the end of round 3 and at the end of round 7, you will have the opportunity to tell the other group members how many tokens you are planning to invest in the remaining rounds in the group project. This announcement is not binding, i.e. you are not required to actually invest the announced amount. Your announcement and other members' announcements will be shown to all group members before the next round begins.

End of the Experiment

At the end of the experiment your payout will be converted at an exchange rate of 1 token = 0.25 Euro. Before you receive your payment, we will ask you to complete a questionnaire. You will receive an extra payment of 5 Euro for answering these questions. Please take your time to answer the questionnaire. The payment will be made only after all participants have completed the questionnaire.

After having read and understood these explanations and the summary on the next page, please press NEXT. Please raise your hand if you have questions.

Summary

The experiment is played in groups of 4 players (2 purple players, 2 orange players) and consists of 10 rounds.

	Purple player	orange player
Endowment	40 tokens	40 tokens
Round budget per round	5 tokens	3 tokens
Total round budget	50 tokens	30 tokens
Loss if target value not reached	20 tokens	40 tokens

Investments: In each round you can invest tokens from the round budget in a group project. Tokens that you do not invest in the group project will be added to your private wealth.

Private wealth: initial endowment + tokens from the round budgets you did not invest in the group project

Target value: 80 tokens

Payment: Payment (if target value reached within 10 rounds) = Private wealth.

Payment (if target value not reached within 10 rounds) = Private wealth–loss.

Appendix 3: Descriptive Statistics

Table 5 provides an overview of basic descriptive statistics for the variables sex, age, cognitive abilities, risk preferences, positive reciprocity, negative reciprocity and political attitude. Randomization across treatments with respect to these variables was successful (Wilcoxon rank sum tests, $p > 0.1$ for all treatments and variables).

Risk preferences were elicited by providing subjects with five interdependent hypothetical choices between a lottery and a safe option. I followed the more time-efficient "staircase" procedure suggested by Falk et al. (2016), which results in a measure for risk preferences on a scale from 0 to 31, with higher values indicating more risk-lovingness. Cognitive abilities were elicited using a test by Toplak et al. (2014) consisting of four questions, with more questions answered correctly indicating higher cognitive abilities. Inspired by Dohmen et al. (2009), the measures for positive and negative reciprocity are based on an individual's agreement with a statement on a scale from 0 ('does not apply at all') to 10 ('totally applies'). The statements for positive reciprocity cover willingness to return a

Table 5 Descriptive statistics

	BASE	VUL	WLT	WLT-VUL
Sex (% female)	0.5	0.59	0.55	0.63
Age (average)	23.7	24.25	24.27	24.05
Cognitive abilities (average; 1–4)	1.84	1.7	1.48	1.84
Risk preferences (average; 1–32)	12.28	11.86	12.61	12.58
Positive reciprocity (average; 0–10)	7.5	7.2	7.17	7.34
Negative reciprocity (average; 0–10)	5.08	5.18	4.98	4.92
Political attitude (average; 1–10)	4.2	4.07	4.52	4.05

favor, the one for negative reciprocity the willingness to take revenge for a serious wrong. To measure political attitude, subjects were asked to describe their political position on a scale from 1 (right) to 10 (left).

Appendix 4: Determinants of Individual Contribution Behavior

The OLS regression model depicted in Table 6 confirms the results presented in Sect. 5.4. In successful groups (model (1)) the coefficients for more vulnerable types in *VUL,* rich types in *WLT* and rich/less vulnerable types in *WLT-VUL* are positive and significant, indicating higher contributions of these types relative to subjects in *BASE*. The coefficients of the other types are negative and also significant. Subjects' ex-ante preferences regarding their own type have a positive and significant effect on individual contributions.

In failing groups (model (3)) the contributions of all types deviate negatively from *BASE* (remember that in the case of failing groups contributions were the highest in *BASE*). Note also the sharp drop in the value of R^2 from 0.552 in model (1) to 0.149 in model (3), which emphasizes the strong differences in the contribution behavior of types, and thus their explanatory power in successful groups. Ex-ante preferences have no significant effect on individual contributions in failing groups.

Models (2) and (4) include several variables from the post-experimental survey that might affect individual contribution behavior: risk preferences, cognitive abilities, preferences for positive and negative reciprocity, sex and political attitude. There is, however, no evidence that any of these variables affects individual contributions in successful groups. This is also true for failing groups with the exception of positive reciprocity. An additional regression using first round contributions as dependent variable also finds no significant effect of these variables.

Table 6 OLS regression—determinants of individual contribution behavior

DV: ∑ individual contributions	Successful groups		Failing groups	
	(1)	(2)	(3)	(4)
Type: more vuln. (*VUL*)	4.469[+]	3.147	−5.849	−4.661
	(2.209)	(2.288)	(3.864)	(4.220)
Type: less vuln. (*VUL*)	−3.693*	−3.995*	−6.888*	−10.000**
	(1.514)	(1.839)	(2.329)	(2.006)
Type: poor (*WLT*)	−5.775***	−6.136***	−7.002[+]	−1.797
	(1.090)	(1.054)	(3.504)	(2.191)
Type: rich (*WLT*)	5.553***	4.619**	−5.177	−7.504
	(1.170)	(1.306)	(5.013)	(4.392)
Type: poor/more vuln. (*WLT−VUL*)	−4.786**	−4.604**		
	(1.437)	(1.476)		
Type: rich/less vuln. (*WLT-VUL*)	3.866*	3.716*		
	(1.621)	(1.733)		
Ex-ante preferences y_i own type	2.419**	2.530**	−0.206	4.847
	−0.847	(0.894)	(4.248)	(2.619)
Risk		0.054		−0.223
		(−0.062)		(−0.192)
Cognitive ability		−0.673		−0.723
		(−0.543)		(−0.708)
Positive reciprocity		0.061		2.135*
		(−0.284)		(−0.872)
Negative reciprocity		0.004		−0.313
		(−0.198)		(−0.823)
Sex		0.211		0.093
		(−1.038)		(−1.331)
Political attitude		0.108		0.131
		(−0.227)		(−0.539)
Constant	15.558***	14.784***	16.661[+]	−5.333
	(1.793)	(2.988)	(8.649)	(6.416)
Observations	132	119	36	32
R-squared	0.552	0.595	0.149	0.474

Standard errors clustered on the group level, clustered standard errors in parentheses; [+] $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: lower *n* in models (2) and (4) due to cognitivity ability variable being coded as missing if subjects did not answer the respective questions within a given time limit

Appendix 5: Beliefs

This section complements the analysis provided in Sect. 5.5 with additional results on the role of subjects' beliefs about group contribution and threshold achievement.
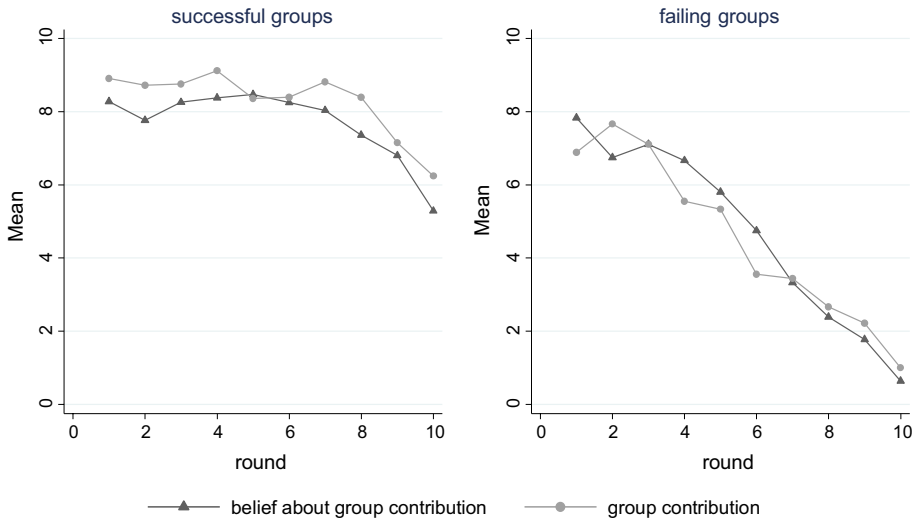
Fig. 6 Group contribution and beliefs about group contribution, by group success

Beliefs About Group Contribution

From the literature on conditional cooperation we know that a large fraction of people makes their contributions contingent on the contributions of others or their expectations regarding others' contributions (Fischbacher et al. 2001; Gächter 2006). The results of the panel model presented in Sect. 5.5 show that in this experiment subjects also contribute conditionally. In addition, there is a strong, positive and significant correlation between subjects' beliefs about group contribution and individual contributions in the respective round (Spearman's $r = 0.505$, $p = 0.000$).

Subjects' beliefs about group contribution tend to be quite accurate, although there is a significant difference between beliefs and actual contributions (Wilcoxon signed rank test, $p = 0.000$). More precisely, beliefs were on average equal to 7.05 tokens per round ($sd = 3.29$) compared to an average actual group contribution of 7.49 tokens per round ($sd = 3.01$). Figure 6 shows how group contributions and the respective beliefs evolved over time in successful and failing groups. Note that beliefs in the first round were not significantly higher in successful groups (8.28 tokens) than in failing groups (7.93 tokens; Wilcoxon rank-sum test, $p = 0.121$). However, group contributions in the first round were significantly larger in successful groups (8.91 tokens) than in failing groups (6.89 tokens) (Wilcoxon rank-sum test, $p = 0.000$). Remember that groups needed an average contribution of 8 tokens per round to reach the threshold by round 10. While only 4 out of 33 successful groups (12.1%) contributed less than the required amount in the first round, this share was significantly larger in failing groups, where 5 out of 9 groups (55.5%) failed to contribute the required sum in the first round. This can be considered evidence for a signaling function of the first round of contributions since they demonstrate subjects' willingness to contribute. Figure 6 also shows that groups did a good job of updating their beliefs. The correlation between group contribution in $t-1$ and subjects' beliefs about group contributions in t is strong and highly significant (Spearman's $r = 0.493$, $p = 0.000$).

Beliefs About Threshold Achievement

The results of the panel model presented in Sect. 5.5 show that subjects were hesitant in adjusting their beliefs about the likelihood of reaching the threshold unless it was definitely out of reach. In that case, subjects adjusted their beliefs and contributions immediately and drastically. The following complements this analysis.

Initially, subjects were in general very optimistic about the threshold being reached. On a scale from 1 (very unlikely) to 7 (very likely), subjects assessed the likelihood on average with 5.23 behind a veil of ignorance and with 5.19 in round 1 after being informed about their type. Notably, members of successful groups were significantly more optimistic in the first round about threshold achievement than members of failing groups (5.29 vs. 4.83; Wilcoxon rank-sum test, $p = 0.018$). Also, final treatment differences in success rates were already reflected in initial beliefs: Subjects were most optimistic about reaching the threshold in *WLT-VUL* (5.4 behind the veil of ignorance, 5.45 in round 1) and least optimistic in *VUL* (4.89 both behind the veil of ignorance and in round 1). The difference is weakly significant (Wilcoxon rank-sum tests, $p = 0.016$ behind the veil and $p = 0.013$ for round 1) for these two treatments and insignificant for all other treatment comparisons.

These results lend support to the hypothesis that threshold beliefs affect group success, even though subjects do not adjust their beliefs continually between rounds. Instead, initial beliefs seem to play a crucial role, in particular because they are correlated with first round contributions, which in turn have a signaling effect as has been shown in the first part of this section.

References

Bagnoli M, McKee M (1991) Voluntary contribution games: efficient private provision of public goods. Econ Inq 29:351–366

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:878–894

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

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

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

Barrett S, Dannenberg A (2014b) Negotiating to avoid "Gradual" versus "Dangerous" Climate change: an experimental test of two prisoners' dilemmas. In: Cherry TL, Hovi J, McEvoy D (eds) Towards a new climate agreement: conflict, resolution, and governance. Routledge, London, pp 61–75

Benedick RE (1998) Ozone diplomacy. Harvard University Press, Cambridge

Bernard M, Reuben E, Riedl A (2014) Fairness and coordination: the role of fairness principles in coordination failure and success

Bicchieri C (2005) The grammar of society: the nature and dynamics of social norms. Cambridge University Press, Cambridge

Bochet O, Page T, Putterman L (2006) Communication and punishment in voluntary contribution experiments. J Econ Behav Organ 60:11–26

Brick K, Visser M (2015) What is fair? An experimental guide to climate negotiations. Eur Econ Rev 74:79–95

Brown TC, Kroll S (2017) Avoiding an uncertain catastrophe: climate change mitigation under risk and wealth heterogeneity. Clim Change 141:155–166

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

Carlsson F, Kataria M, Krupnick A et al (2013) A fair share: burden-sharing preferences in the United States and China. Resour Energy Econ 35:1–17

Carraro C, Siniscalco D (1998) International Institutions and Environmental Policy: International Environmental Agreements: Incentives and Political Economy. Eur Econ Rev 42:561–572

Croson R, Marks MB (2000) Step returns in threshold public goods: a meta- and experimental analysis. Exp Econ 2:239–259

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

DARA (2012) Climate Vulnerability Monitor 2nd Edition: A Guide to the Cold Analysis of a Hot Planet. DARA and the Climate Vulnerable Forum, Madrid

Dohmen T, Falk A, Huffman D, Sunde U (2009) Homo reciprocans: survey evidence on behavioural outcomes. Econ J 119:592–612

Falk A, Becker A, Dohmen T et al (2016) The preference survey module: a validated instrument for measuring risk, time, and social preferences. IZA Discuss Pap No. 9674:

Finus M (2008) Game theoretic research on the design of International Environmental Agreements: insights, critical remarks, and future challenges. Int Rev Environ Resour Econ 2:29–67

Fischbacher U (2007) z-Tree: Zurich Toolbox for ready-made economic experiments. Exp Econ 10:171–178

Fischbacher U, Gächter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. Econ Lett 71:397–404

Freytag A, Güth W, Koppel H, Wangler L (2014) Is regulation by milestones efficiency enhancing? An experimental study of environmental protection. Eur J Polit Econ 33:71–84

Fuessel HM (2010) How inequitable is the global distribution of responsibility, capability, and vulnerability to climate change: a comprehensive indicator-based assessment. Glob Environ Chang 20:597–611

Gächter S, Renner E (2010) The effects of (Incentivized) belief elicitation in public goods experiments. Exp Econ 13:364–377

Gächter S (2006) Conditional cooperation: behavioral regularities from the lab and the field and their policy implications. CeDEx Discuss Pap 2006-03:

Gampfer R (2014) Do individuals care about fairness in burden sharing for climate change mitigation? Evidence from a lab experiment. Clim Change 124:65–77

Greiner B (2015) Subject pool recruitment procedures: organizing experiments with ORSEE. J Econ Sci Assoc 1:114–125

Hayward T (2012) Climate change and ethics. Nat Clim Chang 2:843–848

Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? Behav Brain Sci 33:61–135

IPCC (2014) Climate change 2014: Synthesis Report. Geneva

Konow J (2000) Fair shares: accountability and cognitive dissonance in allocation decisions. Am Econ Rev 90:1072–1091

Konow J (2003) Which is the fairest one of all? A positive analysis of justice theories. J Econ Lit 41:1188–1239

Konow J (2010) Mixed feelings: theories of and evidence on giving. J Public Econ 94:279–297

Lange A, Löschel A, Vogt C, Ziegler A (2010) On the self-interested use of equity in international climate negotiations. Eur Econ Rev 54:359–375

Marks MB, Croson RTA (1999) The effect of incomplete information in a threshold public goods experiment. Public Choice 99:103–118

Milinski M, Sommerfeld RD, Krambeck H-J et al (2008) The collective-risk social dilemma and the prevention of simulated dangerous climate change. Proc Natl Acad Sci 105:2291–2294

Molina M, McCarthy J, Wall D, et al (2014) What we know: the reality, risks, and response to climate change. Am Assoc Adv Sci

Nikiforakis N, Noussair CN, Wilkening T (2012) Normative conflict and feuds: the limits of self-enforcement. J Public Econ 96:797–807

Nordhaus WD (2018) Climate change: the ultimate challenge for economics. Nobel Prize Committee

Rapoport A, Suleiman R (1993) Incremental contribution in step-level public goods games with asymmetric players. Organ Behav Hum Decis Process 55:171–194

Rauchdobler J, Sausgruber R, Tyran J-R (2010) Voting on thresholds for public goods: experimental evidence. Finanz/public Financ Anal 66:34–64

Reuben E, Riedl A (2013) Enforcement on contribution norms in public good games with heterogeneous populations. Games Econ Behav 77:122–137

Ringius L, Torvanger A, Underdal A (2002) Burden sharing and fairness principles in international climate policy. Int Environ Agreements 2:1–22

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

Stern N (2007) The economics of climate change: the stern review. Cambridge University Press, Cambridge, UK

Stiglitz JE (2015) Overcoming the Copenhagen failure with flexible commitments. Econ Energy Environ Policy 4:29–36

Tavoni A, Dannenberg A, Kallis G, Löschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. Proc Natl Acad Sci 108:11825–11829

Toplak ME, West RF, Stanovich KE (2014) Assessing miserly information processing: an expansion of the cognitive reflection test. Think Reason 20:147–168

UNFCCC (2014) Decision CP.20: Lima Call for Climate Action

UNFCCC (2021) Party Groupings. https://unfccc.int/party-groupings. Accessed 24 Sep 2021

United Nations (1992) United Nations Framework Convention on Climate Change