ERF 31ST Annual Conference

YOUTH

DEMOGRAPHIC DIVIDEND & MIGRATION IN MENA:
CHALLENGES & OPPORTUNITIES IN UNCERTAIN TIMES

reshaping the future

2025

Climate Change Negotiations Under the Shadow of History

A Computable General Equilibrium Analysis

Sheheryar Banuri,
Ha M. Nguyen
Ernest J. Sergenti

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

Climate change negotiations under the shadow of history

Sheheryar Banuri, Ha M. Nguyen, and Ernest J. Sergenti[1]

December 2024

DRAFT: Please do not cite without permission

Abstract

Climate change is a global problem which requires unprecedented levels of collective action to solve. We ask the following question: do historical factors and the debate over climate justice help or hurt climate change negotiation? While it is important to be cognizant of how the current situation arose, the shadow of history constitutes a form of sunk cost that has little bearing on the need for mitigation globally. Using a lab experiment with subjects assigned to rich and poor countries, we examine how subjects decide how much to contribute to a mitigation fund and who among them should bear more of the costs. A key feature of our design is that the need for mitigation is triggered based on the previous actions of our subjects, who undertook these actions without knowledge of their impact on the need for mitigation. We conduct two conditions, a baseline where the cause for mitigation (past actions) is not revealed to subjects and a treatment (the shadow of history) where this cause is revealed to the subjects. In both conditions, subjects negotiate how much to contribute to a mitigation fund. While we find evidence that total group contributions increase only slightly under the shadow of history, the *distribution* of those contributions changes markedly. When made aware of the historical causes of the climate problem, poor countries significantly reduce their contributions, while rich countries contribute slightly more – offsetting the reduction from the poor countries. Moreover, the slight increase in total contributions to the mitigation fund lowers the probability of disaster, raising total expected earnings. Poor countries see their welfare increase as they contribute less to the fund yet reap the benefits of higher total contributions, as they suffer more from climate disasters. Regarding policy implications, "naming and shaming" rich countries for their previous economic activities thus appears to be a successful negotiations tactic for poor countries.

[1] Banuri: School of Economics, University of East Anglia and Hughes Hall, University of Cambridge (e-mail: s.banuri@uea.ac.uk); Nguyen: IMF (HNguyen7@imf.org); Sergenti: World Bank (esergenti@worldbank.org). The authors have no relevant or material financial interests that relate to the research described in this paper. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors.

1. Introduction

Climate change is a big challenge for humanity. Temperatures are rising, with the global average temperature already about 1.2 Celsius higher than pre-industrial levels. Droughts, wildfires, and massive storms are occurring more frequently with devastating effects. Rising temperatures are also leading to a rise in world sea levels. Taken together, these can lead to biodiversity loss, food and water scarcity, and an increase in disease prevalence. Together, different manifestations of climate change bring devastating effects on the global economy and global livelihoods.

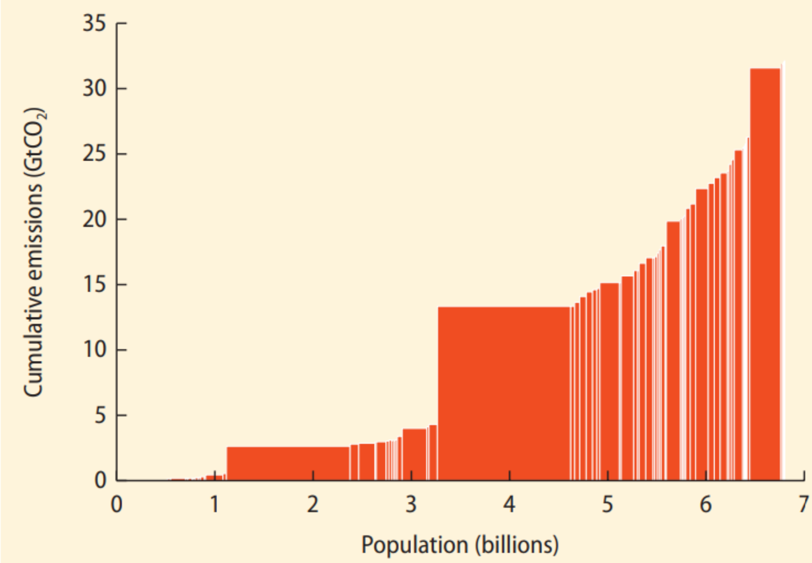
To reduce the causes of climate change and limit the increase in the average global temperature, humanity is confronting a dauting social dilemma to quickly reduce greenhouse gas (GHG) emissions such as carbon dioxide and methane. Reducing GHGs is challenging for several reasons. First, our economic and energy systems currently rely heavily on fossil fuels (coal, oil, and natural gas), which are major sources of GHG emissions. Transitioning to cleaner alternatives, such as renewable energy sources, requires significant investment, infrastructure development, and policy changes. These large upfront costs can be financially challenging for businesses, individuals, and governments, especially in low-income countries.

Second, climate change is a global problem that requires international cooperation and coordination. GHG emissions carry a negative externality: the social cost of GHG emissions, which includes pollution and intensifying climate change to the entire world, is much larger than the private cost of carbon. Similarly, reducing GHG emissions carries a positive externality: while the cost of investment to reduce GHG emissions is borne by the country or firm that implements the investment, the benefit is reaped by the whole world (by reducing the risks of climate change). For most countries, the private benefit of *unilateral* climate change mitigation is smaller than its private cost. Thus, for most countries, it does not make economic sense (in terms of cost and benefits) to *unilaterally* reduce GHG emissions, yielding a social dilemma. Therefore, collective action from countries worldwide is needed to achieve meaningful emission reductions.

Negotiating and implementing global agreements to reduce GHGs are thus at the core of humanity's strategy to fight climate change. Nevertheless, negotiations are a complex and challenging process due to differences in the priorities, interests, and capacity of countries.

A core issue is the inequality of GHG emissions. Rich countries historically emitted much more GHG than poor countries, with the poorest 1 billion people historically emitted less than 1% of GHG (Figure 1). Sadly, poor countries and the poor population within a country are the most exposed to climate change. Their income sources, such as from agriculture, outdoor services, and construction, are more vulnerable to climate change. And they do not have as much capacity as rich countries to adapt to and cope with the impacts of climate change. Because of this "climate inequality", poor countries have been demanding "climate justice". They argue, because rich countries historically emitted much more carbon dioxide, they have the responsibility to compensate the poor countries for their past emissions, or to assist the poor countries with adaptation and the energy transition, both financially and technologically.

Figure 1: The poorest 1 billion people historically emitted less than 1% of GHG



Source: World Bank (2020).[2]
Note: Cumulative population ranked by income on the horizontal axis; cumulative carbon emissions on the vertical axis. Each rectangle represents a country. GtCO2 = gigatons of carbon dioxide.

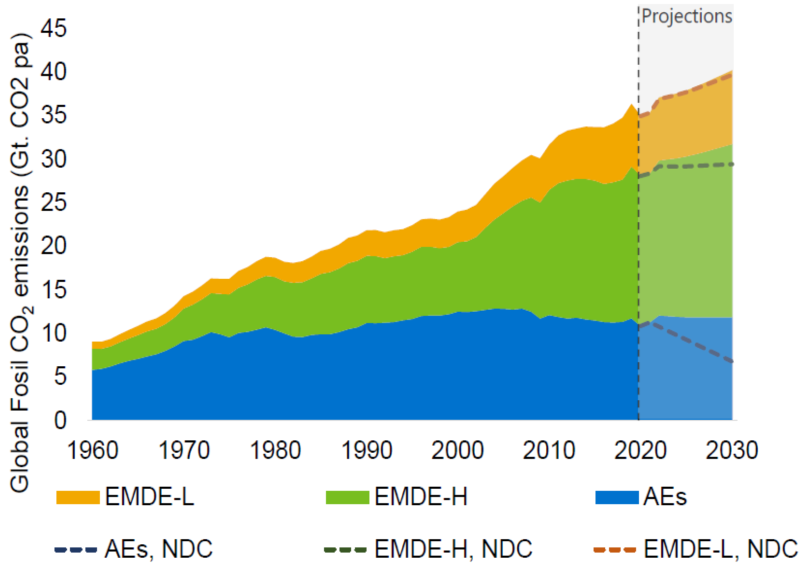
Poor countries have had some success in securing commitment by rich countries. At the 15th Conference of Parties (COP15) of the UNFCCC in Copenhagen in 2009, developed countries committed to a collective goal of mobilizing USD 100 billion per year by 2020 (including loans and grants) for climate action in developing countries. The goal was formalized at COP16 in Cancun and the total amount was increased to USD 300 billion at the most recent COP29 in Baku. Moreover, at COP27, an agreement was reached to provide "loss and damage" funding for vulnerable countries hit hard by climate disasters. However, commitments have not been followed through by actions. Rich countries failed to meet their climate financing target of USD 100 billion to the poorer countries per year by 2020. And the "loss and damage" fund as of December 2023 had only $700 million, which could cover less than 0.2% needed.[3]

Some rich countries, while acknowledging past emissions, argue that the responsibility for emissions reduction should be more evenly shared, taking into account current levels of emissions. In fact, since around the year 2000, poorer countries have been responsible for more emissions than rich countries, and that trend is forecast to continue (Figure 2). Rich countries can also plausibly claim that they used the best technologies available in the past and were not aware that GHG emissions would cause climate change. It is not just emissions that allowed rich countries to develop economically. Innovation, effort, and strong institutions also played a major part in the process. As such, some argue that all countries should have equal obligations and responsibilities in addressing climate change, regardless of their level of development, and that developing countries should also take actions to reduce their emissions and contribute financially to global climate efforts.

[2] World Bank (2020), p. 56, figure 6.3. https://openknowledge.worldbank.org/bitstream/handle/10986/22787/9781464806735.pdf

[3] https://www.theguardian.com/environment/2023/dec/06/700m-pledged-to-loss-and-damage-fund-cop28-covers-less-than-02-percent-needed

Figure 2: Historical GHG emission by country group



Source: Global Carbon Budget (2021), UNFCCC (2021a), World Bank (2021) and IMF staff (IMF Staff Climate Notes 2021/005).[4]
Note: NDC = nationally determined contributions; AEs = advanced economies; EMDE-H and EMDE-L = higher-income and lower-income emerging and developing economies.

This paper asks the following question: do historical factors and the debate over climate justice help or hurt climate change negotiation. This is a critical question because the debate over climate justice based on historical emissions could be contributing to impasses in international negotiations and the delayed action in solving the collective action problem to reduce global GHG emissions.

There are theoretical points for both arguments. On the one hand, the debate over climate justice is important to poor countries who feel that past historical emissions are unfair. Any negotiation of global actions would not be possible without resolving the issue of historical responsibility. On the other hand, it is not clear if "name and shame" facilitates or impedes global negotiations over collective action to mitigate climate change.

To answer this all-important question, we use a behavioral lab experiment framework. We find that facing the exact same mitigation problem, groups of subjects who are made aware of the cause of the problem (i.e., unequal distribution of historical emissions) contribute slightly more than groups not knowing this history – although this finding is not statistically significant. This results in a lower likelihood of climate disaster and higher total economic earnings – with poor countries seeing an absolute increase in their expected earnings.

Poorer participants in the treatment where they can associate climate change to past activities and emissions, especially from the rich participants, reduce their mitigation contribution significantly compared to the baseline where they do not connect climate change to past activities and emissions. Richer participants on the hand increase their contributions when they were made aware of the linkage between past activity emissions and climate change. Their increase in contribution offset the poor's decrease in contributions.

[4] https://www.imf.org/-/media/Files/Publications/Staff-Climate-Notes/2021/English/CLNEA2021005.ashx

2. Literature review

Our paper contributes to a very nascent lab experiment literature examining the role of historical emissions on climate change mitigation. The closest work is Kline et. al. (2018). Emphasizing that current rich countries are wealthier than poor countries for the most part because rich countries emitted more greenhouse gases in the past, Kline et. al. 2018 designed an experiment that attempts to isolate the effects of perceived historical responsibility on the outcome of climate negotiations from the wealth differences of the parties. They do this with what they refer to as the "compound climate dilemma" – combining an economic development phase with a subsequent mitigation phase, similar to the Milinksi et. al. (2008) disaster game.

Before analyzing the Kline et. al. model in detail, it's important to describe the main features of the influential Milinksi et. al. (2008) model, commonly referred to as the "disaster game".[5] In the disaster game, Milinski et al. (2008) employ a threshold public goods game. 180 undergraduate students in Germany were anonymously placed into 30 groups of 6 subjects. Subjects played the game privately in the computer lab, and each subject started with €40, for a total group endowment of €240. Groups were told that they faced the risk of devastating climate change with varying probabilities of disaster (90%, 50%, and 10%) and, if disaster struck, they would lose their entire remaining endowment. To prevent disaster from occurring with certainty, groups needed to contribute at least €120, i.e., half of the group's total initial endowment. The game was played for 10 rounds, with subjects able to contribute €0, €2, or €4 each round. All contributions made to the disaster mitigation fund were lost, i.e., non-refundable. After each round, all subjects were shown a listing of contributions by player, so they could see what strategies other subjects in the group were following. *Subjects were not allowed to communicate with one another* – a major difference from our research design, which we describe below.

The authors report findings for the 3 treatments: probability of disaster of 90%, 50%, and 10%. With a probability of disaster of 90%, 5 of the 10 groups achieved the target of €120 after 10 rounds of play, and the other 5 groups got very close to the target. The average group contribution for all 10 groups was €113. With the 50% and 10% probability-of-disaster treatments, by contrast, only one group out of 10 reached the target with the 50% treatment, and none of the groups receiving the 10% treatment reached the target. Interesting, groups that received the 10% treatments still made substantial contributions of on average €73, which represented more than 60 percent of the target. This is surprising because a game-theoretic analysis with rational preferences would have predicted that subjects should contribute nothing, given that the expected payoff of doing nothing is €36 (0.1 * €0 + 0.9 * €40) compared to a strategy of contributing one's fair share to the mitigation target of €20, which would lead to a certain payoff of €20. The authors offer as one possible explanation that subjects were highly risk adverse, but it seems more likely that another mechanism is required to explain the finding.

An ingrained sense of fairness could be one such explanation. Presenting results by round over time, the authors note that almost all groups started in round 1 by contributing the fair-share amount of €2. Sustained contributions of €2 each over 10 rounds would have allowed groups to meet the threshold of €120. Without communication, the fair-share contribution of €2 would seem an obvious focal point (Schelling 1960) for subjects to start with. However, as play continued over subsequent rounds, some subjects contributed €0, which other subjects noticed and modified their behavior. This led to a decline in total group contributions, especially for

[5] See Andrews, Delton, and Kline (2024) who provide a rich description of the model, as well as many other climate-related lab experiments.

groups receiving the 10% and 50% probability-of-disaster treatments. Interestingly, especially with the 90% treatment, some subjects increased their contributions to €4 to compensate for the free riders and help the group reach the threshold. But this did not always occur or to a sufficient extent, such that some groups just missed the target. The authors postulate that more altruistic players may have been reluctant to reward the perceived unfair past behavior of the more-selfish free riders, even if that entailed a personal loss for themselves.

The authors and others (e.g., Andrews, Delton, and Kline 2024) note a few take-aways from this experiment. First, a high probability of disaster is required for groups to meet the threshold target. Second, in this high-probability-of-disaster situation and in the absence of communication, subjects converge on the fair-share contribution as a focal point. Third, subjects appear to behave irrationally, at least initially, wasting a lot of money when the threshold was not met – especially with the 10% and 50% treatments.

Turning to the Kline et al. 2018 model, the authors add an initial economic development phase, where subjects harvest their endowments before playing the Milinksi et. al. (2008) disaster game. In the economic development phase, subjects were confronted with a common pool resource dilemma (Ostrom 2002, Dietz et. al. 2003). Over 10 rounds, they could harvest between $0 and $4, and after each round they learned how much each other subject had harvested. Furthermore, *subjects were informed that the more they harvested, the greater the probability of a subsequent climate disaster and the higher the threshold to avert such a disaster.* Thus, harvesting represents wealth accumulation by emitting greenhouse gases and using up a common carbon budget. Subjects could harvest a maximum of $40, with a potential group maximum of $240.

Note that this is the key difference between the design in Kline et al. (2018) and our design. In Kline et al. (2018), in their development phase, subjects know that their activities in the development phase (i.e., harvesting) have direct impacts on climate change. In our design, subjects do not know that their economic activities and wealth accumulation during the first phase will have an impact on the climate change mitigation problem that they may need to solve during the second phase. In this way, our design arguably more realistically captures the actual history of emissions and wealth accumulation: countries did not know for a long time that GHG emissions caused climate change.

In the disaster or mitigation phase in Kline et al. (2018), subjects retained the wealth they accumulated during the economic development phase. Subjects were also presented with a group probability of disaster and a threshold amount that the group could contribute to avoid disaster. In case of disaster, subjects would loss all their remaining endowment. The threshold amount was set at 53 cents per $1 harvested, i.e., 53% of the group endowment. And the probability of disaster increased as a step function as total group endowments increased. The mitigation phase lasted 10 rounds. Subjects could contribute between $0 and $4 each round to meet the threshold; and, after each round, subjects were informed of the contributions made by other subjects. As with the Milinski et. al. 2008 design, subjects were not allowed to communicate with one another.

With this initial study, Kline et. al. 2018 report that subjects did not restrain themselves much during the first economic development phase, harvesting on average $31.30 per subject out of a possible maximum of $40. Moreover, in the mitigation stage, only 7 out of 12 groups met the threshold of 53% of the group endowment to avert disaster, with an average contribution rate across all 12 groups of 52%, just below the target. This last result is interesting for several reasons. On the one hand, even though 5 groups did not meet the target, the fact that the average contribution was 52% means groups got very close to achieving the mitigation goal.

Without communication, this shows groups exhibited a high level of cooperation – with the threshold perhaps serving as a focal point. On the other hand, missing the target by a small margin is very wasteful in the disaster game setup. If subjects knew that the target would be missed, a more rational strategy would have been to contribute nothing and take their chances with the risk of disaster on their full endowment.

The above results may conflate wealth differences with responsibility. As such, Kline and co-authors ran another placebo experiment, in which subjects just played the second mitigation phase of the compound climate dilemma game. The endowments, threshold, and probability of disaster that the new 12 groups of 6 subjects faced were pre-determined, by randomly matching the new groups with the post economic development conditions of groups who had played the combined economic development and mitigation phases in the previous experiment. Kline et al. (2018) report that, with this placebo experiment, differences in endowments among group members did not prevent groups from reaching the mitigation threshold. All 12 groups met the target, with an average contribution of 54%. This confirms previous work on inequality (see Milinski, Röhl, and Marotzke 2011), which has shown that people are more willing to contribute more to a common goal if the reason for their relative richness is considered as random (e.g., Kameda et al. 2002, Cappelen et. al. 2007). Hence, Kline et. al. 2018 argue that the results they find in the non-placebo experiment were due to groups contributing too little to meeting the mitigation target where the climate dilemma was self-created, even when holding inequality constant.

In a second experiment, Kline et. al. 2018 modify the first phase of the compound climate dilemma. Here they strive to create larger wealth differences and a greater sense of unfairness or injustice by allowing some subjects more time to harvest wealth. Instead of giving all players the full 10 rounds to harvest wealth, as was the case in the first experiment, here only 3 of the 6 players are able to harvest during the full 10 rounds. These subjects are the "early developers". The other 3 players, i.e., the "late developers", must watch during the first 5 rounds of the economic development phase as the early developers harvest and may only harvest wealth during the last 5 rounds. As with the first experiment, the economic development phase is followed by a mitigation phase where all subjects play the disaster game. As with the first study, the authors also conducted a placebo experiment to attempt to control for inequality effects.

The authors argue that in this second experiment there are two ways for wealthier subjects to help the group prevent disaster during the mitigation stage. In addition to contributing more to meeting the threshold in the mitigation phase, wealthier subjects (early developers) can choose to harvest less during the economic development stage – making disaster easier to avert and less likely to happen. In fact, the authors find just that, compared to the first experiment, in which subjects harvested 78 percent of the maximum allowable amount during the first 5 rounds, early developers restrained themselves somewhat, harvesting only 65% of the maximum amount. The authors report a similar comparison for the second 5 rounds as well. By contrast, the late developers harvested 85% of their allowable amounts during the final 5 rounds.

Furthermore, in the mitigation game, groups comprised of early and late developers had a harder time averting disaster (only 4 out of 12 with an average contribution of 49%) than groups in the placebo experiment (10 out of 12 with an average contribution of 52%). Within the combined early and late developer groups, while early developers contributed *more* of their endowment (55%), late developers contributed much less, only 42% of their (smaller) endowment – leading most of the combined groups to miss the target. In post-experiment interviews, late developers cited the unfair set-up of the game as a main reason for contributing much less.

Our paper is also related to a larger literature that examines whether unequal endowments hinder cooperation to reach agreement in modified versions of the disaster game. Note that this is a related but different question to the one we examine here. We ask if the ex-post realization of unequal historical emissions hurts or helps climate change negotiation.

Tavoni et. al. 2011 build on the work of Milinski and co-authors to examine the impact of inequality. They employ a similar disaster game set-up. Groups of 6 subjects, who start with an initial endowment of €40 each, must contribute at least €120 over 10 rounds to avoid disaster, i.e., losing their remaining endowment. Unlike Milinski et al. (2008), Tavoni and co-authors use only one probability of disaster, namely the 50% probability. Also, a major aspect of their experimental design is to have the computer program make contributions for subjects during the first 3 rounds. In a control treatment, the computer contributes €2 per player (the fair-share amount that would achieve the target if sustained over all 10 rounds), while, in a treatment designed to create inequality among the subjects, the program makes contributions of €0 for 3 players and contributions of €4 for the other 3 players. The total group contribution is the same in both treatments, but the authors argue that, in the "inequality treatment", two groups of rich and poor are formed: those who contributed nothing during the first 3 rounds (rich) and those who contributed €4 per round (poor).

Tavoni et. al. 2011 find that the inequality treatment made it more difficult to prevent disaster. Whereas 50 percent of groups receiving the control treatment avoided disaster, only 20 percent of those receiving the inequality treatment did. Moreover, the authors also repeated the experiment allowing players to communicate and make non-binding commitments. In these cases, although success rates increased – with 70 percent of control groups meeting the target and 50 percent of the inequality treatment groups meeting the target – inequal groups still were less successful. These results suggest that inequality tends to reduce the likelihood of collaboration within groups. However, one important caveat is the way Tavoni et al. 2011 created their poor and rich groups. "Poor" subjects were poor because the computer program forced them to contribute to the mitigation target during the first 3 rounds. Likewise, "rich" subjects were rich because the program did not make them contribute. "Poor" and "rich" subjects were not created outside of the disaster game. Hence, "poor" subjects could have felt that they had already contributed to the mitigation target and thus felt less obligated to contribute in the subsequent 7 rounds – reducing the success rate in the inequality treatment.

Milinksi with another set of co-authors (Milinski, Röhl, and Marotzke, 2011) also examined the issue of inequality within groups. As with Tavoni et al. 2011, the set-up was similar to the Milinski et. al. 2008 disaster game, with groups of 6 subjects needing to contribute at least €120 over 10 rounds to avoid losing their remaining endowments. Differences with the Milinski, Röhl, and Marotzke (2011) experiment included subjects having 2 pots of endowments: a liquid pot that could be used as a source of funds to contribute to meeting the mitigation target and an illiquid pot that represents infrastructure, etc. that could not be used to meet the target but would be destroyed in the case of disaster. The initial endowment for poor subjects was €20 in the liquid pot and €30 in the illiquid pot and double those amounts for the rich subjects: €40 in the liquid pot and €60 in the illiquid pot. The authors implemented 3 treatments: groups with all poor subjects, groups with all rich subjects, and groups with an equal mix of rich and poor subjects.

Milinski, Röhl, and Marotzke (2011) report that inequality did not make it harder for groups to reach the threshold, with mixed poor and rich groups contributing about the same as all-rich groups, roughly just enough to meet the threshold. Interestingly, all-poor groups were less successful – perhaps because meeting the €120 target for them would have required that all 6 subjects contribute the entire €20 from their liquid pot. Comparing the results with Tavoni et. al 2011, where inequality did make reaching the target harder, Andrews, Delton, and Kline (2024) note that the Tavoni et. al. 2011 study used a probability of disaster of 50% while the Milinski, Röhl, and Marotzke study used a probability of 90%. As noted above with the original Milinksi et. al 2008 study, groups were more successful at meeting the threshold when disaster seemed more likely, i.e., when the probability of disaster was 90% compared to 50%. Andrews, Delton, and Kline (2024) also note the difference in how subjects were assigned as rich or poor. With the Tavoni et. al. 2011 study, subjects were assigned based on forced contributions during the first 3 rounds of the game, while with the Milinski, Röhl, and Marotzke (2011) study, subjects were randomly assigned. Other work has shown that people are more willing to contribute more to a common goal if the reason for their relative richness is considered as random (e.g., Kameda et al. 2002, Cappelen et. al. 2007).

Burton-Chellew, May, and West (2013) also use a disaster game to investigate unequal endowments – and add an analysis of differences in vulnerability to disaster. As with the original Milinski et al. (2008) experiment, the authors run their experiments on groups of 6 subjects with a total endowment of 240 tokens (1 token = £0.50) and a mitigation target of 120 tokens. They investigate four treatments. In the first treatment, all 6 subjects are given 40 tokens each. In the rest of the treatments, 2 subjects are given 80 tokens each and the remaining 4 subjects are given 20 tokens each. With these 3 "unequal" treatments, the authors modify the probability of risk: rich subjects face a higher risk of disaster, poor subjects face a higher risk, or both types of subject face the same level of risk. The authors argue that the case in which the poor subjects face more risk than rich subjects is most consistent with the real world. In fact, they find that disaster was rarely prevented for unequal groups when the poor had a higher risk of losing everything. In those cases, only 1 out of 8 groups (12.5%) succeeded in preventing disaster, while in the other 3 treatments (equal and unequal) 75% of groups prevented disaster. Hence, inequality by itself did not lead to a reduction in the success rate. It was only when inequality was combined with the rich having less exposure to risk that the target was less likely to be met as the richer subjects did not contribute as much to meeting the mitigation threshold.

Brown and Kroll (2017) also found that inequality did not matter for groups with unequal endowments who also faced a threshold of uncertain size and an uncertain risk of disaster. Consistent with Barret and Dannenberg (2012, 2014), Brown and Kroll (2017) found that uncertainty in the size of the threshold lowered contributions substantially, preventing all groups from meeting the threshold, while uncertainty with the risk of disaster lowered contributions only by a bit and depressed success somewhat. Inequality did not appear to matter.

Andrews, Delton, and Kline (2024) offer some summary observations for the literature on inequality in general. While on balance inequality does not appear to matter, successful cooperation is less likely when subjects face differing incentives (the rich face less of a risk than the poor as in Burton-Chellew, May, and West, 2013) or there is a potential disagreement about who is responsible for the current predicament (Tavoni et. al. 2011, Kline et. al. 2018).

3. Experimental design

The experiment consists of a baseline and a single treatment, which varies the amount of information subjects have about the triggering of the disaster (simulating a negative climate related event). Our subjects are undergraduate students at the University of East Anglia using the LEDR lab subject pool. Treatments are randomized across sessions, with 212 subjects participating in the entire experiment. Subjects were paid in "tokens," which were exchanged for GBP at the rate of 0.03 tokens per GBP. On average, subjects earned 22 GBP overall.

The experiment begins with subjects engaging in a real effort task to earn their endowment for the session. The effort task was a version of the coding task (Lévy-Garboua, Masclet, and Montmarquette, 2009; Erkal, Gangadharan, and Nikiforakis, 2011), which generated the endowment. Subjects were given 4 practice rounds, where subjects were allowed to freely engage in the task itself for no compensation. Following these practice rounds, subjects were then informed whether they were randomly assigned to be a "Type X" or a "Type Y" player. These types correspond to poor and rich countries respectively. Endowments were set to be 250 tokens for poor countries (Type X players) and 1,000 tokens for rich countries (Type Y players). Subjects were informed as to their own type, and the endowment level of the other player type as well (so as to provide information on inequality). This was true for both the baseline and the treatment.

Subjects were asked to perform the coding task but were given a set target of words to code in order to generate the endowment. Subjects were randomly assigned to rich and poor (based on their seating assignments). Subjects in the role of rich countries were provided a target of 10 words to decode in six rounds (with each round lasting 45 seconds) to earn an endowment of 1,000 tokens. Subjects in the role of poor countries were provided a target of 5 words to decode in three rounds to earn an endowment of 250 tokens. Hence, subjects in the rich country role were tasked with producing twice as much effort as poor countries but received four times the endowment. This was done to induce inequalities between countries in the two roles, where subjects are randomly assigned to rich or poor (highlighting the role of luck) and rich countries tasked with doing more (highlighting the role of effort) that contributed to the inequality. The main purpose behind inducing both luck and effort as contributing to the inequality was to simulate real-world differences between rich and poor countries, and also giving each a reason to act in an uncooperative fashion. The poor could highlight the luck aspect, while the rich could highlight the greater work effort aspect.

Note that the targets were chosen such that it would require effort to attain the endowment, but not too difficult so as to induce attrition (since subjects that were unable to achieve the target are then unable to contribute in the subsequent game). Of our 212 subjects, only 3 subjects were unable to complete the task and, thus, were subsequently dropped from the analysis.

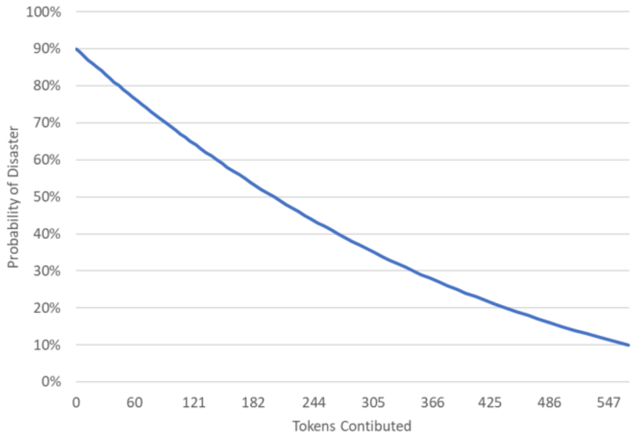
Once the coding task is completed, subjects are then asked to undertake a bonus coding task. In this task, subjects are given a low piece rate per word decoded (either 1 token per word for poor countries, or 2 tokens per word for rich countries), and can undertake the task for a maximum of 14 rounds (with each round lasting 2 minutes each). Importantly, this task is completely optional, and subjects are free to quit this task at any time. Hence, subjects have an incentive to continue this task, but the incentive is not the same across all subjects. The reason for this bonus coding task is to induce variation in their earnings, which (as explained below) is a key determinant of whether each group engages in the "disaster phase".

Once the bonus coding task is completed, subjects then engage in 10 rounds of the "main phase" of the experiment. Here, subjects are randomly assigned to groups of four, with each group containing two rich and two poor countries. Subjects are informed that in some rounds a "disaster" is triggered which has a 90% chance of occurring. If a disaster occurs, each group member loses an equal amount of 200 tokens. This means that the disaster (simulating a climate change triggered event) imposes an equal cost on each subject (and is akin to a public bad). However, subjects can pay tokens to reduce the probability of the disaster occurring. This is done via tokens contributed to a group account. The relationship between the probability of a disaster occurring and the contributions follows the formula:

$$y = \frac{2}{\left(0.03 + 0.004(p)\right)^2} + \frac{2}{\left(0.06 + 0.08(p)\right)^2} - 1751.29$$

Where y is total amount of tokens contributed to the group account, and p is the probability of a disaster occurring. This relationship is computed based on the parameters used in the experiment, and assuming a coefficient of relative risk aversion (CRRA) of 1.5 which is consistent with estimates for the UK population. As the formula is rather complicated, subjects were also provided with an online calculator which allowed them to compute the probability of disaster for different amounts of tokens contributed. In this manner, the relationship between tokens and probability was clarified for the subjects. Importantly, however, the probability of disaster drops significantly with initial contributions and then drops at declining rates with greater contributions to a minimum of 10% (Figure 3).

Figure 3: Relationship between tokens contributed and probability of disaster



Subjects are given two minutes to anonymously chat with their group members to coordinate. Chat is in the form of free-form text, so subjects post messages anonymously to their group members. Furthermore, subjects are allowed to chat even when they cannot contribute (as when there is no possibility of a disaster), to mitigate boredom while they wait for other groups to finish deciding on their contributions. Note that any identifying messages are strictly forbidden, and at no point did subjects ever reveal their identity.

The most critical aspect of this experiment is the manner in which the disaster is triggered. Subjects' activity in the bonus coding rounds directly contributes to the disaster being triggered in the following way: subjects earn tokens in the bonus coding rounds at different piece rates, with rich countries earning 2 tokens per word decoded, while poor countries earn 1 token per word decoded. In the main phase of the experiment, groups are randomly formed each round, and the total earnings from the bonus coding phase are added up and averaged over the four group members. If this average is greater than the average for the entire session, the disaster game is triggered. However, if this average is lower than the average for the entire session, the disaster game is not triggered. Each subject played the disaster game at least once. By setting up the disaster trigger in this fashion, subjects' effort in the bonus coding round directly contributes to triggering the disaster. Importantly, this aspect of the bonus coding phase is not revealed to the subjects during the coding phase, meaning that subjects undertake the bonus coding phase with only the knowledge of the piece rate. Subjects have an incentive to decode as many words as they can, but the marginal benefit of each token is low, and they are free to quit the task at any time. Indeed, we find that most subjects quit the task early.

The control group is then informed that during the main phase, in some rounds the disaster game is triggered, but in other rounds the disaster game is not triggered. Importantly, subjects are not informed about what triggers the disaster game. The treatment simply informs the subjects how the disaster game is triggered. This means that subjects are able to link the disaster game (which will likely lower their earnings) to past actions of subjects in the group. That is, group members effort in the bonus coding task directly contributes to the triggering of the disaster game. Moreover, given that the piece rate for rich subjects was twice that of poor subjects, rich subjects are twice as likely than poor subjects to be responsible for the triggering of the disaster phase. However, as all subjects are given the option to end the task early, rich subjects can plausibly deny that their actions directly contributed to the disaster game being triggered. In other words, this information adds tension within the group and introduces reasons to not contribute to disaster mitigation by both rich and poor subjects.

Finally, at the end of each round when the disaster game is triggered the following information is provided to each of the subjects: their contributions, the total group contribution, and the probability of disaster.

4. Results

Comparisons between the baseline and the treatment (shadow of history) allow us to address three questions: Does focusing on the source of the triggering of climate change affect contributions to climate change mitigation? Furthermore, does the treatment affect both rich and poor countries, or do the effects differ across subject types? Finally, is the probability of disaster greater when the treatment provides information on the source of climate change?

We first examine the impact of the treatment (providing information on history) on contributions to the mitigation fund. The base regression specification focuses only on the treatment effect, while additional specifications add controls for player type (rich = 1), the round (from 1 to 10), gender, age, educational status, income level, and clarity of instructions.

Model I of Table 1 and Figure 4 demonstrate that the treatment has a statistically insignificant and positive effect on contributions overall ($p<0.10$). This means that when information on the source of the triggering of the disaster game is available (that is, greater subject effort in the previous bonus coding stage yields an increase in the likelihood of facing a disaster) contributions to the mitigation fund increased by 6 tokens on average (a very small amount). Across all models, the treatment effect is similarly small (6 to 8 tokens) and not statistically significant at the 10% level. Model II adds a dummy variable for player type and shows that Rich countries contribute 64.42 tokens more on average across all conditions ($p<0.01$). Furthermore, the round variable (ranging from 1 to 10 for the 10 rounds in the experiment) does not show a significant effect, indicating that contributions do not change as subjects gain more experience playing the game over time.

Figure 4: Contributions to group fund

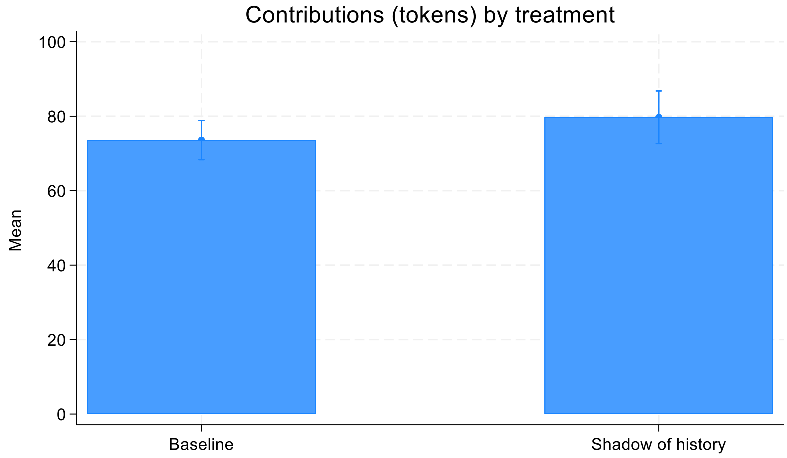


Table 1: Contributions to group fund

Dependent Variable: Contributions to group fund (Tokens)	I	II	III	IV
Treatment: Shadow of history	6.131	6.532	6.961	8.391
	(9.05)	(7.77)	(7.87)	(8.37)
Country type (1 = Rich)		64.42***	64.83***	68.08***
		(7.68)	(7.76)	(8.20)
Round		0.599	0.602	0.608
		(0.61)	(0.61)	(0.61)
Gender (1 = Female)			6.596	5.689
			(7.89)	(7.66)
Age (in years)			0.398	0.292
			(0.56)	(0.49)
Education status (1 = second year)				-11.830
				(13.02)
Education status (1 = third year)				-24.02**
				(10.53)
Education status (1 = Masters)				-2.743
				(9.73)
Education status (1 = PhD)				11.650
				(22.60)
Income (5 = Higher than others)				-0.034
				(4.57)
Clarity of instructions (5 = Clear)				-3.316
				(4.41)
Constant	67.46***	31.29**	18.120	38.840
	(12.84)	(12.44)	(20.41)	(30.57)
Observations	1048	1048	1048	1048
R-squared	0.002	0.203	0.208	0.227
P-value	0.499	0.000	0.000	0.000

Note: OLS specifications with individual level clustered standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

The next table and Figure 5 shows the results of the treatment on our two types of countries: Rich or Poor. Recall that subjects were randomly assigned to the role of either a rich or poor country. Rich countries received a much higher endowment (1,000 tokens) compared to poor countries (250 tokens). Both rich and poor countries faced the same probability of disaster, which imposed a cost of 200 tokens on each country. The results show a clear drop in contributions by poor countries when the shadow of history is present: contributions by subjects in the role of Poor countries reduced their contributions to the mitigation fund by 12.93tokens (model I), relative to a baseline contribution of 63.47 tokens ($p<0.05$). Rich countries, by contrast increased their contributions in the shadow of history, by 26.04 tokens (model III) over a baseline contribution of 70.10 tokens on average ($p<0.1$). Overall, these results show that the reduction in contributions is mainly coming from subjects in the poor country role, who are observed reducing their contributions by about 5%, while subjects in the rich countries increase their contributions by about 2.5% on average compared to their endowment of about 1,000 tokens. Nevertheless, we cannot rule out whether these drop in contributions are significantly different across player type. Table 2 displays these results.

Table 2 demonstrates that the reduction in contributions by subjects assigned to the poor country role is more than offset by the contributions of subjects assigned to the rich country role. Figures 6 and 7 show the evolution of contributions by round, as subjects gain more experience playing the game. The main result is that little changes across rounds, indicating that subjects' behavior does not change substantially as they gain more experience with the disaster game. Below we differentiate patterns among the poor and rich countries.

Figure 6 displays average contribution to the mitigation fund by subjects in the poor country role in the baseline and in the shadow of history treatment. Two patterns are immediately clear. First, on average, in the shadow of history treatment, poor countries contribute less. The red line representing the contribution in the shadow of history treatment is lower than the blue line, representing the contribution in the baseline. Second, towards the latter rounds where participants become more experienced, the difference between the treatment and the baseline become slightly larger.

We next focus on contributions over time by the rich in both the baseline and in the shadow of history treatment (Figure 7). On average, the rich contribute more under the shadow of history treatment than in the baseline. The red line, representing the contribution in the shadow of history treatment, is higher than the blue line, representing the baseline. However, there is some volatility across rounds.

Figure 5: Contributions to group fund by player type

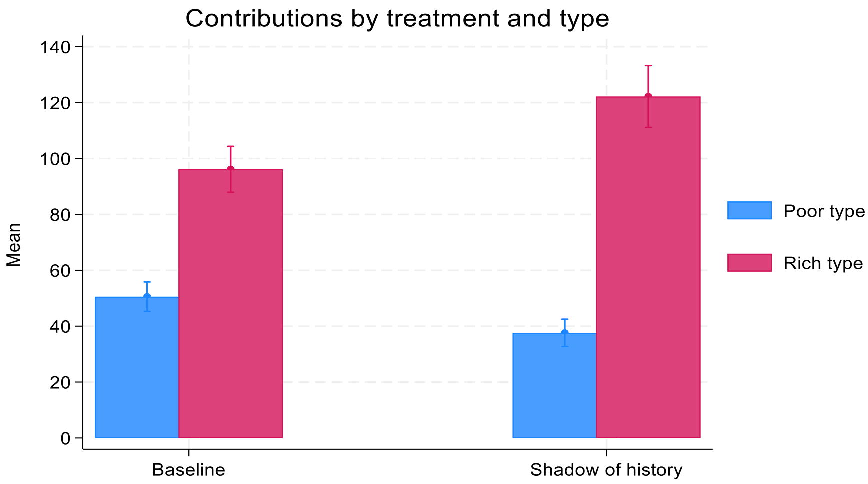


Table 2: Contributions to group fund by player type

Dependent Variable: Contributions to group fund (Tokens)	I	II	III	IV
	Type: Poor		Type: Rich	
Treatment: Shadow of history	-12.93**	-15.85**	26.04*	25.52*
	(6.42)	(7.32)	(13.93)	(14.22)
Round		-0.224		1.348
		(0.70)		(0.94)
Gender (1 = Female)		-3.032		13.240
		(6.45)		(13.16)
Age (in years)		0.053		1.066
		(0.21)		(1.85)
Education status (1 = second year)		-0.786		-25.190
		(14.45)		(16.49)
Education status (1 = third year)		-8.090		-45.07**
		(8.11)		(17.97)
Education status (1 = Masters)		-1.988		-18.440
		(8.27)		(22.06)
Education status (1 = PhD)		-13.880		-1.895
		(12.97)		(40.33)
Income (5 = Higher than others)		2.175		-5.889
		(3.43)		(9.12)
Clarity of instructions (5 = Clear)		-1.396		-5.931
		(3.84)		(11.36)
Bonus earnings		-0.526***		-0.101
		(0.16)		(0.19)
Constant	63.47***	85.19**	70.10***	104.6*
	(10.50)	(33.54)	(19.57)	(56.27)
Observations	522	522	526	526
R-squared	0.023	0.089	0.026	0.100
P-value	0.047	0.028	0.065	0.203

Note: OLS specifications with individual level clustered standard errors in parentheses. * 10%, ** 5%, *** 1% significance level.

Figure 6: Contributions to the mitigation fund by round by the poor countries

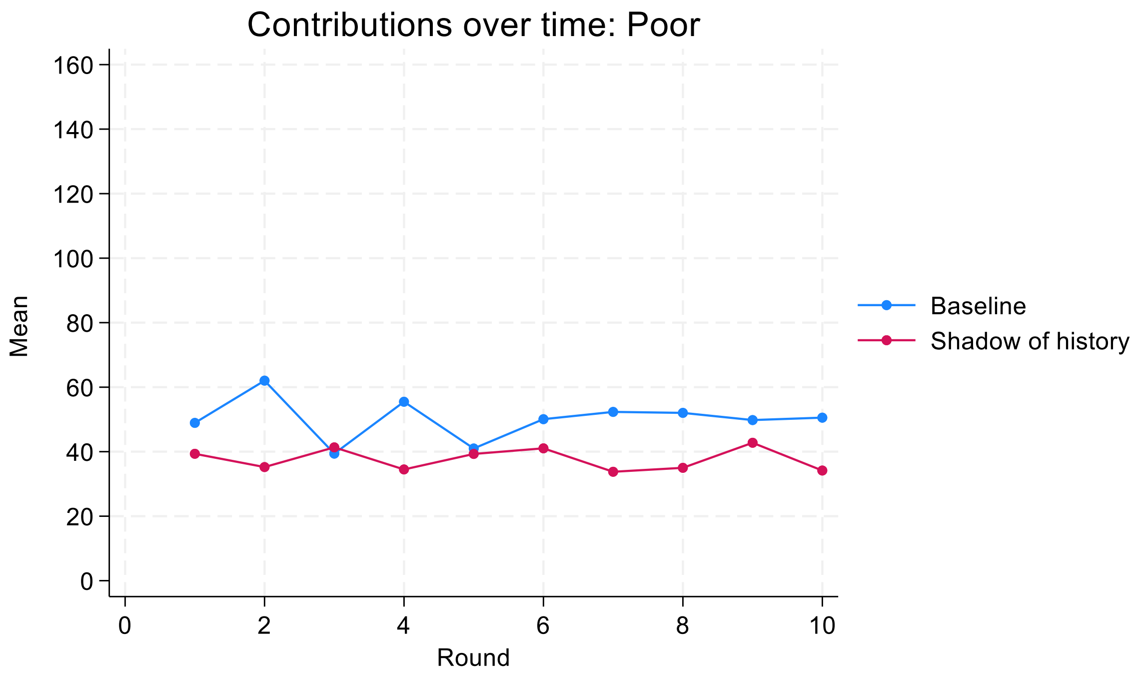
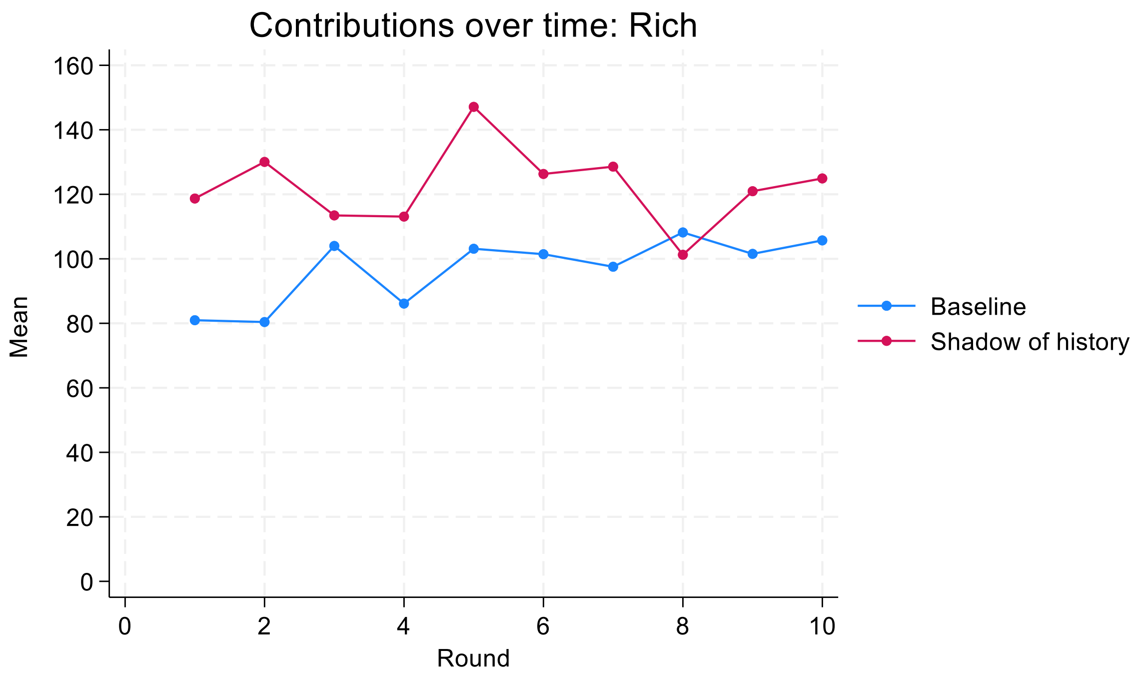


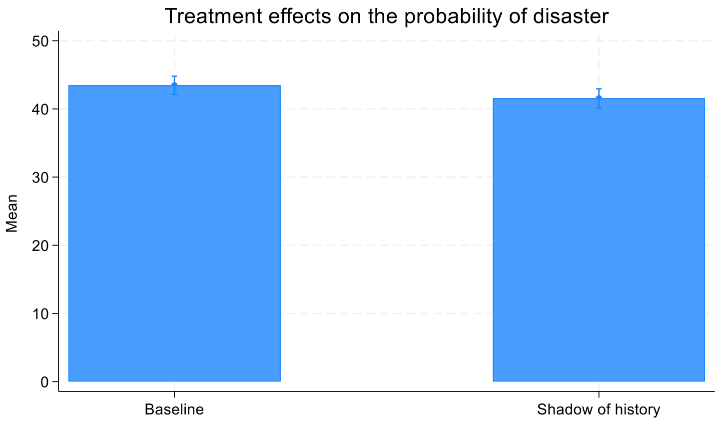
Figure 7: Contributions to the mitigation fund by round by the rich



Ultimately, our findings have implications for climate change negotiations. While the total contribution to the mitigation fund remains roughly the same under the Shadow of History, the distribution of those contributions changes – with richer countries shouldering more of the burden than poorer countries compared to the Baseline condition.

Moreover, given the slight increase in contributions to the mitigation fund under the Shadow of History treatment, the probability of disaster falls somewhat (Figure 8) raising the average expected earning for all subjects. As poor countries benefit more from a reduction in the likelihood of a disaster (as they suffer more harm from the disaster), their welfare increases as the total contribution to the mitigation fund increases. And, if the total contribution increases while their contribution falls, poor countries benefit even more.

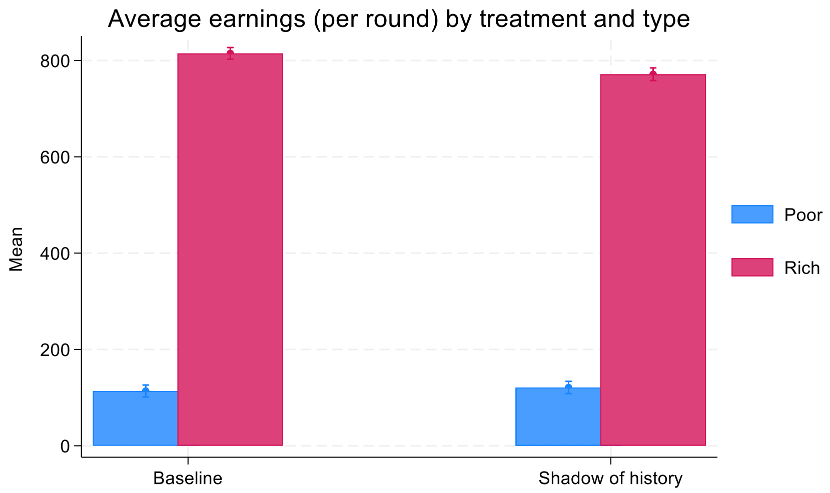
Figure 8: Treatment effects on the Probability of Disaster



As we show in Figure 9, earnings for subjects in the role of poor countries increase from 106 tokens in the baseline to 114 tokens under the shadow of history, an increase of 7 tokens (p=0.57). For subjects in the role of rich countries, earnings are significantly lower, going from 858 tokens on average in the baseline, to 815 tokens under the shadow of history, a drop of 43 tokens (p<0.01). This lends support to policies implemented by poor countries to highlight the historical responsibility of rich countries by "naming and shaming" previous economic activities that have led to the planet being its current situation – leading to higher welfare for the poorer countries.

That said, reducing the probability of disaster even further is in the interests of both rich and poor countries. Therefore, policies to encourage both groups of countries to contribute more to climate mitigation initiatives would benefit both groups of countries but would be especially beneficial to poor countries – suffer more from climate disasters.

Figure 9: Average Earnings (per round) by treatment and type



5. Conclusions

Climate change is a global problem that requires unprecedented levels of collective action to solve. Negotiating and implementing global agreements to reduce GHGs is at the core of humanity's strategy to fight climate change. Negotiations are a complex and challenging process due to differences in the priorities, interests, and the capacity of countries.

A core issue facing negotiators is the inequality of GHG emissions. Rich countries historically emitted much more GHG than poor countries. But, since around the year 2000, poorer countries have been responsible for more emissions than rich countries, and that trend is forecast to continue.

In this paper, we examined whether focusing on the historical source of the climate change problem affects parties' willingness to contribute to climate change mitigation. Facing the exact same mitigation problem, we find that while total group contributions to a mitigation fund are largely unchanged when subjects were made aware of the historical source of the problem, the *distribution* of those contributions changes markedly.

Namely, poor countries contribute much less when made aware of the historical responsibility of rich countries, while rich countries contribute slightly more – offsetting the decrease in contributions by the poor.

While these results are similar to results reported by other studies, we argue that our experimental design better captures the process of wealth accumulation and historical responsibility of rich countries. This is a key difference between our design and that of others, such as Kline et al. (2018). In Kline et al. (2018), subjects know that their activities in an initial development phase will have a direct impact on climate change. In our design, subjects do not know that their wealth accumulation during the first phase will have an impact on the climate change problem that they must mitigate during the second phase.

Turning to the average earnings of countries, the average expected earnings for all subjects increases under the Shadow of History treatment. This because of the slight increase in contributions to the mitigation fund, which lowered the probability of disaster somewhat. As poor countries benefit more from a reduction in the likelihood of a disaster, their welfare increases as the total contribution to the mitigation fund increases. Also, as the total contribution increases while their contribution falls, poor countries benefit even more. Earnings for subjects in the role of poor countries increased from 106 tokens in the baseline to 114 tokens under the shadow of history, while the earnings of subjects in the role of rich countries were significantly lower, falling from 858 tokens on average in the baseline to 815 tokens under the shadow of history – a drop of 43 tokens.

This last point highlights the importance of negotiations. The vast majority of experiments in this literature do not allow subjects to communicate or negotiate. Another contribution of our novel design is to allow just that – negotiations. If we as scholars are interested in understanding the mechanisms underlying international negotiations, we should make actual negotiations part of our experiments.

In future work, we plan to examine more closely the text recordings we have of subjects' actual negotiations during our experiments using textual analysis. We are hopeful that this will permit us to identify more fully the mechanisms underlying the all-important international climate change mitigation negotiations process.

Finally, turning to policy implications, the above results lend support to negotiation tactics used by poor countries to highlight the historical responsibility of rich countries – by "naming and shaming" rich countries for their previous economic activities that helped bring about the climate crisis humanity currently faces. As rich countries contribute more to GHG emissions mitigation, the welfare of poorer countries increases. That said, reducing the probability of disaster even further is in the interests of both rich and poor countries. Therefore, policies to encourage both groups of countries to contribute more to climate mitigation initiatives would benefit both groups of countries but would be especially beneficial to poor countries, who suffer more from climate disasters.

References

Andrews, Talbot M., Andrew W. Delton, and Reuben Kline. 2024. *Climate Games*. Ann Arbor, MI: University of Michigan Press.

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

Barrett, S. & Dannenberg, A. 2014. "Sensitivity of Collective Action to Uncertainty about Climate Tipping Points." *Nature Climate Change.* 4(1): 36-39.

Brown, Thomas C., and Stephan Kroll. 2017. "Avoiding an uncertain catastrophe: climate change mitigation under risk and wealth heterogeneity." Climatic Change, 141 (2): 155-166.

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

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden. 2007. "The pluralism of fairness ideals: An experimental approach." *American Economic Review* 97 (3): 818-827.

Dietz, Thomas, Ostrom, Elinor, and Paul C. Stern. 2003. "The Struggle to Govern the Commons." *Science* (302) 1907–1912.

Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis. 2011. "Relative earnings and giving in a real-effort experiment." *American Economic Review* 101 (7): 3330-3348.

Kameda, T., Takezawa, M., Tindale, R. S., & Smith, C. M. 2002. "Social sharing and risk reduction: Exploring a computational algorithm for the psychology of windfall gains." *Evolution and Human Behavior*, 23(1): 11–33.

Kline, Reuben, Nicholas Seltzer, Evgeniya Lukinova, and Autumn Bynum. 2018. "Differentiated responsibilities and prosocial behaviour in climate change mitigation." Nature Human Behavior 2 (9): 653-61. https://doi.org/10.1038/s41562-018-0418-0

Lévy-Garboua, Louis, David Masclet, and Claude Montmarquette. 2009. "A behavioral Laffer curve: Emergence of a social norm of fairness in a real effort experiment." *Journal of Economic Psychology* 30(2): 147-161.

Milinski, M., Sommerfeld, R. D., Krambeck, H.-J., Reed, F. A. & Marotzke, J. 2008. "The collective-risk social dilemma and the prevention of simulated dangerous climate change." *Proc. Natl Acad. Sci.* 105: 2291–2294.

Milinski, M., Röhl, T. & Marotzke, J. 2011. "Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets." *Climatic Change* 109: 807–814.

Ostrom, E. et al. 2002. The Drama of the Commons (National Academies Press, Washington DC).

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.