ERF Working Papers Series

Climate Change Negotiations Under the Shadow of History

Sheheryar Banuri, Ha M. Nguyen, and Ernest J. Sergenti

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

Working Paper No. 1810
December 2025

CLIMATE CHANGE NEGOTIATIONS UNDER THE SHADOW OF HISTORY

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

Working Paper No. 1810

December 2025

We thank many seminar participants at the University of East Anglia, the World Bank, the IMF, and at the annual conferences of the following organization: the Midwest Political Science Association (April 2024), the American Political Science Association (September 2024), and the Economic Research Forum (April 2025), especially Robert Sugden, Colin Kuehl, Kevin Carey, Vladimir Klyuev, Hala Abou Ali, Roberta Gatti, and Nadir Mohammed. 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 and do not necessarily represent the views of the World Bank and the IMF, their Executive Directors, or the countries they represent.

Send correspondence to:
Sheheryar Banuri
University of Cambridge
s.banuri@uea.ac.uk

[1] Nguyen: IMF (HNguyen7@imf.org); Sergenti: World Bank (esergenti@worldbank.org).

First published in 2025 by
The Economic Research Forum (ERF)
21 Al-Sad Al-Aaly Street
Dokki, Giza
Egypt
www.erf.org.eg

Abstract

Climate change is a global challenge requiring unprecedented levels of collective action. In this context, this paper asks: do appeals to historical responsibility facilitate or hinder collective action? This paper uses a simple lab experiment simulating climate mitigation bargaining between high- and low-income countries. A key design feature is that the need for mitigation is triggered based on historical actions that were undertaken without knowledge of their impact on the environment (and hence, the need for mitigation). Two treatment arms were conducted, a baseline where the cause for mitigation (past actions) is not revealed, and a treatment – "the shadow of history" – where the historical origins of the problem are made explicit. In both conditions, negotiations take place regarding contributions to a mitigation fund (i.e., collective action). Results show that revealing the shadow of history marginally increases average contributions, but the distribution of those contributions changes markedly. When made aware of the historical causes of the climate problem, low-income countries significantly reduce their contributions, while high-income countries contribute more – offsetting the reduction. Critically, the overall welfare of low-income countries increases, while it decreases for high-income countries. Moreover, results from textual analysis of chat data show greater tension when historical responsibility is made explicit, with more negative sentiment and adversarial conversations. These results suggest that appealing to historical responsibility appears to be a successful negotiations tactic for poor countries.

Keywords: Climate Change Negotiations, Historical Responsibility, Collective Action, Bargaining, Inequality
JEL Classifications: C91; D63; Q54; H87

ملخص

إن تغير المناخ يشكل تحديًا عالميًا يتطلب مستويات غير مسبوقة من العمل الجماعي. وفي هذا السياق تتساءل هذه الورقة: هل النداءات إلى المسؤولية التاريخية تسهل العمل الجماعي أم تعيقه؟ تستخدم هذه الورقة تجربة معملية بسيطة تحاكي المساومة على التخفيف من آثار تغير المناخ بين البلدان المرتفعة والمنخفضة الدخل. إن إحدى السمات الرئيسية للتصميم هي أن الحاجة إلى التخفيف تنشأ على أساس الإجراءات التاريخية التي تم اتخاذها دون معرفة تأثيرها على البيئة (ومن ثم الحاجة إلى التخفيف). تم إجراء ذراعين للعلاج، خط الأساس حيث لم يتم الكشف عن سبب التخفيف (الإجراءات السابقة)، والعلاج – "ظل التاريخ" – حيث يتم توضيح الأصول التاريخية للمشكلة. وفي كلتا الحالتين، تجري المفاوضات بشأن المساهمات في صندوق التخفيف (أي العمل الجماعي). وتظهر النتائج أن الكشف عن ظل التاريخ يؤدي إلى زيادة طفيفة في متوسط المساهمات، ولكن توزيع تلك المساهمات يتغير بشكل ملحوظ. وعندما يتم توعية البلدان المنخفضة الدخل بالأسباب التاريخية لمشكلة المناخ، فإنها تخفض مساهماتها بشكل كبير، في حين تساهم البلدان المرتفعة الدخل بشكل أكبر – مما يعوض التخفيض. والأمر الحاسم هنا هو أن الرفاهة العامة في البلدان المنخفضة الدخل ترتفع، في حين تنخفض في البلدان المرتفعة الدخل. وعلاوة على ذلك، تظهر نتائج التحليل النصي لبيانات الدردشة توتراً أكبر عندما يتم توضيح المسؤولية التاريخية، مع المزيد من المشاعر السلبية والمحادثات العدائية. وتشير هذه النتائج إلى أن اللجوء إلى المسؤولية التاريخية يبدو تكتيكاً ناجحاً للمفاوضات بالنسبة للبلدان الفقيرة.

1. Introduction

Does the focus on debates about historical responsibility improve welfare outcomes for poor countries? In recent years poor countries have pointed to the historical emissions of rich countries in contributing to climate change. At the same time, the need for global cooperation for climate change mitigation is imminent. Hence, does the focus on responsibility and culpability in negotiations make it more difficult for countries to cooperate to mitigate climate change? Furthermore, does this affect negotiations behavior? And finally, how does this focus on historical responsibility improve the welfare of poor countries? In this paper, we use a lab experiment that directly models issues of historical responsibility, and show that (1) including debates about historical responsibility has no impact on overall cooperation, (2) these debates reduce cooperation by poor countries, but rich countries compensate for these reductions; (3) these debates negatively affect sentiments and make negotiations more adversarial and contentious; and (4) they have no impact on aggregate welfare, but importantly increase the welfare of the poor while reducing the welfare of the rich.

Climate change is a major 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.[2]

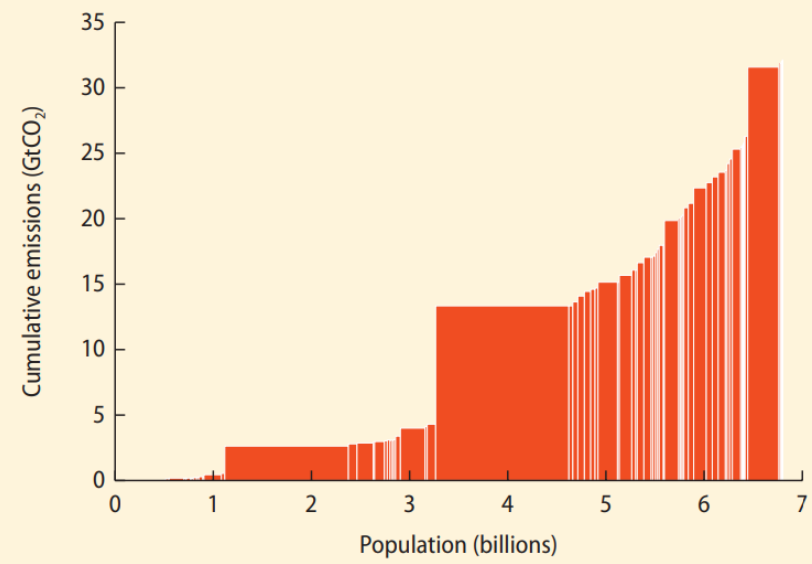
Negotiating and implementing global agreements to reduce GHGs are thus at the core of humanity's strategy to fight climate change. Nevertheless, negotiations are a complex and challenging process due to differences in the priorities, interests, and capacity of countries. This paper investigates whether invoking historical responsibility helps or hinders international cooperation on climate mitigation. Using a laboratory experiment simulating negotiations between high- and low-income countries, we compare two treatment arms or conditions: one where participants are unaware of the historical causes of climate damage, and another where past emissions are explicitly linked to current climate risks. We find that while overall contributions to a shared mitigation fund do not significantly increase when history is revealed,

[2] Reducing GHGs is challenging for two main 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 impose a negative externality: the social cost of GHG emissions—through pollution and the intensification of climate change—far exceeds the private cost of carbon. Likewise, reducing GHG emissions generates a positive externality: while the cost of investment is borne by the country or firm undertaking it, the benefits—such as reduced climate risks—are shared globally. As a result, for most countries, the private benefit of acting alone on climate mitigation is smaller than the private cost. Therefore, from a national cost-benefit perspective, unilateral mitigation efforts may not appear economically justified, which makes collective international action essential.

the burden of contributions shifts: low-income participants contribute less, and high-income participants contribute more. This redistribution does not reduce collective mitigation, but it does increase negotiation tension and identity salience, as shown by sentiment analysis of participant chat data. Our findings suggest that appeals to historical responsibility can reshape burden-sharing in climate negotiations without undermining collective action, but at the cost of heightened tensions and adversarial relationships.

A core issue is the inequality of GHG emissions. Rich countries historically emitted much more GHGs than poor countries, with the poorest 1 billion people historically emitted less than 1% of GHGs (Figure 1). However, 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. One argument is that, 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 both financially and technologically with adaptation (to the effects of higher temperatures and sea levels) and the transition towards more renewable forms of energy (see Fanning and Hickel, 2023; Climate Action Network International, 2024).

Figure 1. The poorest 1 billion people historically emitted less than 1% of GHGs



Source: Hallegatte et al. (2016).[3]

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.

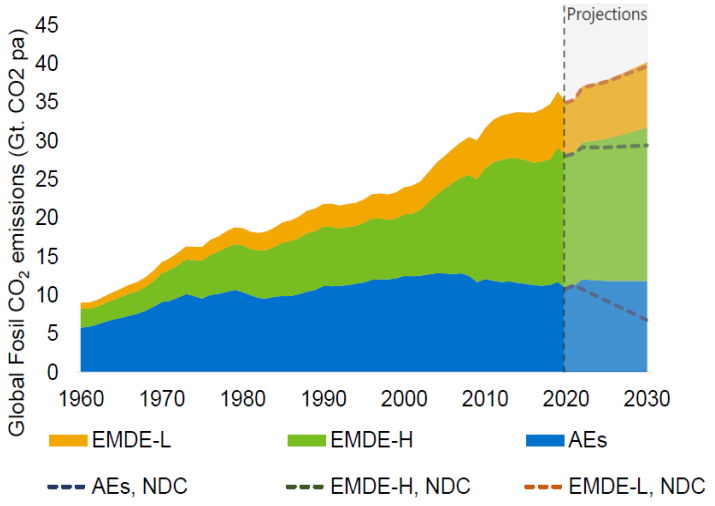
Poor countries have had some success in securing commitments by rich countries. At the 15th Conference of Parties (COP15) of the UNFCCC in Copenhagen in 2009, developed countries committed to a collective goal of mobilizing $100 billion per year by 2020 (including loans and grants) for climate action in developing countries. The goal was formalized at COP16 in

[3] Hallegatte et al. (2016), p. 193, figure 6.3.

Cancun and the total amount was increased to $300 billion at the most recent COP29 in Baku. Moreover, at COP27 in Sharm el-Sheikh, an agreement was reached to provide "loss and damage" funding for vulnerable countries hit hard by climate disasters. However, concrete actions have not followed these commitments. Rich countries failed to meet their climate financing target of $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.[4] As of March 2025, the fund has increased to only $768 million (www.frld.org).

Another argument, while acknowledging past emissions of rich countries, is that the responsibility for emissions reduction should be more evenly shared, considering current emissions levels. In fact, since around the year 2000, emerging market and developing economies (EMDEs) have been responsible for more emissions than advanced economies (AEs), 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 role in the process. As such, some argue that all countries should have 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. For example, the Paris Agreement from COP21 requires that all parties prepare, communicate, and maintain successive Nationally Determined Contributions (NDCs), reflecting their highest possible ambition. While acknowledging different national circumstances, it emphasizes that all countries should contribute to global efforts to combat climate change.

Figure 2. Historical GHG emission by country group



Source: Global Carbon Budget (2021), UNFCCC (2021a), World Bank (2021) and IMF staff (Black et al., 2001).

Note: NDC = nationally determined contributions; AEs = advanced economies; EMDE-H and EMDE-L = higher-income and lower-income emerging market and developing economies.

[4] See Lakhani (2023)

There are theoretical points for both arguments. On the one hand, the debate over historical responsibility 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 whether "naming and shaming" facilitates or impedes global negotiations over collective action to mitigate climate change.

This paper asks the following question: do historical factors and the debate over historical responsibility help or hurt climate change negotiation? This is a critical question because the debate over historical responsibility 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. To answer this, we use a lab experiment: Subjects engage in a real effort task that earns them an endowment but also generates emissions. Subjects are incentivized to exert effort but are unaware of the impact their actions have in later stages. Furthermore, subjects are of two types: randomly assigned to play the role of rich countries (with greater endowments and greater incentives to emit), and of poor countries (with lower endowments and lower incentives to emit). Next, subjects are assigned to groups and participate in a collective action problem. They are informed that there is a high probability of a disaster, which reduces the earnings of all members of the group but that they can pay to reduce that probability. Subjects then negotiate over contributions made by group members. The baseline condition carries no explicit linkage between past effort and current emissions, while the treatment (the "shadow of history") explicitly links past behavior to current outcomes. Note that the only change between treatment and control is the presence of information on past behavior linked to current outcomes.

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. Subjects playing the role of poor countries reduce their mitigation contribution significantly when made aware of historical responsibility, compared to the baseline. Rich countries, on the other hand, increase their contributions when they are made aware of the linkage between past activity emissions and climate change. Their increase in contributions offsets the poor's decrease in contributions. Results from textual analysis of the chat data among subjects show an increase in tensions during negotiations under the shadow of history condition compared to negotiations under the baseline control condition – with a country's "type" becoming more salient and the sentiment value of the chat becoming more negative under the shadow of history condition. Our results show that negotiation tactics employed by poor countries to highlight historical responsibility of rich countries (i.e., by "naming and shaming" rich countries for their previous economic activities) might not inhibit collective action overall but might change the distribution of contributions in favor of the poor, who suffer more from climate disasters. Moreover, we find that the average welfare of the poor (i.e., expected earnings of subjects in the role of poor countries) increases under the shadow of history, while those of the rich decline, highlighting the effectiveness of such negotiation tactics.

Our paper contributes directly to the literature on climate change mitigation and historical responsibility (Milinski et al., 2008; Andrews, Delton, and Kline, 2024; Tavoni et al., 2011; Kline et al., 2018; among others). We extend this literature by endogenizing inequality using effort tasks and stochastic elements, rather than imposing exogenous income distributions, directly modeling relevant historical emissions dynamics. Furthermore, our design also captures how historical actions (undertaken without explicit knowledge about their consequences on climate) influence present day cooperative behavior. Our paper implements actual bargaining and negotiations, and reports on how historical responsibility debates influence such behavior. We show that highlighting historical responsibility changes the distribution of cooperation and welfare between the rich and the poor but yields overall collective outcomes unchanged.

2. Literature review

Our paper contributes to a growing experimental literature on climate change mitigation, particularly the role of historical responsibility and inequality. The foundational study by Milinski et al. (2008)—known as the "disaster game"—examines cooperation in a threshold public goods context[5], where groups must meet a collective contribution target to avoid climate disaster. If a disaster strikes, they lose their entire remaining endowment. One important limitation of their study is that *subjects were not allowed to communicate with one another* – a major difference from our research design described below. They find that only under high probabilities of disaster (e.g., 90%) do groups consistently reach the target. Lower probabilities of disaster (10%, 50%) reduce contributions. The authors offer as one possible explanation that subjects were highly risk averse, but it seems more likely that another mechanism is required to explain the finding – such as an ingrained sense of fairness. Presenting results by round over time, the authors note that some subjects contributed less than their fair shares, which other subjects noticed, modifying their behavior. This led to a decline in total group contributions, especially for groups receiving the 10% and 50% probability-of-disaster conditions. Interestingly, especially with the 90% condition, some subjects increased their contributions to compensate for the free riders and help the group reach the threshold.

Our study also contributes to the broader literature on how inequality influences cooperation in climate dilemmas. Tavoni et al. (2011) find that imposed inequality—introduced through pre-assigned contributions to a mitigation fund during the first three rounds of a 10-round experiment—reduced overall cooperation. However, communication among participants partly mitigated this effect. Importantly, the artificial nature of the inequality assignment may have diminished the perceived responsibility among the "poorer" participants. That is, these participants may have believed they had already fulfilled their contribution to the mitigation goal, leading to lower willingness to cooperate in the remaining seven rounds—ultimately reducing the success rate in the inequality treatment.

[5] See Andrews, Delton, and Kline (2024) who provide a rich description of Milinksi et al. (2008), as well as many other climate-related lab experiments.

By contrast, Milinski, Röhl, and Marotzke (2011) report that randomly assigned inequality did not harm cooperation, especially under high disaster probabilities. 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 point to differences 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 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).

Finally, Burton-Chellew et al. (2013) show that inequality becomes problematic only when paired with asymmetric risk exposure—i.e., when "poorer" participants face greater risk. In such cases, "richer" participants contributed less, and groups often failed to reach the mitigation target. Similarly, Brown and Kroll (2017) find that inequality alone does not reduce cooperation, but uncertainty about threshold levels does. Consistent with Barrett and Dannenberg (2012, 2014), uncertainty undermines coordination more than inequality does.

In sum, Andrews, Delton, and Kline (2024) synthesize the literature on inequality and cooperation by concluding that inequality per se is not the key obstacle to cooperation. Rather, differences in perceived responsibility, incentives, or exposure to risk—especially when unfairly distributed—tend to undermine collective action.

Kline et al. (2018) extend this line of research by modeling inequality and perceived responsibility together. To do that, they introduce a two-stage framework: an "economic development" stage with a common pool resource dilemma (Ostrom 2002, Dietz et. al. 2003), followed by a mitigation stage akin to the disaster game. As with the Milinski et al. (2008) design, subjects were not allowed to communicate with one another. Crucially, in their design, subjects understand that economic activity in the development phase (i.e., harvesting) raises the mitigation threshold. They find 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, many groups fall just short of the mitigation target in the second phase. Without communication, this shows groups exhibited a high level of cooperation – with the threshold perhaps serving as a focal point (Schelling, 1960). On the other hand, missing the target by a small margin is a very wasteful use of resources in the disaster game setup. If subjects knew that the target would not be met, a more rational strategy would have been to contribute nothing and take their chances with the risk of disaster on their full endowment. Furthermore, subjects know that their activities in the development phase have direct impacts on climate change (which is a key difference with our design, where subjects are unaware of their impacts).

As Kline and co-authors note, their results may conflate wealth differences with responsibility. A placebo version—where inequality was randomly assigned rather than self-created—showed that groups were more successful in mitigation. This might suggest that perceived responsibility for emissions plays a role – or it could just confirm the Milinski, Röhl, and Marotzke (2011) results that people are more willing to contribute to a common goal if the reason for their relative richness is considered as random. In a second experiment, Kline et al. (2018) modify the first phase of the compound climate dilemma. With this experiment, 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, only 3 of the 6 players may 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.

Kline et al. (2018) 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, during the disaster game stage of the experiment, groups comprised of early and late developers had a harder time averting disaster than groups in the placebo experiment. In post-experiment interviews, late developers cited the unfair set-up of the game as a main reason for contributing much less. Although their results suggest that historical responsibility could play a part in the breakdown of cooperation, the Kline et al. (2018) study suffers from the same limitations as the Tavoni et al. (2011) study: namely, the artificial nature of how inequality and historical responsibility are assigned. Our paper explicitly overcomes these limitations.

3. Experimental design

The experiment consists of a baseline and a single treatment, which varies the amount of information subjects receive about the triggering of the "disaster game" (simulating a negative climate related event). The overall experiment is conducted in three distinct phases: (1) endowment generation, (2) resource extraction; and (3) negotiations. We explain each phase in turn.

Endowment generation phase

The experiment begins with subjects engaging in a real effort task to earn their endowment for the session. The effort task is a version of a coding task commonly used within experimental economics (Lévy-Garboua, Masclet, and Montmarquette, 2009; Erkal, Gangadharan, and Nikiforakis, 2011): Subjects are first randomly assigned to the role of a "Type X" or a "Type Y" player, which correspond to poor ("X") or rich ("Y") countries. This type sets the endowments for subjects from the effort task:

- For poor countries (Type X), subjects earn a fixed endowment of 250 tokens
- For rich countries (Type Y), subjects earn a fixed endowment of 1000 tokens

The endowments were common knowledge such that subjects are informed of their type at the outset but also know that there is another type of player in the session. Next, subjects are asked to engage in the coding task to earn their endowment:

- For poor countries (Type X), subjects are given a target of 5 words in 135 seconds
- For rich countries (Type Y), subjects are given a target of 10 words in 270 seconds

Note that to generate the respective endowment, subjects need to achieve the target in the given timeframe.[6] This design allows subjects assigned to different types to generate different levels of endowment, but also to exert different levels of effort to do so, reflecting the source on inequality, based on a combination of luck (random assignment) and effort (higher target). Subjects in the role of a rich country (Type Y) were tasked with exerting twice as much effort as poor countries but received four times the endowment, reflecting different levels of productivity. This simulates real-world differences between rich and poor countries and provides each type of subject a reason to act in an uncooperative fashion, with poor countries able to highlight the element of luck contributing to inequality, while the rich able to highlight the element of effort in contributing to inequality.

Resource extraction phase

Upon completing the task generating endowments, subjects are then given the opportunity to engage in "bonus" rounds of the coding task. These rounds are entirely optional and at the discretion of each subject to engage in for an (undisclosed) maximum of 14 rounds (with each round lasting 2 minutes). Subjects are provided the following piece rates for effort:

- For poor countries (Type X), subjects are paid 1 token for each word decoded
- For rich countries (Type Y), subjects are paid 2 tokens for each word decoded

Note that the piece rates here are far lower than the amount generated in the main coding task earlier and differ by a factor of 2-to-1 across country types. These piece rates are also common

[6] Note that the targets are such that they require effort but are not too difficult so as to minimize attrition due to poor performance in the task (if subjects did not achieve the target, the game ends). Of the total sample of 212 subjects, only 3 were unable to complete the task and are dropped from the analysis.

knowledge so that subjects are aware of the different rates, but not of how much effort is exerted. This simulates the differences in productivity between the rich and the poor, but also that greater effort generates a higher likelihood of triggering a disaster – a key feature of our experiment. It also simulates the differential historical impacts of rich and poor countries on climate change.

Negotiations phase

Once subjects complete the bonus task, the negotiations phase of the experiment begins. In this phase, subjects are assigned to groups of 4, with each group containing two rich and two poor countries (subjects are aware of this composition). Subjects are informed that in some rounds a "disaster" affecting all members of the group has a 90 percent chance of occurrence (a simulated climate disaster). If the disaster occurs, each group member loses 200 tokens (80 percent of the endowment for poor countries, and 20 percent of the endowment for rich countries). Subjects are informed that they can pay tokens to reduce the probability of experiencing the disaster (into a "mitigation fund") and must discuss how much each country should contribute (i.e., negotiate over contributions). We label this phase as the "disaster game."

The most critical aspect of this experiment is the way the disaster is triggered. Subjects' activity in the resource extraction phase (bonus coding rounds) directly contributes to the disaster being triggered in the following way: subjects earn tokens in the resource extraction phase at different piece rates, with rich countries earning 2 tokens per word decoded, while poor countries earn 1 token per word decoded. Once a group is formed, the average earnings from the resource extraction phase for the entire group are compared with the average earnings for the session. If the group average is higher than the session average, the disaster game is triggered, otherwise the game is not triggered. If the game is not triggered, the group earns the full endowment. If the disaster game is triggered, the group is informed that there is a probability of a disaster, and that they can pay tokens to reduce the probability.

The rationale for this group average comparison with the session average is that groups that extracted more in the resource extraction phase now must face the high probability of a disaster. Furthermore, given the difference in piece rates during the resource extraction phase, rich countries have plausibly contributed more to this outcome than poor countries, though rich countries can also have plausibly chosen to exert less effort and hence claim less culpability. This feature of the experiment captures the relationship between historical resource extraction and the current need for climate mitigation. 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. The baseline does not inform subjects about the conditions that trigger the disaster game (simulating a situation where historical responsibility debates are not relevant for negotiations) while the treatment informs subjects about the conditions that trigger the disaster.

Subjects negotiate over the amount to contribute to the mitigation fund. Subjects are given two minutes to anonymously chat with their group members. Chat is in the form of free form text, so subjects post messages anonymously to their group members.[7] Note that any identifying messages are strictly forbidden, and at no point did subjects ever reveal their identity. The negotiations phase is repeated for 10 rounds, with groups being randomly reformed each round.

Disaster mitigation technology

To develop a group probability-lowering cost function that is non-trivial, we rely on insights from the risk and uncertainty literature. The disaster game as formulated up to this point is essentially a lottery. With a 10 percent probability, subjects keep their endowment, whereas, with a 90 percent probability, they lose 200 tokens from their endowment. Thus, at the start of the disaster game, the expected payoff is 820 tokens [.10 * 1,000 tokens + .90 * 800 tokens] for a rich country and 70 tokens [.10 * 250 tokens + .90 * 50 tokens] for a poor country.

We know from the risk and uncertainty literature, that we can calculate a certainty equivalent for all subjects. The certainty equivalent is the amount of money or tokens received with certainty that would make agents indifferent from the lottery – given subjects' utility function with respect to the above lottery. Assuming a constant relative risk aversion (CRRA) utility function coefficient of 1.5 – consistent with estimates for the UK population (Groom and Maddison, 2019), at 90 percent probability of disaster, the certainty equivalent is approximately 817 tokens for rich countries and 56 tokens for poor countries. Hence, rich countries would (theoretically) be willing to pay up to 183 tokens (1,000 – 817) on average to eliminate the risk of disaster and keep their remaining endowment with certainty. For poor countries, this amount would be 194 tokens. Given that groups are made up of two rich countries and two poor countries, the total group willingness to pay to eliminate the lottery is 754 tokens (out of a total group endowment of 2,500 tokens).

At 80 percent probability of disaster, certainty equivalents are higher and thus the willingness to pay are lower, as the expected payoffs are higher. At 80 percent probability of disaster, rich countries would be willing to pay up to 165 tokens to get out of the lottery, while poor countries would be willing to pay up to 187 tokens – leading to a total willingness to pay for the group of 704 tokens. Hence, to reduce the probability of disaster from 90 percent to 80 percent, the group can contribute 50 tokens (the difference between the willingness to pay at 90 percent probability of disaster and the willingness to pay at 80 percent probability of disaster, 754 – 704). And so on for other probabilities of disaster.

Figure 3 plots the relationship so derived between the probability of disaster and the tokens contributed to the mitigation fund. A contribution of 50 tokens leads to a decline in the probability of disaster to 80 percent, while a contribution of 100 tokens leads to a decline in

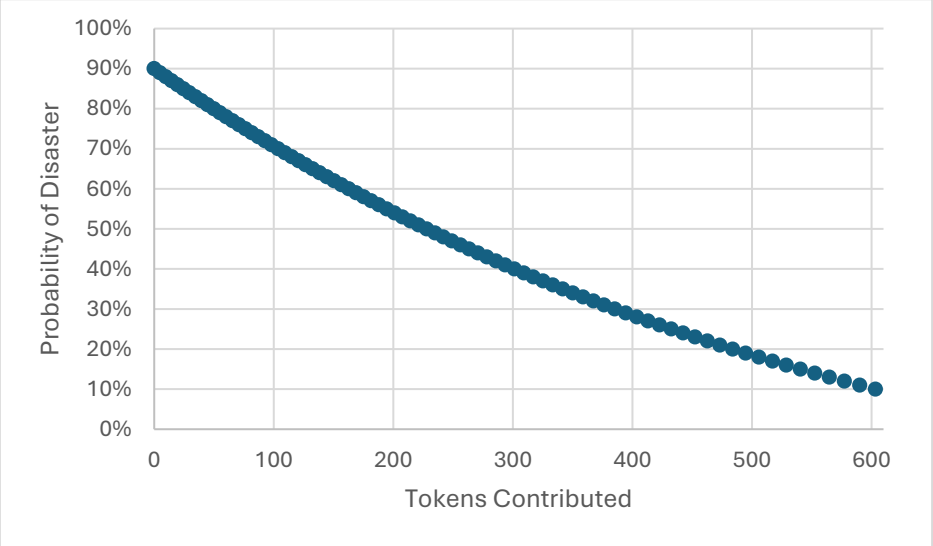
[7]Subjects are allowed to chat even when there is no possibility of a disaster, to mitigate boredom while they wait for other groups to finish deciding on their contributions.

the probability to roughly 70 percent and 200 tokens leads to a decline to roughly 55 percent, etc. As such, the relationship between the amount of tokens contributed to the fund and the decline of the probability of disaster is a convex function. Importantly, the probability of disaster drops significantly with initial contributions and then drops at declining rates with greater contributions to the minimum of 10 percent.[8] Algebraically, given the above setup, the relationship between the probability of a disaster occurring and the contributions to the mitigation fund follows the following formula (where y is the total amount of tokens contributed to the mitigation fund, and p is the probability of a disaster occurring):

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

As the formula is rather complicated, subjects are provided with an online calculator which allows them to compute the probability of disaster for different amounts of tokens contributed.

Figure 3. Relationship between tokens contributed and probability of disaster



Experimental procedures

Within the negotiations phase, subjects are provided information on their own contribution, the total contributions of the group, and the revised probability of disaster. The phase continues for 10 rounds, but only one round is randomly selected to be paid. The subjects were students at the University of East Anglia in the LEDR lab subject pool. Treatments were 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, with sessions lasting between 75 minutes to 90 minutes on average.

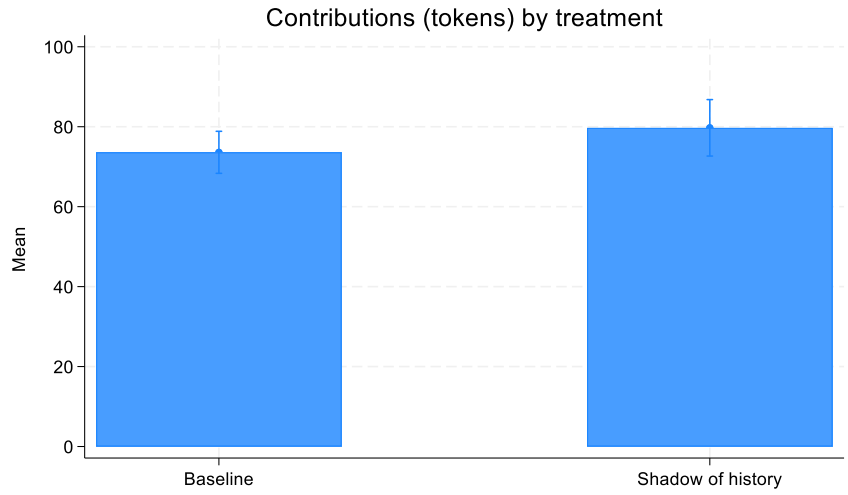
[8] Groups cannot eliminate the risk of disaster completely. The lowest level of risk that groups can achieve is 10 percent. This is because some risk of disaster is always present and permitting a complete elimination of risk would not be realistic - another distinction from our design and previous experiments.

4. Results

Treatment effects on cooperation

We first examine the impact of the treatment (providing information on history) on overall cooperation (contributions to the mitigation fund). Figure 4 presents the average contributions to the mitigation fund by treatment. Table 1 presents results from two-sample t-tests, focusing only on the treatment effect. Tables 2 and 3 augment these findings with results from regression analyses: the base regression specification repeats the analysis focusing only on the treatment effect, while additional specifications add controls for player type (rich = 1), round (from 1 to 10), gender, age, educational status, income level, and clarity of instructions.

Figure 4. Contributions to mitigation fund



From the figure, we observe a nominal increase in contributions in the shadow-of-history treatment relative to the baseline. From the first row of Table 1, subjects contributed 73.6 tokens on average to the mitigation fund under the control and 79.7 tokens in the treatment (two sample t-test $p=0.17$). These results are confirmed with Model I of Table 2 and shown graphically in Figure 4. They demonstrate that the treatment has a positive but statistically insignificant effect on contributions to the mitigation fund overall. When information on linkage between resource extraction and the probability of a disaster is available, average contributions to the mitigation fund increased by 8 percentage points (6 tokens on average). Across all models, the treatment effect is similarly small (6 to 8 tokens) and not statistically significant. 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. Overall, we find that the treatment yields no aggregate change in contribution levels. Naturally, this lack of an increase in contributions also means a lack of a statistically significant reduction in the probability of incurring a disaster. In the baseline, the average probability of disaster faced by groups is 43.5%, while the average probability of disaster in the shadow of history

treatment is 41.5%, a statistically insignificant reduction of 2 percentage points (*p=0.25*). In other words, reinforcing historical culpability does not affect cooperation levels overall, nor does it substantially reduce the probability of a disaster.

Table 1. Contributions to the group mitigation fund

	Mean Control	Mean Treatment	Mean Difference	Standard Error	t value	p value
All Countries	73.591	79.722	6.131	4.482	1.350	0.172
Poor Countries	50.538	37.602	-12.935	3.659	-3.550	0.001
Rich Countries	96.139	122.175	26.036	7.003	3.700	0.000

Note: Two-sample t test with unequal variances

We next turn to the impact of the treatment on cooperation by country type. Table 3 and Figure 5 show the results of the treatment on contributions to the mitigation fund by rich and poor countries. From the figure, we note a clear and marked decrease in contributions by poor countries (*p<0.01*), and an increase in contributions by rich countries (*p<0.01*).[9] Results from regressions (Table 3) confirm the patterns: contributions by subjects in the role of poor countries reduced their contributions to the mitigation fund by 12.9 tokens (model I) (*p<0.05*). Rich countries, by contrast, increased their contributions in the shadow of history, by 26.0 tokens (model III) (*p<0.10*). Overall, these results show that the average increase in contributions is mainly coming from rich countries (27 percentage point increase), while poor countries reduce their contributions (26 percentage point decrease). Note also that we find no systematic increase or decrease in contributions by round (*p=0.75* and *p=0.15* for poor and rich countries, respectively), indicating that contributions remain stable over the course of the negotiations phase.

[9] Results from the two-sample t-tests by country type are also presented in second and third rows of Table 1 for the poor and rich types, respectively.

Table 2. Contributions to the group mitigation 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	73.59***	37.82***	25.080	47.23*
	(5.26)	(6.07)	(15.79)	(28.19)
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.*

Figure 5. Contributions to group fund by player type

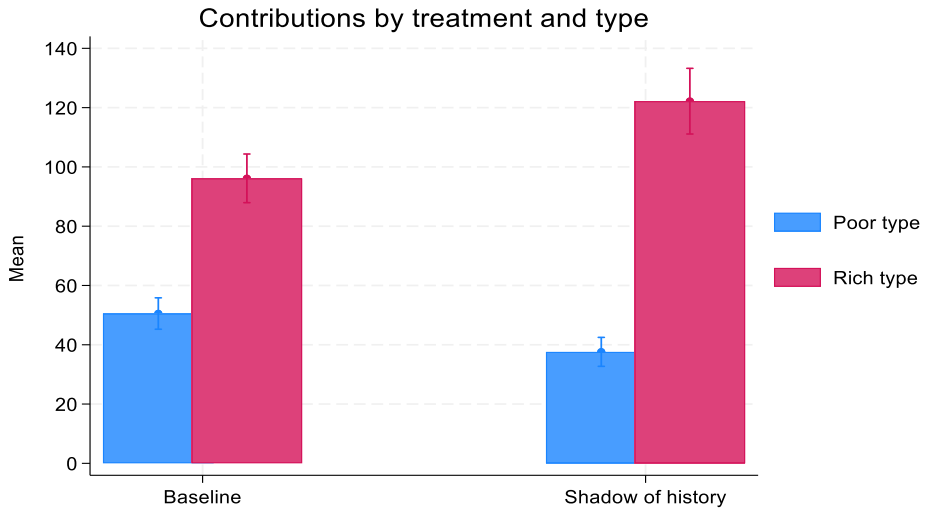


Table 3. Contributions to the group mitigation fund by player type

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	50.54***	69.34**	96.14***	130.1**
	(4.80)	(29.27)	(7.93)	(54.23)
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.*

Treatment effects on negotiations

The next question we ask is what effects the treatment had on negotiations between country types. For this, we turn to analysis of the chat data during negotiations when a disaster was triggered. We present two sets of results from our analysis of the chat data. First, Figures 6 and 7 present Word Clouds of the most frequently used words during negotiations – under the control and under the shadow of history treatment, respectively. From Figure 6, we see that under the control condition, subjects tend to use more positive language: "tokens", "yes", "contribute", "good", and "yeah" are the top five most frequently used words. By contrast, from Figure 7, we see that under the shadow of history condition, words such as "good" and "yes" appear less frequently. More important, now the most frequently used word is "type" – suggesting that subjects' endowments and issues related to inequality and responsibility for being in the disaster game are more salient during the negotiations.[10] "Type" is followed by "tokens", "contribute", "round", and "yeah" as the top five most frequently used words under the shadow of history condition.

[10] The experiment instructions referred to rich countries as "Type Y" and poor countries as "Type X", hence the frequent use of the word "type" to discuss contributions based on subject roles.

Figure 6. Word cloud of chat text under baseline negotiations

Figure 7. Word cloud of chat text under shadow of history negotiations

Our second set of results of the chat data relies on the more formal methods of sentiment analysis (see for example a seminal work by Tetlock, 2007).[11] The sentiment measure ranges from -1 (most negative) to 1 (most positive). It is calculated by using the number of positive and negative words present in the chat, scaled by the total word count. Formally, the sentiment of a chat message k is expressed as follows:

$$sentiment_k = \frac{Number\ of\ Positive\ Words_k - Number\ of\ Negative\ Words_k}{WordCount_k}$$

From Table 4, we see that the average sentiment of chat messages under the baseline condition was 0.087, while under the shadow of history condition, it was 0.054. Hence, the sentiment value is 0.33 points lower, or close to 40 percent lower (38.3% lower) ($p<0.05$). These results indicate that subjects displayed more negative sentiments during their negotiations under the shadow of history condition than under the control condition, implying more difficult negotiations.

Table 4. Sentiment analysis

Dependent Variable: Sentiment (n = 4,040)	
Treatment: Shadow of history	**-.0333** (se .0136; p-value 0.014)

Treatment effects on welfare

Finally, we focus on the effects on welfare by country type, using expected earnings as the outcome metric. As the average contribution to the mitigation fund increases only slightly under the shadow of history (Table 2, model 1: p=0.50), the probability of disaster declines only marginally (a reduction of two percentage points: p=0.25). However, to identify the effect of the treatment on welfare, we compute the expected earnings for each subject, which is given by the formula:

$$\pi_i = p(E_i - GC_i - C_d) + (1-p)(E_i - GC_i)$$

[11] Sentiment is calculated using the polyglot package in Python, which builds on the work of Chen and Skiena (2014).

where π_i is the expected earning for subject i, p is the probability of a disaster, $(E_i - GC_i - C_d)$ is the subject earnings in the event of a disaster, and $(E_i - GC_i)$ are subject earnings in the event of no disaster. E_i is the subject endowment level, GC_i is the tokens contributed to the mitigation fund, and C_d is the cost of the disaster.

For all subjects taken together, we find that expected earnings decline – by 8 tokens (or by just 1.7 percent), from 467 to 461 tokens – but that this effect is not statistically different from zero (p=0.71). This is because the decline in expected earnings from the increase in group contributions to the mitigation fund more than offsets the benefits to expected earnings from the slightly lower probability of disaster. See Table 5, which presents results from two sample t-tests.

However, although average contributions to the mitigation fund remain roughly the same under the treatment condition, the distribution of these contributions change – with richer countries shouldering more of the burden than poorer countries compared to the baseline. As poor countries benefit more from a reduction in the likelihood of a disaster (as they suffer relatively more harm from the disaster), their welfare increases as the total contribution to the mitigation fund increases. And, given that the total contribution to the mitigation fund increases under the shadow of history, while their individual contribution falls, poor countries benefit even more. These results are borne out in Table 5, which shows a positive and statistically significant (p<0.01) increase in the expected earnings of poor countries from 113 tokens in the baseline condition to 129 tokens in the treatment, a 16-token (14 percent) increase.

For rich countries, by contrast, they contribute more under the shadow of history but do not receive a much higher benefit from the small decline in the probability of disaster. As such, their expected earnings fall from 817 to 795 tokens, a 22-token (3 percent) decline in expected earnings (p<0.01).

Table 5. Expected earnings

	Mean Control	Mean Treatment	Mean Difference	Standard Error	t value	p value
All Countries	468.649	460.697	-7.952	21.430	-0.350	0.711
Poor Countries	112.985	129.107	16.121	3.582	4.500	0.000
Rich Countries	816.525	794.921	-21.605	5.619	-3.850	0.000

Note: Two-sample t test with unequal variances

In sum, our results show that debates about historical responsibility do not significantly impact cooperative behavior overall. But the distribution of welfare changes substantially. This means that debates about historical responsibility are a useful tool for developing countries. Focusing on historical responsibility allows them to reduce their cooperative burden while increasing their welfare. At the same time, rich countries also recognize the role of historical culpability and increase their contributions and by doing so reduce their overall welfare.

5. Conclusion

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 priorities, interests, and the capacity of countries. A core issue facing negotiators is the inequality of GHG emissions. Rich countries historically emitted many more GHGs 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 average 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. Poor countries contribute much less when made aware of the historical responsibility of rich countries, while rich countries contribute more – offsetting the decrease in contributions by the poor. Importantly, the welfare of poor countries increases, while that of the rich countries decline – as the poor benefit both from a lower probability of disaster and lower contributions to the mitigation fund, leading to higher expected earnings. Text analysis of the chat data from subjects' negotiations indicate that country type becomes more salient under the shadow of history condition compared to the control condition – and the sentiment value became more negative meaning more contentious negotiations overall.

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. This strategy allows the poor to alleviate their contribution burdens and improve their welfare. 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 relatively 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.

Black, Simon, Ian Parry, James Roaf, and Karlygash Zhunussova. 2021. "Not Yet on Track to Net Zero The Urgent Need for Greater Ambition and Policy Action to Achieve Paris Temperature Goals" *IMF Staff Climate Note 2021/005*.

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.

Chen, Y. and Skiena, S. 2014. "Building sentiment lexicons for all major languages," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Short Papers), pp. 383-389.

Climate Action Network International. 2024. "US$4 trillion owed to Global South by Global North due to the climate crisis". https://climatenetwork.org/2024/09/20/us5trillion-owed-to-global-south-by-global-north-due-to-the-climate-crisis/

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.

Fanning, A.L., Hickel, J. 2023. "Compensation for atmospheric appropriation." *Nature Sustainability* 6, 1077–1086.

Groom, Ben and David Maddison Pr. 2019. "New Estimates of the Elasticity of Marginal Utility for the UK". *Environmental and Resource Economics*. 72 (4): 1155–1182.

Hallegatte, Stephane, Mook Bangalore, Laura Bonzanigo, Marianne Fay, Tamaro Kane, Ulf Narloch, Julie Rozenberg, David Treguer, and Adrien Vogt-Schilb. 2016. "Shock Waves: Managing the Impacts of Climate Change on Poverty*". Climate Change and Development Series*. Washington, DC: World Bank.

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

Lakhani, Nani. 2023. "$700m pledged to loss and damage fund at Cop28 covers less than 0.2% needed", *The Guardian*.

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.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard University Press.

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.

Tetlock, Paul C. 2007. “Giving Content to Investor Sentiment: The Role of Media in the Stock Market.” *The Journal of Finance*, 62(3), 1139–1168.