Differentiated Responsibilities and Prosocial Behavior in Climate Change Mitigation: Behavioral Evidence from the United States and China

Reuben Kline[*1], Nicholas Seltzer[†2], Evgeniya Lukinova[‡3] and Autumn Bynum[§1]

[1]Department of Political Science & Center for Behavioral Political Economy, Stony Brook University
[2]Department of Political Science, University of Nevada, Reno
[3]New York University, Shanghai

Abstract

The recent Paris agreement has increased optimism that climate change might be successfully mitigated through international agreement. However, the commitments of countries are unenforceable. Therefore domestic political will, including on the part of citizens to make regular sacrifices, will be required in order for countries to meet these commitments. Understanding prosocial behavior in climate change mitigation is thus more important than ever. Many behavioral studies model the mitigation dilemma using public goods games: But, because wealth creation in a carbon-based economy inevitably leads to the appropriation of the global climate commons, climate change and its mitigation actually constitute a dual, interdependent social dilemma. Acknowledging the interdependence of this dilemma is necessary to capture "common but differentiated responsibilities," the equity principle that underlies international climate negotiations. To do so, we introduce the *compound climate dilemma*: a new behavioral game that expands the public goods approach to the mitigation dilemma by combining it with a preceding common pool dilemma. To explore the implications of the compound climate dilemma for prosocial behavior, we conduct experiments in the United States and China, the world's two largest emitters of carbon. Though the pattern of prosocial behavior is virtually identical across the two countries, the introduction of differentiated responsibilities nonetheless has a deleterious effect on successful mitigation.

[*]reuben.kline@stonybrook.edu
[†]naseltzer@gmail.com
[‡]jlukinova@gmail.com
[§]autumn.d.b@gmail.com

Wealth creation in a carbon-based economy inexorably leads to the appropriation of the global climate commons and thus, in the post-industrial revolution era, wealth embodies historical responsibility for the consequences of anthropogenic climate change. This interdependence between climate change and the cost of its mitigation is reflected in the equity principle—common but differentiated responsibilities (CDR)—that underlies international climate negotiations. But, the existence of such interdependence is also the root cause of disputes between developed and developing countries in such negotiations. Existing social dilemma models of climate change negotiations focus only on the mitigation dilemma in isolation, and therefore cannot capture this interdependence between responsibility for climate change and the cost of its mitigation. Responsibilities with respect to climate change are differentiated precisely because of this interdependence. Thus, only by recognizing that anthropogenic climate change and its mitigation constitute an interdependent social dilemma—what we call the *compound climate dilemma*—can we investigate behavior with respect to this important equity criterion. In this study, we put forth the first model of the compound climate dilemma, and use experiments conducted in the United States and China to investigate the implications of such a model.

There is a sense of cautious optimism since the recently concluded United Nations Framework Convention on Climate Change (UNFCCC) negotiations in Paris, in which nearly 200 countries agreed to implement specific steps to curb greenhouse gas emissions. Indeed, commitments under Paris do reflect CDR. However, these commitments—so called Intended Nationally Determined Contributions (INDCs)—are neither binding nor enforceable, and, moreover, are subject to re-negotiation every five years. Thus, in the near and medium term, everyday citizens will become increasingly critical in determining whether these commitments are met. They must be willing to routinely make material sacrifices in order to achieve these goals. Whether and to what extent they are willing to do so may depend on their prosocial preferences in light of differentiated responsibilities arising from different historical trajectories of economic development.

Despite the importance of the climate change problem, several scholars have recently argued that the issue has not been given the attention it deserves in the political science literature (Javeline, 2014; Keohane, 2015). This study contributes to the literature on the politics of climate change in several important ways, including, most importantly, by developing a new behavioral model that explicitly recognizes the compound nature of the climate change dilemma. We then use this model to experimentally investigate the effects of explicitly incorporating differentiated responsibilities on prosocial behavior in the compound climate dilemma. Our results highlight the importance of CDR in both international climate negotiations and individual behavior.

CDR is as old as the international climate negotiations themselves. Indeed, in 1992 it was included in the very first document issued by the UNFCCC, the so-called Rio Declaration on Environment and Development (UNCED, 1992):

> In view of the different contributions to global environmental degradation, States have *common but differentiated responsibilities*. The developed countries acknowledge the responsibility that they bear in the international pursuit of sustainable development in view of the pressures their societies place on the global environment and of the technologies and financial resources they command. [Emphasis added.]

CDR recognizes that anthropogenic climate change and its mitigation constitute a compound and interdependent social dilemma in which the differences in countries' capacities (i.e. the "financial resources they command") are endogenous to differential historical responsibility (i.e.,"different contributions to global environmental degradation"). These two crucial and fundamentally intertwined aspects of CDR correspond to important ethical principles that have emerged in the literature: "ability to pay" and "causal responsibility" (Hayward, 2012; Page, 2008). The effect of ability to pay on mitigation behavior has been thoroughly explored (Tavoni, Dannenberg, Kallis and Löschel, 2011; Wang, Fu and Wang, 2010; Milinski, Röhl and Marotzke, 2011; Burton-Chellew, May and

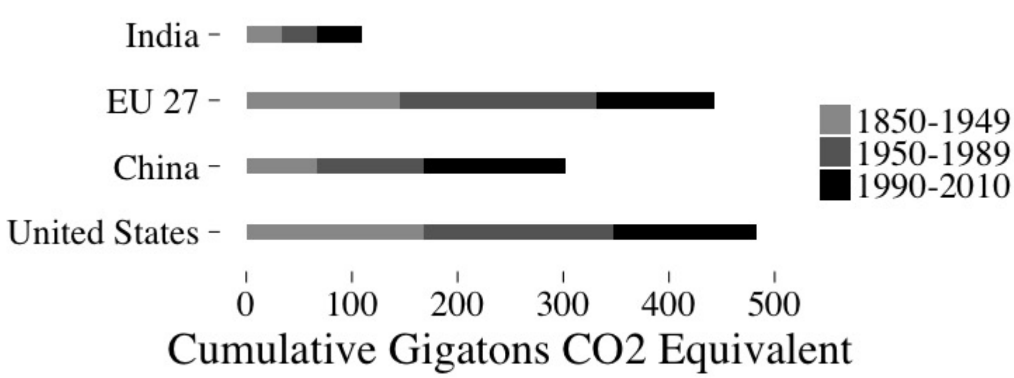
West, 2013; Vasconcelos, Santos, Pacheco and Levin, 2014; Chakra and Traulsen, 2014), *but all such studies treat ability to pay as exogenous*, and therefore are unable to capture the endogenous nature of responsibility for the severity of climate change and the cost of efforts to mitigate it. We argue that endogenous responsibility arising from the compound nature of the climate dilemma is the key to understanding both the promise and perils of the CDR, a principle which is at the core of all multilateral climate agreements successfully negotiated under the auspices of the UNFCCC–e.g., the distinction between Annex I and non-Annex I countries in the Kyoto Protocol; the Warsaw Mechanism for Loss and Damage; and the more recent Paris agreement.

Most of the literature on cooperation in climate change mitigation correctly casts mitigation as a public goods dilemma involving the incentive to free-ride in the production of the public good of mitigation. Thus, game theoretic treatments of the problem often employ some variant of a public goods game—either a linear public goods game (Hasson, Löfgren and Visser, 2010; Barrett and Dannenberg, 2012, 2014) or a type of threshold public goods game known as the collective risk social dilemma (Milinski, Sommerfeld, Krambeck, Reed and Marotzke, 2008; Tavoni et al., 2011; Bynum, Kline and Smirnov, 2016). In order to explore the implications of CDR, we must recognize that climate change is a more complex social dilemma—one that is not exclusively concerned with mitigation. Rather the severity of climate change, and thus the cost of mitigation, is endogenous to historical carbon emissions. As recognized by CDR, causal responsibility for climate change cannot be decoupled from a nation's history of industrialization (Baer, 2006; Tavoni et al., 2011).

We embed the notion of differentiated responsibilities into our experimental design by exogenously manipulating historical opportunities for development. Because it is also well-established that citizens of countries tend to reveal preferences for burden-sharing arrangements that are aligned with self-interest vis-a-vis historical responsibility for climate change and current levels of wealth (Brick and Visser, 2015; Carlsson, Kataria, Krupnick, Lampi, Löfgren, Qin and Sterner, 2013; Lange, Löschel, Vogt and Ziegler, 2010), we conduct this experiment in both the United States and China, currently the world's two largest emitters of greenhouse gases.

Figure 1 shows the dramatic differences in historical carbon emissions for a number of the world's key economies and largest historical carbon emitters, including the United States and China. The vast cross-national differences in historical carbon emissions reveal widely differential historical responsibility for climate change. For example, China's cumulative emissions through 2010 do not even reach the level of U.S. emissions through 1990. This wide disparity in historical responsibility represents a major hurdling in solving the climate change mitigation dilemma. These disparities present a challenge precisely because climate change and its mitigation constitute a compound social dilemma in which increased appropriation of the climate commons implies increased responsibility for the gravity of the mitigation dilemma.

Figure 1: **Historical Emissions by Key Emitters**



Historical emissions data show that wealthier countries, as a result of their long history of industrialization, have appropriated the global climate commons to a greater extent than poorer countries. Currently, wealthier countries' emissions represent the bulk of the stock of carbon emissions. Developing countries, especially China, are engaging in "catch-up" growth and, as a result, are increasing their share of carbon emissions. Figure adapted from http://www.economist.com/blogs/graphicdetail/2013/11/daily-chart.

The behavioral model we introduce here links endogenous causal responsibility with ability to pay, thus capturing the duality of the social dilemma that gives rise to the CDR principle and makes it crucial in the context of international attempts at climate change mitigation. This model allows us, for the first time, to investigate prosociality in both phases of the compound climate dilemma, permitting the observation of two potential types of prosocial behavior: attenuated appropriation of the climate commons and increased contributions to mitigation. Because developed countries—the "early developers" in our experiment—are the most historically responsible for climate change, these two behaviors are broadly what CDR calls on the citizens and governments of industrialized countries to do. Behavior reflecting such prosociality organically emerges among experimental participants in *both* the United States and China and is broadly consistent across the two countries. Despite significant prosocial behavior being exhibited across both of our samples, when we compare the results from experimental treatments using our compound social dilemma to identically, but exogenously, parameterized mitigation dilemmas, we find that the compound nature of the social dilemma does make it significantly more difficult to resolve.

Nonetheless, our results allows for some hope that the CDR principle underlying the commitments countries made under Paris may in fact be subject to a similar understanding across nations, even those whose national interests strongly diverge with respect to historical accountability for climate change. This divergence is reflected in the controversial listing of China as a non-Annex I country—therefore not subject to emissions reductions under the Kyoto Protocol—while the United States is listed in Annex I, and therefore subject to emissions reductions.

Climate Change and its Mitigation as a Compound Social Dilemma

Climate change mitigation is a public good, contributions to which are often modeled behaviorally using a threshold public goods game known as the collective risk social dilemma (Dreber and Nowak, 2008; Milinski et al., 2008), or other standard public goods games. The collective risk social dilemma has become prominent in other disciplines, but aside from Bynum, Kline and Smirnov (2016) is still largely unknown in the political science literature. Here we extend it to capture the important features of the the compound climate dilemma and the CDR equity principle. The collective risk social dilemma is a threshold public goods game of loss avoidance. Players in small groups are endowed with an exogenous initial wealth level, portions of which can be contributed—over multiple periods—toward an exogenous threshold value. If the group's collective contributions meet or exceed this threshold, the group avoids a collective loss of earnings and each member retains the un-contributed remainder of their endowment. If the group's collective contributions fail to reach the threshold value, the members face some exogenously given and commonly known probability of climate change-induced "catastrophic economic loss" (loss of the entirety of their retained earnings). Thus, as in all public goods games, individuals have an incentive to free-ride on the contributions of others, even if free-riding is not group welfare-maximizing. (Unlike in linear public goods games, threshold public goods games do generally have equilibria characterized by non-zero contributions. See appendix for more information about the equilibria in the compound climate dilemma.)

But, as we touched on above, responsibility for anthropogenic climate change in a global, carbon-based economy is a function of the cumulative appropriation of the "global climate commons." In a carbon based economy this appropriation is a necessary byproduct of wealth-generating activities (Baer, 2006; Tavoni et al., 2011). This recognition—that responsibility for climate change mitigation is held in common but is differentiated—is inextricably linked to a recognition that anthropogenic climate change and its mitigation actually constitute a unique form of compound social dilemma comprising a mitigation dilemma in which the parameters are endogenous to the level of appropriation in a common pool resource dilemma. This interdependent social dilemma is what we call the *compound climate dilemma*.

In order to capture the notion of common *but differentiated* responsibilities, what is missing from purely public goods treatments of climate change mitigation relates to the fact that the Earth's atmosphere is capable of absorbing only a finite amount of carbon [1] before anthropogenic climate change results in catastrophic economic losses (Schellnhuber, 2006). Therefore, appropriation of the global climate commons is a common pool resource dilemma (Dietz, Ostrom and Stern, 2003). In the global climate commons, as in all common pool dilemmas, each individual actor gains through wealth generating activities that appropriate the common pool. This appropriation, while enriching those who engage in it, also increases the probability of economic loss from climate change and raises the cost of any successful mitigation effort—i.e., the threshold (Council of Economic Advisers, 2014). Collective action in the domain of climate change therefore constitutes a compound, interdependent social dilemma: the degree to which the climate commons have been appropriated determines the cost of climate change mitigation *and* the severity of the consequences should mitigation efforts fall short. In this way, the three key parameters of the collective risk social dilemma—initial wealth, collective risk, and the threshold value—are in reality endogenously determined. In current conceptions of the collective risk social dilemma (e.g., Milinski et al., 2008; Milinski, Röhl and Marotzke, 2011; Tavoni et al., 2011) each of these three parameters has been subject to experimental manipulation, but they are all invariably treated as exogenous. In our reconceptualization of the mitigation dilemma, each of these three parameters is endogenously determined through subjects' own appropriation behavior in a preceding common pool resource dilemma. This novel design therefore captures the interdependence of the climate change dilemma and allows us to behaviorally investigate the CDR principle.

This study is the first to introduce endogenous endowments into the collective risk social dilemma. Previous studies have endogenized endowments in other public goods games, with mixed results. Some studies find no difference between exogenously and endogenously generated endowments (Cherry, Kroll and Shogren, 2005; Clark, 2002), whereas other studies find that cooperation is increased when endowments are earned in

[1] Technically, the crucial quantity of interest is a "carbon-dioxide equivalent" (Gohar and Shine, 2007) but here we refer to carbon for simplicity's sake.

linear public goods games (Harrison, 2007; Spraggon and Oxoby, 2009) and "best-shot" public goods games (Kroll, Cherry and Shogren, 2007).

In the compound climate dilemma there are two ways in which actors can display prosocial behavior: by limiting their appropriation of the global climate commons and/or by shouldering a greater share of the burden of mitigation. In addition to being behavioral embodiments of the CDR principle, both of these behaviors have analogues in existing climate agreements. To manage the global climate commons, the Kyoto protocol mandated emissions cuts for developed countries but explicitly exempted developing countries to allow for "catch-up" growth (Baer, 2002). With respect to mitigation and adaptation, both the Cancún Adaptation Framework (UNFCCC, 2010) and the Warsaw International Mechanism for Loss and Damage (Mechler, Bouwer, Linnerooth-Bayer, Hochrainer-Stigler, Aerts, Surminski and Williges, 2014) would obligate wealthy countries to build capacity in the poorest and most vulnerable countries and to compensate vulnerable developing countries for economic loss and damage associated with climate change.

We simulate the real world strategic conditions that give rise to the compound climate dilemma. To do so, we employ student samples in both the U.S. and China. In studies such as this, it is natural to be concerned about external validity. In particular, some may be skeptical about what we can learn about climate negotiations from experiments with student samples. First, it is important to note that experiments, in order to achieve internal validity, always sacrifice external validity to some extent. With our incentivized controlled experiment we gain significant internal validity compared to observational studies. Precisely because of the multiple sources of endogeneity, it is virtually impossible to separate the many potential motivations that exist for any given negotiating position. Here we systematically isolate important conceptual factors and can observe their *causal* effect. Moreover, in the Post-Paris world, the attitudes and behaviors of the mass public will become ever-more important if the ambitious goals set out in the INDCs are to be met.

There is now a large number of studies of international bargaining, conflict and negotiation that have used student samples in laboratory experiments (e.g., Hafner-Burton, LeVeck, Victor and Fowler, 2014; McDermott, Cowden and Koopman, 2002; Tingley, Wang et al., 2010; Tingley, 2011; Tingley and Walter, 2011). Second, while student subjects do differ from policymaking elites in many ways, Hafner-Burton et al. (2014) find that students and policy elites behave in similar ways. Moreover, norms (our focus here), by their very nature, are widely shared within a culture and the behavioral norms that have been internalized by international negotiators must be to some extent a function of the culture of the country they represent. Even if citizens may differ from the elites involved in climate negotiations, the preferences of average citizens are important because the elites will ultimately need the support of the mass public to reach an effective agreement (Bechtel and Scheve, 2013; Tingley and Tomz, 2013). This support, we believe, will become ever more important for meeting INDC obligations under the Paris agreement.

Most importantly, once we have introduced the compound climate game, we are interested in whether we can uncover any general patterns of prosocial behavior in either or both of the phases of the climate game dilemma. Previous research on cross-national differences in climate change equity principles are not grounds for optimism. Using both surveys (Carlsson et al., 2013; Lange et al., 2010) and incentivized experiments (Brick and Visser, 2015), studies have consistently shown that a country's citizens and negotiators tend to support equity principles that favor their own country's interests. For example, Brick and Visser (2015) show that "the use of the historical and future polluter-pays rules by American and Chinese participants is consistent with material self-interest." We think it is possible that the compound climate dilemma—which makes the link between wealth generation and responsibility for climate change salient and explicit—could result in a convergence toward equitable behavior that is consistent with the CDR principle. In sum, we are interested in the effect of this interdependence of the social dilemma on prosocial behavior–both appropriation and contribution behavior–and whether this interdependence has a salubrious or deleterious effect on successfully overcoming the mitigation dilemma.

Experimental Design

To model the compound climate dilemma with differentiated responsibilities, we begin with two between-subjects treatments, each with 12 randomly and anonymously constituted groups of 6 players. We conduct these identical set of treatments in both the United Sates and China. The compound climate dilemma is characterized by a scenario in which the parameters of the mitigation dilemma (endowments, thresholds, and collective risk) are determined by, and therefore endogenous to, behavior in a preceding appropriation dilemma. In the *Endogenous-Undifferentiated Responsibilities* (*End-Undiff*) and *Endogenous-Differentiated Responsibilities* (*End-Diff*) treatments, appropriation behavior in a preceding common pool resource dilemma phase determines the players' initial wealth, the threshold value, and the collective risk in a subsequent collective risk social dilemma mitigation phase. This will allow us to measure prosocial behavior in each of the two interdependent social dilemmas that together constitute compound climate dilemma. To determine what effect this interdependence has on successfully resolving the mitigation dilemma, we conduct two exogenous treatments in the United States: *Ex1* and *Ex2* serve as exogenous responsibility control conditions for *End-Undiff* and *End-Diff* respectively. These control treatments lack a preceding common pool resource dilemma—and therefore the feature of endogenous responsibility—but are otherwise identical in that the key parameters in the mitigation dilemma are exogenously assigned based on parameters inherited from their corresponding *End-* conditions. In total, across the 6 conditions, 432 undergraduate students participated in our study.

In *End-Undiff*, responsibility is endogenous and undifferentiated: in each of $t = 1, \ldots, 10$ periods each participant $i = 1, \ldots, 6$ in group $j = 1 \ldots 12$ can choose an appropriation level

$a_{ij,t} \in \$[0,1,2,3,4]$ from a common pool of resources, R. $A_{ij} = \sum_{t=1}^{10} a_{ij,t}$ is the total appropriation for i in j and constitutes their endowment in the next phase of the game. The total appropriation of group j is $\mathbb{A}_j = \sum_{i=1}^{6} \sum_{t=1}^{10} a_{ij,t}$.

In *End-Diff* responsibility is endogenous, but differentiated: in the appropriation phase three randomly selected members of each group (late developers, or "*lates*") are unable to engage in appropriation in the first 5 periods, while the other half of the group (early developers, or "*earlys*") are free, as are all subjects in *End-Undiff*, to appropriate the common pool in all 10 periods.

In both *End-Undiff* and *End-Diff* the next phase of the game–the mitigation phase–comprises 10 periods of a collective risk social dilemma, with the threshold for group j, $\mathbb{T}_j$, equal to $\mathbb{A}_j * 0.53$. Participant ij's endowment is A_{ij}. In each of $s = 1, \ldots, 10$ periods, each player chooses a contribution level, $c_{ij,s} \in \$[0,1,2,3,4]$. $C_{ij} = \sum_{s=1}^{10} c_{ij,s}$ is each participant's total contribution toward the threshold, and $\mathbb{C}_j = \sum_{i=1}^{6} \sum_{s=1}^{10} c_{ij,s}$ is each group's aggregate contribution. Recall that in a collective risk social dilemma, if $\mathbb{C}_j < \mathbb{T}_j$ then catastrophic economic loss occurs with a specific, commonly known probability (p). In the experiment, this economic loss is represented by the loss of the group members' remaining endowments. If $\mathbb{C}_j \geq \mathbb{T}_j$ then loss is avoided and ij's payoff is $\pi_{ij} = A_{ij} - C_{ij}$. If $\mathbb{C}_j \leq \mathbb{T}_j$, then the (expected) payoff is $\pi_{ij} = (1 - p) \times (A_{ij} - C_{ij})$

If $\mathbb{C}_j < \mathbb{T}_j$, all members of group j will face a collective risk that they lose their remaining endowments. Here p is increasing as a function of $\mathbb{A}_j$(Hansen, Sato, Kharecha, Beerling, Berner, Masson-Delmotte, Pagani, Raymo, Royer and Zachos, 2008). Climate risk is monotonically increasing in the degree of appropriation of the climate commons, but there are likely to be tipping points in which the climate system moves rapidly from one equilibrium to another (Alley, Marotzke, Nordhaus, Overpeck, Peteet, Pielke, Pierrehumbert, Rhines, Stocker, Talley et al., 2002). We capture these tipping points with a step function that gives us the values in Table 1. Because *End-Diff* contains 25% fewer player-rounds in which appropriation can take place, maximum possible appropriation is reduced by 25% relative to *End-Undiff*, and the mapping is scaled down proportionally.

Table 1: **Collective risk is step-wise increasing in the amount appropriated**.

$\mathbb{A}_j$: *End-Undiff*	$\mathbb{A}_j$: *End-Diff*	Collective Risk
0-60	0-45	0.167 (2/12)
61-120	46-90	0.5 (6/12)
121-180	91-135	0.75 (9/12)
181-240	136-180	0.917 (11/12)

In order to uncover the effect of endogenous responsibility on performance in the mitigation dilemma we must compare it to an otherwise identical scenario that lacks such a feature. In the *Exogenous* control treatments (*Ex1* and *Ex2*) there is no common pool resource dilemma. At the outset of the mitigation dilemma, the conditions are otherwise identical because the important parameters in the collective risk social dilemma are exogenously determined by parameterizing each group in the *Ex1* (*Ex2*) condition with A_{ij}, $\mathbb{A}_j$, p, and $\mathbb{T}_j$ from a group in the *End-Undiff* (*End-Diff*) condition. Thus, when the mitigation dilemma begins the difference between the *Exogenous* and *Endogenous* conditions is only the shared history of appropriation behavior and the historical responsibility for the size of the threshold and the magnitude of the collective risk. Table 2 summarizes the important features of the four conditions we describe above.

The experimental sessions followed standard economics protocols: no deception, payment of real money and full information about the decision-to-payoff mapping. The experiments were conducted at large, public universities in the Northeastern United States and Northeastern China. Further details of the experimental sessions and protocol can be found in the supplementary appendix.

Hypotheses

In this study, we explicitly recognize the compound climate dilemma, and operationalize this concept in a novel experimental framework. As discussed above, the compound climate dilemma allows for two distinct types of prosocial behavior: attenuation of one's

Table 2: **Summary of Treatments**.

Treatment [Site(s)]	Appropriation Phase	Parameters in Mitigation Phase	Mitigation Phase
End-Undiff [U.S. & China]	Each of the 6 players in each group can appropriate between $0 and $4 in each of 10 periods.	Individual endowment equal to amount total amount appropriated. Threshold equal to 53% of group's total endowment. Collective risk depends on total amount appropriated (see Table 1).	Using their endowment, each of the 6 players in each group can contribute between $0 and $4 in each of 10 periods. If sum of all contributions meets or exceeds the threshold, players retain remainder of their endowment. If contribution less than threshold, the collective risk that group loses their earnings is based on group's appropriation amount.
End-Diff [U.S. & China]	Each of the 3 early developers in each group can appropriate between $0 and $4 in each of 10 periods. Each of the 3 late developers can appropriate only in periods 6 through 10.	Same as *End-Undiff*	Same as *End-Undiff*.
Ex1 [U.S. only]	No appropriation phase.	Each group inherits parameters (endowments, threshold and collective risk) from a randomly matched group in *End-Undiff*.	Same as *End-Undiff*.
Ex2 [U.S. only]	No appropriation phase.	Each group inherits parameters (endowments, threshold and collective risk) from a randomly matched group in *End-Diff*.	Same as *End-Undiff*.

appropriation in the common pool dilemma, and increasing one's contributions in the mitigation dilemma. We want to investigate how the interdependence of these two social dilemmas affects behavior in each of the two constituent dilemmas. The CDR in effect says that early developers should both limit their appropriation *and* increase their contributions. At the same time, we expect that two interrelated social dilemmas should be harder to solve than one, especially because the embedded endogenous responsibility for the severity of the mitigation dilemma makes the interdependent nature of the climate change dilemma particularly salient. With these considerations in mind, we lay out the following four hypotheses:

Hypothesis 1 *Early developers in* End-Diff *will appropriate less, relative to both Late developers in* End-Diff *and all subjects in* End-Undiff*.*

Hypothesis 2 *Early developers in* End-Diff *will contribute proportionally more, relative to both Late developers in* End-Diff *and all subjects in* End-Undiff*.*

Because there are a number of studies showing that citizens' decision making is biased toward behavioral equity norms that favor their own countries' positions in international climate negotiations, we conduct our experiment in both the United States and China. This allows us to investigate behavior with respect to CDR principles across subjects in the world's two largest emitters of carbon. Despite China currently having a larger flow of carbon emissions, as Figure 1 demonstrates, the U.S. and China have quite different historical stocks of emissions. There are a number of studies which find that both the mass public and elites of a country tend to behaviorally reveal preferences for equity principles that favor their own country's interests (Brick and Visser, 2015; Carlsson et al., 2013; Lange et al., 2010) As a result, we may also expect to observe different behavior with respect to differentiated responsibilities in the game.

In particular, in our framework, early developers who attenuate their appropriation and increase their proportional contributions are revealing a preference for ability to pay and polluter pays principles. On the other hand, when early developers contribute their "fair share" (53%) or less or late developers contribute less than their fair share, then they reveal preferences for more equal burden sharing arrangements.

Hypothesis 3 *With respect to appropriation and contribution choices, Chinese subjects will be more likely than their American counterparts to favor differentiated responsibilities, i.e. that early developers appropriate less and contribute more. Conversely, American subjects should be more predisposed to favor equal burden sharing arrangements.*

Even if we do observe prosocial behavior in both the appropriation and contribution phases, we expect the compound nature of the climate mitigation dilemma to make successful resolution more difficult than it would otherwise be in a simple dilemma. Because a compound social dilemma is likely more difficult to solve than an exogenous one, we expect that success in the *Endogenous* conditions will be lower than that of the *Exogenous* conditions, even if, by design, the initial parameters of the collective risk game are identical.

Hypothesis 4 *Endogenous responsibility will decrease success in the mitigation dilemma, relative to the standard scenario in which responsibility is exogenous.*

Results

We first analyze the prosociality of appropriation behavior, followed by a similar analysis of contribution behavior, while also comparing and contrasting the behavior of the Chinese and American samples.

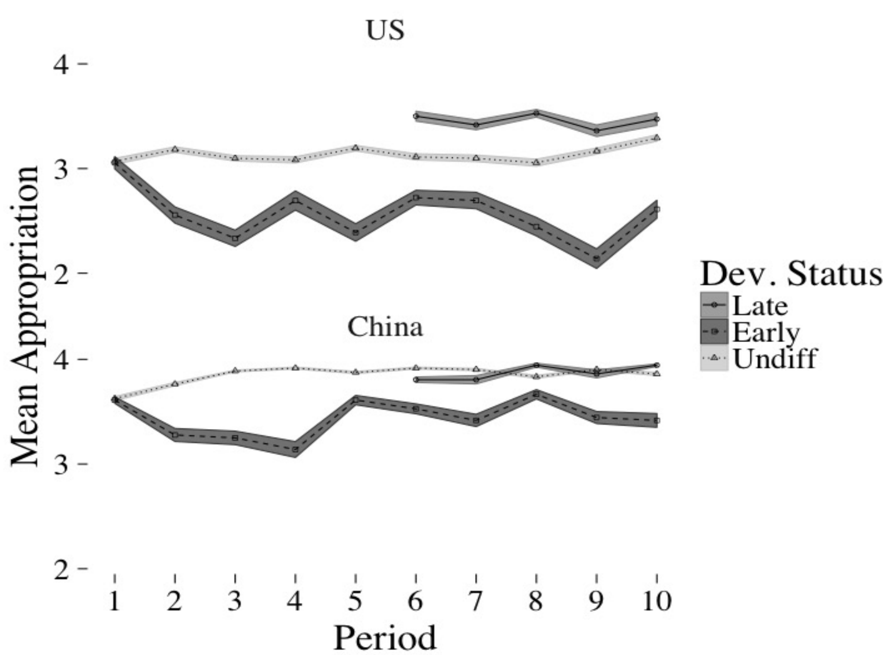
In order to investigate the effects of common but differentiated responsibilities on prosocial behavior in the compound climate dilemma, we exploit the randomly assigned *early/late* developer distinction within the *End-Diff* treatment and compare the behavior of each of those groups with the behavior of the (un-differentiated) participants in *End-*

Undiff. We first analyze differentiated responsibilities in appropriation behavior, and then we turn to contribution behavior.

Appropriation Behavior

Figure 2 shows the mean appropriation across rounds by development status in the U.S. and China. In both panels, the figure indicates large and significant differences in appropriation levels that are driven by the subjects' status in the experiment. In the U.S., the earlys in *End-Diff* appropriate considerably less compared to both the lates in *End-Diff* and all subjects in *End-Undiff*. This large attenuation of appropriation by earlys in *End-Diff* is replicated with our Chinese sample. In fact, the overall pattern of the results is very similar across the two samples, indicating a shared understanding of differentiated responsibilities that is conditional on one's development status.

Figure 2: **Average appropriation by period and development status, U.S. and China:** Error bars represent 95% confidence intervals. Across both countries, early developers in *End-Diff* appropriate much less than late developers and less than those in *End-Undiff*.



We use Tobit regressions to estimate treatment effects with respect to total individual appropriation. Tobits are used because they allow us to take into account the fact that the appropriation amounts are constrained between 0 and 4 in each period—i.e. they are "censored." To take into account the non-independence of behavior within groups, we cluster standard errors at the group level.

Figure 3 presents the average treatment effects estimated from three such bivariate regressions with intercepts each for both the US and China (full results of the regressions can be found in the appendix). The coefficients in each of these regressions is the estimated difference in appropriation between two groups: in (a) it is the difference (in the last 5 periods) between lates in *-Diff* and all subjects in *-Undiff*; (b) plots the difference (in the last 5 periods) between the earlys and the lates in *-Diff*; and (c) is the difference (across all 10 periods) between the earlys in *-Diff* and those in *-Undiff*.

Figure 3: **Differences in Appropriation by Country and Development Status:** Ranges represent 95% confidence intervals of differences. Earlys appropriate significantly less than both lates and those in *Undiff*. Lates do not appropriate significantly more than those in *Undiff*.

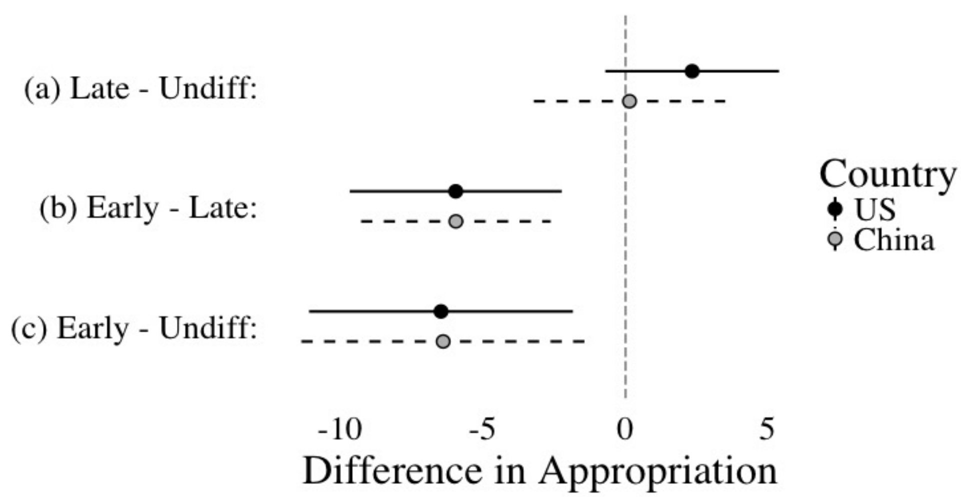


Figure 3 shows that in both the US and China, earlys in *End-Diff* appropriate significantly less—$6.47 in the US and $6.40 in China (out of a maximum of $40)—than participants in *End-Undiff* over the course of the ten periods of appropriation. Likewise, in terms of appropriation across periods 6 through 10, in both the US and China we find that earlys in *End-Diff* appropriate significantly less than lates (in both samples, $5.96 out of a maximum of $20). Again focusing on the final 5 periods, if we compare all participants in *End-Undiff* to lates in *End-Diff* we find that the latter group does not extract significantly more. This finding of no difference is likely due to ceiling effects, especially in China: given the near-maximal appropriation of the subjects in *End-Undiff*, it is virtually impossible for the lates in *End-Diff* to appropriate significantly more.

Despite these broadly similar patterns across the two countries, we do find one important set of differences between them: across each randomly assigned development status, Chinese subjects appropriated significantly more than U.S. subjects. Using two-sided Kruskal-Wallis (KW) tests, we find these differences are highly significant: For earlys [n=24; p=0.006]; for lates [n=24; p=0.002]; and those in *End-Undiff* [n=24; p=0.001]. If the Chinese subjects are pre-disposed to take the perspective of late developers and therefore attempt to engage in "catch-up" growth, this could explain these differences. Still, the virtually identical set of treatment effects (see figure 3) demonstrates that participants in both countries are nonetheless inclined toward similar interpretations of differentiated responsibilities that are consistent with CDR, and in particular those relating to economic growth and causal responsibility.

In sum, given the significantly reduced appropriation by the earlys, our results are consistent with hypothesis 1. On the other hand, and in contrast to most previous findings, because the appropriation behavior of the U.S. and Chinese subjects was virtually identical, our results do not support hypothesis 3's prediction of cross-national differences.

Contribution Behavior

What is the effect of differentiated responsibilities on prosociality in contribution behavior? To answer that question, we again turn to the early/late developer distinction, and its contrast to the *End-Undiff* condition. We hypothesize that earlys in *End-Diff* will contribute proportionally more than lates, simply as a result of their privileged status and the implications such privilege has for differentiated responsibilities.

Indeed, in China, just as in the U.S., early developers contribute proportionally more than late developers. In China the difference—0.546 versus 0.397—is highly significant [N=24; p = 0.001, Kolmogorov-Smirnov (KS) test]. The difference in the U.S. is somewhat smaller (0.555 versus 0.424) but still highly significant [N=24; p = 0.004, KS test]. The relationship is clear and strong: in both samples, early developers, when compared to late developers, contribute a significantly higher proportion of their endowment toward their group's threshold. This difference could be because the earlys are unusually pro-social in their contribution behavior, because the lates are particularly selfish in theirs, or even some combination of both. To sort out what is driving this difference, we can compare each of them to the *End-Undiff* condition. This pattern of behavior is again consistent across countries.

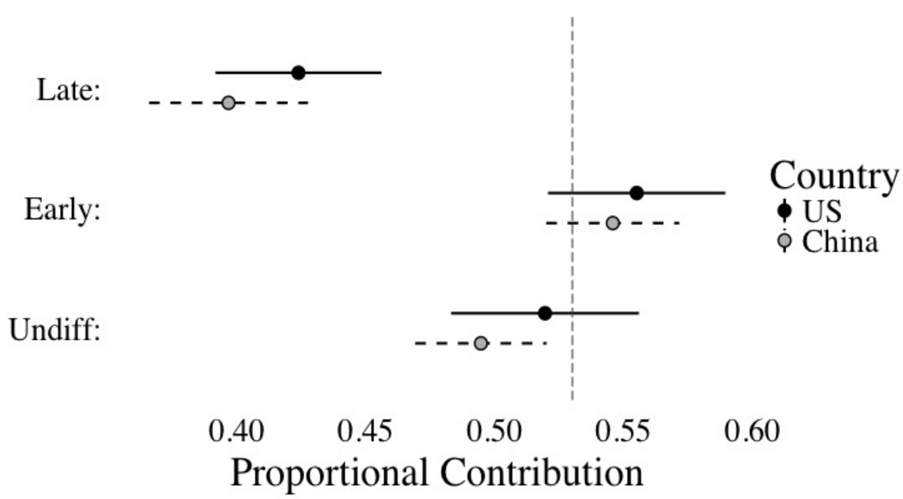
First, in both the US [0.555 versus 0.519; N=24; p=0.058, KS test] and China [0.546 versus 0.495; n=24; p=0.017, KS test] earlys contribute significantly more than those in *End-Undiff*. Though this difference is only marginally significant for the U.S., it is highly significant it we combine the data from the two countries. Moreover, the result is based on group level data, so the test is powered with only 24 observations.

Second, in both the US (0.424 versus 0.519; N=24; p=0.000, KS test] and China [0.397 versus 0.495; N=24; p=0.004, KS test] lates are contributing significantly *less* than their counterparts in *End-Undiff*. In both the US and China, the earlys contribute significantly (proportionally) more than the lates in *End-Diff*.

These results show that although early developers are significantly more prosocial than the undifferentiated participants in *End-Undiff*, their prosociality is more than offset by the increased selfishness of the late developers, relative to those in *End-Undiff*. This would seem to imply that the introduction of differentiated responsibilities has a net deleterious effect on cooperation in the mitigation dilemma. In the next subsection, we will re-visit the effect of the compound climate dilemma on performance in the mitigation dilemma.

For each country and development status, figure 4 plots the mean proportional contributions with their 95% confidence intervals. These results again highlight the similarities in behavior between participants in the two countries: when we compare within development categories, the proportional contribution rates are not significantly different across countries. This graphical display of the results also reinforces the claim that the prosociality of the earlys is more than offset by the lates' under-conributions (relative to their responsibility to the dilemma and also relative to participants in *End-Undiff*). This is an important result that has potential implications for understanding and perhaps even improving climate change mitigation efforts.

Figure 4: **Average Proportional Contributions by Country and Development Status** Ranges represent 95% confidence intervals. Dashed vertical line indicates "fair-share" contribution of 0.53. Lates contribute less than both earlys and those in *-Undiff*.



Overall, the results with respect to contribution behavior tend to support hypothesis 2's prediction that earlys will contribute proportionally more than other participants. They did contribute much more than the lates, but not significantly more than those in *-Undiff*. Thus, the disparity in contributions between earlys and lates seems more due to reduced contributions by lates rather than increased contributions by earlys. Contribution behavior is quite similar across countries, and therefore again is not consistent with hypothesis 3.

Behavior in both the appropriation and mitigation phases of the dilemma, and across both countries, is clearly and consistently affected by our experimental manipulations of development status. These results strongly demonstrate that, despite the difficulties introduced by the compound climate dilemma, we observe prosocial behavior in terms of both attenuated appropriation and increased contributions by the early developers. At the same time, the early developers in some cases demonstrated increased selfish behavior. The question remains, then, as to what is the overall effect of the compound nature of the dilemma on success in solving it? To answer this question, we turn to comparisons of the *End-* conditions and the *Ex-* conditions conducted in the U.S.

Success and Performance in the Compound Climate Dilemma

The results we have presented thus far demonstrate that we do in fact observe a fair amount of prosocial behavior, in terms of both appropriation and contribution levels. Here we investigate whether there was sufficient prosocial behavior to overcome the challenges of the jointly constituted compound climate dilemma. To do so, we compare the *End-* conditions we conducted in the U.S. and analyzed above, to the exogenous–*Ex-*– conditions described above.

To begin we give an overview of the behavior across the two sets of treatments. Table 3 displays the endowments, thresholds, contributions, and expected earnings efficiency for each of the groups in the *End-Undiff* and *Ex1* treatments for the sessions conducted in the United States. Efficiency is calculated as the proportion of the maximum aggregate payoffs for each scenario. This maximum is reached through full appropriation combined with contributions sufficient to just cover the threshold (this behavior would also constitute one of many Nash Equilibria in this game. See appendix for a more thorough discussion of the equilibria of our game). These values are 112.8 for the *-Undiff* treatments and 84.6 for the *-Diff* treatments. Despite the fact that group-level endowments, thresholds and collective risk are, by design, identical between the respective *End-* and *Ex-* conditions, table 3 also shows success, contributions, and efficiency to be lower in the *End-* conditions compared to the *Ex-* conditions.

Table 3: **Summaries of *-Undiff* and *-Diff* Conditions**

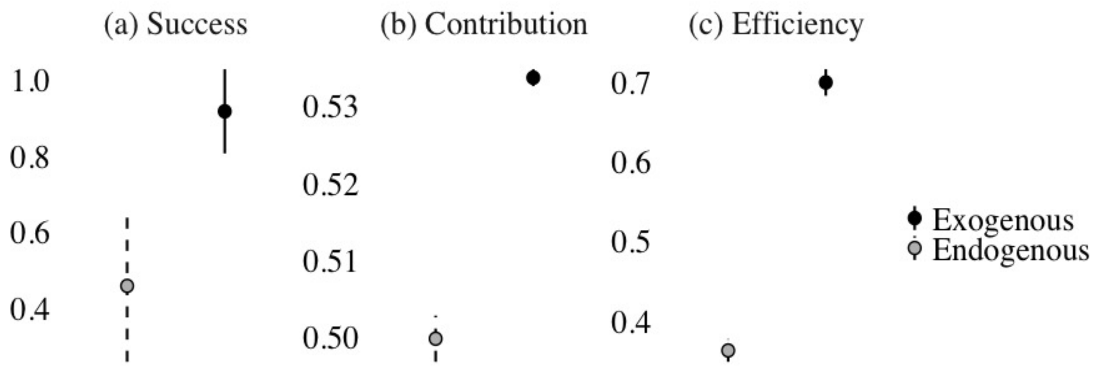
Condition	Endow.	Thresh.	Coll. Risk	Success		Contributions		Efficiency	
				End-	Ex-	End-	Ex-	End-	Ex-
-Undiff	188	100	0.83	7/12	12/12	97	101	0.49	0.77
-Diff	129	68	0.81	4/12	10/12	64	69	0.35	0.63

But, is there a significant difference between the *End-* and *Ex-* conditions in terms of their performance in the mitigation dilemma? Figure 5 displays the difference between the two combined *End-* treatments and the two combined *Ex-* treatments, using three different performance metrics: (a) success [in terms of proportion of groups meeting the threshold]; (b) [proportional] contributions to the threshold, and; (c) [earnings] efficiency. As figure 5 demonstrates, regardless of the metric used, participants in the *-End* treatments were significantly less successful in solving the mitigation dilemma.

As figure 5 indicates, there is strong evidence that the incorporation of endogenous responsibility makes successful mitigation less likely, behavior which is consistent with hypothesis 4. Given that two interdependent social dilemmas would likely be more difficult to solve than a single, isolated one, this result is not entirely surprising. On the other hand, given that we did observe a significant amount of prosocial behavior in both phases of the dilemma, these reduced levels of success are somewhat of a puzzle. Part of the answer to that puzzle is that while the early developers do tend to be more prosocial in both phases of the compound dilemma, the late developers tend to be *much* less so.

Our final result does show one mechanism through which the chances of success in the compound climate dilemma might be increased. Given the symbolic importance of earlys' contributions in *End-Diff* in signaling that early developers are "taking the lead",

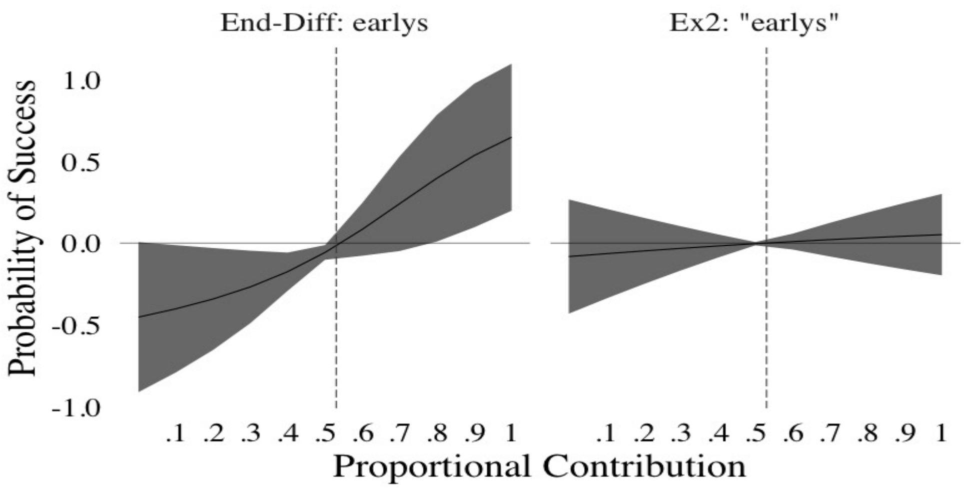
Figure 5: **Effect of Endogenous Responsibility on Three Performance Metrics:** (a) success [in meeting the threshold]; (b) [proportional] contributions, and; (c) [earnings] efficiency. Point estimates and 95% confidence intervals indicate that groups in the endogenous responsibility conditions perform significantly worse than those in the exogenous conditions.



the rate of such contributions could affect the likelihood of success in resolving the mitigation dilemma. Using logistic regression with standard errors clustered at the group level, we find that in *End-Diff* the marginal effect of proportional contributions by earlys (compared to lates) significantly increases the odds that a group is successful in meeting the threshold [n=72; Odds Ratio=840.3, p=0.026]. Conversely, there is no significant effect of giving by "earlys" in *Ex2* [n=72; OR=2.56, p=0.649]. Note, however, that because only four groups in *End-Diff* met the threshold the corresponding estimate is subject to a large degree of uncertainty, with the 95% confidence interval for the estimate of the OR being (2.2, 319,282.8). Thus these estimates should be treated with some caution, but on the whole they indicate that increased proportional contributions by early developers are a potentially important source of successful cooperation.

Figure 6 graphically displays the key result from these estimations. We plot the marginal effects of additional proportional contributions on the likelihood of a group's success, i.e., the additional effect on the probability of success that giving by an early/"early" has over

Figure 6: **Marginal Effect of Proportional Contributions on the Probability of Group Success.** The solid black line represents the estimate of the marginal effect. Effects are calculated for each proportional contribution level marked on the y-axis. The shaded area represents the 95% confidence interval. Dashed vertical line indicates "fair-share" contribution (0.53). Results show that increased contributions by earlys in *End-Diff* (left panel) significantly increase the chance of success, while similar behavior by "earlys" in *Ex2* has no such effect.



and above the baseline effect of that same level of proportional giving by a late/"late". In the left panel we plot the marginal effect for earlys in *End-Diff*, and in the right panel we plot this same quantity for "earlys" in *Ex2*. Recall that within the experiment, this latter group was not differentiated in *Ex2*. However, because there is no

Results show that, compared to giving the 0.53 proportion, maximum contributions by early in *End-Diff* increase the probability of success by nearly 0.5, and minimum contributions *decrease* the likelihood of success. At the same time, minimum contributions by earlys Results for *Ex2* show that a maximum "early" contribution does not significantly increase the probability of success. Together, these results demonstrate that the signaling effect of early giving positively affects success rates.

These results with respect to our compound climate change dilemma replicate findings by (Tavoni et al., 2011) in the context of the simple mitigation dilemma, who showed that non-binding "pledges" by (exogenously) wealthier participants increased the likelihood of meeting the threshold. But, because participants in *Ex2* who were assigned the *End-Diff* earlys' endowments—the "earlys"— are, from a standpoint of relative wealth at least, identical to the true earlys in *End-Diff*, then this results implies that there is a very real symbolic value in proportional giving by early developers in the compound climate change dilemma.

Discussion

Though public goods games are most often used to model the *mitigation* phase of the climate change dilemma in isolation, current approaches fail to recognize the interdependent appropriation dilemma, and therefore fail to account for the compound nature of the climate change dilemma. This study is the first to broaden the mitigation dilemma's ontology to include these crucially important linkages—what we have dubbed the "compound climate dilemma."

Indeed the dual nature of the social dilemma was recognized by a number of the study participants, many of whom noted that their strategy included attempting to limit their appropriation to keep their group's probability of loss low. One good example of such a quote is the following: "I tried to appropriate enough to keep below a certain amount according to what the rest of the group was appropriating. I wanted to aim for 120 max, but then it went over quite quickly, so I tried to aim for below 180, but unfortunately, [that] did not happen." (See appendix for additional quotations from participants.)

As in the real-world, our results show that the compound nature of the dilemma is a considerable obstacle to successful mitigation efforts. In the seminal study that introduced the collective risk social dilemma, Milinski et al. (2008) find that higher levels of collective risk facilitate cooperation. In our study, despite the fact that average levels of collective risk—0.83 in *-Undiff* and 0.81 in *-Diff* are nearly as high as in the highest collective risk condition (0.90) in Milinski et al. (2008), endogenously-created risk does not seem to do so. Using three distinct measures of performance in the mitigation dilemma we find that casting the mitigation phase as a compound social dilemma reduces successful mitigation efforts. The implication of these findings is that endogenous responsibility complicates the dilemma, even at high levels of collective risk. This finding has important implications for mitigation behavior because it would seem to indicate that the increasingly dire predictions by climate scientists will not necessarily spur political action.

Still, the prosocial behavior we observe in both phases of the dilemma and, importantly, across countries, provides some grounds for optimism as well: experimental subjects tend to organically arrive at behavioral arrangements which resemble notions of equity consistent with the CDR principle. Earlys in *End-Diff*—advantaged with greater opportunity for wealth creation—are not only willing to forego additional wealth-creation that would have exacerbated climate change but also to provide a proportionally higher rate of contribution towards mitigation efforts. Our results are based on student samples taken from universities in the U.S. and China—currently the world's two largest emitters of carbon.

Because these two countries are key players in international climate negotiations, the behavior of their citizens could prove to be the determining factor in success of the climate change dilemma. However, as demonstrated in figure 1, their development trajectories, historical emissions, and historical responsibility for climate change are quite distinct. Nonetheless—and contrary to previous studies which find that cross-national differences in behavior align with the interests of the subjects' own countries (Carlsson et al., 2013; Lange et al., 2010; Brick and Visser, 2015)—revealed preferences for equity are nearly identical across the two countries. It therefore seems that the compound nature of the dilemma, when made salient, can perhaps lead to a convergence in the norms that govern behavior in the compound climate dilemma, even in countries with disparate historical emissions trajectories. The spirit of this behavior mirrors the CDR principles that are embedded in all important international agreements on climate change: the Kyoto Protocol, the Cancún Adaptation Framework, the Warsaw Mechanism, and, most recently, the

INDCs committed to by nearly 200 countries in Paris in 2015.

Larger (proportional) contributions by earlys are significantly more likely than those by lates to lead to success in the mitigation dilemma, suggesting that—as envisioned by the CDR principle—there is an important leadership role to play by "early developers" in the real world. Given this behavioral tendency by the earlys, the question is then why were groups in *End-Diff* the least successful of all the treatments? Our design cannot offer a definitive answer, but a clue lies in the fact that the lates in this treatment are significantly under-contributing compared to their share of the threshold, perhaps because they perceive that the explicit historical advantage offered to the earlys is unfair. This is borne out by several post-treatment open-ended responses by lates. For example, the following quote from a late developer in *End-Diff* captures this sentiment perfectly: "I decided not to contribute any because I felt that the individuals who were able to extract more money in the first round [earlys] should contribute more because I started with a disadvantage." This sentiment, of course, has important implications for reaching an effectual international climate agreement in the face of endogenous and differentiated responsibilities.

Therefore, our results also help explain why climate change mitigation efforts continue to have limited success. In the post-Paris world, more needs to be done to address the concerns of the developing nations in negotiating process, and developed countries should take the lead in implementing their ambitious INDCs under the Paris agreement. Recent developments in China and Brazil—with both countries making more or less unilateral commitments to reduce emissions even before Paris—are welcome because they demonstrate that perhaps late-developing countries can indeed make significant contributions despite their status as late-developers. On the other hand, even after the Paris agreement, India's official commitment to emissions reductions is explicitly conditional on "additional means of implementation to be provided by developed country parties." The Indian position makes clear that developing countries still feel strongly about developed country obligations implied by differential responsibilities which arise within the compound climate dilemma. This will undoubtedly be a point of major contention even after the acknowledged success in Paris.

The results we present based on our model of the compound climate dilemma, as well as the model itself, are enlightening in a number of distinct ways. Still, it is important to point out some of the limitations of our study, all of which represent opportunities for further investigation of the compound climate dilemma.

First, though we omit explicit communication (an important feature of the real world) the results in figure 6 are similar to those in Tavoni et al. (2011) who do allow for cheap talk communication. In our study, leadership is demonstrated by signaling cooperation with contributions rather than cheap talk. We see our study as complementary: we included features, like endogenous responsibility, not included in earlier work and, to keep the experiments parsimonious, excluded others. We omitted communication in order to focus on the compound nature of the climate dilemma. Undoubtedly, though, communication would increase cooperation and pro-sociality, as it nearly invariably has been shown to do across a broad range of social dilemmas. There are a number of opportunities for investigating the effects of communication in the context of the interdependence of the appropriation and mitigation phases of the dilemma. For example, it would be interesting to see how communication in the appropriation dilemma affects behavior in the mitigation dilemma. This is the sort of important question that cannot be conceived of, let alone answered, when we focus only on the mitigation dilemma.

Second, our experimental design treats the appropriation phase and the mitigation phase as sequential and discrete dilemmas while in fact they are overlapping. Furthermore, from the outset, the participants have complete information about the precise relationship between appropriation and the parameters in the collective risk dilemma. Therefore, from the outset, subjects know that their appropriation decisions in the first game will determine the parameters for the subsequent climate game. Though somewhat artificial, this information environment resembles what scholars often refer to as the "post-

1990" period, with 1990 being seen as the date at which the anthropogenic nature of climate change was sufficiently evident (Allen, 2003). Though this date is admittedly arbitrary, it does coincide with the first IPCC report.

Third, in our design, the difference between *Ex-* and *End-* conditions is not only that the latter feature endogenously generated parameters in the contribution phase of the game, but also that the participants in the latter have built up a shared history during the appropriation phase which they carry with them into the contribution phase. The shared history, rather than the endogeneity per se, could be driving the differential success between in the exogenous treatments. Given our design, we cannot definitively rule out that explanation, but in any case, such shared history is an important feature of the real-world compound climate dilemma. Perhaps the most plausible way in which a shared history of appropriation might affect behavior in the contribution phase would be through levels of appropriation serving as signals of selfishness: participants may suspect that fellow group members who heavily appropriated the common pool would be likely to be more selfish with their contributions as well. However, we find no evidence that appropriation levels predict success at the group level. Again, further investigation along these lines is needed.

Another feature of the shared history in *End-Diff* is the explicit early/late developer distinction, which is not present in *Ex2*. As with shared history more generally, this is an important feature of the real world climate dilemma—the distinction is even codified in the UNFCC as Annex I and non-Annex I countries—and one we thought important to preserve. While the explicit nature of the distinction in *End-Diff* might confound the behavioral differences between it and *Ex2*, the more important relationships for our results are those between the early and late developers within *End-Diff* and between them and the subjects in *End-Undiff*. These comparison are not affected by this feature of the design.

Although our model captures key features of the climate change dilemma in a novel way, care must be exercised when extrapolating the behavior we observe in experiments to the real-world analogues they are meant to embody (Levitt and List, 2007). First, the behavior of elites may differ from those of university students—the population from which we drew our samples for this study. Though there is some evidence that policy elites and students behave in similar ways (Hafner-Burton et al., 2014), there are are clearly many sources of potential behavioral differences between these two populations. Even so, one need not assume that our experiments are meant to exclusively model the behavior of elite negotiators in order for our results to be informative. A better understanding of citizens' behavior in climate change mitigation is important because successful mitigation must ultimately involve sacrifice on the part of the mass public (Tingley and Tomz, 2013) who must provide a mandate to their negotiators in order for an effective agreement to be reached and, ever-more important in a post-Paris world, implemented. Thus, an understanding of the revealed preferences for equity in the domain of historical responsibility for climate change is an important, if only a first, step. Survey research has shown that individuals across many countries agree on basic principles of fairness such as extrinsic reciprocity and causal responsibility (Bechtel and Scheve, 2013; Tingley and Tomz, 2013).

Still, individual decision-making processes are not the same as those underlying the behavior of groups. Thus, future research should include experiments in which appropriation and contribution decisions reflect group deliberation rather than isolated individual decisions—an approach which would also permit an investigation of per capita emissions rights. Other avenues of investigation might include asymmetric collective risk—to model differential vulnerability to climate change—and asymmetric rates of return to contributions in the collective risk social dilemma—to model differential efficacy. Though our results display consistency in behavior across participants in the United States and China, our design could also be applied to measure behavior in virtually any other country in the world, and in that way it could be determined to what extent the behavioral norms we observe here are, or are not, truly general.

The current design is limited in a couple of other ways. First, the interdependence between causal responsibility and economic capacity/ability to pay is only part of the equity dimension that impacts debates over anthropogenic climate change: intergenerational tradeoffs must also be made (Jacquet, Hagel, Hauert, Marotzke, Röhl and Milinski, 2013). Second, the starkness of our design, which makes the link between wealth and responsibility explicit and deterministic, differs in a fundamental way from climate science which is permeated by uncertainty. Incorporating a probabilistic, rather than deterministic, number of contribution rounds into our framework could capture the fundamental uncertainty about the onset of catastrophic climate change.

In summary, we conceptualize the climate change dilemma as a dual, interdependent social dilemma, and in so doing, our results offer valuable insights into behavioral norms of fairness in the climate change social dilemma. Additionally, insofar as our design captures the CDR principle it offers a simple yet robust framework for future studies to investigate this important tenet of international climate negotiations. Clearly, for the future of the planet, more research into the politics of climate change is needed. Our compound climate game offers a fruitful new avenue of such investigation.

References

Allen, Myles. 2003. "Liability for climate change." *Nature* 421(6926):891–892.

Alley, Richard B, Jochem Marotzke, William Nordhaus, J Overpeck, D Peteet, R Pielke, R Pierrehumbert, P Rhines, T Stocker, L Talley et al. 2002. *Abrupt climate change: inevitable surprises*. National Academy Press.

Baer, Paul. 2002. Equity, greenhouse gas emissions, and global common resources. In *Climate change policy: A survey*, ed. John O. Niles Stephen H. Schneider, Armin Rosencranz. Island Press: Washington, DC pp. 393–408.

Baer, Paul. 2006. *Adaptation: who pays whom*. MIT Press, Cambridge Massachusetts.

Barrett, Scott and Astrid Dannenberg. 2012. "Climate negotiations under scientific uncertainty." *Proceedings of the National Academy of Sciences* 109(43):17372–17376.

Barrett, Scott and Astrid Dannenberg. 2014. "Sensitivity of collective action to uncertainty about climate tipping points." *Nature Climate Change* 4(1):36–39.

Bechtel, Michael M and Kenneth F Scheve. 2013. "Mass support for global climate agreements depends on institutional design." *Proceedings of the National Academy of Sciences* 110(34):13763–13768.

Brick, Kerri and Martine Visser. 2015. "What is fair? An experimental guide to climate negotiations." *European Economic Review* 74:79–95.

Burton-Chellew, Maxwell N, Robert M May and Stuart A West. 2013. "Combined inequality in wealth and risk leads to disaster in the climate change game." *Climatic change* 120(4):815–830.

Bynum, Autumn, Reuben Kline and Oleg Smirnov. 2016. "Passive non-participation versus strategic defection in a collective risk social dilemma." *Journal of Theoretical Politics* 28(1):138–158.

Carlsson, Fredrik, Mitesh Kataria, Alan Krupnick, Elina Lampi, Åsa Löfgren, Ping Qin and Thomas Sterner. 2013. "A fair share: Burden-sharing preferences in the United States and China." *Resource and Energy Economics* 35(1):1–17.

Chakra, Maria Abou and Arne Traulsen. 2014. "Under high stakes and uncertainty the rich should lend the poor a helping hand." *Journal of theoretical biology* 341:123–130.

Cherry, Todd L, Stephan Kroll and Jason F Shogren. 2005. "The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab." *Journal of Economic Behavior & Organization* 57(3):357–365.

Clark, Jeremy. 2002. "House money effects in public good experiments." *Experimental Economics* 5(3):223–231.

Council of Economic Advisers, White House. 2014. "The cost of delaying action to stem climate change.".
URL: *www.whitehouse.gov/sites/default/files/docs/the_cost_of_delaying_action_to_stem_climate_change.pdf*

Dietz, Thomas, Elinor Ostrom and Paul C Stern. 2003. "The struggle to govern the commons." *science* 302(5652):1907–1912.

Dreber, Anna and Martin A Nowak. 2008. "Gambling for global goods." *Proceedings of the National Academy of Sciences* 105(7):2261–2262.

Gohar, LK and KP Shine. 2007. "Equivalent CO2 and its use in understanding the climate effects of increased greenhouse gas concentrations." *Weather* 62(11):307–311.

Hafner-Burton, Emilie M, Brad L LeVeck, David G Victor and James H Fowler. 2014. "Decision maker preferences for international legal cooperation." *International Organization* 68(04):845–876.

Hansen, James, Makiko Sato, Pushker Kharecha, David Beerling, Robert Berner, Valerie Masson-Delmotte, Mark Pagani, Maureen Raymo, Dana L Royer and James C Zachos. 2008. "Target atmospheric CO2: Where should humanity aim?" *arXiv preprint*

arXiv:0804.1126 .
URL: *http://arxiv.org/ftp/arxiv/papers/0804/0804.1126.pdf*

Harrison, Glenn W. 2007. "House money effects in public good experiments: Comment." *Experimental Economics* 10(4):429–437.

Hasson, Reviva, Åsa Löfgren and Martine Visser. 2010. "Climate change in a public goods game: investment decision in mitigation versus adaptation." *Ecological Economics* 70(2):331–338.

Hayward, Tim. 2012. "Climate change and ethics." *Nature Climate Change* 2(12):843–848.

Jacquet, Jennifer, Kristin Hagel, Christoph Hauert, Jochem Marotzke, Torsten Röhl and Manfred Milinski. 2013. "Intra-and intergenerational discounting in the climate game." *Nature climate change* 3(12):1025–1028.

Javeline, Debra. 2014. "The most important topic political scientists are not studying: adapting to climate change." *Perspectives on Politics* 12(02):420–434.

Keohane, Robert O. 2015. "The Global Politics of Climate Change: Challenge for Political Science." *PS: Political Science & Politics* 48(01):19–26.

Kroll, Stephan, Todd L Cherry and Jason F Shogren. 2007. "The impact of endowment heterogeneity and origin on contributions in best-shot public good games." *Experimental Economics* 10(4):411–428.

Lange, Andreas, Andreas Löschel, Carsten Vogt and Andreas Ziegler. 2010. "On the self-interested use of equity in international climate negotiations." *European Economic Review* 54(3):359–375.

Levitt, Steven D and John A List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *The journal of economic perspectives* pp. 153–174.

McDermott, Rose, Jonathan Cowden and Cheryl Koopman. 2002. "Framing, uncertainty, and hostile communications in a crisis experiment." *Political Psychology* 23(1):133–149.

Mechler, Reinhard, Laurens M Bouwer, Joanne Linnerooth-Bayer, Stefan Hochrainer-Stigler, Jeroen CJH Aerts, Swenja Surminski and Keith Williges. 2014. "Managing unnatural disaster risk from climate extremes." *Nature Climate Change* 4(4):235–237.

Milinski, Manfred, Ralf D Sommerfeld, Hans-Jürgen Krambeck, Floyd A Reed and Jochem Marotzke. 2008. "The collective-risk social dilemma and the prevention of simulated dangerous climate change." *Proceedings of the National Academy of Sciences* 105(7):2291–2294.

Milinski, Manfred, Torsten Röhl and Jochem Marotzke. 2011. "Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets." *Climatic Change* 109(3-4):807–814.

Page, Edward A. 2008. "Distributing the burdens of climate change." *Environmental Politics* 17(4):556–575.

Schellnhuber, Hans Joachim. 2006. *Avoiding dangerous climate change*. Cambridge University Press.

Spraggon, John and Robert J Oxoby. 2009. "An experimental investigation of endowment source heterogeneity in two-person public good games." *Economics letters* 104(2):102–105.

Tavoni, Alessandro, Astrid Dannenberg, Giorgos Kallis and Andreas Löschel. 2011. "Inequality, communication, and the avoidance of disastrous climate change in a public goods game." *Proceedings of the National Academy of Sciences* 108(29):11825–11829.

Tingley, Dustin H. 2011. "The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining1." *International Studies Quarterly* 55(2):521–544.

Tingley, Dustin H and Barbara F Walter. 2011. "Can cheap talk deter? An experimental analysis." *Journal of Conflict Resolution* 55(6):996–1020.

Tingley, Dustin H, Stephanie W Wang et al. 2010. "Belief updating in sequential games of two-sided incomplete information: An experimental study of a crisis bargaining model." *Quarterly Journal of Political Science* 5(3):243–255.

Tingley, Dustin and Michael Tomz. 2013. "Conditional cooperation and climate change." *Comparative Political Studies* p. 0010414013509571.

UNCED. 1992. Report of the United Nations Conference on Environment and Development , Rio de Janeiro (A/CONF.151/26/Annex.1). Technical report New York: United Nations.

UNFCCC. 2010. COP16: Report (addendum 1), Cancún 2010 (FCCC/CP/2010/7/Add.1). Technical report Bonn: UN Framework Convention on Climate Change.
URL: *http://unfccc.int/resource/docs/2010/cop16/eng/07a01.pdf*

Vasconcelos, Vítor V, Francisco C Santos, Jorge M Pacheco and Simon A Levin. 2014. "Climate policies under wealth inequality." *Proceedings of the National Academy of Sciences* 111(6):2212–2216.

Wang, Jing, Feng Fu and Long Wang. 2010. "Effects of heterogeneous wealth distribution on public cooperation with collective risk." *Physical Review E* 82(1):016102.

Supporting Information for: Differentiated Responsibilities and Prosocial Behavior in Climate Change Mitigation: Behavioral Evidence from the United States and China

Contents

1 Equilibrium Analysis

For exposition, we focus our theoretical analysis on *End-Undiff*. However adjusting for the different cutoff levels in the probability of loss function, the equilibria in *End-Diff* are analogous to those we describe here for *End-Undiff*. Our collective risk phase is equivalent to a threshold public goods game (threshold game), thus we can look to the literature on such games for guidance. Unlike linear public goods games, threshold games can support Pareto efficient Nash equilibria [1], so long as the threshold value is certain and common knowledge. If the threshold is subject to uncertainty, then this can lead to Pareto-inefficient equilibria [2, 3]. In our game, though the threshold is uncertain before the common pool resource dilemma, it is commonly known and certain before the start of the collective risk social dilemma; therefore, we can treat that part of the game as a threshold game with a certain threshold. In threshold games with certain thresholds, there are two types of equilibria [4, 5]: one which is socially inefficient (zero contributions) and an efficient one (contributions are minimally sufficient to meet the threshold). Since the game is played over 10 rounds with 6 players, the efficient equilibrium may be arrived at through many possible vectors of allocations. From [6] we know that there is a set of efficient equilibria, one of which is an "all fair-sharer equilibrium."

If the probability of loss were constant and independent of appropriation amounts, then maximal appropriation would be the unique and unconditional equilibrium strategy. However, because of our step function mapping appropriation to probabilities of loss, we must also take this probability into account when finding the equilibrium. Given the discontinuous nature of the function, it is certainly irrational to take anything less than the maximal amount in each category of collective risk: 60, 120, 180 and 240. If the group appropriation is in between these amounts, then there must be an incentive to appropriate either more or less. For example, if the group extraction is 59, then there is an incentive to appropriate one unit more, because it has no effect on the probability of loss. Given each of the four equilibrium appropriation amounts, we can determine whether the rational contribution strategy is to contribute nothing and/or to contribute up to the threshold. For appropriations of 60 and 120 and their correspondingly low probabilities of loss, the utility maximizing group contribution is zero, which yields per capita expected utilities of \$8.33 and \$10 respectively, whereas contributing to meet the threshold yields expected utilities of \$4.7 and \$9.4. Conversely, for group appropriations of 180 and 240, the probabilities of loss are high enough that contributions up to the threshold nets a higher expected utility (\$14.1 and \$18.8) than zero contribution (\$7.5 and \$3.32 respectively). Nonetheless, zero contribution—despite its lower expected utility—

at these two higher probabilities of loss is also an equilibrium because no player would benefit from unilaterally deviating from a zero-contribution strategy.

To summarize, we have six classes of equilibria. The step function that maps appropriation to probability of loss increases the number of equilibria by inducing two types of equilibria at each of the higher levels of appropriation. For the two lowest levels of appropriation (which, we should note, never occur in our data), it is only rational for all members of the group to take on the collective risk, as the probabilities of loss are relatively low. For the two highest levels of appropriation it is also rational for all members of the group to take the risk associated with zero contributions, but given the greater collective risk, it is also rational for all members of the group to contribute just up to the threshold. As we can see from the analysis, the introduction of endogenous responsibility changes the nature of the equilibria compared to the standard collective risk social dilemma.

2 Supporting Analyses

2.1 Group by Group Summaries

2.1.1 U.S.

Tables S1 and S2 display the endowments, thresholds, contributions, initial inequalities and expected earnings efficiency for each of the groups in the *End-Diff* and *Ex2* treatments for the sessions conducted in the United States. Efficiency is calculated as the proportion of the maximum aggregate payoffs for each scenario. This maximum is reached through full appropriation combined with contributions sufficient to just cover the threshold (this behavior would also be a Nash Equilibrium—see appendix for a more thorough discussion of the equilibria of our game). These values are 112.8 for the *End-Diff* and *Ex2* treatments and 84.6 for the *End-Undiff* and *Ex1* treatments. As a result of our matched design, the group level endowments, thresholds and inequality (Gini) are identical for each of the matched pairs of groups within the *-Diff* and *-Undiff* conditions. Despite these equivalencies, note that the levels of contributions, the groups' success in meeting the threshold, and their earnings efficiency vary considerably. This suggests that endogenous responsibility has an effect independent from and in addition to the effect of inequality per se.

2.1.2 China

As in Tables S1 and S2 above, tables S3 and S4 display the endowments, thresholds, contributions, initial inequalities and expected earnings efficiency for each of the groups in the *-Diff* and *-Undiff* treatments respectively. The sole difference in the formatting of the two sets of tables is that the exogenous treatments are not included here because they

Table S1: **Endowments, Thresholds, Contributions & Inequality: Symmetric (*-Diff*) Treatments, United States**

Group	Endow.	Gini	Thresh.	Pr(Loss)	Contributions		Met Threshold		Efficiency	
					End-	Ex-	End-	Ex-	End-	Ex-
1	145	0.13	77	0.75	77	77	Yes	Yes	0.60	0.60
2	165	0.16	87	0.75	85	88	No	Yes	0.18	0.68
3	172	0.12	91	0.75	50	92	No	Yes	0.27	0.71
4	175	0.19	93	0.75	96	94	Yes	Yes	0.70	0.72
5	176	0.18	93	0.75	98	95	Yes	Yes	0.69	0.72
6	176	0.08	93	0.75	96	94	Yes	Yes	0.71	0.73
7	193	0.04	102	0.92	100	103	No	Yes	0.07	0.80
8	196	0.10	104	0.92	106	105	Yes	Yes	0.80	0.81
9	203	0.09	108	0.92	109	110	Yes	Yes	0.83	0.82
10	203	0.08	108	0.92	107	108	No	Yes	0.07	0.84
11	225	0.02	119	0.92	126	122	Yes	Yes	0.88	0.91
12	228	0.04	121	0.92	118	121	No	Yes	0.08	0.95
Mean	188	.10	100	0.83	97	101	7/12	12/12	0.49	0.77

were not conducted in China.

There are several differences between tables S1 and S2 on the one hand and tables S3 and S4 on the other. All of these differences are related to the fact that the Chinese subjects appropriated substantially more—as reflected in the group by group endowments listed in the respective tables—of the common pool than subjects in the U.S.: 231 versus 188 (out of 240) in *End-Diff* and 161 versus 129 (out of 180) in *End-Undiff*. As a result, average thresholds and probabilities of loss were higher in China, whereas efficiency, success in meeting the threshold and inequality were all lower. Despite the fact that inequality was quite low in the Chinese sessions, the groups struggled to meet the threshold, with just 5 successful groups in *End-Diff* (compared to 7 in the U.S.) and only 1 successful group in *End-Undiff* (compared to 4 in the U.S.). This provides further evidence that differential endogenous responsibility, rather than inequality per se, is the more important stumbling block to cooperation. Still the number of groups meeting the threshold was *not* significantly lower in China than in the U.S. [n=48; p=0.23, two-sided FE test], nor is there a significant difference between earnings efficiency [n=48; p=0.21, two-sided KW test] or proportional contributions [n=48; p=0.18, two-sided KW test]. However, groups in the Chinese sessions did have significantly lower Gini coefficients than their U.S. counterparts [n=48; p=0.03, two-sided KW test] as well as significantly higher appropriation levels [n=48; p=0.004, two-sided KW test].

Table S2: **Endowments, Thresholds, Contributions & Inequality: Asymmetric (-*Undiff*) Treatments, United States**

Group	Endow.	Gini	Thresh.	Pr(Loss)	Contributions		Met Threshold		Efficiency	
					End-	Ex-	End-	Ex-	End-	Ex-
1	101	0.20	54	0.75	51	54	No	Yes	0.15	0.56
2	105	0.14	56	0.75	47	56	No	Yes	0.17	0.58
3	113	0.14	60	0.75	53	57	No	No	0.18	0.17
4	114	0.17	60	0.75	64	63	Yes	Yes	0.59	0.60
5	119	0.01	63	0.75	64	65	Yes	Yes	0.65	0.64
6	120	0.09	64	0.75	62	59	No	No	0.17	0.18
7	131	0.23	69	0.75	41	72	No	Yes	0.27	0.70
8	132	0.16	70	0.75	67	72	No	Yes	0.19	0.71
9	138	0.19	73	0.92	70	73	No	Yes	0.07	0.77
10	148	0.21	78	0.92	79	78	Yes	Yes	0.82	0.83
11	154	0.18	82	0.92	82	82	Yes	Yes	0.85	0.85
12	170	0.17	90	0.92	87	91	No	Yes	0.08	0.93
Mean	129	0.16	68	0.81	64	69	4/12	10/12	0.35	0.63

2.2 Additional Information on Appropriation Analysis

The results presented below in tables S5 and S6 contain further information on the tobit regressions used to estimate the results presented in the main text in figure 3.

2.3 Success in the Mitigation Dilemma

Number of groups meeting the threshold:

- U.S. *End-Undiff*: 7 of 12
- U.S. *End-Diff*: 4 of 12
- U.S. *Ex1*: 12 of 12
- U.S. *Ex2* 10 of 12

- China *End-Undiff*: 5 of 12
- China *End-Diff*: 1 of 12

However, the use of a dichotomous outcome measure of meeting the threshold masks considerable variation in the *degree* to which the groups were unsuccessful in meeting the

Table S3: **Endowments, Thresholds, Contributions & Inequality: End-Diff Treatment, China**

Group	Endow.	Gini	Thresh.	Pr(Loss)	Contributions	Met Threshold	Efficiency
1	219	0.05	116	0.92	120	Yes	0.88
2	223	0.03	118	0.92	120	Yes	0.91
3	224	0.06	119	0.92	101	No	0.09
4	224	0.03	119	0.92	118	No	0.08
5	225	0.04	119	0.92	119	No	0.08
6	233	0.02	124	0.92	118	No	0.08
7	235	0.01	125	0.92	125	Yes	0.98
8	236	0.01	125	0.92	120	No	0.09
9	236	0.01	125	0.92	128	Yes	0.96
10	238	0.01	126	0.92	53	No	0.14
11	239	0.003	127	0.92	129	Yes	0.98
12	239	0.003	127	0.92	121	No	0.09
Mean	231	0.02	122	0.92	114	5/12	0.44

threshold. Another way of measuring success is to look at the distance of each group from their threshold. A number of groups in *End-Undiff* and *End-Diff* fell short of the threshold by substantial amounts, including by $41 and $73 in *End-Undiff* in the U.S. and China respectively, and $28 and $39 in *End-Diff* respectively. Figure S1 displays the average deviation from the threshold for each of our six treatments. Figure S1 starkly shows that only in the exogenous conditions did the groups tend to consistently meet the threshold.

An alternative measure of success is relative earnings efficiency, as reported in Tables 3 and 4. Using this measure of success in the U.S. sessions, we find that groups in the *End-* conditions are more successful than those in the *Ex-* conditions [n=48; p=0.003, Kruksal-Wallis (KW) test]. This result remains even when we consider only the *-Undiff* treatments [n=24; p=0.022, KW test] or only the *-Diff* treatments [n=24; p=0.037, KW test]. Thus, again, we find that endogenous responsibility reduces success. Conversely, as was the case with our previous measure of success, there is no relationship between inequality and success when measured by earnings efficiency, either overall [n=48; Pearson's $r = -0.066$, p=0.655], in the *Ex-* conditions [n=24; Pearson's $r = -0.108$, p=0.615], or in the *End-* conditions [n=24; Pearson's $r = -0.059$, p=0.786]. In our data, inequality (ability to pay) is not driving outcomes in the collective risk social dilemma, but responsibility and asymmetric development opportunities are.

Table S4: **Endowments, Thresholds, Contributions & Inequality: End-Undiff Treatment, China**

Group	Endow.	Gini	Thresh.	Pr(Loss)	Contributions	Met Threshold	Efficiency
1	116	0.1	62	0.75	60	No	0.17
2	132	0.11	70	0.75	63	No	0.20
3	161	0.23	86	0.92	88	Yes	0.86
4	161	0.16	86	0.92	84	No	0.08
5	164	0.16	87	0.92	71	No	0.09
6	165	0.16	88	0.92	87	No	0.08
7	168	0.17	89	0.92	88	No	0.08
8	171	0.16	91	0.92	89	No	0.08
9	171	0.16	91	0.92	87	No	0.08
10	171	0.15	91	0.92	86	No	0.08
11	175	0.17	93	0.92	54	No	0.12
12	179	0.17	95	0.92	93	No	0.08
Mean	161	0.16	86	0.89	79	1/12	0.17

Table S5: **Tobit regression results: appropriation behavior, United States.**

	(a) Late − Undiff	(b) Early − Late	(c) Early − Undiff
Outcome	$\sum_{t=1}^{10} a_{ij,t}$	$\sum_{t=6}^{10} a_{ij,t}$	$\sum_{t=6}^{10} a_{ij,t}$
Predictor	*End-Diff* late	*End-Diff* early	*End-Diff* early
Coeff. (s.e.)	2.3 (1.6)	−6.0 (1.9)	−6.5 (2.4)
$p - value$	0.136	0.002	0.007
Intercept (s.e.)	18.9 (1.3)	19.2 (1.6)	26.0 (1.9)
$p - value$	0.000	0.000	0.000

2.4 Analysis of Proportional Contributions by Relative Wealth

Table S7 displays mean proportional contributions by treatment and by *wealth*, i.e. above or below the group's median endowment. Those above the median contribute significantly more than those below the median only in *End-Diff*—and this is the case both in the U.S. [n=24; p=0.058, two-sided Kolmogorov-Smirnov (KS) test] and China [n=24; p=0.017, two-sided Kolmogorov-Smirnov (KS) test], with the difference being even larger in China. Strangely, in the U.S., those *below* the median contribute significantly more than those above in *End-Undiff*. No significant differences are found in the other treatments. These results are consistent with our analysis in the text of contributions by earlys and lates. These results are basically equivalent to those in the main text presented in figure 4.

Table S6: **Tobit regression results: appropriation behavior, China.**

	(a) Late − Undiff	(b) Early − Late	(c) Early − Undiff
Outcome	$\sum_{t=1}^{10} a_{ij,t}$	$\sum_{t=6}^{10} a_{ij,t}$	$\sum_{t=6}^{10} a_{ij,t}$
Predictor	*End-Diff* late	*End-Diff* early	*End-Diff* early
Coeff. (s.e.)	0.1 (1.7)	−6.0 (1.7)	−6.4 (2.5)
$p - value$	0.936	0.001	0.013
Intercept (s.e.)	24.4 (1.7)	19.2 (1.6)	35.6.0 (1.9)
$p - value$	0.000	0.000	0.000

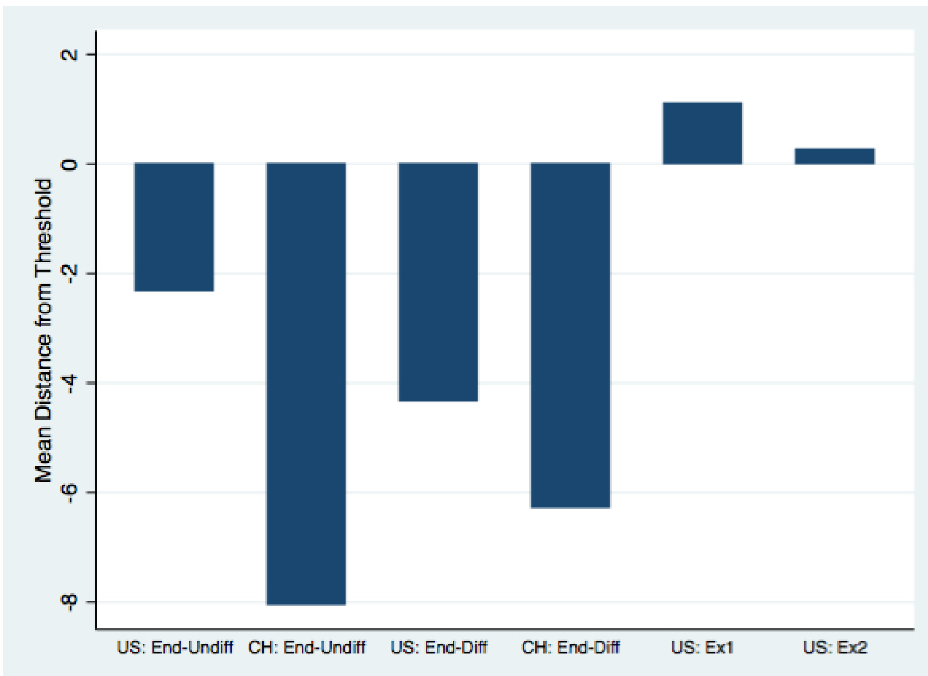


Figure S1: Mean Group Distance from Threshold by Treatment

2.5 Contribution, Appropriation and Inequality Dynamics

A focus on aggregate contributions at the treatment level obscures the dynamics of the appropriation and contribution processes, and their joint effect on inequality at the level of the group. In the main text, Figures 3 and 5 display average appropriation behavior by treatment and period for the U.S. and China respectively. Here, we analyze contribution decisions by period across treatments. Due to the restrictions on appropriation in *End-Diff* the thresholds were systematically lower, and thus the levels of contributions necessary to meet the threshold are proportional to this lower threshold. We therefore normalize the contributions by the size of each group's thershold, contributions normalized in this

Table S7: Proportional Contributions by Treatment, Wealth, and Development Status. The p-values in the rightmost column are derived from two-sided KS tests for equality between two distributions, *a* and *b*: (a) is the distribution for subjects who were above their group's median endowment and (b) is the distribution of those below the group's median.

Treatment	**Status**	**Proportion**	$p(a = b)$
End-Undiff (US)	above median (a)	0.49	0.058
End-Undiff (US)	below median (b)	0.56	
End-Undiff (China)	above median (a)	0.48	0.166
End-Undiff (China)	below median (b)	0.50	
Ex1 (US)	above median (a)	0.52	0.403
Ex1 (US)	below median (b)	0.56	
End-Diff (US)	above median (a)	**0.53**	**0.058**
End-Diff (US)	below median (b)	**0.43**	
End-Diff (China)	above median (a)	**0.54**	**0.017**
End-Diff (China)	below median (b)	**0.40**	
Ex2 (US)	above median (a)	0.54	0.166
Ex2 (US)	below median (b)	0.48	

fashion are plotted in Figure S2. First, focusing only on the U.S. treatments, *proportional* contributions do not differ drastically, but the small differences are consistent with the differential success across the four treatments. Though there are significant differences in the final outcome, the trends for each of the treatments do not look radically different. Contributions in the exogenous treatments (*Ex1* and *Ex2*) started off at a substantially higher level, than the endogenous treatments, and this would appear to be a principal reason for their greater success. Also, we note the drop-off in contributions in the final period, especially in *Ex2* and *Ex1*. This is a well known phenomenon in linear public goods games [7], but the collective risk social dilemma is in effect a *threshold* public goods game. The steep drop-off thus reflects a surfeit rather than a dearth of cooperation: in *Ex1*, two-thirds of the groups met the threshold *before* the final period, with three meeting the threshold in the 8th period. *Ex2* is characterized by similar, but less pronounced, behavior. In contrast, only two groups in *End-Undiff* and one group in *End-Diff* managed to meet the threshold before the 10th period. Nonetheless, aside from the interpretation of the end-game effects, the pattern of contributions is similar to more standard public goods games: the game begins with relatively high levels of contributions, which steadily decline, until the final few periods in which there is a more dramatic decline in contributions. These "end-game" effects are not present in the data from the two treatments conducted in China. In fact, in the *End-Diff* treatment in China, we observe a fairly sharp *increase* in contributions in periods 9 and 10, and more or less constant contributions in *End-Undiff* across the final

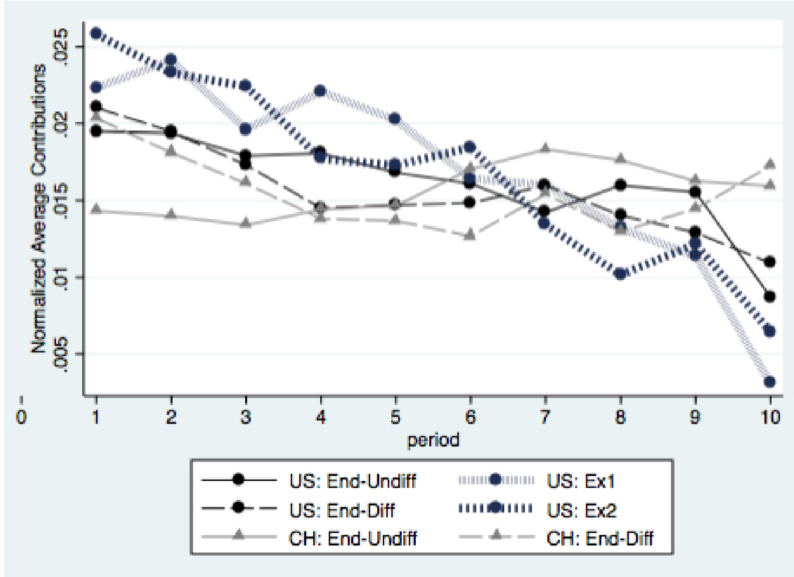


Figure S2: Normalized Contribution by Treatment and Round

few periods. This is likely because no groups met the threshold early in this treatment, and so many were making last-ditch efforts to do so, though in the end only a single group succeeded. In *End-Undiff* we observe more or less constant contributions across the final few periods.

The joint effect of the group's appropriation and contribution decisions is manifested in the group's degree of wealth inequality (in the *Endogenous* treatments; in the *Exogenous* treatments, of course the initial wealth is exogenously determined). We measure inequality using the well-known Gini coefficient, which we calculate for each group's initial endowments entering the contribution phase of the game. We then calculate the round-by-round Gini by calculating the coefficient of the retained endowment at the end of each contribution round. The calculations were carried out using the `egen_inequal` package in Stata. Figure S3 plots this dynamic Gini coefficient. Again, we first focus on the U.S. treatments. By design, *End-Undiff* and *Ex1* commence at the same level of inequality, as do *End-Diff* and *Ex2*. However, while the plots for *End-Undiff* and *Ex1* closely track each other throughout, *End-Diff*'s Gini quickly increases (at a rate similar to that of *End-Undiff* and *Ex1*) while *Ex2*'s Gini increases at a much slower rate. Both increase across the course of the game much less than in *End-Undiff* and *Ex1*, and this is because of the fact that the wealthy individuals in *End-Diff* and *Ex2* gave proportionally more than the wealthy individuals in *End-Undiff* and *Ex1*, and this results in a relatively smaller increase in inequality. In the Chinese sessions, inequality in *End-Diff* begins at a level comparable to the U.S. treatments, but remains relatively flat. *End-Undiff* on the other hand, while the

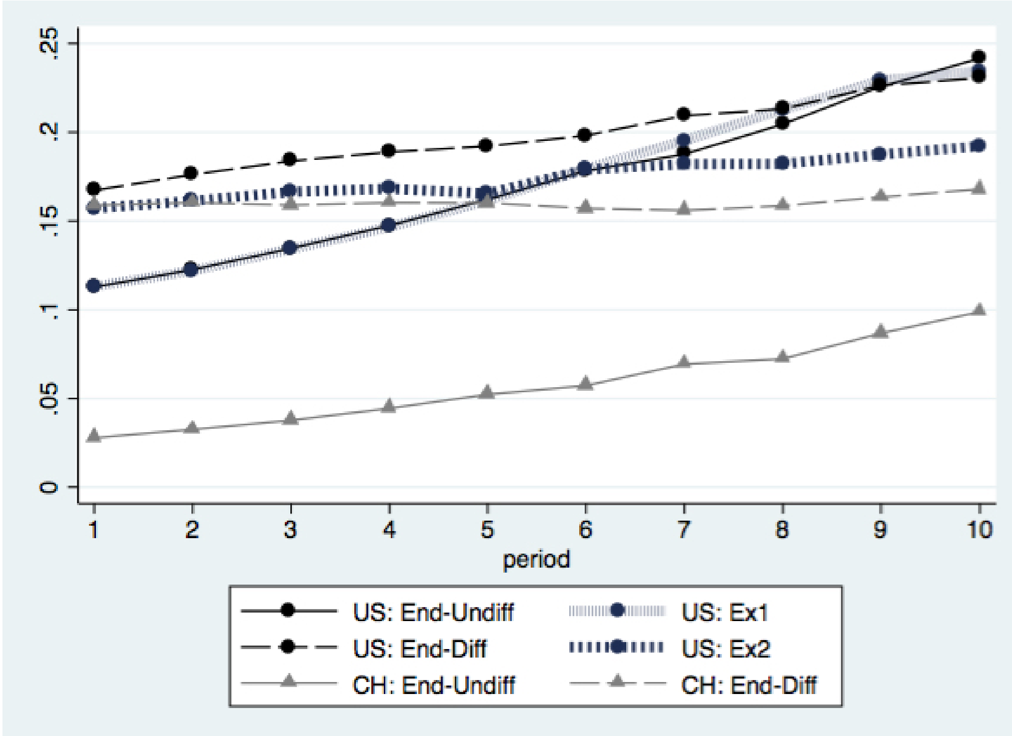


Figure S3: Average Gini Coefficient by Treatment and Round

initial inequality is quite low, it increases quite sharply across the 10 periods.

Exogenous inequality was first introduced into the collective risk social dilemma in [**?**] and [**?**]. In the design of the latter study, the equal treatments are truly equal (initial endowments of 40 for each group member) and as a result have a Gini coefficient of 0. The unequal treatments feature exogenously given endowments of: 28, 28, 28, 40, 40, 40—yielding a Gini coefficient of 0.09. They find that exogenous inequality has a deleterious effect on cooperation in the collective risk social dilemma. Given the endogenous nature of our inequality, we cannot ensure direct comparability with the inequality in [**?**]. However, our design ensures equivalence—in terms of the inequality of endowments—between our *End-* and *Ex-* conditions. Inequality is held constant between each group in a given *End-* condition and the matched group in its corresponding *Ex-* condition, allowing us to isolate the effects of endogenous responsibility and asymmetric opportunity. Overall, the *-Undiff* conditions have a lower ineqaulity (0.10) than the *-Diff* conditions (0.16). Unlike tavoni, we find no effect of inequality per se. Using the group-by-group Gini coefficients in **??**, we find no relationship between inequality and success in meeting the threshold. This result obtains whether we pool all 48 groups together [n=48; Pearson's r = -0.029, p=0.844], we look only at those in the *Ex-* conditions [n=24; Pearson's r =0.074 , p=0.733], or in the *End-* conditions [n=24; Pearson's r =-0.095, p=0.658].

2.6 Self-Reports of Appropriation and Contribution Strategies

At the end of each session the experimental subjects were asked to give open-ended responses regarding their appropriation and contribution strategies. There were of course a wide variety of responses, ranging from the terse or insouciant to the very thoughtful and calculated.

We asked the following open-ended questions:

- How did you decide how much to extract from the Common Pool Account during the Economic Development Game?
- How did you decide how much to extract from the common pool resource during the Economic Development Game?
- How did you decide how much to contribute to the Climate Account during the Climate Game?
- What did you think other players in your group would do?
- Did you expect that your group would make the threshold?
- If you were going to play the game again, would you do anything differently?

When asked about their appropriation strategies, many subjects noted that their strategies involved trying to keep the group below a certain probability of loss level. This indicates that they voluntarily limited their appropriation in order to benefit the group and/or reduce the likelihood that they lose their own retained earnings in the event that the group failed to contribute enough to meet the threshold. This is a potentially promising attitude, and perhaps indicates that if the potential for catastrophic loss is made salient enough to the public, that individuals may indeed be willing to sacrifice for the common good. Here, we enumerate a few responses in this vein:

- "Based on the trend of the common pool appropriation of others, I could see how quickly others were taking out the largest amount (4$) as the rounds continued. At first I wanted to keep small amounts to keep the probability of loss as low as possible, but others did not think of that it seems. Once the probability of loss was quickly 75%, I decided to shrug it off and take out the last amount of money available to stay in the < 180 range."
- "I decided to take $2 each round in order to have some to give back during the second round while not trying to take to much to lessen the probability of loss."

- "I tried to appropriate enough to keep below a certain amount according to what the rest of the group was appropriating. I wanted to aim for 120 max, but then it went over quite quickly, so I tried to aim for below 180, but unfortunately, did not happen. I started high and quickly decreased the amount I appropriated as the game went on. Going from $4 in the beginning to $0, $1, and $2 trying to stay below the 180 line."

- "I was trying to keep the group below the $180 appropriation mark (so if we didn't meet the threshold we would at least have a 25% chance to keep) so I appropriated my share of $30"

In terms of contribution strategies, many of the players elaborated what amounts to a strategy of conditional cooperation in terms of contributions. Perhaps conditional cooperation in the international arena could be the key to solving the climate change mitigation problem. Below are a few select quotes from the participants who were conditional cooperators:

- "I saw no1 [sic] contributing in my group so I thought it would be better to let a dice chance my endowment than lose the money by putting in the climate account when there was a strong chance threshold would not be met."

- "Same method as before, then I noticed Player 7 wasn't contributing anything, so in the last round, I didn't contribute anything and interestingly, neither did other people. Hence we lost by 1 and now we are all going to lose our money, because we are not going to get a 12."

- "Based on each player's contribution after each round and how much each player started with, I tried to see how much they were willing to sacrifice. If they sacrificed a certain fraction of their endowment then I would too. If they sacrificed only a small amount then I would not contribute because I had much less than the other players did."

3 Materials and Methods

3.1 Experimental Details

The experimental sessions were fully computerized and held at a public university in the Northeast United States (between April and November 2013), and a university in Northeast China (in May 2014). All aspects of the procedures and treatments are identical across both sites, except the instructions were translated into Chinese and the appropriation, contribution and payoff schedules were adjusted for the local currency at a rate of 4 yuan per dollar (which is roughly the purchasing power parity rate of 3.5). In our analysis of the Chinese data we normalize these amounts (i.e. divide them by four) so that decisions in both sites are comparable.

The sample consists of students recruited from the general undergraduate population. In the U.S. sessions, participants were paid a $5 show-up fee and additionally compensated according to the payoffs described in the text, with average payoffs of approximately $13, and a range of $5-$45. In China, the show-up payment was 20 yuan ($3.20); average payments were $8.50 with a range of $3.20 to $26. When a group failed to meet the threshold, the experimenter rolled a 12-sided die in full view of all participants to determine whether the group suffered a catastrophic economic loss according to the group's endogenously determined probability of loss outlined in Table **??**. When the die roll indicated such a loss, all group members lost their remaining endowments, and thus received only their show-up fee. Participants were privately paid in cash immediately following the completion of the experimental session. Each session consisted of two groups of six participants each. There were 12 groups in each treatment for a total of 288 participants in the United States, and 144 participants in China. *Endogenous* (*End-Diff* and *End-Undiff*) sessions lasted for about one and a half hours on average, whereas *Exogenous* (*Ex2* and *Ex1*) sessions lasted about an hour.

Subjects were all provided with hard copies of written instructions, and the instructions were read aloud by an experimenter. All of the experimental treatments, when described to the participants, were framed in the context of climate change and its mitigation. Although somewhat atypical for experimental studies of cooperation in general, this approach is consistent with previous studies employing the collective risk social dilemma design, the majority of which have been contextually framed in this way. This is important for the external validity of these fairness norms, which may differ in distinct but strategically equivalent contexts. Along with the instructions, comprehension questions were included, and were overwhelmingly returned with the correct answers. As part of the instructions, participants were briefed on the basic mechanisms behind anthropogenic climate change. In all sessions in both sites the group membership and individual behavior were anonymous, but all players' actions were common knowledge and a summary of contribution and appropriation decisions by each player were given at the conclusion of each period. Players were identified only by a number randomly assigned from 1-6 (Group A) or 7-12 (Group B). Players were informed that group membership was fixed throughout the experiment. The common pool resource dilemma was described to the subjects as the "economic development" phase, and the collective risk social dilemma was referred to as the "contribution" phase. Participants interacted anonymously through the software z-Tree [10] at computer terminals that were physically separated by dividers. The experiments were approved by the appropriate human subjects review boards and written informed consent was obtained from all subjects (in both countries) prior to beginning the experiment.

3.2 Additional Methodological Details

Our baseline treatment—Endogenous-Undiffmetric Opportunities (*End-Undiff*)—is meant to capture this fundamental link between the current level of wealth of nations and their past greenhouse gas emissions and therefore directly capture responsibility for the degree of difficulty in (i.e., the cost of) mitigating the risk of climate change induced catastrophic economic loss, should it occur. As explained in more detail in the *Text*, we do so by combining a common pool resource dilemma with a collective risk social dilemma in which the initial wealth levels, the probability of loss and the threshold are all a function of individual behavior in a common pool resource dilemma. This directly captures the link between wealth and the severity of the climate change problem. The second endogenous treatment we conduct—Endogenous-Diffmetric Opportunities (*End-Diff*)—captures a second dimension of causal responsibility, i.e. the fact that, due to different histories of economic development, some countries have had greater opportunity to appropriate the climate commons than others. The countries with greater historical opportunities, if they avail themselves of them, would then be more causally responsible for any economic losses due to climate change, should they occur. However, such actors may in fact be willing to limit their wealth creation in an effort to limit their impact on the climate while allowing the historically disadvantaged actors to engage in catch-up growth.

In addition to the endogenous nature of the inequalities, our design also endogenizes—as a function of the total group appropriation in the common pool resource dilemma—both the contribution threshold and the probability of loss should contributions not be sufficient to meet the threshold. For both treatments, we chose the threshold to be 53% of the total group appropriation. We chose such an unusual number in order to make it less straightforward to choose their strategies in the collective risk social dilemma as a function of their appropriations in the common pool resource dilemma. We wanted to retain a complete information environment while not having a "fair" strategy be too obvious. Endogenizing the probability of loss was also important, as a greater concentration of greenhouse gases increases the likelihood of catastrophic climate change [8]. However, though this function is monotonically increasing, it is likely not *strictly* monotonically increasing, and is likely more akin to a step function, as there may be tipping points in which the climate abruptly moves from one equilibrium to another [9]. Our function mapping total group appropriation into a probability of loss was designed to capture this reality, albeit roughly.

In *End-Undiff* we calibrated our experimental parameters to be as consistent as possible with [6] and [11]. Nonetheless, it was impossible, given the endogenous nature of the inequalities, to get them to match perfectly with those from [11]. Therefore, in order to isolate the effect of *endogenous* inequality in *End-Undiff* and *End-Diff* we instead needed to construct our own control conditions with *exogenous* inequality that is identical to the inequalities in the two endogenous treatments. We conducted the endogenous treatments first, and then we used the individual appropriation choices—and by extension the group's aggregate choices—to parameterize each group in each of the exogenous treatments by matching it with a group in the endogenous treatments. This keeps the group-by-group inequalities, aggregate wealth, threshold levels and probabilities of loss constant between groups. As a result, the only difference between the groups is the *source* of these inequalities, and their historical genesis.

This is important because causal responsibility is a key stumbling block in international climate negotiations. In [11] there is no link between wealth levels and the gravity of the climate change problem. In the *unequal* treatments, all participants began the game with an equal endowment (40 €). Players were *randomly* chosen to be a "low" or "high" contributor. High contributors were forced to contribute 4 € for the first three rounds, leaving them with 28 € to begin the active phase, whereas low contributors cannot contribute anything through the first three inactive rounds and therefore begin the active phase with their full endowment of 40 €. This is actually a very clever way of going about inducing inequality. Despite the fact that the high contributors have already made contributions in the unequal conditions, the threshold value is unaffected by these as it is identical to the threshold in the equal conditions. There are good reasons for this, in particular comparability. However, the implication of this design choice is that the contributions and the past history have actual effect on the threshold, and therefore there is no link between wealth and the gravity of the climate change problem (the probability of loss) or the costs required to overcome it (the threshold). With our exogenous treatments being parameterized by the wealth levels in the endogenous treatments we are able to preserve comparability while still allowing the probability of loss and the threshold to be a function of group wealth levels.

3.3 Sample Instructions (for *End-Undiff*)

Note: To ensure maximum comparability with previous results from [11], we have included some of the language used in their instructions, especially with respect to the mechanisms behind anthropogenic climate change.

Thank you for agreeing to participate in our experiment today that involves collective action in a climate game.

Before we begin, please take a moment to turn off all cell phones and other devices that could interrupt the experiment. To make this experiment a success, please do not talk to the other participants. If you have a question, please raise your hand and wait for an experimenter to come to you.

Rules and Payment. In this experiment, you can earn money based on your decisions in a collective decision making environment. This money is in addition to your show up fee of $5, which is not affected by any decisions you make within the game.

Please, listen attentively as the rules of the experimental game are read aloud to you. You will be able to refer to this written copy of the rules at any time during the game should you need clarification. However, please, make sure you understand the rules of the game before we begin. Should you have any questions, please raise your hand. At the end of the instructions, you will find several comprehension questions. Please answer these questions to the best of your ability. We will then go over the answers as a group, to ensure that everyone understands the game.

Climate Change. Today, we will introduce you to a game simulating the causes and consequences of climate change. Global climate change is seen as a serious environmental problem faced by mankind. CO2 originates from burning of fossil fuels like coal, oil, or natural gas in industrial processes. Therefore, while industrialization creates wealth, a byproduct of this wealth is the production of greenhouse gases. CO2 is a *global* pollutant (i.e., each quantity unit of CO2 emitted has the same effect on the climate regardless of the location where the emission has occurred).

In the following game, any money you extract from the common pool resource (your endowment generated by your choices in the first phase of the game) represents real money you have earned in industries that emit CO2. Any money you return in the second phase of the game represents a reduction in the effect of CO2 emissions that correspond to an equal reduction in your personal endowment (reduction in real earnings from playing the game today).

Rules of Play. In total, six players, including yourself, are assigned to your group for the duration of the game. Every player faces the same decision-making problem.

You will be making your decisions anonymously. To guarantee this, you will be assigned a player number for the duration of the game. This number is distinct from your payment number that is listed on the green card you were given. There are two groups of six in the room today, but each group is playing independently. If your group consists of players 1-6, you are in group 1; if your group consists of players 7-12 you are in group 2. While the players in your group are anonymous, their decisions are not. You will not know who you are playing with, but you will know the individual decisions of group members you are playing with. You will see the decisions of anonymous players displayed in a table after each round. (See Table S2 [Figure S4])at the end of the instructions for an example.)

During the course of the experiment, you will first play 10 rounds of an "Economic Development Game" followed by 10 rounds of a "Climate Game." However, these games are dependent on one another, as the nature of the "Climate Game" depends on the group's decisions in the "Economic Development Game."

During the "Economic Development" phase of the game, each player will be able to extract money from the common pool resource. For each of ten rounds, you may choose to extract \$0, \$1, \$2, \$3, or \$4. This money represents the money a developing nation earns from industrial production that leads to CO2 output. This money also represents your real potential earnings from participating in today's experiment. (These earnings are in addition to your guaranteed \$5 show-up fee.)

During the "Climate Game" portion of the game (second phase of the game), each member of the group will start with a personal endowment equal to that they extracted from the "Economic Development" portion of the game (earnings from Game 1 carry over to Game 2). Your group will be given a target threshold you must meet to avoid a probable loss of earnings. The amount of this threshold is determined by the total amount your group extracts in the extraction phase (the "Economic Development" game). During the "Climate Game" phase of the experiment, you will choose how much money from your private account you will invest in a "Climate Account" to prevent the possibility of catastrophic climate change—represented by the economic loss of your earnings in today's game.

Your total earnings at the end of the game are determined only by the extraction and contribution amounts you individually make during the two phases of the game. Thus, if the threshold is met, and Player 1 has an endowment of \$30 and Player 2 \$10, Player 1 will earn \$30 and Player 2 will earn \$10. However, the probability that you earn the amount you kept in your private account (your endowment) is determined by the collective extractions and contributions of the group. If your group fails to meet the climate account threshold in the "Climate Game" you face a probable loss of your total earnings based on the amount the group extracted in the "Economic Development Game." We will roll a 12-sided die corresponding to your probability of loss (see Table S[S8]) to determine your earnings.

Table S8: Probability of Loss if Climate Threshold is NOT Met

Total Group Extraction in Economic Development Game	Probability of Loss	Numbers on Die that Lead to LOSS of Entire Endowment	Numbers on Die that Allow Players to KEEP Entire Endowment
\$0-\$60	2/12 (≈17%)	1, 2	3, 4, 5, 6, 7, 8, 9, 10, 11, 12
\$61-\$120	6/12 (50%)	1, 2, 3, 4, 5, 6	7, 8, 9, 10, 11, 12
\$121-\$180	9/12 (75%)	1, 2, 3, 4, 5, 6, 7, 8, 9	10, 11, 12
\$181-\$240	11/12 (≈92%)	1, 2, 3, 4, 5, 6,7, 8, 9, 10, 11	12

More detailed instructions for each round of play are given below. Please, follow along with the instructions carefully, and raise your hand if something is unclear. Please, do not ask questions of others participating in the experiment.

Economic Development Game: Phase I

In the "Economic Development Game," each group of 6 players begins with a common pool of resources equal to \$240. In each of the "Industrialization" rounds of the game, the 6 players will be simultaneously asked "How much do you wish to extract from the common pool?" The possible choices of extraction values are \$0, \$1, \$2, \$3 or \$4 per round. After each player has chosen the extraction amount for the round, they will see a table displaying the sum total amount each player has extracted from the climate account as well as the remaining total that can be extracted from the climate account. While you will know the individual total extractions for each player, you will not know the identity of the players in your group. Figure S5 on the following page is a screenshot of the screen you will see each round during the economic development phase.

Given that a maximum of \$4 can be extracted in each of ten rounds, each player can choose to extract up to a total of \$40 from the common pool. Any amount remaining in the common pool is not allocated to any of the players. The total amount that each player has extracted over the 10 rounds will then become their endowment for the "Climate Game", which is described next. For example, if a player were to choose to extract \$2 each round, then their endowment for the climate game would be \$20. Because players cannot extract any more resources from the climate account in the second phase of the game, this amount represents their maximum possible earnings for the entire course of the game.

Economic Development Game: Phase I

The “Climate Game” will also be played out over exactly 10 rounds. Each player begins the “Climate Game” with an endowment equal to the total amount (over all 10 rounds) they extracted from the common pool in the “Economic Development” portion of the game. In each round of the climate game, you can invest $0, $1, $2, $3, or $4 per round in an attempt to protect the climate and to prevent dangerous climate change. Among other things, dangerous climate change will result in a probable significant economic loss, which will be simulated in this experiment by loss of players personal endowments. The figure below (Figure S6) is a screenshot of the screen you will see each round during the climate phase of the game.

Your group’s threshold in the climate game represents the total amount of earnings your group needs to give back to the climate account to avoid a probable loss. The probability of loss is based on the total extraction made by the group from the economic development phase of the game (see Table S2 [S8] below). This threshold will be the equivalent of 53% of the total extraction of your entire group from the “Economic Development” phase of the game. Thus, if your group extracted $100 of the $240 common pool resource, you would need to collectively return at least $53 (100 x 0.53 = $53) to the climate account over the next 10 periods of the game to avoid a probable loss.

As in the first round of the game, you will play ten rounds to make decisions regarding the climate account. However, in each round of the “Climate Game,” the amount you select will be subtracted from your total endowment. (For example, if you have an endowment of $20 in round 1, and you choose to contribute $2 to the global climate account, you will be left with $18 in round 2.)

The probability of loss (based on group’s “Climate Game” extractions), in the event the threshold is <u>**not**</u> met, can be seen in Table S[S8].

If the threshold *is* met, each group member will keep the entirety of their remaining private endowment (the total amount extracted minus the total amount contributed toward the threshold). **If the threshold value is <u>not</u> reached**, dangerous climate change will occur with a probability that increases as the total extraction of the group increases and is displayed in the **Table S2 [S8]** above (also visible on the white board in the lab). In the event that the threshold is not met, at the end of the game we will roll a die in your presence to determine your earnings. If both groups fail to meet the threshold, the results of a single die role will apply to both groups. If the die indicates a loss, each player will lose all the money left in their account and no one will be paid anything (other than the show-up fee). If the die indicates a win, each player will be paid the entirety of the endowment they earned at the end of the Climate Game (in addition to the show-up fee). The payments will be made anonymously, so no other player will know how much you earned.

Comprehension Questions

Please, answer the following questions so we can make sure you understand the game before beginning. When you have completed the questions, raise your hand and someone will be by to check your answers. Once everyone has completed the comprehension questions, the game can begin.

Question 1: Assume you are in a group in which the total extraction is 200. Now let's assume that your group, after 10 rounds has contributed a total of 100.

What would the threshold for your group be?

1. 53
2. 86
3. 106
4. 200

Based on your answer to question 1, did your group meet the threshold? YES NO

Question 2: Now let's say that you are in a group with a threshold of 150, and your group meets the threshold (meaning that collectively the group contributed at least $150). You have a retained endowment of 20 and player B has a retained endowment of 30.

How much will you be paid (earn)? ____________

1. $0
2. $10
3. $20
4. $30

What would the threshold for your group be? ____________

1. 53
2. 86
3. 106
4. 200

Question 3: Now let's say that in the "Economic Development" phase of the game your group extracted $150. This resulted in a "Climate Game" threshold of $80, but your group only contributed $65. You currently have a personal endowment of $20. (Remember you can refer to the probability table to calculate your probability of loss.)

With what probability will you get to keep all $20 of your remaining endowment?

1. 17%
2. 50%
3. 75%
4. 92%

With what probability will you lose your entire endowment?

1. 17%
2. 50%
3. 75%
4. 92%

For the following questions, please, answer True or False:
Question 4: You can potentially increase your private endowment in both the "Economic Development" and "Climate" phases of the game? **T F**

Question 5: The amount you take from the "common pool resource" in the "Economic Development" phase of the game is your total endowment at the beginning of the "Climate Game." **T F**

Question 6: If the target threshold is met in the "Climate Game," all group members can keep their individual earnings. **T F**

References

[1] Palfrey, Thomas and Howard Rosenthal (1984) Participation and the provision of discrete public goods: a strategic analysis. *Journal of Public Economics* 24(2): 171–193.

[2] Dannenberg, Astrid and Löschel, Andreas and Paolacci, Gabriele and Reif, Christiane and Tavoni, Alessandro (2014) On the Provision of Public Goods with Probabilistic and Ambiguous Thresholds. *Environmental and Resource Economics*, 1-19.

[3] McBride, Michael (2006) Discrete public goods under threshold uncertainty. *Journal of Public Economics* 90(6): 1181-1199.

[4] Croson, Rachel TA and Melanie Beth Marks (2000) Step returns in threshold public goods: A meta-and experimental analysis. *Experimental Economics* 2(3):239–259.

[5] Cadsby, C Bram and Elizabeth Maynes (1998) Gender and free riding in a threshold public goods game: Experimental evidence. *Journal of Economic Behavior & Organization* 34(4):603–620.

[6] Milinski M, Sommerfeld RD, Krambeck HJ, Reed FA, and Marotzke J (2008) The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc Natl Acad Sci USA* 105(7):2291–2294.

[7] Ledyard, John O. (1995) "Is there a problem with public goods provision." *The handbook of experimental economics.*

[8] Hansen J et al. (2008) Target atmospheric co2: Where should humanity aim? *Open Atmosph Sci J* 2:217–231.

[9] Committee on Abrupt Climate Change, *et al.*, *Abrupt Climate Change: Inevitable Surprises* (National Academies Press, 2002).

[10] Fischbacher, Urs. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*. 10(2), 171-178.

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

Figure S4:

TABLE 1: Extraction Table (shown after each round)

Your player number:	6
Player 1 extraction	2
Player 2 extraction	3
Player 3 extraction	1
Player 4 extraction	2
Player 5 extraction	3
Player 6 extraction	3
Total extraction	14

Your player number (if it is 1-6 you're in group 1, if it is 7-12 you're in group 2)

This is the sum total player 5 has taken out of the common pool in the previous rounds. You will not know the identity of these players, only the decisions of the players in your group.

This is a running total of the entire amount your group has extracted in ALL rounds of play up until this point.

Figure S5:

FIGURE 1: Extraction Page from the Economic Development Game

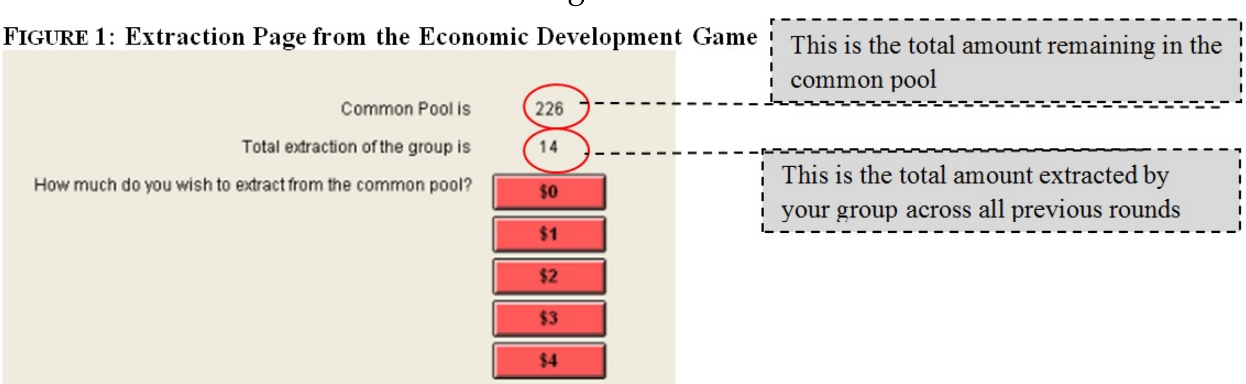


Figure S6:

FIGURE 2: Contribution Page from the Climate Game

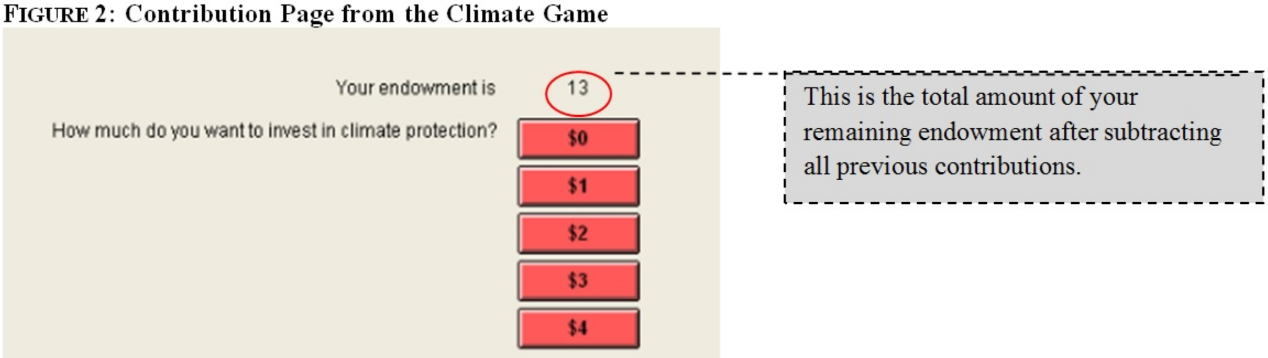


Figure S7:

TABLE 3: Contribution Table (shown after each round)

Your player number:	3
Player 1 contribution	10
Player 2 contribution	22
Player 3 contribution	19
Player 4 contribution	18
Player 5 contribution	9
Player 6 contribution	0
Threshold:	86
Total contribution toward threshold:	78
Probability of loss if threshold not met (percent):	75

Your player number (if it is 1-6 you're in group 1, if it is 7-12 you're in group 2)

This is the sum total player 5 has contributed to the climate account across all previous rounds.

Threshold your group must meet in order to avoid (probable) loss of remaining endowments.

Total amount group has contributed to threshold across previous rounds.

Percentage probability that your group will lose all remaining endowments if the threshold is not reached after 10 rounds.