Punishment, Inequality, and Heterogeneous Risk in Threshold Public Goods Games: Tackling the Climate Change Social Dilemma

Ty Robbins
University of California, Santa Barbara
trobbins@umail.ucsb.edu

October 12, 2016

Abstract

International efforts to stave off adverse climate change have been lengthy in deliberation and largely unsuccessful. While emission reductions are costly for the mitigating individual but benefit the global population, additional factors (varying risk susceptibility to climate change, heterogeneous mitigation efforts, and inequitable pollution shares) have inhibited the ability to reach a binding environmental agreement. I experimentally characterize the climate change social dilemma and evaluate how heterogeneous environmental impacts and unequal endowments affect the propensity to avoid catastrophic climate change. Introducing a punishment mechanism to alleviate the collective bargaining problem, I identify the external factors and intrinsic preferences that impede cooperation. Inequality and delayed contributions negatively affect successful provision, while higher levels of collective-risk increase the probability of threshold attainment. A consensual punishment mechanism incentivizes cooperation in low-risk and heterogeneous groups, overcoming the collective action problem. Social preferences yield guilt that increases contributions while risk aversion negatively impacts threshold attainment in a game with strategic uncertainty.

JEL Classification: C71, C73, C91, C92, D81, H41, Q54
Keywords: threshold public goods, risk aversion, inequality aversion, climate change, peer punishment, cooperation

*I thank Gary Charness, Marco Casari, Alessandro Tavoni, and Philipp Wichardt for their help in the design of my experiment. I’d also like to express my gratitude to the participants of the UCSB Environmental Lunch group, the ESA Meeting 2013 in Santa Cruz, and the 5th Experimental & Behavioral Economics Conference 2014 in Santa Barbara for helpful comments and suggestions. Lastly, I thank the UCSB Economics department for funding.
1 Introduction

A future devoid of dangerous climate change constitutes a global public good with costs incurred by emission-reducing nations and a universal benefit. Disagreements exist as to distributional responsibility in tackling this global dilemma, resulting in an under provision of the public good (Barrett, 2007). Though international meetings extend discussion and boost cooperation, the question persists on whether heterogenous agents can collectively reduce global emissions to avoid the ubiquitous risk of dangerous climate change.

Framed as a “collective-risk social dilemma” (Milinski, Sommerfeld, Krambeck, Reed and Marotzke, 2008), a group of agents must often cooperate to provide a public good where failure to do so may harm all individuals. Many examples can be found in public policy and even in cinematic plots, from building levees for flood prevention to vaccinating all citizens to prevent the zombie apocalypse. All it takes is a weak link in the chain of social cooperation to undermine the vested efforts of the willing and leave the global population susceptible to risk. In the context of climate change, this social dilemma is prominent in the effort to reduce emissions to maintain a habitable climate. Rational agents may desire a future without catastrophic changes, but the magnitude and timing of emission reductions continue to be conditionally voluntary. Reasons for reneging on prior commitments include free riding off the investments of other nations, scientific uncertainty in climate thresholds, and the inequity of developing countries reducing emissions while developed countries benefit from significant historical emissions.

The theoretical and experimental literatures have identified the impact of group size and repeated play in providing public goods, but the climate change dilemma poses a richer environment. In this paper I investigate the influence of varying risk susceptibility to climate change, heterogeneous mitigation efforts, and inequitable pollution shares on contributions to a threshold public good. High risks of financial loss can incentivize prosocial contributions, but when the risk is only as high as the necessary average investment (or lower), groups fail to reach a common target (Milinski et al., 2008). Assuming homogeneous risk susceptibility to economic loss, however, negates the reality that nations will experience heterogeneous impacts from dangerous climate change. Endowment inequality creates an additional barrier to threshold attainment. Highly-endowed agents often fail to signal cooperation early enough to achieve social efficiency (Tavoni, Dannenberg, Kallis and Löschel, 2011). My experiment models the interaction of heterogeneous risk and endowment inequality on group contributions in the climate game. In groups of four, participants contribute to a climate fund over tens rounds to reach a target threshold. Assigned risk factors denote the individual probability of losing net endowments if a group threshold is not met. Each group has two “rich” and two “poor” players controlled by random predetermined play to induce endowment inequality. Without a salient threat of economic loss, heterogeneity in risks and endowments across group members will deter cooperation relative to the homogeneous baseline.

Complementing the external factors that influence behavior in social dilemmas, intrinsic preferences have only recently been modeled in public-choice research (Ostrom, 1998; Teysier, 2012). Agents endowed with different levels of risk and inequality have varying attitudes about their economic standing relative to their peers. While risk and inequity aversion have been incorpo-
rated in sequential games, I examine the impact of social preferences on cooperative behavior in a multi-period public goods setting. Contributions in the threshold public goods game minimize the probability of economic loss and signal prosocial behavior, suggesting the existence of altruism. Individuals who encourage redistributive policies to reach the social optimum also benefit from their own actions, either directly satisfying their desire for fairness or indirectly influencing cooperation among the group. The reciprocation or deflection of social responsibility to contribute toward the public good is thus contingent on both heterogeneous endowments and preferences. A primary focus of this paper is to isolate the external and internal factors that simultaneously impact social cooperation in a collective-risk dilemma. A follow-up questionnaire helps identify the motives of individuals who disproportionately contribute to avoid dangerous climate change.

While a joint high probability of losing net earnings frequently induces collective action, the added dimensions of heterogeneous risk, endowments, and intrinsic preferences complicate public good provision. Communicating a non-binding intended contribution increases the probability of threshold attainment (Tavoni et al., 2011), but cheap talk with layers of heterogeneity may necessitate a credible threat. I introduce a consensual punishment mechanism to incentivize cooperation. At the conclusion of the multi-period threshold game, regardless of threshold attainment, players may choose to collectively punish an individual group member. This instrument inhibits unilateral penalties and emulates regional or global economic sanctions levied against free riders.

In this paper I find that a consensual punishment mechanism is only effective when perceived risk of economic loss is low among homogeneous or heterogeneous agents. Complementing Milinski et al. (2008), coordination issues are overcome if perceived risk is high enough in homogeneous treatments. Nations in reality develop their own climate change susceptibility beliefs based on political agendas and scientific reports. With a wide spectrum of risk beliefs among countries, I find that punishment may be the “great equalizer” that eliminates total contribution differences between homogeneous and heterogeneous groups. Rich players within successful groups significantly reduce, but do not eliminate, the endowment inequality gap. Large early contributions enable rich players to signal cooperation and increase the likelihood of threshold attainment by overcoming trust issues commonly held by poor players. Punishment reduces a coordination problem in heterogeneous groups by holding low-risk players accountable and incentivizing early cooperation.

Incorporating social preferences alongside the previous external factors into a random-effects model, while a player’s relative wealth fluctuates every round, both guilt and envy influence individual contributions. Supported by the follow-up questionnaire, guilt is ten times more influential than envy as players avoid material advantages even when faced with heterogeneous risks. Risk aversion is a proxy for distrust in a game with strategic uncertainty and negatively impacts poor player contributions. Aggregating intrinsic preferences, group composition can further influence cooperation. Comparable to the individual analysis, higher levels of collective guilt boost cooperation while increasing levels of mean risk aversion continue to decrease levels of trust and negatively impact contributions. Although intrinsic preferences influence individual contributions, high risk in homogeneous treatments negates the impact of social preferences as a substantial risk of economic loss is the ultimate free riding deterrent. In heterogeneous groups with a complex coordination problem, envy plays a significant role in decreasing contributions while greater variation in aggregated risk aversion further stunts cooperation.
The remainder of this paper is organized as follows: Section 2 presents a thorough literature review on the varying components of the experiment and my contribution to the literature. After reporting the results of traditional public goods games, I exhibit how threshold mechanisms and collective risk have altered the direction of public goods research. Section 3 introduces the experimental design, a procedure for eliciting and quantifying aversion preferences, and the threshold public goods game that simulates dangerous climate change. I also delineate the intricacies of a punishment mechanism implemented in half of the treatments to incentivize cooperation. Section 4 presents my hypotheses. External factors (inherited risk susceptibility & endowment inequality) and intrinsic preferences (risk and inequality aversion) impact both individual and group behavior. My hypotheses account for how these factors, in conjunction with punishment possibilities, influence cooperation between the different treatments. Section 5 reports trends among the estimation of aversion instruments. Experimental findings are presented in Section 6, both summary statistics in the threshold public goods game and a random-effects regression analysis testing for the significance of external and internal factors across treatments. Section 7 concludes.

2 Literature Review

2.1 Public Goods Games

In a traditional public goods game (PGG), homogenous agents have the option to voluntarily contribute personal endowments to attain a public good. Public good costs and benefits are typically linear in the contribution amount and there is no inherent risk when failing to supply the good, other than potentially losing prior investments and a lack of the public amenity. The threshold PGG varies slightly, incorporating a contribution target that needs to be met or exceeded to supply the public good. Some models incorporate refunds of private investment in the event that the threshold target is not met (making it less risky to contribute), while others utilize a rebate system that returns funds contributed over the necessary threshold (Ledyard, 1995). Within the global game of climate change and emission reductions, no such refund or rebate system exists that can dually compensate over/under efforts to curb green house gases (GHGs), so this paper will also preclude from such possibilities. In lieu of receiving a traditional public good payoff for successful cooperation, when agents fail to achieve the targeted number of avoided emissions (or contributions toward the public good), all agents face a probabilistic risk of losing remaining endowments not invested in the group account. In reality, contributing to the climate public good by reducing emissions may yield continuous benefits before and up to the threshold, including health gains and a reduced risk of dangerous climate change. In an effort to maintain transparency and simplicity for experimental participants, however, I focus on discrete benefits (i.e., avoiding the risk of losing remaining private endowments) that only apply with threshold attainment.1

1Asch, Gigliotti and Polito (1993) examined contribution levels for discrete public goods distributed with a provision point mechanism relative to a continuous public good that returned a constant fraction of group contributions to all subjects for all contribution levels. Though free riding is a dominant strategy in the continuous case and not in the provision of discrete public goods, the authors found that contribution levels were not significantly different.
Theoretical and empirical research for the public goods problem have produced varying equilibria in games with and without a threshold. In a classic PGG without a target threshold, the dominant strategy to maximize individual payoffs is to free ride off the contributions of others, resulting in a socially inefficient outcome. In a survey of experimental research, Ledyard (1995) finds that in repeated linear game settings with small groups, results converge to the zero contribution equilibrium theoretically predicted. Though the noncooperative equilibrium is rational, recent empirical evidence finds positive initial contributions ranging from 40 to 60 percent of period endowments in a rich variety of models. Utility maximization theory is unable to explain the existence of social preferences, which include inequality aversion, selfishness, and other preferences that differ from material interests (Fehr and Schmidt, 1999; Charness and Rabin, 2002). Analyzing strategic decisions and learning in repeated two-stage contribution games, Muller, Sefton, Steinberg and Vesterlund (2008) show that though contributions are initially positive and decrease over time, experience generates smaller declines in contribution levels between stages in repeated games.

Croson and Marks (2000) explore devices to correct the noncooperative equilibrium by introducing a provision point (threshold) mechanism, intended to increase the costs of free riding and induce cooperation. In traditional public goods games individual free riding marginally impacts the level of public good provision. Using a provision point mechanism, given that a threshold contribution target must be met, individual deviations toward the free rider equilibrium could result in a total lack of public good provision. Multiple theoretical equilibria exist in these games: a set of efficient equilibria where the public good is provided when the threshold is exactly met and a set of inefficient equilibria where provision fails to meet the threshold and the public good is not provided. Complementing Isaac, Schmidt and Walker (1989), Croson and Marks find that the inclusion of the threshold mechanism significantly induces higher contributions to the public good. Successful provision, however, varies according to the size of the threshold relative to group wealth, as relatively smaller targets are easier to attain and result in greater provision rates. Bagnoli and McKee (1991) obtain a high rate of threshold attainment (nearly 90%), but had low threshold targets relative to group wealth (about 23%). As the International Panel on Climate Change (Metz, Davidson, Bosch, Dave and Meyer, 2007) has called for a 50% reduction in current GHG levels to reduce the risk of dangerous climate change, I adopt this provision point as the target threshold in my experiment.

2.2 Collective Risk

Failure to reach a target threshold in the climate game results in a probabilistic risk among all agents of losing remaining net endowments. Embedding this risk factor differs from the traditional setting where failure to reach a threshold merely leads to non-provision of the public good. Incorporating a homogeneous risk factor into their experimental setup of avoiding the public bad of dangerous climate change, Milinski et al. (2008) conclude that a strategy to solve the collective-risk social dilemma is to convince agents that failure to reach the target contribution threshold will result in significant individual financial loss. Modeling 10 groups for each treatment of 10, 50, and 90 percent risk probabilities of losing net endowments when a target threshold is not met, the study found 5 of the 10 groups in the 90% treatment successfully collected the target sum, while the other
5 groups marginally failed. Of the 50 and 10 percent treatments, one and zero groups respectively achieved the target, suggesting the severity of potential risks and economic ruin may induce cooperation. Santos and Pacheco (2011) also found that decisions within small groups under high risk scenarios increase the coordination rate. Large-scale cooperation, they conclude, is difficult to achieve and collective action problems may be better solved with a combination of decentralized local agreements focusing on region-specific issues. If coordination cannot be achieved with high risk factors and small groups, experimental results maintain external validity in the context of the world’s nations tackling dangerous climate change.

While incorporating risk is critical to model the provision of climate public goods, nations are unequally susceptible to catastrophic impacts (Metz et al., 2007). Commonly cited consequences of dangerous climate change include the West Antarctic ice sheet collapse and a resulting sea level rise. All nations whose economies are connected through international trade will feel an adverse effect from this event, but citizens in Kentucky do not face equal risks of losing everything compared to those who live in small island states, like Fiji. These latter nations are highly susceptible to extreme loss in the event of catastrophic environmental change. Fisher, Isaac, Schatzberg and Walker (1995) do not vary individual risk factors, but instead vary valuation of a public good among subjects finding that group contributions increase relative to the homogeneous valuation baseline. These results may be driven by the common finding that contributions rise as valuations rise, not proving that valuation heterogeneity exclusively overcomes collective action problems. Fischbacher, Schudy and Teyssier (2012) find that heterogeneity in the return to public goods negatively affects unconditional contributions. Instead of heterogeneous valuations in the traditional public goods setting, games involving collective risk require modeling heterogeneous risk probabilities on threshold attainment. Dividing group members according to high and low risk susceptibility, this paper models heterogeneity in the global climate game and explores solutions to incentivize social cooperation.

2.3 Endowment Inequality in PGGs

Emission reduction agreements are hindered by economic inequality between developed and developing countries. Technology used to decrease emission production is costly and developing nations find it difficult to finance such investments. Developed nations introduced a Green Climate Fund to redistribute aid to poor countries for green investment but significant delays inhibit the program. Distributional responsibility for cutting emissions and sharing costs continues to be a key tipping point in the development of an international climate change treaty.

Relaxing assumptions from earlier work by Warr (1983) that found group contributions to a public good should be invariant after income redistribution, Bergstrom, Blume and Varian (1986) show that income redistribution away from noncooperators may actually increase group contributions to the public good. An experiment by Chan, Mestelman, Moir and Muller (1996) reveals

---

2 International environmental agreements (IEA) are typically addressed in a single group setting (Barrett (1997); Asheim, Froyn, Hovi and Menz (2006)).

3 Unequal risks among affected regions depend on geographic location, ecological conditions, prior preparation for extreme events, and past investments (Ostrom, 2010).

4 Heterogeneous risk probabilities can also be interpreted as asymmetric adaptation capabilities.
that while mean group contributions do increase with redistribution, individual contributions significantly vary between rich and poor players. Inconsistencies between the theoretical model and experiment are possibly driven by notions of fairness. Incorporating social preferences, behavioral models predict that higher income individuals contribute a larger share of their endowments than do low income individuals, though empirical support for this claim is mixed (Fehr and Schmidt, 1999; Charness and Rabin, 2002). Anderson, Mellor and Milyo (2008) find that common knowledge of each individual’s relative endowment reduces contributions for all participants in public goods games. Reuben and Riedl (2011) also find that free riding is frequent and steadily increases over time comparably in heterogeneous and homogeneously endowed groups. Integrating variation in both income and preferences, Chan, Mestelman, Moir and Muller (1999) conclude that inequality in one dimension has a strong positive impact on public good provision while heterogeneity in both dimensions simultaneously yields a smaller but still significant effect. Conflicting evidence of contribution levels in the presence of inequality necessitates further empirical investigation.

Modeling endowment inequality in the climate game with homogeneous risk among group members, Tavoni et al. (2011) augmented Milinski et al. (2008) to evaluate how equity concerns between rich and poor individuals affect group contributions in the attainment of a climate threshold. As the 2009 Copenhagen Climate Change Conference introduced a pledge system for emission reductions, Tavoni et al. induced coordination by including an option to communicate non-binding intended contributions. They found that early signaling by rich agents increased the probability of meeting the target threshold. While communication may alleviate collective action problems, cheap talk might ensue, requiring a stronger mechanism to provide credible threats. Ostrom (2010) notes that persistent communication and updated monitoring, without relying on preexisting levels of trust, are important devices needed to solve the collective action problem. This paper’s experiment takes a similar approach to Chan et al. (1999) by incorporating two types of inequality, endowment and risk heterogeneity, to model the climate game.

2.4 Risk Aversion versus Inequality Aversion

Climate change modeling has sparked debate regarding a simplifying assumption that exploits a single parameter (the elasticity of marginal utility of consumption) to capture (i) risk aversion, (ii) intertemporal substitution preferences, and (iii) spatial inequality aversion (Dasgupta, 2007; Dietz, Hope and Patmore, 2007; Nordhaus, 2007). My experiment distinguishes risk and inequality aversion to model behavior in cooperative games. Traditional efforts to measure risk aversion present experimental subjects with a set of pairwise choices between lottery distributions containing the same mean but different variances, associating preferences based on the individual’s decisions. Kroll and Davidovitz (2003) argue that when the less unequal state is preferred, the subject could be considered inequality and risk averse, rather than exclusively one or the other. Carlsson, Daruvala and Johansson-Stenman (2005) investigate the determinants of individual risk (holding inequality constant) and inequality aversion (holding risk constant). They find that inequality averse subjects are also more risk averse (and vice versa), and that both factors vary significantly with sex, field/major, and political preference. Building off of Kroll and Davidovitz (2003), Magdalou, Dubois and Nguyen-Van (2009) do not find a significant correlation between the two aversion parameters. Teyssier (2012) finds that risk aversion in a sequential public goods game is negatively correlated with contribution levels of leading movers, while advantageous inequity averse
second movers tend to free ride less and cooperate more than others. While contributions may be simultaneously influenced by risk and inequality aversion profiles, a collective-risk threshold PGG introduces a separate psychological risk factor of non-attainment. This paper elicits risk and inequality aversion, distinguishing the influence of intrinsic preferences on social cooperation.

2.5 Punishment Mechanism

Experimental studies have found that costly options to punish free riders greatly incentivize sustained contributions toward public good provision (Ostrom, Walker and Gardner, 1992; Fehr and Gächter, 2000). Agents who contradict socially acceptable behavior are retaliated against even if tangible benefits from costly punishment is negligible. While the existence of peer-to-peer sanctions can induce near efficient cooperation, varying the effectiveness of punishment (the factor by which punishment reduces a punished player’s income) below relative income thresholds may not be able to prevent the cooperation failure (Nikiforakis and Normann, 2008). Complicating matters is the disparate use of linear versus non-linear punishment that can incorporate variable fine-to-fee sanction schedules (Casari, 2005). The fine-to-fee ratio reflects the punished player’s reduced income relative to the punisher’s fee to punish. When punishment reduces the punished player’s income by a certain percentage in non-linear studies, punishment effectiveness becomes convex in the target’s income, making comparisons between treatments difficult.

Higher levels of punishment effectiveness have also been found to increase the propensity to punish (Anderson and Putterman, 2006; Carpenter, 2007). Subjects in these studies who wish to impose punishment on a pair of individuals may opt to only punish the player whose fine-to-fee ratio is greatest, all else equal, maximizing sanctions given costly punishment. Examining this player’s decision strategy would fail to reveal the demand for punishing both agents, permitting only partial preference identification. Studies have also identified a propensity for contributors to punish defectors even when the fine-to-fee ratio is one (Falk, Fehr and Fischbacher, 2005; Sefton, Shupp, Walker, Dawes, Glomm, Gächter, Ockenfels and Putterman, 2005; Nikiforakis and Normann, 2008). In this scenario punishment does not reduce income differences between individuals and may instead indicate the desire to sanction particular actions. Casari (2005) exhibited that the fine-to-fee punishment ratio must be constant for all agents to credibly identify the factors that induce punishment decisions. Successful collective action also depends partly on the types of individuals that comprise the group, shown in Ones and Putterman (2007), where homogeneous and heterogeneous group formation by punishment proclivity helped predict the differences in contributions to a public good.

Bochet, Page and Putterman (2006) paired punishment with communication, concluding that the paired mechanisms do not significantly increase homogeneous group contribution levels relative to communication alone. They also detected the existence of perverse punishment, that is, punishment being directed at individuals whose contributions were higher than average. Punishment modeled with endowment heterogeneity generally induces stable cooperation in both homogeneous and heterogeneous groups (Visser and Burns, 2006; Reuben and Riedl, 2011). Prediger (2011) finds that heterogeneous groups punish less often and at smaller magnitudes, yielding higher group contributions. Casari and Luini (2009) used a consensual rule mechanism to influence cooperation where punishment was only carried out if a coalition of two or more agents chose to
sanction an individual. Consensual punishment induced higher cooperation and lower rates of punishment than autonomous punishment. The authors contend that the lower threat of punishment under the consensual treatment provides stronger incentives for cooperation as this mechanism censors out perverse punishment, effectively blocking more than 70 percent of attempts to punish strong cooperators while only 10 percent of requests to punish free riders was blocked. Gächter and Herrmann (2006) find that autonomous punishment without constraint may cripple cooperation and circumvent the gains from punishment that intend to credibly threaten free riders. Improved institutions beyond autonomous peer-to-peer punishment seem to exist and warrant closer examination.

In the realm of threshold PGGs, particularly with collective risk, there appears to be a dearth of research that incorporates a punishment mechanism to generate collective action. This paper contributes to the literature by enriching the experimental environment and identifying the elicited individual preferences and external factors that dually inhibit and induce cooperation.

3 Experimental Design

All sessions of the study were carried out in the Experimental and Behavioral Economics Laboratory at UC Santa Barbara with 216 subjects recruited from an ORSEE research pool, programmed and conducted with the experiment software z-Tree (Fischbacher, 2007). At this undergraduate level, subjects have little training in expected utility and public goods games. Fifteen experimental sessions were run, involving between 8 and 16 participants per session (depending on show-ups) who earned an average of $12.20 for about 50 minutes of their time.

In each treatment subjects first played three independent games (described below) followed by a threshold public goods game for 10 rounds. In this last game participants were randomly assigned into four-person groups that were held constant for the duration of play. Instructions appeared on a subject’s computer monitor before each independent game, followed by a short set of control questions to check understanding. Any questions of misunderstanding were answered privately. The experimental design avoided giving subjects feedback related to their own and group earnings in the sub-games before the threshold PGG, restricting any incentive to alter behavior in future games based on prior performance. Subjects had common information of the repeated play format with constant group members in the threshold PGG. At the conclusion of this fourth game, a questionnaire was given to elicit beliefs and socioeconomic data. Experimental Tokens (ETs) were used in all games, with a conversion rate boldly stated in each game’s instructions. Show up fees were $5 with the chance to increase payoffs depending on individual and group play in the subsequent games. To incentivize maximum effort, subjects were informed at the beginning of the session that final payoffs would be contingent upon the outcome from one randomly selected game of the games played. Final individual payoffs were distributed privately at the end of the session.

---

5 Constant groups in the threshold PGG allows for reputation effects. International environmental agreements typically hold constant the group of nations making collective decisions as there is rarely a change in the number of countries taking part in negotiations.

6 Participants were advised that one of their peers would roll a die at the conclusion of the experiment to determine
3.1 Risk Elicitation

In the three independent games played before the threshold PGG, I quantitatively measure risk and inequality aversion. Assuming that a subject’s preferences can be approximated by a particular expected utility model, risk aversion is traditionally measured with a lottery game (Holt and Laury, 2002). Regression analysis and qualitative interpretation are improved with point-estimates of aversion parameters whereas Holt's method only supplies an interval estimate of risk aversion. To avoid the analytical problems of interval estimates, I instead employ the method developed in Charness and Genicot (2009) to measure risk aversion. Individuals are endowed with 100 tokens and can invest any portion of this amount in a risky asset that yields a payoff of 2.5 times the amount invested if successful (50% probability), always retaining net endowments not invested. The more an individual invests in a risky asset the less risk-averse she is relative to her peers. Assuming constant relative risk aversion and eliciting the investment decision, I then calculate each individual’s point-estimate risk aversion parameter. Robustly measuring risk aversion, Holt’s method was utilized during the followup questionnaire to compare against the point-estimates elicited. Subjects sequentially chose between a set of pairwise lotteries, where a given lottery differs from its counterpart in the spread of potential payments and the lottery mean (Beckman, Formby, Smith and Zheng, 2004).

3.2 Inequality Aversion

I adopt the approach from Blanco, Engelmann and Normann (2011) and Dannenberg, Riechmann, Sturm and Vogt (2007) to measure inequality aversion, using the strategy method to record preferences in a modified dictator game (MDG) and an ultimatum game (UG) described below. Utilizing both games allows for the identification of advantageous and disadvantageous inequality aversion. Differentiating the magnitude of relative inequality and risk aversion will be useful in order to identify the main drivers that both inhibit and support collaborative efforts in the supply of public goods (or avoidance of public bads). Three independent games (Risk, MDG, UG) are presented to subjects prior to introducing the more complicated threshold PGG. In this fashion there is a natural progression of complexity in the sub-games which may help filter out errors in understanding and decision making. I employ the model developed in Fehr and Schmidt (1999) to parameterize aversion preferences in the two-player UG and MDG games. An advantage of this model is that aversion preferences elicited can be applied to the threshold PGG, assuming that aversion pref-
erences in one game are a simple monotonic transformation of the thresholds in a separate game (Teyssier, 2012). Fehr and Schmidt’s utility function is given by:

\[ U_i(x_i, x_j) = x_i - \alpha_i \max[x_j - x_i, 0] - \beta_i \max[x_i - x_j, 0] \quad \text{for } i \neq j, \]

(1)

where \( \alpha \) measures disadvantageous inequality aversion and \( \beta \) measures advantageous inequality aversion. Assumptions in this model include \( \alpha_i \geq \beta_i \) and \( 0 \leq \beta_i < 1 \). The first of these conditions imparts the assumption that an individual suffers more disutility from being at a material disadvantage than at an advantage relative to her counterpart. The latter condition rules out the existence of individuals who take pleasure in being better off than others (\( \beta_i < 0 \)), while \( \beta_i < 1 \) departs from the implausible event that an individual would give up a dollar or more to reduce their advantage relative to player \( j \).

Deriving disadvantageous inequality aversion, \( \alpha \), participants are introduced to a two-stage ultimatum game whose focal point is the division of a pie worth 20 tokens between two individuals, a proposer and a responder. In the first stage the proposer offers an integer share of the pie, \( s \), of which the responder accepts or rejects in the second stage. The outcome of this offer is \( 20 - s \) for the proposer and \( s \) for the responder, if the proposal is accepted, and zero for each participant otherwise. Unaware of their possible role assignment later in the experiment, all participants made choices in each of the two roles: (i) proposers chose an integer share \( s \in \{0, 20\} \) to offer the responder and (ii) responders chose to accept or reject each of the twenty-one possible “Proposer-Responder” distributions of tokens (20-0, 19-1, ..., 0-20). If this game is chosen for final payoffs at the end of the experiment, participants are randomly paired and randomly assigned one of the two roles, at which point the actual proposed offer and responder’s decision to this offer are compared to determine payoffs. Applying the strategy method to capture the responder’s contingent decision set, the minimum accepted offer yields information to calculate near point-estimates of \( \alpha \). Following Blanco et al. (2011), suppose that \( s'_i \) is the lowest offer that individual \( i \) hypothetically chooses to accept, thus \( (s'_i - 1) \) represents the highest offer they would reject. Assuming well-behaved preferences such that there exists a single point where the individual switches from rejecting a set of offers to accepting the rest, a responder will be indifferent between accepting an offer \( s_i \in [s'_i - 1, s'_i] \) and receiving zero payoff from a rejection. Further assuming the proposer offers no more than half the pie, Fehr and Schmidt’s utility function (1) yields: \( U_i(s_i, 20 - s_i) = s_i - \alpha_i(20 - s_i - s_i) = 0 = U_i(0, 0) \). Solving for \( \alpha_i \), the estimate for disadvantageous inequality aversion is

\[ \alpha_i = \frac{s_i}{20 - 2s_i}. \]

(2)

For the purposes of estimation, I set \( s_i = s'_i - \gamma \), where \( \gamma = 0.5 \). Individuals who repeatedly re-

---

12Section 5 reports estimates for aversion parameters and tests the relevancy of Fehr and Schmidt’s assumptions. Empirical evidence in the literature suggests a positive correlation between advantageous and disadvantageous inequity aversion, which will also be explored with the data set.

13See Game 3 in the Appendix for a screen shot.

14Of the 216 participants, only 8 offered more than half of the pie. The estimation of \( \alpha \) is not affected by these offers but instead impacts the point at which a switch from rejection to acceptance occurs.

15To estimate \( \alpha \), the choice of a parameter adjustment factor (\( \gamma \)) equal to 0.5 is indeed arbitrary to identify an offer,
jected low proposed offers were characterized as increasingly disadvantageous inequality adverse and consequently had higher values of $\alpha$. Exploring the implications of equation (2), the offer must be $s \in [0, 10]$ as I assume the proposer offers no more than half of the pie. Since rational responders accept an equal share offer then $s_i' \leq 10$ and division by zero does not occur. Extreme values of $\alpha_i$ materialize if the responder never switches from their initial decision in the first set of pairwise choices. Individuals who reject every feasible offer less than half of the pie accept only if $s_i \geq 10$, allowing us to infer at most that $\alpha_i \geq 4.5$. I cautiously assign these individuals $\alpha_i = 4.5$ with no further information to make a better estimation of their preferences. On the other side of the spectrum, I never observe a switching point for individuals who accept every offer ($s_i' = 0$) and assign them $\alpha_i = 0$. This characterization assumes the nonexistence of subjects who derive utility from being at a disadvantage relative to others.

Eliciting advantageous inequality aversion, $\beta$, the modified dictator game posits an initial endowment of 20 tokens for the dictator who decides how much of this total she is at most willing to sacrifice for an equitable distribution of payoffs between herself and the recipient. In this setting, participants make choices in the lone role of the dictator. A list of twenty-one pairwise payoff decisions are listed and the participant chose their preferred payoff distribution in each case. The left choice is always a distribution of (20,0) for the Dictator-Recipient payoff and the right choice contained equal payoffs ranging from (0,0) to (20,20). If this game is chosen for final payoffs at the end of the experiment then: (i) participants are randomly paired and randomly assigned one of the two roles, (ii) one of the twenty-one pairwise payoff vectors is randomly chosen, and (iii) the decision of the dictator determines payoffs. Again employing the strategy method to measure aversion preferences, I identify the point at which the dictator switches from the (20,0) unequal distribution to the equitable payoff distribution. Per Blanco et al. (2011), if an individual switches from the unequal payoff vector of (20,0) to the egalitarian outcome at $(x_i', x_i')$, then I can infer that they prefer the payoff (20,0) over $(x_i' - 1, x_i' - 1)$. Together these two reference points relate the individual’s threshold for sacrificing a (20,0) outcome in favor of an equitable one. Since payoffs are based on integer values, there must exist an egalitarian payoff vector, $(x_i, x_i)$, that renders the individual indifferent between this outcome and (20,0). Using Fehr and Schmidt’s utility function (1) again, it must be that $U_i(x_i, x_i) = x_i = 20 - 20\beta_i = U_i(20, 0)$ for some $x_i \in [x_i' - 1, x_i']$ and $x_i' \in \{1, 20\}$. Solving for $\beta_i$, the estimate for advantageous inequality aversion is

$$\beta_i = 1 - \frac{x_i}{20}. \tag{3}$$

$s_i$, that is between the individual’s maximum offer rejected and minimum offer accepted. If I assume that $\gamma$ is normally distributed between the continuum of values in [0,1], then I can justify the choice for $\gamma$. A sensitivity analysis inserting $\gamma = 0.1, 0.3, 0.7, 0.9$ for robustness yielded no substantial qualitative differences in the regression analysis.

16See Game 2 in the Appendix for a screen shot.

17Similar to the UG game, this characterization assumes individuals have well-behaved preferences such that they have a unique switching point from the payoff (20,0) to the egalitarian payoff. Experimentally I found that a number of participants routinely switched between these payoffs more than once. These participants may not have well-behaved preferences, possibly suffered from fatigue, or did not completely understand the logic of the game. For these participants, I calculated three values for their switching point (a minimum, average, and maximum) based on their decisions. The estimation of these individuals’ aversion parameters is imperfect and I take care to explore possible implications later in the paper by analyzing subject behavior with (i) the full pool of participants and (ii) restricting analysis to those individuals with well-behaved preferences.
Similar to the justification for $\alpha$, I set $x_i = x_i' - \gamma$, where $\gamma = 0.5$. In accordance with Fehr and Schmidt I assume $\beta \in [0, 1]$, however, the two endpoints warrant discussion. Individuals who choose the equitable option for each of the 21 decisions (i.e. forgo a payoff of (20,0) for all choices) have a strict aversion to advantageous inequality and $U_i(0, 0) > U_i(20, 0)$, which implies that $\beta_i > 1$. No switching point is ever observed for these participants and it is possible that they are willing to sacrifice in excess of $1 to reduce inequality by $1. As in Blanco et al. (2011), I cautiously assign these participants $\beta_i = 1$. Other subjects for whom a switching point is unobserved include those that never deviate from the (20,0) choice. At the extreme this suggests $U_i(20, 0) > U_i(20, 20)$ and $\beta_i < 0$, thus they may be willing to sacrifice funds to increase inequality. These individuals are assigned $\beta_i = 0$ since I do not observe a switching point and cannot further divulge their unique preferences.

### 3.3 Threshold Public Goods Game

The threshold public goods game (TPGG) modifies the experimental setup developed by Milinski et al. (2008) and amended by Tavoni et al. (2011), randomly dividing participants into groups of four (constant for the game) whose aim is threshold attainment after 10 rounds of play. Conducted in an environment of complete information, subjects within each group were attributed a commonly known unique endowment and risk factor bundle that defined their initial standing in the TPGG. Though a player’s endowment and risk profile were known to all group members, players did not know each other’s identities and instead were assigned a Player ID Number (P1-P4) maintained for the duration of the game. Player ID Numbers and associated endowment/risk profiles were displayed constantly on the computer terminal throughout the game. Constant IDs allowed group members to identify each other exclusively by endowment, risk, and contribution profile, enabling the formation of reputations. Each player was endowed with 40 experimental tokens (ETs) at the start of the game.

Modeling endowment inequality within all treatments, I adopt the approach of Tavoni et al. (2011) and subject all group members to three inactive contribution rounds that force half of the subjects (2) to contribute 4 ETs per round to the collective fund while the other half (2) are forced to contribute nothing. “Rich” players are characterized by starting round 4 with 40 ETs in their individual account whereas “poor” players begin with 28 ETs. The target contribution threshold was set at 80 ETs and for the sake of comparability all treatments started the active phase (rounds 4-10) with 24 ETs in the collective fund. A threshold of 80 ETs constitutes 50 percent of the aggregate group endowment of 160 ETs, reflecting the Intergovernmental Panel on Climate Change’s suggestion of achieving a 50% reduction in emissions relative to prior levels.\(^{18}\) There are no framing effects in this experiment as I distinctly avoid using any verbiage related to the climate or climate change. Previous studies relate the group fund as a “climate account,” potentially biasing individual behavior.

In four of the six treatments, homogeneous risk factors were distributed to all group members that imposed a collective $\frac{1}{3}$ or $\frac{2}{3}$ risk of losing private net endowments when failing to reach the target threshold by the game’s conclusion. All other treatments were assigned heterogeneous risk

\(^{18}\)Contributions to the climate fund and reductions in emissions are synonymous in this experiment.
Table 1: Threshold PGG Treatment Breakdown

<table>
<thead>
<tr>
<th>Treatments</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endowment Inequality*</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Punishment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homogeneous low ((\frac{1}{3})) risk</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homogeneous high ((\frac{2}{3})) risk</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneous risk**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Two players from each group are “poorly” endowed; two players are “richly” endowed.

**Two players from each group have high-risk factors (\(\frac{2}{3}\)); two players have low-risk factors (\(\frac{1}{3}\)).

See Section 3.3 for more details.

Factors amongst group members, whereby two players inherited a \(\frac{1}{3}\) (and the other two a \(\frac{2}{3}\)) risk susceptibility of losing private net endowments when the threshold was not met. Varying risk factors characterize the different levels of dangerous climate change that individuals and nations may be vulnerable to, depending on geographical location, adaptive capacity, etc. Homogeneous risk groups maintain a single dimension of inequality in endowments, whereas heterogeneous risk groups have an equal distribution of rich/poor and low/high risk types. In each active period of the TPGG (rounds 4-10), subjects were simultaneously asked for a 0, 2, or 4 ET investment from their private account to the group fund. At the conclusion of each round, individual contributions to the group account were revealed by Player ID with risk/endowment type, as were total past contributions by individual, aggregate group contributions in the current round, and aggregate group contributions for all rounds up to the present.

End game payoffs were calculated according to whether the collective 80 ET threshold was met or exceeded after 10 rounds, including punishment results if relevant. In the no-punishment treatment when the threshold was met after 10 rounds of contributions, a subject retained any private endowments not invested in the public good. If the threshold was not met, remaining private

---

19The chosen distribution of risk/endowment types being equal within heterogenous groups (one individual of each risk/endowment profile per group) may have an impact on the results of the TPGG. This experiment attempts to model inherent heterogeneity in the extreme case of one profile type per group, but it could be argued that experimental results may be conditional on group type distributions. A simple uniform distribution of types within each group was chosen to model the heterogeneity in wealth and risk susceptibility, exhibited by the multitude of member types present during international climate agreements.

20Isaac et al. (1989) and Croson and Marks (2000) suggest avoiding the word “contribution” (framing effect) and instead phrase as “allotments” or “allocations” to the public fund.

21This subset of possible contributions from total endowments reflects the gradual process in cutting back emissions as opposed to a discontinuous bevy of emission reductions in a single round which is difficult to accomplish with known technology constraints. Further, integer contributions are imposed to help identify altruists, fair sharers, and free riders (Milinski et al., 2008).

22Tavoni et al. (2011) abstract away from revealing the aggregate group contribution for all rounds up to the present and instead allow players to calculate the total amount on their own. In a game with strategic uncertainty and complicated behavioral interactions, I remove the possibility of individual calculation errors to ensure that decision strategies are based on complete and accurate information, rather than potential mistakes in arithmetic.
endowments were at risk of being lost with respect to the relative risk factor assigned in their treatment.

Incorporating a mechanism to induce cooperation, most experiments introduce an autonomous form of punishment. Within this experiment, autonomous punishment might be integrated where participants at the conclusion of the 10 round game simultaneously have a decision whether to buy punishment points to decrease the payoffs of fellow group members who deviate from their personal norm of expected contributions. Following recent studies (Fehr and Gächter, 2000; Casari and Luini, 2009), at a private cost of one ET per punishment point purchased, an individual can decrease the earnings of another group member by a constant three ETs. If an individual received multiple punishment points across group members, their reduction in earnings would reflect the cumulative amount of punishment points received. In the event that the target threshold is at a minimum achieved, each player’s individual payoff would be:

$$\pi^A_i = \max\{e_i - \sum_{t=4}^{10} c_{it} - \sum_{j \neq i} p^i_j - \sum_{j \neq i} p^j_i, 0\}$$

(4)

where:

\(e_i\equiv\) net endowment after 3rd inactive round for person \(i\) for \(i = 1, \ldots, 4\)

\(c_{it}\equiv\) contribution by person \(i\) in round \(t\)

\(p^j_i\equiv\) amount of punishment points \(i\) buys to hurt \(j\) \(p^j_i \in 0, 1, \ldots, 7\).

If the threshold target is not met, then subject payoffs (remaining private endowment net punishment) are in jeopardy of being depleted according to the risk factors assigned in the treatment.

Instead of autonomous punishment, I introduce a consensual punishment mechanism (Casari and Luini, 2009), where simultaneous punishment requests are only carried out when two or more

---

23Contrary to other punishment studies that allow sanctions after each round, I implement end-game punishment. In the complex climate change game, retributive actions will typically be taken in the event that a collective threshold is not met. It would not be practical to punish noncooperators after each round since in the real world the level of emission reductions undertaken is an imperfect measure that only can be quantified after a certain number of lagged periods, justifying end-of-game punishment.

24See discussion in Section 2 for the impact of different fine-to-fee effective punishment ratios.

25This setup assumes that potential punishment will happen before each individual’s risk die is rolled, in the event that the target threshold is not met. This is the preferred approach (as opposed to imposing punishment after an individual’s risk die is rolled), given the possibility that an agent wishing to punish another may find themselves unable to do so if the targeted agent has lost the game of chance (unfavorable risk outcome) and all remaining private endowments. In this scenario, because the punisher cannot punish the targeted individual, the outcome of the game would not capture one’s desire and the magnitude by which they punish another, excluding this important behavioral choice from the results. Additionally, if one agent wanted to punish another agent and can only do so after a die has been rolled (with a favorable result for the risk taker), punishment levels may be more vindictive than the initial desired punishment allocation for game behavior, given that the proposed punishable agent not only escaped positive contributions during the TPGG but also avoided their inherent risk of losing remaining private funds. One can easily imagine that the punisher may be left fuming at the result of the game and the punishee’s favorable die roll, willing to make even more costly punishment decisions that is not exclusively based on participation in the TPGG.

26I cap punishment possibilities to detract away from highly endowed individuals having the capacity to punish more than poorly endowed individuals (Fehr and Gächter, 2002; Reuben and Riedl, 2011).
members from a group assign punishment points to a particular individual.\textsuperscript{27} There only needs to be an agreement in the decision to punish, not the magnitude of punishment. When the reductions are carried out and the threshold has been met, payoffs are:

\[
\pi^c_i = \max\{e_i - \sum_{t=4}^{10} c_{it} - \sum_{j \neq i} K(j)p^j_i - 3K(i)\sum_{j \neq i} p^j_i, 0\}
\]  \hspace{1cm} (5)

where:

\[
K(i) = \begin{cases} 
1 & \text{if } (\sum_j I_{i,j}) \geq 2 \\
0 & \text{otherwise}
\end{cases}
\]

and \(I_{i,j} = 1\) if agent \(j\) wants to punish agent \(i\) (ie, \(p^j_i > 0\)).

If a consensus to punish cannot be reached, punishment requests will bear no costs on the punisher and the non-punished individual will maintain their net private endowment that remains after any successful punishment requests of their own. In the event that total punishment reductions exceed the punished player’s remaining funds, the punished player is left with a zero payoff and no rebate will be given to punishers for points allocated beyond the punished player’s remaining funds.

If the threshold is not met, each agent additionally faces a treatment assigned probabilistic risk \((r_i)\) of losing remaining net private payoffs, resulting in expected earnings of

\[
E(\Pi^C_i) = r_i(0) + (1 - r_i)\pi^c_i = (1 - r_i)\pi^c_i. \hspace{1cm} (6)
\]

### 3.4 Questionnaire

The questionnaire distributed at the conclusion of the four games collected socioeconomic information (gender, age, number of siblings, college major, political party affiliation, tuition source, etc.) and elicited beliefs regarding the responsibilities of the rich and poor to contribute toward a public good. In addition to fairness beliefs and trust, I also gauged the influence of risk and predetermined play on an individual’s contribution choices. Subjects were asked to identify the main driver for their cooperative behavior in the threshold PGG (predetermined endowment inequality, cumulative group investments starting in round 4, monetary self-interest, fairness considerations, or achievement of the targeted threshold). Select summary statistics are presented in Table 10.

\textsuperscript{27}A consensual punishment mechanism parallels governing bodies responsible for imposing sanctions. When punishments are enforced after agreed upon by a majority of members, these sanctions would be carried out by those who vote for them and thus incur the cost of sanctioning themselves. Outside of the environmental arena, examples abound where economic sanctions are imposed upon nations that deviate from the realm of acceptable behavior. A group of nations would vote on whether to impose sanctions and if successful the member nations voting in favor would be responsible for the sanctions imposed. Of course, unilateral sanctions are possible (along the lines of autonomous punishment), but I abstract away from this possibility in the experiment and focus on developing a mechanism that could be implemented along the global scale. (See Section 2.5 for more information.

\textsuperscript{28}Contrasting Casari and Luini (2009), this setup does not prevent the formation of reputation during the game and allows delayed punishment.
4 Hypotheses

Collective action is difficult to sustain in multi-period public goods games. Though empirical evidence reveals non-zero contributions in contrast to the dominant strategy of free riding, cooperation in repeated play games tapers off rendering the socially efficient outcome unachievable. Acknowledging general findings from the literature regarding decreasing (stable) contributions in threshold PGGs without (with) punishment, heterogeneous asymmetries along the dimensions of endowment and risk further complicate cooperation.

Group Behavior, External Factors

Hypothesis 1a: Across non- or pro-punishment treatments, increased homogeneous risk will increase group contributions while heterogeneous risk will impede cooperation.

Hypothesis 1b: Punishment will lead to higher group contributions and greater threshold attainment, relative to their non-punishment counterparts.

While high levels of homogeneous collective risk have induced cooperative behavior in experimental games (Milinski et al., 2008), heterogeneous risk factors coupled with endowment inequality is conjectured to hinder cooperative inclinations when punishment is not available. Along the single dimension of endowment inequality with homogeneous risks, Tavoni et al. (2011) showed that collective efforts to reach a target threshold are impaired if no coordination-inducing mechanism is introduced. As punishment has been shown to have a positive impact on group behavior, I predict it will increase cooperation among low-risk treatments and align interests in heterogeneous risk groups.

Individual Behavior, External Factors

Hypothesis 2a: Poorly endowed individuals will give relatively less than richer group members, across like risk treatments. Conditional cooperation by “poor” players is anticipated if “rich” group members signal willingness to contribute in early rounds.

Hypothesis 2b: Identically endowed players with high-risk factors will systematically contribute more than group members with low-risk factors.

Hypothesis 2c: Punishment possibilities will serve as a coordination-inducing mechanism to further close the contribution gap between differently endowed participants.

Although both endowment and risk inequality figure to impact individual contributions, I hypothesize that the former asymmetry will be the more salient factor influencing choices. Given the immediate specification delegating two individuals poorer in endowment than the remaining group members, this attribute figures to be the defining influence in early periods for individuals with high discount rates and who lack forward-looking tendencies. Large initial endowments and high risk factors will induce risk-averse behavior among “rich” participants who have a signifi-
cant incentive to protect their private account, leading these individuals to contribute a relatively greater share of their endowment to threshold attainment. Poor subjects with lower initial endowments will conditionally cooperate if significant early contributions are made by those with higher endowments, signaling an intent to reduce distributional endowment inequality. In punishment treatments, available sanctions may further increase the gap in contributions among poor and richly endowed individuals given the mechanism’s credible threat.

I anticipate that equally endowed individuals with high-risk factors will contribute relatively more than identically endowed low-risk individuals, given their greater susceptibility to economic ruin. It remains an empirical question whether lowly-endowed high-risk subjects contribute a greater relative share of their initial endowment compared to highly-endowed low-risk subjects.

A punishment mechanism will induce greater responsibility for rich subjects to close the inequality gap, as fairness concerns can only be solved with mirrored inequality in distributional contributions. Incorporating only punishment may improve upon simultaneous punishment and communication mechanisms by making threats credible and diluting the existence of cheap talk (Bochet et al., 2006). In heterogenous risk groups, a consensual punishment mechanism should increasingly help eliminate anti-group behavior relative to homogeneous treatments.

**Individual Behavior, Internal Factors**

**Hypothesis 3:** High risk aversion will induce larger contributions to the group account across equally endowed individuals. High levels of inequality aversion motivate cooperation, but combined with endowment inequality will impact “rich” and “poor” players differently.

Though I randomly assign endowment and risk factor inequality among participants, inequality and risk aversion will also impact decision-making. Risk aversion should induce individuals with high risk factors to contribute more to the public good. Disadvantageous inequality aversion should negatively impact contribution levels of poorly endowed subjects and to a greater degree among individuals with low risk factors. Examining implications from Fehr and Schmidt (1999), this experiment is able to test the validity that subjects with high levels of advantageous inequality aversion are more likely to cooperate in the public goods game.

**Group Behavior, Internal Factors**

**Hypothesis 4:** Larger mean levels of risk or inequality aversion among group members will positively impact total contributions, but a greater variance between group members will inhibit threshold attainment.

While individuals marginally impact the final outcome, group composition and collective intrinsic preferences can greatly influence cooperation. Assigned risk and endowment profiles are explicit barriers to cooperation, however, groups have a better chance of overcoming these obstacles if they share similar risk and inequality aversion preferences. Larger group means for these intrinsic parameters should impact group contributions and threshold attainment similar to individual measures. Greater variation in these measures should negatively impact group cooperation.
5 Instrument Check

This section reports trends in the estimation of aversion instruments (see Sections 3.1 and 3.2 for experimental design), identifies suboptimal decisions among participants, and analyzes how these results compare to those of prior research.

5.1 Risk Aversion Estimation

Endowed with 100 tokens and permitted to invest any portion of this amount in a risky asset that returned 2.5 times the amount invested if successful, the average investment amount was about 46 tokens. The estimated risk aversion parameter ranged from 0 to 32.48, representing risk-loving preferences and relative “extreme” risk aversion.29 The average 46 token investment is associated with a risk aversion parameter equal to 0.755. As noted earlier, the endpoints of potential investment do not allow for explicit point-estimation of the risk aversion parameter. Of the 216 participants, 26 chose to invest the full 100 token endowment and 1 chose to invest nothing. These extreme choices were coded in accordance with suggestions from Charness and Genicot (2009). Further, 71% of participants chose to invest 50 or fewer tokens.

5.2 Inequality Aversion Estimation

In the ultimatum game, used to estimate disadvantageous inequality aversion (\(\alpha\)), proposers offered a share of the 20 token pie which the responder could accept or decline. The proposers’ mean offer was 44% of the pie, in line with an average offer of 40% in Blanco et al. (2011). About 57% of the proposers offered the even (10,10) split, while eight individuals proposed offers greater than half of the pie.30 A mere 3% of offers were consistent with the profit-maximizing subgame perfect equilibrium of offering nothing or 1 token. Employing the strategy method by presenting each of the 21 possible Proposer-Responder allocations, I sought to find each individual’s switching point from rejection to accepting the proposed offer. The average switching point occurred in between allocations (14,6) and (13,7), with participants being amenable to offers that exceeded

---

29 This estimation procedure produces a range of risk aversion estimates that is strictly contingent on the specified utility function. Qualitatively, a one-unit increase in the estimated risk aversion parameter does not yield interpretive value beyond being more risk averse relative to another individual. Similar risk elicitation experiments (see Holt and Laury (2002)) produce their own arbitrary range of estimated risk preferences that do not always correlate with the slew of alternate methods. Specifically, my chosen procedure to derive point-estimates for risk aversion does not correlate well with Holt’s interval elicitation method, whose method I also used in the end-of-experiment questionnaire. A main reason for this deviation is that Holt’s method restricts possible risk aversion into 11 distinct values while the Charness and Genicot (2009) point-estimate method produced 28 separate risk aversion values determined by the amount individuals invested (i.e., there were 28 distinct investment choices among the participants). Comparing the two methods, it is possible that no systematic relationship exists as an increase in the point-estimate risk aversion for Person \(i\) may not find itself in a higher Holt risk aversion category, but instead assigned to the same category given the non-comparable cutoff points for this method. When testing for correlation between the two methods, any possible association (positive, negative, none) may be anticipated for the above reasons. With this insight, I restrict my analysis to point-estimates for risk aversion and avoid interval estimates.

30 Offers from the ultimatum game have been used to derive advantageous inequality aversion (see Fehr and Schmidt (1999)), however, Blanco et al. (2011) points out certain restrictions and thus I defer to the MDG game to estimate this preference more aptly.
about 33% of the pie. Roughly 76% of participants would accept an offer less than the egalitarian (10,10) payoff distribution.

In the modified dictator game to measure advantageous inequality aversion (β), players decided between a selfish (20,0) allocation that benefited the dictator and an increasing equitable payoff vector (x_i, x_j). The mean switching point occurred just after (9,9) and 31% of participants switched to the equal payoff at or before (9,9). The modal switching point was at (10,10), a level 31% of participants chose as their personal threshold between choosing the selfish payoff and the egalitarian outcome. Eleven (5%) individuals never deviate from the (20,0) option and thus do not exhibit any advantageous inequality aversion (β_i = 0). On the opposite side of the spectrum 7 (3%) participants chose the equitable outcome for each of the 21 decisions and are characterized as extremely averse to advantageous inequality (β_i = 1). A total of 7 (3%) participants switch to the equitable choice only when it is costless (i.e., at the (20,20) payoff). These results are comparable to the findings in similar dictator games. Blanco et al. (2011) find a mean switching point around (11,11). They report that 8% of players only switch to the egalitarian outcome when it is costless and 10% never switch from the (20,0) option, preferring this outcome over (20,20).

Across aversion estimates, there is wide berth of heterogeneity among subjects. Over 87% of subjects display aversion to both advantageous and disadvantageous inequality (β > 0 and α > 0), while only 2% can be characterized as purely stoic (β = 0 and α = 0). Exploring the independence between the parameters, advantageous and disadvantageous inequality aversion are not correlated (Spearman correlation test, p=0.533). Figure 1 corroborates this test and exhibits a wide distribution among the subject population. Further, I find that 65 of 216 participants (about 30%) violate Fehr and Schmidt’s assumption that α_i ≥ β_i. These individuals are represented by points to the left of the line “α = β” in Figure 1. In addition to a lack of correlation between the two inequality aversion parameters, I also find no significant correlation between risk aversion and advantageous inequality aversion (Spearman test, p=0.224) or disadvantageous inequality aversion (Spearman test, p=0.182), consistent with prior empirical findings.

6 Results

In this section, I first present aggregate summary statistics and trends from the threshold public goods game, then report observable differences in contribution levels by treatment, period, and player type. The following sub-section is devoted to statistical tests that establish the impact of external factors (risk, inequality, and punishment) on dynamic behavior. Lastly, I delve into the impact of intrinsic preferences on group cooperation and investigate my hypotheses.

---

31 Among the ultimatum and modified dictator games, non-unique switching points occurred in 12% of cases, less than the 15.3% observed in Blanco et al. (2011). Holt and Laury (2002) present a similar format to the UG and find that 18.9% of participants do not have well-behaved preferences. See Footnote 17 for handling these participants.

32 This result violates the assumptions by Fehr and Schmidt (1999) that the two factors are positively correlated, but Teyssier (2012) and Blanco et al. (2011) have since exhibited the lack of correlation in empirical studies.
6.1 Summary Statistics from TPGG

6.1.1 Individual Mean Contributions

Individual mean contribution trends in the aggregate setting of all treatments (Table 2) center around the symmetric equilibrium of 2 token contributions in rounds 4-8. Inducing inequality in rounds 1-3, recall that participants were subjected to three inactive contribution periods, forcing half of group members to give 4 tokens while the other half were forced to contribute nothing. By construction individual mean contributions in the three inactive periods was exactly 2 tokens and I refrain from posting summary statistics from predetermined play. Rounds 9 and 10 see a drop in contributions as group totals approach the established threshold. Refining analysis to treatments grouped by the existence of punishment (last two columns), larger individual mean contributions occur in punishment treatments up to and including round 7. In the later rounds (8-10) the trend between these two comparison groups is reversed as participants in the punishment treatment scale back contributions more quickly as they approach the threshold.

Investigating mean individual contributions by treatment reveals similar results. Treatments without punishment options (T1, T3, & T5) generally under contribute relative to their punishment counterparts in the initial active contribution rounds (4-7). Homogeneous high-risk groups (T3 & T4), however, contribute similar amounts irrespective of the existence of punishment. I return to this point following proper testing for statistical differences. Across non-punishment treatments, homogeneous low-risk individuals contribute relatively less than those in high-risk or
Table 2: Mean Individual Contributions (Threshold PGG)

<table>
<thead>
<tr>
<th></th>
<th>Aggregate</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round_4</td>
<td>2.08</td>
<td>1.72</td>
<td>2.22</td>
<td>1.83</td>
<td>2.44</td>
<td>1.83</td>
<td>2.44</td>
<td>1.80</td>
<td>2.37</td>
</tr>
<tr>
<td>Round_5</td>
<td>2.11</td>
<td>1.67</td>
<td>2.11</td>
<td>2.39</td>
<td>2.28</td>
<td>1.89</td>
<td>2.33</td>
<td>1.98</td>
<td>2.24</td>
</tr>
<tr>
<td>Round_6</td>
<td>2.09</td>
<td>1.78</td>
<td>2.22</td>
<td>2.33</td>
<td>2.17</td>
<td>1.72</td>
<td>2.33</td>
<td>1.94</td>
<td>2.24</td>
</tr>
<tr>
<td>Round_7</td>
<td>2.10</td>
<td>1.56</td>
<td>2.11</td>
<td>2.28</td>
<td>2.28</td>
<td>1.72</td>
<td>2.22</td>
<td>2.00</td>
<td>2.20</td>
</tr>
<tr>
<td>Round_8</td>
<td>2.01</td>
<td>1.94</td>
<td>2.00</td>
<td>2.11</td>
<td>1.83</td>
<td>2.22</td>
<td>1.94</td>
<td>2.09</td>
<td>1.93</td>
</tr>
<tr>
<td>Round_9</td>
<td>1.73</td>
<td>1.56</td>
<td>2.06</td>
<td>1.83</td>
<td>1.94</td>
<td>1.89</td>
<td>1.11</td>
<td>1.76</td>
<td>1.70</td>
</tr>
<tr>
<td>Round_10</td>
<td>1.64</td>
<td>1.61</td>
<td>1.67</td>
<td>1.78</td>
<td>2.11</td>
<td>1.56</td>
<td>1.11</td>
<td>1.65</td>
<td>1.63</td>
</tr>
<tr>
<td>Rds_4_to_10</td>
<td>1.97</td>
<td>1.69</td>
<td>2.06</td>
<td>2.08</td>
<td>2.15</td>
<td>1.90</td>
<td>1.93</td>
<td>1.89</td>
<td>2.04</td>
</tr>
</tbody>
</table>

N: 216 36 36 36 36 36 36 108 108

T1: Homo (1/3) Risk, Non-Punish
T2: Homo (1/3) Risk, Punish
T3: Homo (2/3) Risk, Non-Punish
T4: Homo (2/3) Risk, Punish
T5: Hetero Risk, Non-Punish
T6: Hetero Risk, Punish

heterogeneous groups. Between homogeneous high-risk (T3) and heterogeneous risk (T5) individuals, those in the former treatment gave larger amounts on average in the initial contribution rounds, potentially highlighting the absence of accountability among group members in the heterogeneous treatment. Across punishing treatments, contribution levels are consistently higher than the symmetric 2 token contribution equilibrium in early active rounds and slightly larger in groups containing high-risk individuals.

Examining coordination during early and late rounds, Figure 2 helps visualize the dynamics of individual behavior across treatments. Bar columns in this figure represent the average number of times group members contributed 0, 2, or 4 experimental tokens (ETs) during early and late rounds, by treatment. For example, in Panel A the initial value of 4.3 represents the mean number of times individuals in Treatment 1 contributed 0 ETs in early rounds (5-7). Among all treatments, free riding increases as the game progresses which can be for one of two reasons: either (i) individuals give up their pursuit of an unattainable threshold and hold on to remaining net private endowments or (ii) contributions are scaled back as groups approach the collective threshold. In homogeneous low-risk groups without punishment (Panel A), relatively high levels of early free riding increase, overwhelming a modest increase in 4 token contributions in late periods. Over half of contributions in low-risk punishment groups (Panel B) were for the 2 token level, while low levels of 0 and 4 token contributions balanced each other out. Homogeneous high-risk treatments (Panels C & D) exhibit nearly identical early and late round contribution trends, enjoying

---

\[33\] Round 4 is excluded from this analysis for two reasons: (i) to compare “early” versus “late” round dynamics, each bin needed to contain an equal number of rounds, and (ii) decisions made in Round 4 more closely reflect inequality or risk aversion preferences as active group interaction has not yet occurred.

\[34\] Note that all bars in a panel sum to 24 which represents the total number of decisions made by four group members in rounds 5-10.
relatively high levels of 2 and 4 token contributions. Coupled with low levels of free riding, these two treatments combined to reach the threshold 83% of the time. Lastly, the punishment mechanism had a visible impact on contributions over time in heterogeneous groups. Without the overt punishment threat (Panel E), a coordination failure occurs with both an increase in free riding and 4 token contributions in late rounds. Punishment possibilities induce cooperation and Treatment 6 (Panel F) enjoys the highest level of early round 4 token contributions. Large early contributions are complemented by increased free riding and decreased 4 token contributions in late periods as players scale back investments with a quickly approaching threshold.

Among all punishment treatments, free riding is mild in early rounds compared to non-punishment groups (p=0.078) as players coordinate investments toward the collective fund. Large contributions regularly decrease in late periods, but only after threshold attainment is nearly guaranteed. Relatively high levels of early free riding is observable across non-punishment treatments, and without the threat of retribution continues into late rounds as large contributions cannot compensate to avoid threshold non-attainment.
Table 3: Mean Group End-of-Round Contribution Totals (Threshold PGG)

<table>
<thead>
<tr>
<th>Aggregate</th>
<th>by Treatment</th>
<th>by Punishment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Round 4</td>
<td>32.33</td>
<td>30.89</td>
</tr>
<tr>
<td>Round 5</td>
<td>40.78</td>
<td>37.56</td>
</tr>
<tr>
<td>Round 6</td>
<td>49.15</td>
<td>44.67</td>
</tr>
<tr>
<td>Round 7</td>
<td>57.56</td>
<td>50.89</td>
</tr>
<tr>
<td>Round 8</td>
<td>65.59</td>
<td>58.67</td>
</tr>
<tr>
<td>Round 9</td>
<td>72.52</td>
<td>64.89</td>
</tr>
<tr>
<td>Round 10</td>
<td>79.07</td>
<td>71.33</td>
</tr>
</tbody>
</table>

N = 54 9 9 9 9 9 9 27 27

T1: Homo (1/3) Risk, Non-Punish  T2: Homo (1/3) Risk, Punish
T3: Homo (2/3) Risk, Non-Punish  T4: Homo (2/3) Risk, Punish
T5: Hetero Risk, Non-Punish  T6: Hetero Risk, Punish

6.1.2 End-of-Round Aggregates

Exploring end-of-round total contributions (Table 3), I evaluate how groups across treatments differ in their pursuit of the collective threshold and Section 6.2 presents statistical testing. Among all treatment groups, the mean total contribution after ten rounds of play was about 79 tokens, marginally missing the threshold on average. Distinguishing by threshold attainment, 43 of the 54 groups (80%) were successful in meeting the target (80 tokens) with an average group contribution of 83.2 tokens, while the 11 groups that failed collected an average of 62.9 tokens after ten rounds. Of the non-attainment groups, 7 of the 11 groups were from non-punishment treatments with an average contribution of 59.4 tokens, while the remaining four from punishment treatments averaged 69 tokens.

Breaking down the temporal momentum of total group contributions aiming for threshold attainment (Table 3), I find similar patterns to those observed among individual contributions. Non-punishment treatments systematically produce lower contribution totals after every round relative to their punishment counterparts, both in the aggregate and in pairwise comparisons. At the end of ten rounds, punishment treatments on average attain the threshold (81.3 tokens) while non-punishment treatments fail (76.9 tokens). Although there is a sizable increase in group contributions after 10 rounds in punishment treatments relative to their non-punishment baselines, in reality, the former achieves the threshold 85% of the time with the latter not far off at 74% achievement. Whether or not a complex punishment mechanism is necessary to induce cooperation to reliable levels will be explored in the coming analysis.

In determining the impact of the punishment mechanism, I examine final contribution totals among the treatments with their punishment counterparts. In homogeneous low-risk treatments, punishment induces a sizable increase in total contributions to the group fund, pushing the average total (81.6 tokens) beyond the targeted threshold. Further, the number of groups failing to reach
the threshold decreased by 67% with the inclusion of the punishment mechanism. When groups were characterized by only high-risk members (T3 & T4), both treatments on average achieve the threshold and punishment is potentially not necessary. Heterogeneous groups, with and without punishment, appear to be plagued by a coordination issue as these treatments on average do not attain the threshold. Delving deeper, however, both treatments bear one group with a very low contribution total, dragging down their respective mean contribution totals. Excluding these possible outliers, the mean contribution totals were 79.3 and 82.0 for the non- and pro-punishment treatments, exhibiting the positive impact of a coordination-inducing mechanism when agents are heterogeneous along two dimensions (risk & endowment). Similar to the homogeneous low-risk treatment, the inclusion of a punishment option decreased the number of groups failing to reach the threshold by 67 percent.

6.1.3 Player-Type Dynamics

Lastly, I examine individual contribution trends by player type (Table 4). Amidst the mix of homogeneous, heterogeneous, and punishment groupings, there were four types of players that varied according to endowment and assigned risk factor: rich/low-risk (P1), rich/high-risk (P2), poor/low-risk (P3), and poor/high-risk (P4). Although I avoid the wording in the actual experiment, “rich” players are those who started round 4 with a 40 token net endowment (forced to contribute nothing during the three inactive rounds) and “poor” players are those who were forced to contribute 4 tokens per inactive round, starting round 4 with a 28 token net endowment. Observing active period (round >3) mean contributions in Table 4, there is a propensity for rich individuals to not only contribute more than the 2 token symmetric equilibrium (p=0.000), but also to contribute more on average than their poor-player counterparts regardless of punishment possibilities (p=0.000 for both low-risk and high-risk types). Such a result suggests that rich players sought to correct the endowment inequality induced at the start of the game, irrespective of the punishment threat. Further, high-risk players unequivocally give more than low-risk players on average (p=0.000). Disaggregating by punishment possibilities, the mechanism imparts greater contributions for most player types (P1-P3) while only player type P4 realizes a decrease in contributions when punishment is present. Testing for significance, low-risk player (P1 & P3) contributions increase (p=0.013 and p=0.078, respectively) with the punishment mechanism. Threatened by punishment, rich high-risk players (P2) have plenty to lose and contribute more (p=0.112) relative to non-punishment treatments, while contributions decrease for poor high-risk players (P4) (p=0.717). These results suggest that at the individual level punishment is effective as a coordination-inducing mechanism when risk is low, but when risk factors are high punishment does not significantly increase contributions to the target fund.

Though rich players contribute more on average across all treatments during active periods (Table 4), the question persists as to whether threshold attainment is characterized by equalizing the contribution burden between rich and poor players within groups. Modeling endowment inequality by inducing predetermined play for the first three rounds, it is constructive to examine the share of total contributions from rich and poor players during the entire game. In addition to varying degrees of risk susceptibility, unequal endowments serve as a barrier to cooperation that can only

35 All p-values are two-tailed unless noted otherwise.
Table 4: Mean Individual Contributions, by Player Type (Threshold PGG)

<table>
<thead>
<tr>
<th></th>
<th>All Treatments</th>
<th>No Punishment</th>
<th>Punishment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Contribution</td>
<td>2.30</td>
<td>2.50</td>
<td>1.35</td>
</tr>
<tr>
<td>N</td>
<td>378</td>
<td>378</td>
<td>378</td>
</tr>
</tbody>
</table>

P1: Rich/Low-Risk   P2: Rich/High-Risk
P3: Poor/Low-Risk   P4: Poor/High-Risk

Note: Summary statistics for active contribution periods only (Round > 3)

be overcome with sufficient redistribution of contribution responsibilities to offset the induced inequality. Scaling up mean contributions for active periods in Table 4 and including predetermined play contributions, I find that although rich players gave more in active periods, they did not fully close the inequality gap for equal burden sharing. Figure 3 highlights this point for all treatments, where poor-player contributions make up 58% of the collective fund while rich players contributed 42% of the group total. Distinguishing by threshold attainment in Panel B, successful groups decrease the inequality gap by about 4 percentage points relative to failing groups (p=0.069). Panels C and D in Figure 3 break down rich and poor contribution shares with the introduction of the punishment mechanism. In both non- and pro-punishment treatments, successful threshold attainment is characterized by reducing the inequality gap beyond the all treatments case (panel A). Without a punishment mechanism (Panel C), successful groups close the inequality gap by over 6 percentage points (p=0.052) relative to non-successful groups. Both failing and successful groups in punishment treatments (Panel D) outperform all other specifications in the quest to equalize contribution shares, however, successful groups do not close the inequality gap any better than failing groups (p=0.592). Never quite achieving full burden sharing, the punishment mechanism successfully reduces the inequality gap which increases the likelihood of threshold attainment.

Signaling cooperation and a willingness to equalize burden sharing, immediate contributions in Round 4 play an important role in successful threshold attainment. Across all treatments (Table 5) rich players, regardless of risk type, contribute over a full token more than poor players in Round 4 (p=0.000). Assessing how larger initial contributions by rich players translate into successful cooperation, I break down the data by non-attainment and attainment. Among rich players (P1 & P2), those in successful groups gave significantly more on average in Round 4 than those in failing groups (p=0.043). Among poor players (P3 & P4), there was no significant difference in contributions between those in successful or failing groups (p=0.313). Applying a simple probit model to ascertain the significance of Round 4 contributions among rich and poor players in reaching the threshold, I further substantiate the importance of rich player contributions (p=0.050) and

Relative to non-attainment groups, the questionnaire revealed individuals from successful groups increasingly preferred higher contributions from rich players in active rounds (p=0.085). Both attainment types agreed that poor players were not exclusively responsible for increasing their contributions relative to rich players (p=0.668). Further breaking down beliefs, within non-attainment groups rich players significantly disagree with poor players that it is their social responsibility to contribute more in active rounds (questionnaire response, p=0.051).
Figure 3

Contribution Share of Rich and Poor

A: All Treatments

B: All Treatments, by Attainment

C: Non-Punish, by Attainment

D: Punish, by Attainment

Note: Red = Rich, Green = Poor; 0 = Non-Attainment, 1 = Attainment
Table 5: Round 4 - Mean **Individual** Contributions, by Player Type (Threshold PGG)

<table>
<thead>
<tr>
<th></th>
<th>All Treatments</th>
<th>Non-Attainment</th>
<th>Attainment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>Contribution</td>
<td>2.74</td>
<td>2.70</td>
<td>1.30</td>
</tr>
<tr>
<td>N</td>
<td>54</td>
<td>54</td>
<td>54</td>
</tr>
</tbody>
</table>

P1: Rich/Low-Risk  P2: Rich/High-Risk  P3: Poor/Low-Risk  P4: Poor/High-Risk

Note: (Non)Attainment pertains to reaching threshold by Round 10

the non-significance of poor player contributions (p=0.308). These results stress the importance of immediate signaling by rich players in Round 4 to influence threshold attainment.

### 6.2 Statistical Tests of Group Contributions

On the surface there appears to be a detectable difference between mean contributions with and without a punishment mechanism. With only nine groups per treatment there is insufficient power when testing for differences between final end-game contribution totals. Since participants contribute personal endowments to a group fund for ten rounds and scale back investments as the total gradually reaches the predetermined threshold, end-game totals among all treatments hover above and below the threshold without significant variation in end-game totals.

Acknowledging the above issues, I test for statistical differences in contributions by round among the treatments. It is constructive to compare mean group-period contributions, by treatment, to establish a set of stylized facts consistent among the data. Mean group-period contributions relate average contributions by group, period, and treatment for all active group play (rounds 4-10). The values of this metric can be derived from Table 2, multiplying each number by four to account for the complete group. Figure 4 tests for differences among all combinations of treatments. The first reported statistic in each column relates two-sided p-values (unless noted otherwise) for statistical differences when comparing low/high-risk homogeneous and heterogeneous groups to their punishment counterparts. Of these comparisons, only the low-risk homogeneous treatment is statistically different than its pro-punishment alternative (p=0.009) which is consistent with the findings reported in the prior section, comparing mean end-of-round contribution totals among treatments. Varying risk and group composition, I assess differences among treatments in the same punishment category. For non-punishment treatments, homogeneous low-risk groups statistically under contribute relative to high-risk and heterogeneous groups.\(^{37}\) High-risk homogeneous groups are not plagued by coordination issues and significantly contribute more on average per period relative to heterogeneous groups (one-sided, p=0.092). With regard to punishment treatments,

\(^{37}\)Comparing non-punishment homogeneous low-risk groups (T1) with heterogeneous groups (T5), I employ a one-sided t-test. The null hypothesis is that though heterogeneity impedes cooperation, the existence of high-risk individuals in the heterogeneous group (T5) relative to uniformly low-risk individual groups (T1) increases the mean contribution level in the heterogeneous group. Similar logic is used with statistics reported alongside an asterisk (*).
varying risk and group composition, no statistical differences exist. This finding suggests that a consensual punishment mechanism may be the “great equalizer” that eliminates the differences in mean contributions among treatments with different group compositions. While perceived risk levels (be they uniformly high, low, or heterogeneous) to dangerous climate change are privately developed in the real world, experimentally I find that a punishment mechanism ensures almost universal threshold attainment regardless of the validity of personal risk beliefs.

Analyzing all active rounds (4-10) abstracts away from particular trademarks of threshold public goods games. Threshold attainment during group play is usually achieved in one of two general ways: (i) group contributions start strong and taper off as the threshold is approached or (ii) group contributions are humble in early periods and ramp up with the looming threat of non-attainment. While it is uncertain if a punishment threat among individuals is relevant during the entire game or instead gains traction in later periods, there exists a consistent downshift in mean group contributions (Table 6) between early active rounds (4-8) and late active rounds (9-10). Failure to partition contributions in this fashion conceals the impact of a punishment mechanism on early cooperation. Including all active periods in Figure 4, there was no statistical improvement in mean contributions among heterogeneous groups (T5 & T6) when including punishment. In Figure 5 I restrict analysis to “early” active round contributions to determine the early impact of punishment across treatments. All statistical differences, both in direction and significance, found in the prior section (see Figure 4) are identical, bar two. I find a significant difference in mean contributions between heterogeneous groups with (T6) and without (T5) punishment options. Looking at all active rounds (4-10) in Table 6, Treatment 6 experiences a large decrease in mean contributions between these early and late rounds while Treatment 5 did not experience a dramatic downshift. The result of this sizable difference is that averaged over all active rounds (4-10), as reflected in Figure 4, there is no significant difference between T5 and T6 contributions, thus no detectable impact of punishment on heterogeneous group behavior. In Figure 5, however, I find that punishment indeed had the intended effect (p=0.057) in overcoming coordination issues among heterogeneous agents, although did not push average group contributions over the target threshold (Table 3). Lastly, when parsing out the strong early round mean contributions (9.02) in Treatment 6 from the significant decrease in the later rounds (4.44), I find that there is no statistical difference in mean contributions between punishment treatments with homogeneous high-risk (T4) and heterogeneous (T6) agents (p=0.726).

Summarizing Section 6.1 and 6.2 results, I arrive at a handful of key points:

- Observation #1: Punishment is only effective when perceived risk is low enough among homogeneous or heterogeneous agents
- Observation #2: Coordination issues (without punishment) are overcome if perceived risk is high enough
- Observation #3: Punishment may be the “great equalizer” by eliminating total contribution differences between homogeneous and heterogenous groups

---

"Early" and "late" active round designations are arbitrary. Results are qualitatively robust for other specifications: (i) early rounds (4-7), late rounds (8-10); (ii) early rounds (5-7), late rounds (8-10) as in Figure 2.
Table 6: Mean Group Period Contributions, by Treatment (Threshold PGG)

<table>
<thead>
<tr>
<th></th>
<th>Rounds 4-8</th>
<th>Rounds 9-10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Contribution</td>
<td>6.93</td>
<td>8.53</td>
</tr>
<tr>
<td>N</td>
<td>180</td>
<td>180</td>
</tr>
</tbody>
</table>

T1: Homo (1/3) Risk, Non-Punish  
T2: Homo (1/3) Risk, Punish  
T3: Homo (2/3) Risk, Non-Punish  
T4: Homo (2/3) Risk, Punish  
T5: Hetero Risk, Non-Punish  
T6: Hetero Risk, Punish

- Observation #4: Individual accountability in heterogeneous groups is heightened with a punishment mechanism
- Observation #5: Rich players significantly reduce, but do not eliminate, the endowment inequality gap in successful groups
- Observation #6: Rich players who signal early cooperation significantly increase threshold attainment

![Group-Period Mean Comparisons, Round>3 (p-values)](image)

Note: p-values are two-tailed t-tests measuring difference in Group-Period mean contributions, by Treatment. Test results denoted by (*) indicate one-tailed t-tests, with the following hypotheses:  
(i) mean(T5) = mean(T1), (ii) mean(T3) = mean(T5), (iii) mean(T4) = mean(T6)

Figure 4

6.3 Regression Analysis

External factors including endowment inequality, risk susceptibility, and punishment have been shown to impact contributions and threshold attainment across treatments, however, I have ne-
neglected the role of intrinsic preferences that also influence behavior. Incorporating inequality and risk aversion measures from the three independent games prior to the threshold game, I investigate the behavioral preferences that both encourage and inhibit cooperation in the collective-risk public goods setting. With the same set of incentives and constraints, assessing the impact of intrinsic preferences helps identify why similar treatment groups fail to cooperate while others effectively reach the collective threshold.

Collecting repeated decisions from a fixed number of individuals and groups, the experimental data is best analyzed in a panel setting. Before incorporating intrinsic preferences, I first consider the impact of external factors (inequality, risk, and punishment) to corroborate findings from the previous sections. Table 7 presents a random-effects model for individual contributions to the group account during active periods (Round > 3). Across all models, lagged group contributions significantly increase an individual's choice to contribute, signaling the typical preference for cooperation as the social norm materializes. Isolating the impact of lagged contributions from other group members (i.e., without player $i$), however, an individual's contribution significantly decreases showcasing a propensity to free-ride that may negate cooperative behavior even when considering the deterrents of punishment and collective risk. Increased risk susceptibility and lower relative endowments each have the predicted impact on behavior, significantly increasing and decreasing, respectively, an individual's contribution during active periods. While the magnitude of these opposing forces indicate that relatively low endowments have a greater impact on contributions than do high risk factors, for individuals with both high risk and low endowments an interaction term reveals a non-significant combined effect. Lastly after controlling for risk and endowment factors, the punishment mechanism had the intended effect as it significantly increased contributions toward the public good, consistent with prior quantitative findings. Increasing risk susceptibility and relative wealth can significantly impact cooperative behavior, but
on average these factors alone do not incentivize ubiquitous threshold attainment (see Section 6.1). With respect to the social planner, this is an important result as it highlights the necessity of a cooperation-inducing mechanism to efficiently attain the targeted threshold.

Table 7: Individual Contributions to Group Account (External Factors)

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Contrib, (t-1)</td>
<td>0.067***</td>
<td>0.154***</td>
<td>0.153***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Dist2Thresh</td>
<td>0.009***</td>
<td>-0.000</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>High Risk</td>
<td>0.381***</td>
<td>0.346***</td>
<td>0.269*</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.10)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Poor</td>
<td>-0.865***</td>
<td>-0.845***</td>
<td>-1.077***</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.10)</td>
<td>(0.34)</td>
</tr>
<tr>
<td>Group Contrib (others), (t-1)</td>
<td>-0.099***</td>
<td>-0.099***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>High Risk*Poor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.464</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Punish</td>
<td></td>
<td>0.266*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>1.487***</td>
<td>1.809**</td>
<td>1.573**</td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.78)</td>
<td>(0.76)</td>
</tr>
<tr>
<td>Period Dummies</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Treatment Dummies</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R2</td>
<td>0.324</td>
<td>0.476</td>
<td>0.474</td>
</tr>
<tr>
<td>N</td>
<td>1512</td>
<td>1512</td>
<td>1512</td>
</tr>
<tr>
<td>Wald</td>
<td>125.868</td>
<td>266.135</td>
<td>302.473</td>
</tr>
</tbody>
</table>

Note: Analysis for active contribution rounds (4-10) only
Robust standard errors in parentheses (clustered at group level)
* p < 0.1, ** p < 0.05, *** p < 0.01

Modeling intrinsic preferences alongside external pressures in Table 8, I increase the dimensions by which the rationale of individual behavior can be identified. Excluding from the output external factors modeled in Table 7 (all of which maintain their signs and noted significances), varying levels of risk and inequality aversion have a marked impact on individual contributions. In a game shaped by contrived collective risk, risk averse tendencies might conceivably bring about increased contributions. In the Model 4 random-effects regression, however, elicited risk aversion preferences are found to negatively impact contributions, albeit at a small magnitude. Though risk averse agents may contribute more in the threshold game to avoid non-attainment, risk aversion has been found to reduce individual contributions in group games with strategic uncertainty (Teyssier, 2012). Not only are agents internalizing their inherited risk susceptibility, they face the risk of heterogeneous group members failing to cooperate. This two-pronged risk dilemma (with opposite
effects) may explain the negative significance of elicited risk aversion. Carrying greater weight in magnitude, both advantageous and disadvantageous inequality aversion (representing guilt and envy) significantly increase individual contributions, with the former parameter nearly ten times as influential. This result stands in stark contrast to those found in Blanco et al. (2011), where Fehr and Schmidt’s inequality-aversion model had little explanatory power in a one-shot simultaneous public goods game at the individual level. In the threshold PGG with constant groups and multiple rounds, players dynamically incur bouts of envy and guilt during the course of the game, thus the saliency of their significant impacts. Punishment continues to increase contributions even when controlling for the wide range of internal and external factors influencing behavior.

Although aversion preferences are shown to influence behavior, anecdotal evidence suggests that these internal factors may impact unequally endowed individuals differently. Under the stress of strategic uncertainty during the multi-round game, risk aversion negatively impacts contributions. After interacting risk aversion with inherited low endowments, however, I find that this interaction is significantly negative while the conditional effect of risk aversion on rich agents is positive without significance (Model 5). Whereas external high-risk susceptibility coupled with poor endowments had no marked impact on contributions (Table 7), higher levels of risk aversion coupled with poor endowments increase distrust among participants leading to decreased cooperation. A possible explanation gleaned from the questionnaire is that the risk aversion measure, as priorly maintained, more closely models distrust between participants. Poorly endowed players exhibited lower levels of trust, anticipating richer group members to not decrease endowment inequality by giving a larger proportional share during active rounds. Highlighted in Table 5, however, signaling from rich players in the first active round (Round 4) reverses this trend as poor players on average increased their contributions in response to the positive signal of cooperation. Although I hypothesized that higher levels of (dis)advantageous inequality aversion would significantly (decrease) increase individual contributions among the (poor) rich, there were no detectable effects. Game dynamics increasing or decreasing the inequality gap, rich and poor designations do not explicitly vary the overall impact of inequality aversion. Accounting for period and treatment dummies (Model 6) marginally increases model fit and maintains sign and significance for all regressors.

Outside of individual preferences marginally impacting game dynamics, group composition and collective intrinsic preferences can significantly influence cooperation. Across all pooled treatments (first column of Table 9), prior round cooperation and the existence of punishment both significantly increase group contributions. Mean and standard deviation measures for aversion preferences reveal random group composition. Comparable to the individual analysis, higher levels of mean advantageous inequality aversion (eAIA) significantly increase group cooperation,

---

39These results are further explored when evaluating questionnaire responses later in the analysis.
40Similarly in a one-shot sequential public goods game, Teyssier (2012) finds that disadvantageous inequality aversion does not impact the first mover, but advantageous inequality aversion does significantly increase the second mover’s contributions.
41The final question on Game 4 in the questionnaire (Table 10) highlights accountability beliefs and trust. This question asks players if they believe “richly” endowed players would contribute a higher share than poor players during active rounds. While rich types (P1 & P2) agreed with this sentiment, poor types (P3 & P4) on average “neither agree or disagree,” suggesting lower levels of trust in rich agents to reduce the endowment inequality gap.
while increasing levels of mean risk aversion continue to decrease levels of trust and negatively impact contributions. Increasing variation among the envy parameter (eDIA) and risk aversion, however, negate high levels of distrust and instead positively contribute to the collective fund.

Table 8: Individual Contributions to Group Account (Internal Factors)

<table>
<thead>
<tr>
<th></th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>eRisk</td>
<td>-0.038***</td>
<td>0.106</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.09)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>eAIA</td>
<td>0.594***</td>
<td>0.435*</td>
<td>0.428*</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.24)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>eDIA</td>
<td>0.064**</td>
<td>0.024</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Punish</td>
<td>0.145*</td>
<td>0.125*</td>
<td>0.272*</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>eRisk*Poor</td>
<td>-0.155*</td>
<td>-0.149*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>eAIA*Rich</td>
<td>0.419</td>
<td>0.433</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(0.48)</td>
<td></td>
</tr>
<tr>
<td>eDIA*Poor</td>
<td>0.068</td>
<td>0.072</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>0.877***</td>
<td>0.730**</td>
<td>0.802</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.33)</td>
<td>(0.72)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>no</th>
<th>no</th>
<th>yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period Dummies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment Dummies</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>External Factors</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>R2</td>
<td>0.487</td>
<td>0.496</td>
<td>0.501</td>
</tr>
<tr>
<td>N</td>
<td>1512</td>
<td>1512</td>
<td>1512</td>
</tr>
<tr>
<td>Wald</td>
<td>280.830</td>
<td>782.542</td>
<td>842.551</td>
</tr>
</tbody>
</table>

Note: Analysis for active contribution rounds (4-10) only
Robust standard errors in parentheses (clustered at group level)
eRisk = elicited risk aversion
eAIA = elicited advantageous inequality aversion
eDIA = elicited disadvantageous inequality aversion
* p < 0.1, ** p < 0.05, *** p < 0.01

Examining each risk treatment in turn (inclusive of punishment possibilities), a clearer story unfolds. Punishment in homogeneous low-risk groups continues to incentivize cooperation (collecting 8 tokens more than the baseline), whereas the mechanism is not effective (or needed) in high-risk groups, consistent with individual level analyses. Increasing levels of risk aversion and its dispersion among group members have significantly opposite effects for low-risk treatments, while these factors maintain the sign of impact but lack significance among high-risk groups. High risk in homogeneous treatments appears to be the preeminent factor that influences cooperation.
both at the individual and group level. Although intrinsic preferences impact individual contributions, group variation among these parameters does not influence collective behavior as the fixation with high-risk susceptibility remains the ultimate free riding deterrent. Among group inequality aversion (IA) measures, mean advantageous IA significantly increases contributions in high-risk groups while mean disadvantageous IA imparts the same impact on low-risk groups. The questionnaire (Table 10) reveals high-risk groups agreed that rich players should be responsible for contributing a greater share of their endowment, thus the significant positive influence of the mean guilt parameter. A similar effect is seen among the low-risk treatment, however, incremental guilt does not significantly impact group contributions when coupled with low stakes of economic loss.

Table 9: Group Contributions per Period

<table>
<thead>
<tr>
<th></th>
<th>All Treatments</th>
<th>Homo 1/3 Risk</th>
<th>Homo 2/3 Risk</th>
<th>Hetero Risk</th>
<th>Pooled Homo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Contrib. (t-1)</td>
<td>0.321***</td>
<td>-0.048</td>
<td>0.322**</td>
<td>0.611***</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.20)</td>
<td>(0.13)</td>
<td>(0.12)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Punish</td>
<td>0.582***</td>
<td>1.196**</td>
<td>0.583</td>
<td>1.093***</td>
<td>0.776***</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.50)</td>
<td>(0.53)</td>
<td>(0.39)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>mean eAIA</td>
<td>1.867**</td>
<td>0.033</td>
<td>3.303***</td>
<td>5.253**</td>
<td>0.904***</td>
</tr>
<tr>
<td></td>
<td>(0.90)</td>
<td>(0.50)</td>
<td>(1.09)</td>
<td>(2.16)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>mean eDIA</td>
<td>0.087</td>
<td>0.453***</td>
<td>-0.346</td>
<td>-1.011***</td>
<td>0.402**</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.15)</td>
<td>(1.67)</td>
<td>(2.16)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>mean eRisk</td>
<td>-0.976*</td>
<td>-6.906***</td>
<td>-1.480</td>
<td>5.033***</td>
<td>-2.453***</td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td>(2.14)</td>
<td>(1.16)</td>
<td>(1.52)</td>
<td>(0.93)</td>
</tr>
<tr>
<td>sd eAIA</td>
<td>1.207</td>
<td>-4.929*</td>
<td>-3.448</td>
<td>0.804</td>
<td>-2.499</td>
</tr>
<tr>
<td></td>
<td>(1.54)</td>
<td>(2.82)</td>
<td>(6.39)</td>
<td>(2.78)</td>
<td>(2.58)</td>
</tr>
<tr>
<td>sd eDIA</td>
<td>0.276**</td>
<td>-0.030</td>
<td>0.733</td>
<td>1.966***</td>
<td>-0.225</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.23)</td>
<td>(2.15)</td>
<td>(0.56)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>sd eRisk</td>
<td>0.547*</td>
<td>4.966**</td>
<td>0.767</td>
<td>-3.117***</td>
<td>1.396***</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(2.49)</td>
<td>(0.78)</td>
<td>(0.96)</td>
<td>(0.52)</td>
</tr>
<tr>
<td>High Risk</td>
<td>1.058***</td>
<td>1.858***</td>
<td>1.966***</td>
<td>1.966***</td>
<td>1.396***</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.17)</td>
</tr>
</tbody>
</table>

Note: Analysis for active contribution rounds (4-10) only
Robust standard errors in parentheses

eRisk = elicited risk aversion
eAIA = elicited advantageous inequality aversion
eDIA = elicited disadvantageous inequality aversion
sd = standard deviation

* p < 0.1, ** p < 0.05, *** p < 0.01
Table 10: Questionnaire Summary (Mean Responses)

<table>
<thead>
<tr>
<th>Select Questions</th>
<th>Response</th>
<th>Player Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Those who began round 4 having invested 0 points (predetermined play forced this decision) should contribute more to the group account in the following 7 rounds than the other players.”</td>
<td>Likert Scale*</td>
<td>P1 2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 2.5</td>
</tr>
<tr>
<td>What would you consider a fair average investment for the seven active rounds for players whose predetermined investment was 0?</td>
<td>0-4 tokens</td>
<td>P1 2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 2.8</td>
</tr>
<tr>
<td>What would you consider a fair average investment for the seven active rounds for players whose predetermined investment was 12?</td>
<td>0-4 tokens</td>
<td>P1 1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 1.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 1.7</td>
</tr>
<tr>
<td>“I was influenced by my own risk when deciding how much to contribute to the group account in Game 4.”</td>
<td>Likert Scale*</td>
<td>P1 3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 3.2</td>
</tr>
<tr>
<td>“I was influenced by other players risk when deciding how much to contribute to the group account in Game 4.”</td>
<td>Likert Scale*</td>
<td>P1 4.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 4.1</td>
</tr>
<tr>
<td>“I was influenced by my own forced predetermined play when deciding how much to contribute to the group account in Game 4.”</td>
<td>Likert Scale*</td>
<td>P1 2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 2.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 2.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 2.4</td>
</tr>
<tr>
<td>“I was influenced by other players forced predetermined play when deciding how much to contribute to the group account in Game 4.”</td>
<td>Likert Scale*</td>
<td>P1 3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 3.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 3.0</td>
</tr>
<tr>
<td>“In Game 4 with new group members, players with a predetermined investment of 0 would contribute more to the group account than other players.”</td>
<td>Likert Scale*</td>
<td>P1 2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 3.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 3.2</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>P1 20.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 20.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 20.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 20.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 20.5</td>
</tr>
<tr>
<td>Sex</td>
<td>Male=1, Female=2</td>
<td>P1 1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 1.6</td>
</tr>
<tr>
<td>Economics courses taken?</td>
<td></td>
<td>P1 1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 1.4</td>
</tr>
<tr>
<td>Frequency attending religious services?</td>
<td>Time Scale**</td>
<td>P1 4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 3.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 4.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 4.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 4.1</td>
</tr>
<tr>
<td>How do you identify yourself?</td>
<td>Party Scale***</td>
<td>P1 3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P2 3.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P3 3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P4 3.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All 3.7</td>
</tr>
</tbody>
</table>

*Likert Scale (1-7): Strongly Agree, Agree, Somewhat Agree, Neither Agree nor Disagree, Somewhat Disagree, Disagree, Strongly Disagree

**Time Scale (1-5): More than once a week, Once a week, Once or twice a month, A few times a year, Never

***Party Scale (1-8): Strong Democrat, Not so strong Democrat, Independent leaning Democrat, Independent
Independent leaning Republican, Not so strong Republican, Strong Republican, None of the above

Player Types

P1: Rich/Low-Risk     P2: Rich/High-Risk     P3: Poor/Low-Risk     P4: Poor/High-Risk
Among heterogeneous risk groups, nearly all group measures impact contributions. To better interpret the significance of these group factors, it is constructive to consider their effects relative to the pooled homogeneous treatment reported in the last column of Table 9. Punishment incentivizes group contributions and adds almost 8 extra tokens in the final collective fund. Increasing levels of mean guilt spur heterogenous group members to contribute more when they are relatively better off than others, however, higher levels of mean envy significantly decrease their contributions in response to many dimensions of inequality. Noted in their relative magnitudes, heterogeneous group contributions are almost five times more influenced by collective guilt (eAIA) than envy (eDIA), qualitatively similar to the individual analyses. With two levels of variety (risk and endowment) in the heterogeneous risk groups, risk aversion no longer proxies as a measure of distrust and higher levels of the group parameter increase contributions. While there were two identically endowed pairs who all shared a common risk factor in homogeneous groups, players in heterogeneous groups were unique in their initial risk/endowment profile and so were not able to form a faux alliance with similarly assigned players, removing trust issues reported in the questionnaire. With four unique players in heterogeneous groups, mean risk aversion impacts contributions in the positive manner anticipated by theory. Among the dispersion parameters, each has an oppositely signed impact on group contributions compared to their influence among pooled homogeneous groups. Increasing disparity in the group envy (eDIA) parameter is beneficial to contributions as group members do not coordinate in pursuing the free riding option. Higher variation among risk aversion decreases group contributions as preferences to avoid risk are no longer aligned.

7 Conclusion

Integrating intrinsic preferences and external pressures in a public goods game, this paper aims to understand the individual and group motivation that stimulate cooperative behavior in a collective-risk social dilemma. Politicians and global think tanks incorporate the influence of risk, uncertainty, and inequality when devising rational expectations and policy suggestions, however, social preferences have improved models of individual and group behavior. While the environmental literature has identified the external factors that impact public good provision, it has failed to control for and test the significance of behavioral preferences that both inhibit and induce cooperative norms. My experiment bridges this gap by constructing a multi-period threshold public goods game that assigns risk and endowment profiles, addresses aversion preferences, and introduces a punishment mechanism to disentangle the internal and external factors that influence cooperation.

Given the multiplicity of equilibria where group members collectively aim for threshold attainment, theoretical predictions of cooperation and behavior are difficult, thereby paving the way for an experiment. Consistent with prior empirical results, contributions in early rounds were optimistically high for all treatments, but cooperation significantly decreased over time in non-punishment groups. The evolution of cooperation was in line with expectations, as homogeneous groups coordinate more than their heterogeneous counterparts. Whereas high-risk groups stabilize near an efficient two token contribution per player-round, low-risk groups systematically coordinate toward a lower contribution average that induces higher rates of threshold non-attainment.
This result suggests that without punishment, coordination issues may be overcome if perceived risk is high enough. The consensual punishment mechanism increased the likelihood of threshold attainment for homogeneous low-risk and heterogeneous groups, but did not have a marked impact on high-risk groups that already had an incentive to cooperate. While perceived risk levels of economic loss are privately derived in the real world, punishment may serve as the “great equalizer” by eliminating total contribution differences between homogenous and heterogenous risk groups. Furthermore, punishment attenuated free riding in early rounds which increased threshold attainment.

Within groups inequality significantly impacted group behavior and final outcomes. Regardless of punishment possibilities, rich players contributed more than the two token symmetric equilibrium and more on average than their poor-player counterparts during active rounds. This result is consistent with participant expectations on social responsibility. Seeking to correct the endowment inequality induced by predetermined play, large early contributions by rich players signal cooperation and propel these groups to successfully reach the target threshold. Rich players contribute more than poor players during active rounds, but the endowment inequality gap is never eliminated even among successful groups. As a critical barrier to cooperation, the inequality gap for equal burden sharing was minimized after introducing the punishment mechanism, increasing the likelihood of threshold attainment.

Beyond the external pressures of risk and inequality, the theoretical impacts of intrinsic preferences are mostly supported by the data. As a player’s relative wealth fluctuates with every round, both advantageous and disadvantageous inequality aversion increase individual contributions in a public goods game where players dynamically incur bouts of guilt and envy. Supported by the follow-up questionnaire, the impact of guilt was 10-20 times more influential than envy as players were more concerned about being at a material advantage than hurting the group account when relatively poor. Risk averse preferences negatively impacted players in a game with both strategic uncertainty and the threat of economic ruin. Controlling for disparate responses between those unequally endowed, however, risk aversion coupled with poor endowments negatively impacts contributions while the conditional effect of risk aversion on rich agents is predictably positive without significance. Risk aversion proxies for distrust among the poorly endowed, while it maintains its traditional role among the richly endowed who fear economic ruin with higher risk factors. Group composition among collectively held intrinsic preferences further impacted cooperation. Higher mean values of advantageous inequality aversion significantly propelled group contributions among all treatments as higher levels of collective guilt aversion induced cooperation. Mean risk aversion continued to have a negative impact on contributions where strategic uncertainty and trust issues may have outweighed the collective aversion to possible economic ruin.

Experimental insights may be extended beyond the environmental realm to general areas of policy interest. Expected utility-maximizers often invest resources to protect themselves from the threat of economic loss, reducing exposure to risk with insurance policies and a diversification of loss-prevention tactics. National security and education are prime examples of collective-risk dilemmas where cooperation is costly to the individual and benefits may only be realized when a common target is reached. Whether the nation seeks a sustainable flow of military recruits or parents sacrifice time and resources to educate their children, a target level of costly cooperation
must be achieved to supply a public good or avoid a common bad. Incorporating external pressures and intrinsic preferences in policy analysis may reveal behavioral motivations that can be exploited to achieve efficient public good provision.

While this paper contributes to the environmental and behavioral literature, a better characterization of the climate change game would allow for further insight and policy recommendations. Gradual impacts from a lack of investment to curb a public bad, in addition to discontinuous catastrophic events, should be modeled when investigating free riding tendencies and cooperation-inducing institutions. Environmental thresholds are also not known with certainty, which could further weaken collective action. Assigning individuals to groups in the experiment ignores the potential barriers to institution formation that could improve the understanding of international environmental agreements. Improvements aside, this study has identified the external factors, internal aversion profiles, and efficiency gains from a consensual punishment mechanism that influence behavior and induce threshold attainment in a collective-risk social dilemma.
References


Appendix

Welcome to the UCSB Experimental & Behavioral Economics Lab!

In today’s experiment, participants will try to **earn as many tokens as possible** in **4 separate games**. Tokens earned will be converted to dollars at the end of the experiment **based on the conversion rate listed in each game**. Of the four games played today, one will be randomly chosen for payment by rolling a die in front of the group at the end of the experiment. **Please try your best in all games!**

**Game 1**

**Please Note:** All games are independent and **do not** affect each other. **Tokens earned in this game are converted at the end of the experiment.**

Please read the rules carefully. If you have a question, please raise your hand.

In this game you are given 100 tokens and asked to choose a portion of this amount (between 0 and 100 tokens, inclusive) that you wish to invest in a risky asset. **Those tokens not invested are yours to keep.**

**The risky investment:** There is a 50% chance that the investment in the risky asset will be successful. If it is successful, you receive 2.5 times the amount you chose to invest; if the investment is unsuccessful, you lose the amount invested.

**How do we determine if the investment is successful?** The roll of a 6-sided die determines the value of the risky asset. You will be asked to choose 3 “success” numbers.

If this game is chosen at the end of the experiment to determine your earnings, **you will roll a die to determine your payoff depending on the “success” numbers you chose.**

Please choose the number of tokens you wish to invest: _____

My 3 success numbers are: __ 1 __ 2 __ 3 __ 4 __ 5 __ 6
Game 2 - Instructions
Please Note: All games are independent and do not affect each other. Tokens earned in this game are converted at the end of the experiment. (2 tokens = $1)

Please read the rules carefully. If you have a question, please raise your hand.

In this game Person A is asked to choose between two possible distributions of tokens for themselves and Person B in 21 different decisions. Person B can only accept Person A’s decisions.

The roles of Person A and Person B will be randomly determined by computer software at the end of the experiment if this game is chosen for final payoffs. Roles will remain anonymous.

Decisions are presented in a chart and will look like the following:

<table>
<thead>
<tr>
<th>Person A’s Payoff</th>
<th>Person B’s Payoff</th>
<th>Decision 1</th>
<th>Person A’s Payoff</th>
<th>Person B’s Payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>Left</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

You make your decision in the role of Person A:
If in this decision you choose Left, you decide to keep 20 tokens for yourself and Person B’s payoff will be 0 tokens. If you choose Right, you and Person B will each earn 5 tokens.

You will choose one distribution (Left or Right) for each of the 21 decisions when the game begins.

If this game is chosen to determine your earnings, the computer will randomly pair you with another participant, assign the roles, and choose one of the 21 decisions. The matching and role assignment remain anonymous. The outcome in the chosen decision determines your earnings.

Note that you will make all decisions as Person A, but the computer might assign you Person B’s role if this game is chosen for payment.

If assigned the role of Person A, you will earn the amount that you have chosen for Person A in the decision selected by the computer and the person paired with you will earn the amount that you have chosen for Person B.

If you are assigned the role of Person B, you will earn the amount that Person A with whom you are paired has chosen for Person B in the decision selected by the computer.
**Game 3 - Instructions**

Please Note: All games are independent and do not affect each other. Tokens earned in this game are converted at the end of the experiment. (2 tokens = $1)

Please read the rules carefully. If you have a question, please raise your hand.

In this game Person A is asked to choose 1 out of 21 possible distributions of tokens for themselves and Person B. Person B knows that A has been asked to make this decision, and may either accept or reject the distribution chosen by A.

If Person B accepts A’s proposal, this payoff choice is implemented. If B rejects the proposal, both receive nothing.

Person B will choose to accept or reject Person A’s proposal for each of the 21 available distributions. For example:

<table>
<thead>
<tr>
<th>Person A’s Payoff</th>
<th>Person B’s Payoff (me)</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>5</td>
<td>Accept</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reject</td>
</tr>
</tbody>
</table>
If this game is chosen to determine earnings, the computer will randomly pair you with another participant after your choices have been made and will assign the roles. Matching and role assignment remain anonymous.

**You will make decisions as if you were Person A and also as if you were Person B.**

If you are assigned the role of Person A, you will earn the payoff you chose for yourself if the Person B you are paired with accepts your offer. Otherwise, both will earn nothing.

If you are assigned the role of Person B, you will earn the payoff that Person A with whom you are paired has chosen for B *only if* you had accepted that particular offer. Otherwise, both will earn nothing.

(Game 3 Screen Shot – Person A’s Choice)
Game 4 - Instructions

Please Note: All games are independent and do not affect each other.
Tokens earned in this game are converted at the end of the experiment. (2 tokens = $1)

Please read the rules of the game carefully. If you have a question, please raise your hand.

In this game participants are randomly split up into groups of 4. You will not know the identities of your group members other than their assigned Player ID Number on your screen. **Your group members and their Player ID Numbers remain constant for the duration of this game.**

This game takes exactly **10 rounds.** At the beginning of each round, each player receives an income of **4 tokens.**

In each round you can invest tokens (like all players) in the effort to **avoid damage** (described on next page). If damage is not avoided, it could lead to participants losing all their tokens at the end of this game.

In each round, all players are simultaneously asked: “**How many tokens would you like to invest in the prevention of damage?**” The possible choices are 0, 2, or 4 tokens.
After deciding how many tokens to invest, the investment of each player is displayed as well as the group total. All tokens invested in the round are credited to the group account for damage avoidance.

At the end of the game (ie after 10 rounds), the computer compares the total invested tokens of all group members with a predetermined amount of **80 tokens**.

**At the end of 10 rounds this amount must be jointly reached to avoid possible damage to each player's individual account.**

This total is achieved when each player in each round invests an average of 2 tokens, for a group total of **8 tokens** per round.

If the necessary total of **80 tokens** is invested into the group account, the group avoided possible damage and each player will retain the points remaining in their individual account at the end of 10 rounds. That is, 40 tokens (you get 4 tokens in each of 10 total rounds) minus what the player has invested in damage avoidance.

If the necessary total of **80 tokens** is **not** invested into the group account by the end of 10 rounds, the group **did not avoid damage** and each player will face their assigned **risk probability** of losing all remaining funds in their individual account for this game.

Players are assigned different risks of losing their remaining individual funds if the necessary token total **is not** met after 10 rounds. Some players face a **1/3 risk** (about 33% chance) of losing remaining individual funds, other players face a **2/3 risk** (about 66% chance) of losing remaining individual funds.

**In this game you are Player # and your assigned risk is (varies by treatment).**

**Example:**

The game in this example includes **4 players**. To avoid damage there must be a total of **80 tokens** invested in the group account by the end of 10 rounds.

**Round 1:** Alice receives an income of 4 tokens and invests 2 of them. All other players make the same decision. After the first round, the account to avoid damages has collected 8 tokens.

**Round 2:** Alice receives an income of 4 tokens and invests all 4 tokens. All other players invest 2 tokens each. After the second round, the account to avoid damages contains 18 tokens.
Rounds 3-10: Alice receives an income of 4 tokens and invests all 4 tokens. All other players invest 2 tokens each. After the tenth round, the account to avoid damages contains 98 tokens.

With a total of 98 tokens in the group account to avoid damage, more than the required 80 tokens, damage is avoided. Each player retains the remaining points in their individual account. Alice retains 2 tokens because she chose to keep only 2 tokens in the first round. Every other round she gave the full amount of tokens possible.

Suppose instead that Alice’s group did not collect the necessary 80 tokens after 10 rounds and that her individual risk is \(1/3\). Alice still contributed the same amount as previously stated.

If this game is chosen for payment, she faces a \(1/3\) chance of losing her remaining 2 tokens and a \(2/3\) chance of keeping them. Other players face their own risk to determine their own payment.

A die will be rolled to determine each individual’s outcome.
(End of Example)

Please note the following two important features in this game:

First, the computer predetermines the decisions of each player in the first 3 rounds (predetermined play). This means that you and your fellow players cannot decide freely how much you want to invest in the first three rounds. Instead, the computer forces you to choose a particular option in the first three rounds. Starting in round 4, you decide freely the amounts you want to invest.

(The following is ONLY for Punishment Treatments)

Second, after all 10 rounds are complete, we proceed to Stage 2 where you may choose to reduce the earnings of other group members who had the same opportunity to invest in avoiding damage.

You may reduce or leave equal the earnings of each of the other group members. Conversely, other group members can lower your earnings as well. Your decision is about distributing “points” to the other three players if you want to reduce their earnings. Though identities are not revealed, you do know each player’s history of investment to the group task of avoiding damage, individual risks, and predetermined investments the computer chose at the beginning. This information will be presented on your screen.

If you want to reduce a player’s earnings (for example Player X), you can request to distribute a number of points from 0 to 7 to Player X. If you do not want to change their earnings, you will choose 0 points for them. Each point Player X receives
reduces their earnings by 3 (THREE) tokens. For the player who gave the point, each point given costs 1 token out of their own account.

Your overall cost is equal to the sum of the points that you have distributed to each one of the other three group members. Your maximum cost possible for distributed points is then 21 tokens (7 tokens times 3 persons). Your cost is zero if you do not distribute points to anybody.

As will be explained, a request to distribute points is not always carried out. For each player, there are two cases:

1. When **only you** have requested to distribute points to Player X, your decision has no effect. In particular, there is no reduction in their earnings and you are not assessed any cost for your request since it was not carried out.
2. When both you and at least one other group member have requested to distribute points to Player X, then your decision to distribute points is carried out. Requests made by others to distribute points to Player X will also be carried out. Thus, there **must be at least two requests** to distribute points to Player X in order for the request to be carried out. **It does not matter if requests made by multiple players are for different amounts.**

At the end of Stage 2 your total earnings will be:

\[
\text{Round 1-10 remaining earnings} - \text{Stage 2 earnings reduction} - \text{cost to distribute points} \\
= \text{Round 1-10 remaining earnings} - (\text{total points received}) \times 3 - (\text{total points successfully distributed}) 
\]

After all decisions are complete, we reveal points successfully distributed and cumulative earnings after the reduction.

**Negative earnings are not possible.** If a player has their earnings collectively reduced by more than they have remaining in their account, they simply earn zero tokens for this game.

After Stage 2 reductions, if 80 tokens were not collected in Rounds 1-10, then we proceed to let you roll the dice to determine if damage hurts you based on your assigned risk. Otherwise, the game is over.
Control Questions

1. How many total tokens would each player have to invest into the group account, on average, to avoid possible damage after 10 rounds?
2. How many rounds can players freely decide their investments into the group account?
3. If after the first 6 rounds your group collected 28 tokens, how many additional tokens would your group need to collect to avoid possible damage after 10 rounds?
4. Assume your assigned risk is 1/3 (about 33% chance) and your group does not collect 80 tokens after 10 rounds. This means that you ______ (1=keep, 2=lose) your remaining funds with a 1/3 chance.
5. (Punishment treatment) What is the maximum number of points you can assign to any ONE player in Stage 2?
6. (Punishment treatment) If you give 2 points to Player X and Player Y gives 1 point to Player X, how many tokens does it cost you?
7. (Punishment treatment) If you give 2 points to Player X and no one else gives points to Player X, how many tokens does it cost you?
8. (Punishment treatment) In Stage 2, if Player X gives you 2 points and Player Y gives you 1 point, how many tokens are taken from your account?