Environmental and Resource Economics (2022) 81:685–710
https://doi.org/10.1007/s10640-021-00645-y

The Role of Non-Binding Pledges in Social Dilemmas with Mitigation and Adaptation

David M. McEvoy[1] · **Tobias Haller**[2] · **Esther Blanco**[2]

Accepted: 7 December 2021 / Published online: 11 January 2022

Abstract

This study presents experimental results on the role that non-binding pledges have on the ability of groups to manage a threat of probabilistic group damages in two separate environments. We focus on an environment where in addition to collective mitigation, agents can work autonomously to protect themselves from the damages if they occur (adaptation). The tension is that mitigation and adaptation investments are strategic substitutes. We test the hypothesis that non-binding pledges are more effective in a world with both mitigation and adaptation strategies, compared to mitigation only. First-period results show that (i) consistent with previous literature, pledges in a mitigation-only environment do not increase average investments in collective mitigation, but (ii) when both mitigation and adaptation opportunities exist, pledges lead to higher investment in collective mitigation, lower investment in adaptation and increased efficiency. Although the average treatment effect disappears over time as the amount pledged decreases, pledges remain significant predictors of mitigation investments over the course of the experiment.

Keywords Social dilemmas · Economic experiments · Behavioral economics · Public goods · Mitigation · Adaptation · Environmental damages

JEL Classification D9 · Q54 · H4 · C92

1 Introduction

Consider a situation in which a group of agents face the threat of environmental and natural resources damages, and the parties involved have two broad strategies they can use to manage the threat. They can work collectively to try to mitigate the root cause of the damage, or they can work autonomously to try and protect themselves from the damages if they occur. The most prominent example is climate change. We could all work cooperatively to jointly mitigate our emissions of greenhouse gases (a type of public good). We could also

---

✉ David M. McEvoy
mcevoydm@appstate.edu

[1] Department of Economics, Appalachian State University, Boone, NC 28608, USA

[2] Department of Public Finance, University of Innsbruck, Universitätsstrasse 15/4, 6020 Innsbruck, Austria

take adaptive measures like installing air conditioning (a type of private good) to protect ourselves from the impending damages from a changing climate. Local pollution, flooding risk or wildfire risk serve as other examples.[1] The collective mitigation strategy provides public benefits and is more cost effective but is susceptible to free-riding behavior. In most applications, at least some adaptation is available for people to privately protect oneself. The individual adaptive strategy is a private good and therefore void of free-riding concerns, but is inefficient in the sense that it only benefits the party taking the action. The tension between the two is that while the optimal approach is the collective solution, free-riding incentives become more pronounced the greater the investment in private protection. Using controlled lab experiments we explore the capacity of pledges to enhance cooperation in a collective-risk social dilemma in which agents have a combination of public (mitigation) and private (adaptation) management strategies.

To the best of our knowledge, this is the first experimental study exploring the interaction between pledges and mitigation-adaptation investment strategies. We begin with a simple model that helps form our main hypothesis that non-binding pledges are more effective in a world with both mitigation and adaptation opportunities as compared to previously studied settings where only mitigation is feasible. We test this hypothesis using controlled experiments in a 2×2 experimental design, in which the two treatment variables are whether communication is allowed in the form of non-binding pledges and whether there are opportunities for adaptation (in addition to mitigation) against the impending damages.

Investments in mitigation provide a public good by incrementally decreasing the probability that the group suffers environmental damages. Investments in adaptation reduce the magnitude of the damage an individual faces if the damage occurs. The implication is that investments in adaptation reduce the expected return on mitigation, and likewise investments in mitigation reduce the expected return on adaptation. In this mitigation-adaptation environment the optimal solution for the group is to invest everything into mitigation to minimize the probability of the damage occurring. However, all individuals invest zero of their resources in mitigation in a non-cooperative Nash equilibrium given self-interested payoff-maximizing preferences. Decades of behavioral research inform us that many players do contribute to public goods and we should expect investments in mitigation to be positive.

Our research is novel in that it explores the role of structured communication in the form of non-binding pledges in the mitigation and adaptation environment. In a simple world of completely self-interested players, non-binding pledges are a specific form of "cheap talk" and therefore have no bearing on theoretical predictions of individual decision making and equilibrium outcomes. Behavioral research, however, has demonstrated that some forms of non-binding communication can have significant impacts on decisions in social dilemma situations. Many previous experimental papers have examined the effectiveness of non-binding communication strategies to facilitate cooperation in public-good games (e.g., Isaac and Walker, 1988; Palfrey and Rosenthal, 1991; Sally, 1995; Bochet et al., 2006; Bochet and Putterman, 2009; Pogrebna et al., 2011; Koukoumelis et al., 2012;

---

[1] Individual actors can coordinate their activities to reduce the causes of pollution (e.g., reducing emissions of particulate matter) and they can also take measures to protect themselves from the pollution itself (e.g., avoiding physical activity when pollution levels are high). Similarly, constructing dikes in case of flooding risk (mitigation) would reduce the risk for a collective while home protection would only benefit the individual (adaptation). Or forest management to reduce wildfire risk benefits a wider collective (mitigation) while property protection benefits only oneself (adaptation).

Oprea et al., 2014). Ostrom et al. (1994), Sally (1995) and Bochet et al. (2006) find that pre-play, free-form communication was the most important factor helping groups resolve social dilemmas.[2]

That said, not all non-binding communication has proven effective. When communication is structured so that individuals cannot communicate freely but are instead restricted to announcing their intended numeric contributions (referred to as "numeric cheap talk") pre-play communication is largely ineffective on average (Chen and Komorita, 1994; Wilson and Sell, 1997; Bochet et al., 2006; Bochet and Putterman, 2009). Numeric cheap talk was effective only when it was coupled with the ability to sanction noncooperative behavior (Bochet and Putterman, 2009). The pledge environment we consider is a form of numeric cheap talk. Users make non-binding *pledges* regarding how many tokens they plan to invest into mitigation. While we do not consider formal punishment strategies, like others (e.g., Bochet et al., 2006; Bochet and Putterman, 2009), we do reveal each individual's pledge and contribution amount to the other group members by their unique subject ID. Therefore, group members can perfectly match decisions with anonymous individuals, which opens the door to reputation building and to informal reciprocal strategies in following periods.[3]

Our research contributes to different branches of the economics literature. First, we contribute to the literature cited in the previous paragraph addressing the potential of communication to increase cooperation in social dilemmas (an early survey of this literature is provided by Crawford, 1998). Second, we contribute to the growing experimental literature on "collective risk social dilemmas", in which groups of players can cooperate to avoid impending damages (Milinski et al., 2008; Tavoni et al., 2011; Barrett and Dannenberg, 2012, 2014, 2016; Blanco et al., 2017, 2020). In particular, a paper by Barrett and Dannenberg (2016) uses a set of experiments to shed light on the "pledge and review" mechanism that the authors motivate by being part of the Paris Agreement on climate change. Their overall finding is that non-binding pledges, even with peer review, are ineffective. While their main result casts doubt on the ability of structured non-binding communication to help resolve social dilemmas, the experimental design makes it difficult to isolate the individual effect of each policy component.[4] Our study complements this literature by using an experimental design that allows us to identify causal effects of non-binding pledges both

---

[2] Although the decision spaces differ, the pledges made in our experiments have commonalities with "oaths" made in previous studies (e.g., Jacquemet et al. 2017, 2019). In experiments using oaths, the subjects have the opportunity to sign a solemn oath stating that they will provide truthful answers. Subjects in our experiments, in contrast, are not told that they are expected to be truthful with their pledges. Indeed, the instructions include the following statements: "Note that this pledge is not binding. That means, you will not be required to invest the same number of tokens you pledged to the group insurance. You can invest more or less than your pledged amount—it is up to you.".

[3] Reciprocal strategies include "conditional cooperation" in which players increase contributions to a public good in response to others increasing their contributions. Fischbacher et al. (2001) was the seminal experimental study exploring conditional cooperation in a public-goods game, and a recent review of the literature shows the results from the original study are highly replicable (Thöni and Volk 2018).

[4] Barrett and Dannenberg (2016) consider a game in which players can make investments in a public account in order to avoid an impending loss. If the group's contributions exceed a threshold, then the probability of the loss occurring decreases, and eventually turns to zero with full contributions. In their experiment, the threshold is endogenously determined in a first stage. In the next stage players submit their intended contributions (i.e., pledges). They explore multiple "review" treatments in which group members award grades to other member's intended contributions or combinations of intended and actual contributions. With this design the researchers cannot disentangle the effect of setting an endogenous threshold with the role of non-binding communication (with reviews) in a social dilemma.

in a mitigation game and in a game with mitigation and adaptation, isolating the effect of pledges from much of the complexities existing in a natural setting.

Our research also adds to the broader experimental literature on the features of social dilemmas that influence cooperation (see Ledyard, 1995; Ostrom, 2006; Chaudhuri, 2011) by considering the interaction of collective and private protection options as strategic substitutes. The substantial body of work on public-good experiments demonstrates that people are more cooperative than a theory of materially self-interested players suggests. There is a thin but growing empirical literature on the interaction of mitigation and adaptation investments in social dilemma games (e.g., Hasson et al., 2010; Blanco et al., 2020). Hasson et al. (2010) run a simple experiment in which subjects can make a discrete investment in either mitigation or adaptation to avoid a public bad, but not both. Their two treatments vary how vulnerable subjects are to the impending harmful event. They find no significant treatment effects and observe that about 25 percent of subjects choose to mitigate over adapt. Blanco et al. (2020) focus attention on the role of heterogeneity in vulnerability to damages in social dilemmas with continuous mitigation (called public insurance) and adaptation (called private insurance) investments. They find that investments in public insurance are lower for those who face lower levels of potential damage. Moreover, subjects use public and private insurance investments as substitutes; that is, they contribute more to private insurance when they expect lower aggregate investments in public insurance. In contrast to the existing literature, our paper focuses on a world with symmetric agents and our design allows us to compare the mitigation-adaptation environment to a mitigation-only environment with and without non-binding communication.

We start our analysis by concentrating on first-period decisions. Pledges increase investments in mitigation, but the effect is significant only when both mitigation and adaptation investments are possible. In contrast, and consistent with the literature, pledges do not change average investments in the standard mitigation-only linear public-good environment. Our panel-data analyses reveal that average treatment effects are fragile but at the same time subjects are very responsive to the pledges made by their group-members. That is, while the effects of non-binding communication wash out on average, it is clear that subjects do not treat the pledges as pure noise and reciprocate by investing more in mitigation when pledges are higher.

In the next section we introduce the mitigation and adaptation game. We then discuss the experimental design, parameter choices and formalize our testable hypothesis. The results section follows, beginning with first-period analysis and then moving into an analysis of the full panel. Finally, we conclude.

2 The Mitigation-Adaptation Game

The baseline mitigation-adaptation game that we consider builds from Blanco et al. (2020). Consider a world of n identical players that each face the possibility of suffering a loss from an impending natural disaster (i.e., a public bad). To manage the expected damage, different investment options are available in which subjects can invest their initial endowment. They can invest in mitigation and in addition, they have the option to invest in private adaptive measures. Investment in adaptation reduces the size of the damage the investor will suffer from the disaster if it occurs. Therefore, adaptation yields private benefits in the form of a reduction in damages. Modeling adaptation as a private good is standard in the literature (Shibata and Winrich, 1983; Kane and Shogren, 2000; Zehaie, 2009; Buob and Stephen, 2011; Marrouch and Chaudhuri, 2011, 2011; Bayramoglu et al., 2018; Lazkano et al. 2016). This is in contrast to mitigation efforts which are pure public goods. The mitigation game is a reduced form of the mitigation-adaptation game whereby only mitigation investments are possible (see 'Appendix'). Player i's expected payoff function for the mitigation and adaptation game is:

$$\pi_i = e_i - m_i - a_i - (1 - \beta M)\big(D - \gamma a_i\big), \tag{1}$$

where e_i is player i's endowment. The variable m_i is player i's choice of how much of her endowment to spend on mitigation where $m_i \leq e_i$ and $M = \Sigma m_i$. The term D is the size of the impending damage from the loss, affecting all group members, and the term $(1 - \beta M)$ is the probability the loss occurs, where β is a positive constant. Without any mitigation efforts (i.e., $M = 0$), each player faces certain damages of D from the group loss. The probability the loss occurs decreases by β for all players for each unit invested in mitigation by anyone. Mitigation, therefore, is a pure public good. The term a_i is player i's investment in adaptation, and the term $\gamma > 0$ captures the marginal effectiveness of adaptation efforts on the reduction of damages for player i.

Throughout we assume that $1 \geq \beta > 0$ and $D > e_i$. We also restrict $(1 - \beta n e_i) \leq 0$ which ensures that players cannot reduce damages beyond zero. For convenience, from here on we will assume the strict equality holds; that is, $(1 - \beta n e_i) = 0$ so that if all n players invest their entire endowments in mitigation, then the probability the loss occurs is driven to zero and they completely avoid the damages. We add the restriction that $(1 - \gamma e_i \geq 0)$ which implies that a single player cannot reduce damages to zero by investing completely in adaptation. Note also that a player's budget constraint implies $m_i + a_i \leq e_i$. By differentiating the payoff function in (1) with respect to both mitigation and adaptation we get the following:

$$\frac{d\pi_i}{dm_i} = \beta\big(D - \gamma a_i\big) - 1, \tag{2}$$

$$\frac{d\pi_i}{da_i} = \gamma(1 - \beta M) - 1. \tag{3}$$

Equation (2) is decreasing in a_i, which illustrates that investment in adaptation reduces the marginal payoff of mitigation efforts. In addition, Eq. (3) is decreasing in M, as investments in mitigation reduce the marginal payoff of adaptation investments. Therefore, mitigation and adaptation investments are substitutes in our model. This relationship is intuitive and ubiquitous in the literature modeling mitigation and adaptation with coalition formation (Bayramoglu et al., 2018; Lazkano et al. 2016; Borrero and Rubio 2021) and without coalitions (Ingham et al., 2007; Zehaie, 2009; Buob and Stephen, 2011; Ebert and Welsch, 2012; Blanco et al., 2020). It is important to note that we chose linear models for the mitigation and adaptation games because they are easily tested in the laboratory (e.g., Blanco et al., 2020). This is a departure from some of the established theoretical literature on mitigation-adaptation games, in which adaptation expenditures directly alter the effective damage functions. In particular, Ebert and Welsch (2012) show that emissions can be either strategic substitutes or complements depending on the specification of the damage function. The implication is that the relationships in Eqs. (2) and (3) are not common to all models of the interaction between mitigation and adaptation.

Since the marginal benefit from mitigation in our model is negative when $a_i = 0$ and decreasing when $a_i > 0$, purely self-interested players maximize their payoffs by investing zero in mitigation with or without adaptation investment. When $M=0$, the right-hand side of Eq. (3) reduces to γ—*1*. Therefore, a payoff-maximizing player will either invest her entire endowment in adaptation if $\gamma > 1$ or invest zero in adaptation if $\gamma < 1$. We will concentrate on the interesting case in which $\gamma > 1$ and so in a Nash equilibrium, $m_i = 0$ and $a_i = e_i$ for all players. A player's equilibrium payoff with adaptation possibilities is $-(D - \gamma e_i)$ which is greater than the equilibrium payoff a player receives without the option to adapt $(e_i - D)$ by the amount $e_i(\gamma - 1) > 0$. The social optimum in the mitigation-adaptation game is achieved when all n players invest fully in mitigation and drive the probability of the damages occurring to zero.

Previous literature on diverse social dilemmas has shown that subjects make decisions that are consistent with complex and diverse motivations beyond simple self-interested profit maximization (e.g., Camerer, 2003; Camerer and Fehr, 2006; Chaudhuri, 2011; Ostrom and Walker, 2003). As in Blanco et al (2020), it is sufficient to incorporate the empirically supported assumption that some individuals derive utility from acting cooperatively (Chaudhuri, 2011) to expect positive mitigation investments. More complex, and perhaps more realistic, utility functions are feasible but not necessary to illustrate this point. In its simplest form, we can extend the payoff function in Eq. (1) into the utility function in Eq. (4a) below, which includes an additional component $f(m_\mathrm{i})$ that captures a player's utility gained by acting cooperatively through investing in mitigation.

$$u_i = \pi_i + f(m_i) \tag{4a}$$

In Eq. (4a), $f'(m_\mathrm{i}) > 0$ and $f''(m_\mathrm{i}) < 0$ which implies that additional utility is gained from contributions to mitigation with decreasing returns, so that the marginal increase in utility from cooperating decreases for higher levels of cooperation by subject *i*. The associated marginal return in utility from mitigation is:

$$\frac{du_i}{dm_i} = \beta(D - \gamma a_i) - 1 + f'(m_i). \tag{5a}$$

Note that the marginal return for investments in adaptation (Eq. 3) remain unchanged. As a result, for a high enough $f(m_\mathrm{i})$, an equilibrium can include positive contributions to mitigation. Prior evidence shows that people desire to be cooperative (Brandts et al., 2004; Goeree et al., 2002; Henrich et al., 2001) and we expect positive contributions to the public good, namely mitigation.

2.1 Mitigation-Adaptation with Pledges

When pledges are available, the players have the opportunity to communicate their intended investment in mitigation before making their binding investment decisions. The game now involves two stages. In the first, each player gets to announce their intended contribution to mitigation. Each player receives all other players' announcements (called "pledges") before making their mitigation and adaptation decisions in a second stage. The pledges are non-binding and therefore have no effect on equilibrium decision making given payoff-maximizing self-interested players.

Our main hypothesis concerns the relative effectiveness of non-binding pledges when players can either invest only in mitigation or both mitigation and adaptation. This is an exploratory hypothesis, based in part on previous literature (e.g., Bochet et al., 2006; Bochet and Putterman, 2009) which suggests that players do not treat pledges as pure noise.

While the addition of non-binding pledges to invest in mitigation has no effect given payoff-maximizing players all with common knowledge, such pledges may influence decisions if some players have non-standard preferences—having objectives that are not limited to just earning money for themselves. Following Bochet and Putterman (2009), non-binding numeric pledges may increase contributions to a public good if some players are "reciprocal" (or conditional cooperators) and if some players have a preference to tell the truth (referred to as "truth-telling" by the authors). Indeed, it is sufficient that subjects believe that reciprocators and truth-tellers are prevalent for non-binding pledges to not be viewed as meaningless cheap talk. A player with a preference for truth-telling may gain (lose) utility from keeping (breaking) their word. Charness and Dufwenberg (2006) find evidence that players try to live up to others' expectations by telling the truth to avoid experiencing guilt.

Suppose that players, to some extent, trust that others will comply with their pledges regarding mitigation investments. Recall that mitigation and adaptation investments are strategic substitutes, where increases in group mitigation reduces the expected return on investing in private adaptation. For positive amounts being pledged, nonbinding pledges could increase others' expectations of group investment in mitigation (given the belief that some players are truth-tellers). For positive reciprocators, this would in turn increase investments in mitigation (at least weakly) which reduces the expected marginal return to adaptation. This, as a result, would reduce one's own marginal incentives to invest in adaption. And with a decrease in investment in adaptation, the expected marginal return to mitigation increases. This can generate a virtuous cycle where pledges, expectations and behavior reinforce one another in the presence of pledges in mitigation-adaptation environments, fostering more cooperative behavior with higher mitigation investment. These conjectures lead to our main hypothesis, which we test using the experimental evidence presented in Sect. 4.

Hypothesis: *When players have the opportunity to make investments in both mitigation and adaptation, non-binding pledges have a stronger positive effect on average mitigation investment compared to when only mitigation investments are feasible.*

3 Experimental Design

In this section we describe the treatments and parameter choices. For parameters, we fixed group size at four ($n=4$), endowments at 10 ($e=10$), impending damages at 20 ($D=20$), $\beta=0.025$ and $\gamma=1.5$. Earnings are reported in "tokens" which were converted into Euros at an exchange rate of 1 at the end of the experiment. To avoid the potential of subjects making negative earnings, we provided each player with a "savings" account of 25 tokens that could not be used for decision making. In total, there were four treatments: mitigation (*M*), mitigation-adaptation (*M&A*), mitigation with pledges (*MwP*) and mitigation-adaptation with pledges (*M&AwP*). Given the parameter choices and Eqs. (2) and (3), a token invested in mitigation (called a "group insurance" in the experiment) reduces the probability of damages occurring for everyone by 2.5%, and each token invested in adaptation (called a "private insurance" in the instructions) reduces the size of the damage by 1.5 tokens for the investor alone.

In *M&A* and *M&AwP* treatments, in the first stage players make a "pledge" regarding how many tokens they will invest in mitigation in the second stage (from zero to 10). The

Table 1 Average investments, pledges and expectations in initial period

	Investment in mitigation	Investment in adaptation	*Pledged* investment in mitigation	*Expected* investment in mitigation by others	*Expected – actual* mitigation by others
M	6.69 (0.420)	–	–	21.83 (0.936)	1.77 (1.235)
M&A	5.44 (0.495)	2.77 (0.486)	–	18.65 (0.882)	2.33 (1.261)
MwP	7.42 (0.337)	–	7.96 (0.316)	23.17 (0.708)	0.92 (0.797)
M&AwP	7.31 (0.443)	1.40 (0.358)	8.00 (0.352)	21.35 (0.838)	−0.58 (0.822)

Each cell contains averages ($n=48$) and standard errors are in parentheses

pledges are revealed to the other group members at the conclusion of the first stage (by randomly assigned subject ID that remains the same throughout the experiment). Therefore, before making investment decisions players will know how many tokens each other player pledged to invest in mitigation alone. We also provide a calculation of the total number of tokens pledged by the group. The players are informed that their pledges are not binding and they will have the chance to invest more or less than their pledged amount (within the zero to 10 token range).

The experiments were conducted at the University of Innsbruck. For each of the four treatments, we ran two sessions of 24 subjects for a total of 48 individuals and 12 groups per treatment. The groups remained fixed during 10 repeated periods of the same game leading to a panel of 480 individual-level observations and 120 group-level observations per treatment. The experiment was programmed in zTree (Fischbacher, 2007) and the instructions were provided on paper and read aloud by the moderator (see Supplementary Materials for instructions). Subjects had to answer a series of control questions testing their understanding of how earnings are determined before moving forward with the decision periods. On average, the experiment lasted under one hour and students earned 18 Euros.

In all four treatments, when making their investment decisions each player also provide an estimate of how many tokens they believe the other group members will invest in group insurance. Following the experimental literature, we incentivize the belief elicitation decision (e.g., Croson, 2007; Neugebauer et al., 2009; Fischbacher and Gächter, 2010; Smith, 2013). A player earns an additional two tokens if their estimate is within two tokens of the actual aggregate investment in the group account by the other three players.

4 Results

4.1 First Period Results

We begin with an analysis of decisions made in the first period, which allows for simple statistical tests because each individual-level and group-level observation is independent. Table 1 contains the average individual investments in mitigation and adaptation for each of the treatments. First, a comparison between average mitigation levels in the mitigation-only treatment (*M*) and the mitigation and adaptation treatment (*M&A*) shows that when subjects have an opportunity to adapt it causes a reduction in mitigation. Subjects on average invest about two thirds of their endowment in mitigation (6.69 tokens) in the baseline mitigation-only treatment, and invest about half of their endowment (5.44 tokens) in the mitigation and adaptation treatment. A pairwise t-test of the means reveals that when subjects have the opportunity to protect themselves from damages through adaptation, they significantly reduce investments in mitigation (6.69 vs. 5.44, $p<0.01$).

When pledges are introduced to the mitigation-only treatment (*MwP*), average investment in mitigation increases (from 6.69 to 7.42 tokens) but the difference is not significant ($p = 0.179$). Therefore, our findings from the first-period data in the mitigation-only treatment are consistent with other studies using numeric cheap talk (e.g., Bochet et al., 2006; Bochet and Putterman, 2009) in which non-binding pledges have an insignificant effect on mitigation investment.

Pledges, however, have a significant positive effect on investments in mitigation when subjects have the opportunity to invest in adaptation as well. The average investment in mitigation increases from 5.44 to 7.31 tokens when subjects make pledges in the mitigation and adaptation treatment (*M&A* vs. *M&AwP*, $p<0.01$). Therefore, we find evidence in support of our main hypothesis. In other words, in a social dilemma situation in which resource users can mitigate and adapt to impending damages, the use of pledges leads to higher provision of a public good and, in turn, efficiency.

We now turn to investments in adaptation. Recall that if all players were purely materially self-interested then they would invest all 10 tokens in adaptation. Our data clearly do not support the predictions from a model of purely self-interested agents. In contrast, investments in adaptation are quite low. In our treatment without pledges (*M&A*), subjects invest an average of 2.77 tokens to adaptation. When pledges are introduced (*M&AwP*), investments in adaptation drop significantly (from 2.77 to 1.40, $p=0.025$). In short, we find that in a mitigation-adaptation environment, non-binding pledges cause a significant decrease in investment in adaptation and a significant increase in investment in mitigation.

The fourth column in Table 1 shows average expectations of the others group members' investments in mitigation (the variable ranging from 0 to 30). When comparing *M* with *M&A*, the expected investments in mitigation go down (from 21.83 to 18.65, $p=0.015$) when subjects also have the opportunity to adapt. When pledges are introduced, they have no significant effect on expectations in the mitigation-only treatments (from 21.83 to 23.17, $p=0.259$). However, pledges do have a significant positive effect on expectations in mitigation when subjects can adapt (from 18.65 to 21.35, $p=0.028$). Recall, that our main hypothesis was developed under the premise that pledges would increase expectations about others' investments in mitigation, that in turn make expected individual investments in adaptation less lucrative. The results support our hypothesis, and we find that expected and actual mitigation increase when players have opportunities to invest in both mitigation and adaptation to avoid damages.

While pledges increase expected mitigation in the mitigation-adaptation environment, it is possible that pledges decrease the accuracy of expectations. The final column in Table 1 illustrates that this is not the case. Pledges reduce the gap between expected and actual mitigation investment in both environments, and therefore improve the accuracy of expectations. With mitigation only, players remain overly optimistic, but the gap decreases by $(1.77–0.92)=0.85$, which is insignificant ($p=0.563$). However, the change in expected-actual mitigation in the mitigation-adaptation environment is $(2.33+0.58)=2.92$, which is weakly significant ($p=0.056$). With mitigation and adaptation, pledges make players slightly pessimistic in which, on average, they expect less mitigation than is actually observed.

Table 2 Average investments, pledges and expectations pooled over the 10-period experiment

	Investment in mitigation	Investment in adaptation	*Pledged* investment in mitigation	*Expected* investment in mitigation by others	*Expected – actual* mitigation by others
M	5.74 (0.143)	–	–	19.00 (0.293)	1.80 (0.259)
M&A	4.49 (0.175)	4.11 (0.191)	–	14.76 (0.368)	1.31 (0.277)
MwP	6.36 (0.124)	–	7.16 (0.106)	20.38 (0.249)	1.31 (0.210)
M&AwP	5.40 (0.187)	3.51 (0.192)	6.89 (0.156)	17.74 (0.418)	1.54 (0.285)

Each cell contains averages ($n=480$) and standard errors are in parentheses

To conclude the first-period analysis we examine the level of compliance with individual pledges. A compliant subject is one who actually invests at least as much as they pledged to invest in mitigation. In *MwP*, 75% of subjects were compliant and in *M&AwP* 81.25% of subjects were compliant. The difference in compliance rates between treatments is not significant ($p=0.459$). Perhaps more surprising than the high binary compliance measure, we find that, on average, groups invest as much in mitigation as they pledged to. This holds for both treatments *MwP* and *M&AwP* ($p=0.241$ and $p=0.206$, respectively).

In summary we find that in the first period, non-binding pledges significantly increase expected and actual investments in mitigation but only when subjects had opportunities to invest in both mitigation and adaptation. Moreover, while pledges were non-binding, over 3/4 of subjects were either compliant or over-compliant in both treatments. Finally, on average, groups invest as much as they pledged to in the first period of decision making. The first-period analysis suggests that the average participant is truthful with their pledge, and believes others will be as well. In a world with mitigation and adaptation, pledges increase provision of the public good and bring groups closer to the social optimum.

4.2 Panel-Data Summary Statistics

In this section we analyze the full panel of decisions made over the 10-period experiment. We start by reporting the same summary statistics shown in Table 1 for the pooled dataset. These statistics can be found in Table 2. When comparing the two summary statistics tables it is clear that over the 10 periods average investments, pledges and expectations about others' mitigation decrease while investments in adaptation increase relative to the first period.

From Table 2, the inclusion of pledges increased average expected mitigation by 20.38–19.00 = 1.38 in the mitigation only environment ($p=0.000$) and by 17.74–14.76 = 2.98 in the mitigation-adaptation environment ($p=0.000$). While both effects are positive and significant, we see a more pronounced positive effect of pledges when adaptation possibilities are present. Moreover, the accuracy of expectations (final column in Table 2) increased due to pledges in the mitigation environment (1.80 – 1.31 = 0.49) while the accuracy decreased due to pledges with adaptation (1.31–1.54 = –0.23), but both effects are insignificant ($p=0.1440$ and $p=0.564$, respectively).

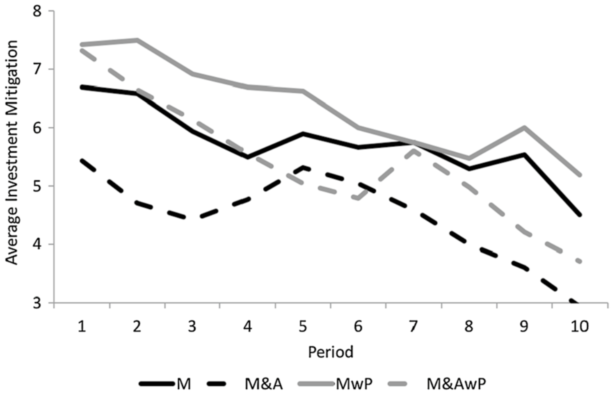


Fig. 1 Average investment in mitigation by treatment over 10 periods

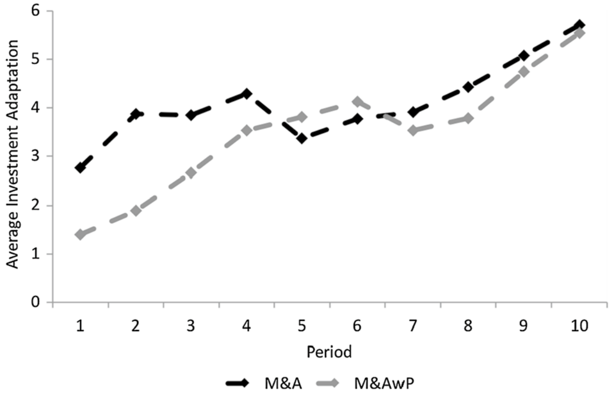


Fig. 2 Average investment in adaptation by treatment over 10 periods

Figure 1 illustrates the average investment made to mitigation for each treatment over the 10 periods. The solid lines are the two treatments with only mitigation investments, the black line is *M* (without pledges) and the grey line is *MwP* (with pledges). Treatment *M* is the closest to the traditional linear public-goods game in the experimental literature. Similar to what others have documented (see Ledyard, 1995), we find that investments in mitigation start at roughly 66% of endowments and decay over the course of the experiment (ending at about 45%). When subjects could make pledges in the mitigation game (*MwP*), investments in mitigation are higher but the gap between the average investments in *M* and *MwP* closes toward the end of the experiment.

The dashed lines are from the two treatments with adaptation investment opportunities; the dashed black line is *M&A* (without pledges) and the dashed grey line is *M&AwP* (with pledges). From a quick visual comparison of the two treatments, pledges appear to have a more pronounced initial effect in the mitigation-adaptation environment compared to the mitigation-only environment but this too dissipates toward the middle of the experiment, only to rebound slightly at the end. Figure 2 traces the average investment made in adaptation over the 10 periods. Again, the dashed black line is from the treatment without pledges (*M&A*) and the dashed grey line is from the treatment with pledges (*M&AwP*). Pledges appear to reduce investment in adaptation in the early periods but the effect vanishes with repeated play.

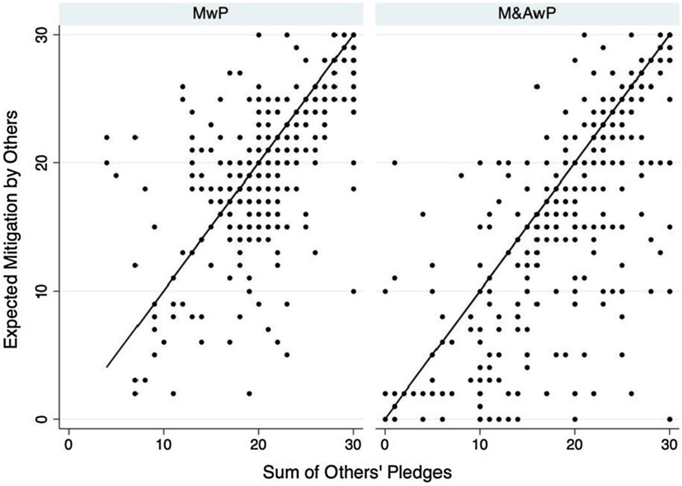


Fig. 3 Relationship between expected mitigation and others' pledges

Figure 3 illustrates the relationship between players' expectations about how much the others will invest in mitigation (y-axis), relative to the amount the other players pledged to invest (x-axis). The relationship is positive in both pledge treatments (*MwP* and *M&AwP*). Comparing the lower portions of the two graphs, it is clear that there are more observations with lower expectations in the treatment with adaptation possibilities. From Fig. 3, it also appears that the correlation between others' pledges and expectations becomes stronger as the amount pledged increases.

4.3 Panel-Data Regression Analysis

We now turn to panel regression models to estimate treatment effects when controlling for repeated group and individual decision making. To begin, we look more carefully at the relationship between expectations and pledges we observe in Figure 3. Using a random effects model, we regress expected investment in mitigation on a dummy variable for treatment (*MwP* is the reference treatment), sum of others' pledges and the interaction between treatment and pledged amounts, while controlling for period effects and clustering the standard errors at the group level. The results in Table 3 confirm that expectations are lower in the environment with adaptation, expectations are positively correlated with others' pledges, and the effect of others' pledges on expectations is statistically equivalent between decision environments with and without adaptation.[5]

Next, we consider investment in mitigation. In the random effects models in Table 4, the dependent variable is investment in mitigation, which is regressed on treatment dummies (*M* is the reference treatment) while controlling for period and clustering standard errors by group. We also add a metric capturing a subject's degree of loss aversion using

---

[5] As a robustness check, we estimated all of our regression models controlling for a player's score on the comprehension quiz. The control was insignificant in all models and did not impact any of the qualitative results.

Table 3 The impact of treatment and pledges on the formation of expectations

M&AwP	−4.465*** (3.582)
Others' pledges	0.640*** (0.110)
M&AwP × Others' pledges	0.114 (0.121)
Period	−0.303*** (0.087)
Constant	8.288*** (2.700)
n	960
$\chi 2$	136.06***

Omitted is the treatment *MwP*. standard errors in parentheses, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

the instrument from Gächter et al. (2007). The variable ranges from zero to six, with higher values indicating a greater degree of loss aversion (i.e., rejecting more gambles to avoid potential losses). We then add additional covariates in order to get a better picture of the role expectations, interactions of expectations with treatments, and previous-period outcomes play in decision making. For ease of comparison with results from the previous section, the first two models in Table 4 include first-period decision making only.

From Model 1 in Table 4 we see that mitigation investments significantly decrease in the *M&A* treatment relative to the *M* treatment. This is consistent with our previous finding using pair-wise t tests that adaptation opportunities reduce investment in mitigation in the initial period. Moreover, from Model 1 we see that while the inclusion of pledges has no effect on average mitigation in the mitigation-only environments, a comparison between *M&A* and *M&AwP* reveals that pledges significantly increase average investment in mitigation when users can invest in adaptation (the p-values for these hypotheses tests are contained in the last row of Table 4).

Model 2 adds the expected investment in mitigation from the other three group members (between 0 and 30) and interaction terms with treatments as additional covariates.[6] The expected investment variable is highly significant suggesting that one's own mitigation investment is highly correlated with expectations about others' investment in mitigation. Note that Model 2 includes the *expectation* rather than the *pledge* by others for two reasons: First, the expectations variable was solicited for all four treatments, while by design, we only have pledges for half of the treatments. Second, given that subjects form their expectations at least partially based on the pledged amount by others, these variables are highly correlated (i.e., Pearson correlation coefficient of 0.7565, $p < 0.01$) and including them together leads to both variables appearing insignificant.[7] Later in this section (Table 5), we analyze mitigation decisions using only the two treatments with pledges and include the aggregate amount pledged by others (while dropping expectations). Finally, we observe that loss aversion does not have a significant effect on subjects' first period decisions.

---

[6] Our empirical approach follows the strategy in Fischbacher and Gächter (2010, page 549) in which contributions are regressed on beliefs about others' contributions in a linear public goods game.

[7] The variance inflation factor diagnostic is close to 20 for these variables.

Table 4 Conditional analyses of mitigation investments

	1st Period		Pooled		
	Model 1	Model 2	Model 3	Model 4	Model 5
MwP	0.743	0.220	0.731	0.310	0.985**
	(0.607)	(2.168)	(0.700)	(0.574)	(0.496)
M&A	−1.244**	−0.473	−1.203	0.285	0.693*
	(0.606)	(1.728)	(0.897)	(0.382)	(0.413)
M&AwP	0.640	2.476	−0.217	−0.426	0.166
	(0.607)	(1.886)	(0.900)	(0.511)	(0.491)
Loss Aversion	−0.0508	0.0273	−0.407***	−0.163**	−0.0805**
	(0.112)	(0.0947)	(0.146)	(0.0660)	(0.0362)
Expected others		0.297***		0.259***	0.220***
		(0.0560)		(0.0162)	(0.0195)
MwP*Expected		0.00456		−3.60e–05	−0.0455**
		(0.0929)		(0.0265)	(0.0229)
M&A*Expected		0.00897		−0.0284	−0.0332
		(0.0817)		(0.0254)	(0.0242)
M&AwP*Expected		−0.0804		0.0262	−0.00914
		(0.0841)		(0.0236)	(0.0239)
Period			−0.242***	−0.0828***	−0.00625
			(0.0396)	(0.0216)	(0.0238)
Damages hit last period					−0.359**
					(0.157)
Mitigation last period					0.382***
					(0.0438)
Constant	6.765***	0.155	7.684***	1.520***	−0.442
	(0.461)	(1.286)	(0.553)	(0.374)	(0.430)
n	192	192	1,920	1,920	1,728
F	3.43***	12.77***	–	–	–
χ^2	–	–	47.12***	1395.45***	4012.25
H_0: M&A = M&AwP	$p<0.01$	$p=0.032$	$p=0.327$	$p=0.102$	$p=0.128$

Standard errors in parentheses, *$p<0.10$, **$p<0.05$, ***$p<0.01$

Models 3–5 in Table 4 are panel regressions. In Model 3 we see that the average treatment effects are no longer present. That is, pledges do not have a significant effect on average mitigation investment in either the mitigation-only or the mitigation and adaptation environments (see bottom row). The period variable is picking up a negative and highly significant time trend. Model 4 adds the expectation of others' investment in mitigation, and this variable is highly significant and positively correlated with mitigation. The negative time trend remains significant in Model 4. The final model includes a lagged variable that equals one if the individual suffered damages in the previous period as well as a lag for own investment in mitigation in the previous period. Unsurprisingly, the variable for last-period mitigation is highly significant and positive. The lagged damage variable is negative and highly significant suggesting that if individuals suffered damages in the previous period, then they are less willing to invest in the collective strategy of mitigation. Note that while the expectation variable remains significant in Model 5, the period time trend is now

Table 5 The role of pledges on mitigation investments

	Model 6	Model 7	Model 8
M&AwP	−0.782 (0.518)	−1.581 (1.086)	−0.444 (0.266)
Loss aversion	−0.419*** (0.111)	−0.403*** (0.117)	−0.248*** (0.065)
Others' pledge	0.200*** (0.022)	0.175*** (0.030)	0.132*** (0.021)
Period	−0.186*** (0.042)	−0.187*** (0.042)	0.010 (0.042)
M&AwP × Others' pledge	–	0.038 (0.041)	–
Damages hit last period	–	–	−0.987*** (0.300)
Mitigation last period	–	–	0.411*** (0.036)
Constant	3.827*** (0.736)	4.348** (0.783)	1.569** (0.631)
n	960	960	864
$\chi 2$	239.70***	239.67***	1088.67***

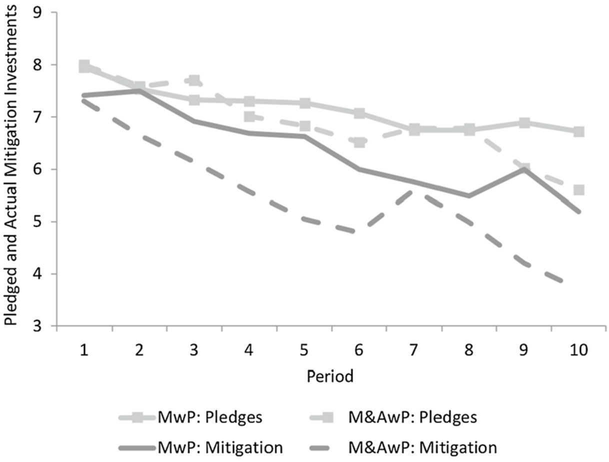
Standard errors in parentheses, $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

insignificant. This suggests that the *period* variable in Models 3 and 4 was, in part, picking up individuals' reactionary responses to whether they were able to collectively prevent the damages from occurring in the previous period. Similarly, while remaining significant, the loss aversion measure loses explanatory power when the two lagged variables are included. However, we can conclude that more loss averse subjects invest less in mitigation.

To round out the analysis on mitigation investments, we concentrate on the role pledges play in treatments *MwP* and *M&AwP*. Table 5 shows estimates from three panel models that are closely linked to Models 4 and 5 from Table 4 except that we include others' pledges in place of expectation of others' investment. From Models 6, 7 and 8, the amount the other group members pledge is positively correlated with own mitigation investment and is highly significant ($p < 0.01$). In Model 6 we again observe a negative and highly significant period time trend, but this dissipates once we include (in Model 8) the lagged variable capturing whether the group suffered damages in the previous period and one's own mitigation investment last period. Again, we observe that loss aversion drives down investments in mitigation. To examine whether the pledges by the other group members have differential effects by treatment, Model 7 includes a control interacting the *M&AwP* treatment dummy with the size of other players' pledges to mitigate. The effect is insignificant, supporting a similar marginal response to pledges on mitigation decisions in the settings with and without adaptation possibilities.

In summary, the treatment effect we observed in Period 1 goes away with repetition of the game. That is, average mitigation levels in *M&AwP* are not higher than in *M&A*. This finding does not imply that pledges do not matter. In contrary, we show that pledges have a significant impact on mitigation decisions. When pledges increase so does mitigation, and when pledges decrease investment in mitigation decreases. This result ties into the economics literature on reciprocity (e.g., Rabin, 1993), and in particular supports a theory of how reciprocal strategies can enhance mitigation efforts in environmental agreements (Nyborg, 2018).

Fig. 4 Average pledge versus average investment in mitigation over 10 periods



4.4 Decreasing Pledges and the Role of Non-Compliance

A main finding from the first-period analysis is that non-binding pledges significantly increase investment in mitigation when players can also invest in private adaptation. Here we explore why the effectiveness of making pledges, on average, unravels as groups of players make repeated decisions. Figure 4 shows the time series of individual pledges and actual investments. The dashed lines are the treatments with adaptation, the dark lines are the investments in mitigation and the light grey lines are the pledged amounts. The gap between average pledges and average mitigation investment appears to widen over time, suggesting increasing non-compliance. It is also important to note that while the gap between the lines widens, both pledges and mitigation appear highly correlated suggesting that pledges are not viewed as pure noise.

The most glaring result from Fig. 4 is that average pledges decrease over time. Since mitigation investment is highly correlated with pledges, it too falls. This begs the question of why players decrease the amount they pledge to mitigation over the course of the experiment. This could be related to compliance behavior. Recall that in the first period, on average, groups complied with what they pledged to do. It is also interesting to note that in the first period players *expected* that their group members would largely comply with their pledges in the *MwP* treatment but not in the *M&AwP* treatment. In *MwP*, the expected contribution was 23.17 compared to the pledged amount of 23.88 ($p=0.431$). In *M&AwP*, players expected a bit more non-compliance; that is, the expected contribution was 21.35 compared to the pledged amount of 24.0 in ($p=0.012$).

To explore the relationship between pledges and compliance we estimate additional regression models for the two treatments with pledges, *MwP* and *M&AwP*, and results are presented in Table 6. In these models the dependent variable is an individual's pledge. Along with the treatment dummy and controls for period, individual and group effects, we regress pledged amounts on compliance variables. One variable we include is a lagged measure of non-compliance (called *lagged non-compliance*). This variable is the difference between the sum of the others' pledges and the sum of the others' actual investment in mitigation in the previous period. Leaning on a body of experimental work documenting that people are conditionally cooperative (e.g., Fischbacher et al., 2001; Frey and Meier, 2004), our prior is that greater non-compliance from others last period reduces the amount

Table 6 Pledges as a function of actual and expected non-compliance

	MwP	M&AwP	Pooled
M&AwP	–	–	-0.024 (0.641)
Lagged Non-compliance	−0.012 (0.029)	0.008 (0.031)	−0.008 (0.024)
Expected Non-compliance	−0.057** (0.026)	−0.124* (0.074)	−0.102** (0.050)
Loss aversion	0.211 (0.188)	−0.493*** (0.181)	−0.144 (0.134)
Period	−0.061 (0.043)	−0.125 (0.092)	−0.078 (0.056)
Damages hit last period	−0.256 (0.301)	−1.170* (0.707)	−0.741*** (0.263)
Constant	7.232*** (0.610)	9.226*** (0.550)	8.160*** (0.651)
n	432	432	864
$\chi 2$	18.84***	81.34***	39.64***

Standard errors in parentheses, $*p<0.10$, $**p<0.05$, $***p<0.01$

individuals pledge this period. We also include a variable called *expected non-compliance*, which is the difference between the sum of the other's pledges and the individual's expectation of the sum of the others' investment. Here we explore whether people condition their pledges on the expected level of compliance of their group members. That is, increases in expected non-compliance in the current period could reduce the amount an individual is willing to pledge. Table 6 contains results from three models, the first two parse the dataset by treatment and the third pools the two treatments with pledges. The models are estimated with subject-specific random effects and robust standard errors clustered at the group level.

Across all three models, we find that pledges decrease significantly when expected non-compliance increases. The effect is weakly significant in both treatments, the coefficient is twice as large in the treatment with adaptation but with lower significance. The relative measures of expected non-compliance between the two treatments are also illustrative. Pooled over all rounds, there is significantly more expected non-compliance on average in the treatment with both mitigation and adaptation (1.108 in *MwP* vs. 2.921 in *M&AwP*, $p<0.01$). To get a better picture of the data over time, Fig. 5 shows the development of pledges and actual investment by others, and own expectations over 10 periods separated in two panels for the respective treatments. The gap between the dashed line and the grey line in each panel is the level of non-compliance (i.e., pledges—investment). The gap between the dashed line and the black line is expected non-compliance (i.e., pledges—expectations). It is clear that players expect their counterparts to comply less when they have both mitigation and adaptation investment strategies available (compared to mitigation only).[8] In short, the measure of actual and expected non-compliance is greater in *M&AwP* than *MwP*, and in turn we observe a more pronounced reduction in pledges over time in the *M&AwP* treatment.

---

[8] When comparing the difference between others' pledges and expected mitigation by treatment, the difference is statistically larger in *M&AwP* compared to *MwP* ($p<0.000$).

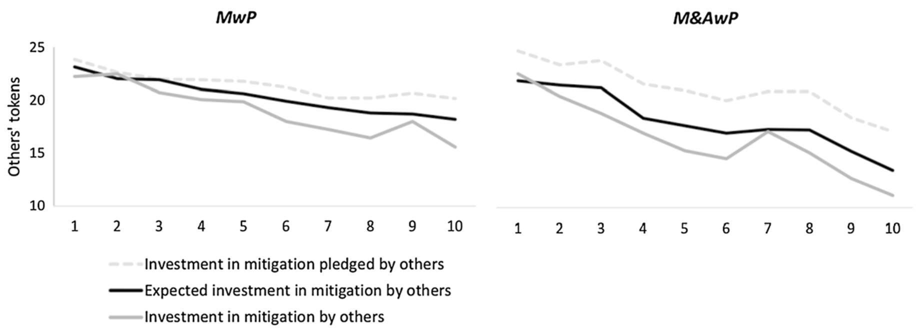


Fig. 5 Pledges, expectations and investment in mitigation by the other group members over time

Table 7 The role of pledges and previous contributions on the formation of expectations

	MwP	M&AwP	Pooled
M&AwP	–	–	−1.168* (0.638)
Others' pledge	0.487*** (0.049)	0.413*** (0.124)	0.434*** (0.069)
Others' mitigation last period	0.357*** (0.080)	0.465*** (0.052)	0.428*** (0.041)
Loss aversion	−0.303* (0.175)	−0.133 (0.339)	−0.232 (0.154)
Period	−0.097** (0.043)	−0.104 (0.079)	−0.094** (0.045)
Constant	3.917* (1.999)	2.025 (2.561)	3.525** (1.501)
n	432	432	864
$\chi 2$	175.89***	303.63***	409.93***

Standard errors in parentheses, *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

From Table 6 we also see that lagged non-compliance is insignificant in all three models, suggesting that players do not condition their pledges on the level of compliance from the previous period. Although we do find that lagged non-compliance and expected non-compliance are highly correlated (Pearson correlation of 0.485, $p < 0.01$), the variance inflation factor (VIF) diagnostic does not suggest a problem of multicollinearity (i.e., $VIFs < 2$) which justifies the inclusion of both. Interestingly, when the group suffered damages in the previous period it decreased the amount players pledged but only when they also had adaptation opportunities. It appears that after a group failed to prevent damages, the pledged group mitigation investments are smaller (pooled model). This negative correlation between damages affecting a group and reduced mitigation pledges is larger in magnitude but smaller in significance when considering only settings where there is an option to privately adapt (*M&AwP*). The coefficient is not significant when only mitigation investments are feasible (*MwP*).

Non-compliance with pledges has important implications for the formation of beliefs about others' behavior. Specifically, in environments where pledges are available, subjects have two sources of information on which they can base their expectations: (i) the pledge and (ii) the actual behavior of others in the previous periods of the game. Table 7 provides evidence that both variables have a highly significant impact on the formation of expectations. The size of the effect of pledges by others, however, is smaller in the *M&AwP* treatment and subjects in this environment base their expectations more heavily on the actual behavior of their peers in the previous period. This may be the case because pledges are perceived as less credible whenever adaptation enables subjects to privately protect from the damage (see footnote 7).

5 Conclusion

This study presents experimental results on the role that non-binding pledges have on managing a threat of probabilistic group damage. A novel feature of our research is that we consider the interaction of two strategies that parties can use to manage the threat. They can work collectively to try to mitigate the root cause of the damage, or they can work autonomously to try and protect themselves from the damages if they occur. While both collective and private strategies are typically available to manage group damages, most of the literature neglects the strategic interaction between the two.

Consistent with our main hypothesis, pledges increase investments in mitigation on average, but the effect is significant only when both mitigation and adaptation investments are possible. Our panel-data analyses reveal that these average treatment effects are fragile. The efficiency gains from pledges all but disappear when controlling for period, subject and group effects. However, the data reveal that players take the pledges as informative and, in addition to actual behavior in the previous period, as an important source of information to form expectations about others' behavior. Investments in collective mitigation are made conditional on the amount pledged by others and we find surprisingly high compliance rates with the nonbinding pledges. We focus on the impact pledges have on expectations of group mitigation and the influence that this in turn has on expected marginal incentives to invest in mitigation. However, other possible mechanisms could be compatible with the observed behavior. For example, in complicated decision environments participants might respond to pledges as a way to simplify their decision rules (e.g. Simonson 1989; Prelec and Herrnstein 1991; Shafir, Simonson, and Tversky 1993).

We interpret these results as supporting two reasons to be somewhat more optimistic regarding the potential of pledges to enhance cooperation than the previous literature suggests. First, the finding from the previous literature that numeric cheap talk is ineffective (Chen and Komorita, 1994; Wilson and Sell, 1997; Bochet et al., 2006; Bochet and Putterman, 2009) might understate the short-term effects that pledges can have when mitigation and adaptation are both feasible strategies to deal with potential losses. This is a feature of many field settings, as discussed in the introduction. Second, while the effects of non-binding pledges wash out on average with repetition and experience, it is clear that subjects do not treat the pledges as pure noise and reciprocate by investing more in mitigation when they expect others to do so as well. Non-binding pledges matter a great deal in terms of formation of beliefs and prediction of behavior. The decrease in the average investment in mitigation is in part triggered by a decrease in the amount players pledge over the course of the experiment. The crucial limitation in the effectiveness of pledges seems to be related to repeated and increasing levels of non-compliance. The expectation of non-compliance by other players increases with repetition in the game, while pledges and investment in mitigation decrease over time. This result highlights the importance of implementing mechanisms to deal with non-compliance when considering pledges as an instrument to enhance cooperation.

The capacity and scope of pledges to enhance mitigation of collective damages is a complex topic of research and requires cumulative evidence from different methods. We interpret our result as pointing to the importance of further field and experimental research that examines the capacity of pledges to enhance public good provision and to identify the conditions under which pledges may be effective. The laboratory provides a controlled environment in which institutions and policy components can be analyzed in a simplified setting in order to isolate causal effects (Smith, 1994; Falk and Heckman, 2009). Lab experiments are useful as a testbed and tend to allow for generalizable inferences about behavior (Snowberg and Yariv, 2021), but results must be interpreted as suggestive evidence rather than being definitive. Our approach is pragmatic, acknowledging that controlled experimentation is just one of many complementary research methods that contributes to the cumulative body of knowledge on this topic.

It is important to keep in mind that the participants in our experiments are randomly and anonymously assigned into groups, with no opportunity for communication beyond announcing their intended numeric mitigation decision. Thus, the setting we investigate is akin to groups of individuals that lack interaction beyond formal, institutional rules. Clearly, the relationship between agents in very local settings and in international negotiations is more complex than the lab environment we consider, and these complexities—including social norms, power dynamics and political economy—could impact the effectiveness of pledges. We consider the results presented most informative for intermediate sized groups such as a community of individuals deciding on constructing dikes in case of flooding risk, forest management to reduce wildfire risk or action against crime. Pledges in such settings may provide relevant information that can help individuals assess whether they want to act cooperatively by investing in group protection or whether they prefer to invest their resources in private protection against the upcoming risk.

This regional level has received less attention in the previous social dilemma literature where much of the emphasis has been on very local settings (e.g. Ostrom 1990) or at international negotiations (e.g. Barrett and Dannenberg 2016). For collective action at the regional level, further investigation into the relationship between pledges, beliefs and actions is warranted. A pre-condition for pledges to help improve public good provision is that the pledges impact the formation of beliefs and predictions regarding others' behavior, and we observe this in our experimental data. Building off of our results, other empirical approaches—including randomized control trials or quasi experimental methods—could help identify whether pledges are effective in more complex, natural settings.

Appendix

The Mitigation Game

In this game each player can invest any of their endowment in a public insurance account that mitigates the probability of the loss occurring. Specifically, player i's expected payoff function is

$$\pi_i = e_i - m_i - (1 - \beta M)D, \tag{A-1}$$

where e_i is player i's endowment. The variable m_i is player i's choice of how much of her endowment to spend on mitigation where $m_i \leq e_i$ and $M = \Sigma m_i$. The term D is the size of the impending damage from the loss, affecting all group members, and the term $(1 - \beta M)$ is the probability the loss occurs, where β is a positive constant. Without any mitigation efforts (i.e., $M = 0$), each player faces certain damages of D from the group loss. The probability the loss occurs decreases by β for all players for each unit invested in mitigation by anyone. Mitigation, therefore, is a pure public good.

Throughout we assume that $1 \geq \beta > 0$ and $D > e_i$. We also restrict $(1 - \beta ne_i) \leq 0$ which ensures that players cannot reduce damages beyond zero. For convenience, from here on we will assume the strict equality holds; that is, $(1 - \beta ne_i) = 0$ so that if all n players contribute their entire endowments to mitigation then the probability the loss occurs is driven to zero and they completely avoid the damages.

The relationship between the parameters is chosen so that the mitigation decision is a social dilemma, with the familiar conflict between what is individually and socially optimal. In particular, the change in player i's expected payoffs from an additional unit of mitigation is

$$\frac{d\pi_i}{dm_i} = \beta D - 1 < 0, \tag{A-2}$$

and therefore all risk-neutral players with self-interested payoff-maximizing preferences make zero contributions to mitigation in a non-cooperative Nash equilibrium. In relation to the broader public goods literature in experimental economics, βD is the marginal per-capita return (MPCR) which is strictly < 1 in this game. It is also the case that

$$\beta nD - 1 > 0, \tag{A-3}$$

which tells us that the joint benefit from a unit of mitigation (the collective reduction in expected damages, which equals βnD) is greater than the individual cost of mitigation (which is constant at 1). Therefore, we have the familiar linear public good game in which no contributions are made in a Nash equilibrium but the social optimum is reached when all n players contribute their entire endowment to mitigation (and avoid all damages). In the non-cooperative Nash equilibrium each player earns $e - D < 0$ and in the social optimal each player suffers zero losses, earning their full endowment e.

The stark difference in mitigation between the non-cooperative Nash equilibrium (invest nothing) and the social optimum (invest everything) is useful to help illustrate the underlying tension in a public-goods game. Decades of behavioral research, however, has demonstrated that people are not completely self interested and tend to make positive contributions in public goods games. The empirical evidence shows that investments in public goods are, on average, higher than Nash predictions but lower than what is socially optimal.

Lab Implementation of the Mitigation game (M)

In this treatment each of the four players make a decision regarding how much of their 10 token endowment they want to invest in mitigation. Each token contributed to mitigation reduces the probability each player incurs the 20 token loss by 2.5%. Therefore, expected damages decrease by $0.025 \times 20 = 0.50$ for each token contributed toward mitigation. Since, $0.50 < 1$ a self-interested payoff-maximizing player contributes zero tokens to mitigation. If instead all four players contribute all 10 tokens to mitigation ($M = 40$) then the probability the damages occur (i.e., $1 - \beta M$) equals zero, which is the social optimum.

While the non-cooperative Nash equilibrium is $M = 0$, a long history of public good experiments inform us that diverse behavioral motives will lead people to contribute tokens to the public account (see research summarized in Ostrom and Walker, 2003; Camerer, 2006; Chaudhuri, 2011). In the experimental literature average contributions tend to start between 40 and 70 percent of endowments (between four and six tokens in our experiment) and decrease over multiple decision periods (Ledyard, 1995).

Lab Implementation of the Mitigation and Adaptation game (*M&A*)

In this treatment players have two investment opportunities. As in M, each token invested in mitigation reduces the probability of the damages occurring for everyone by 2.5%. On the other hand, a token invested in private adaptation reduces the size of the damage by 1.5 tokens for the investor alone. The two investments have competing effects on expected payoffs, as highlighted by the marginal effects presented below. The expected return (i.e., change in profit) for a token invested in mitigation given our parameter choices is

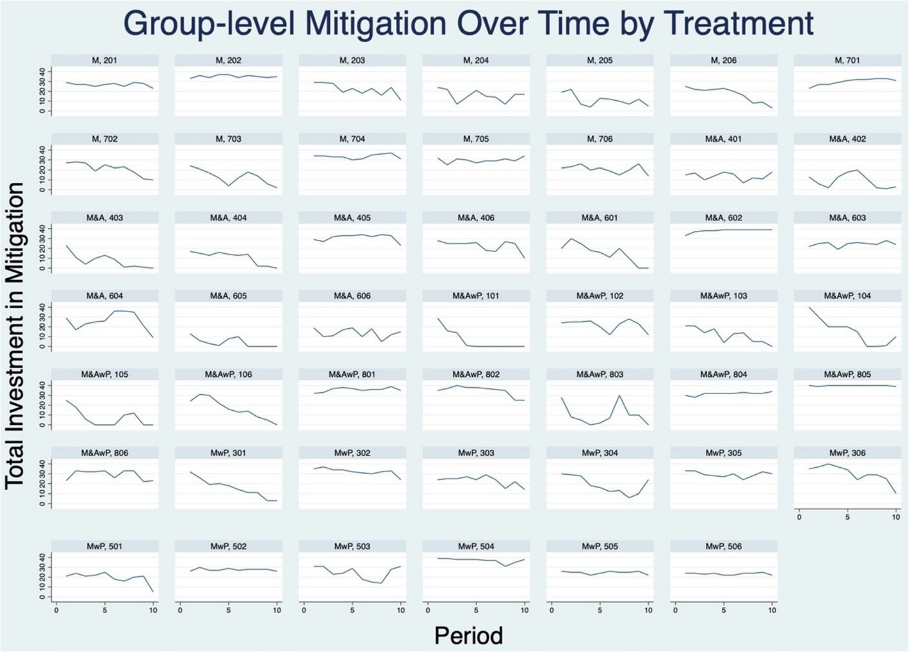
$$\frac{d\pi_i}{dm_i} = 0.025\left(20 - 1.5a_i\right) - 1.$$

Note that the expected change in profit is negative for $0 \leq a_i \leq 10$, and is decreasing as a_i increases. The expected change in profit for a token invested in private adaptation given our parameter choices is

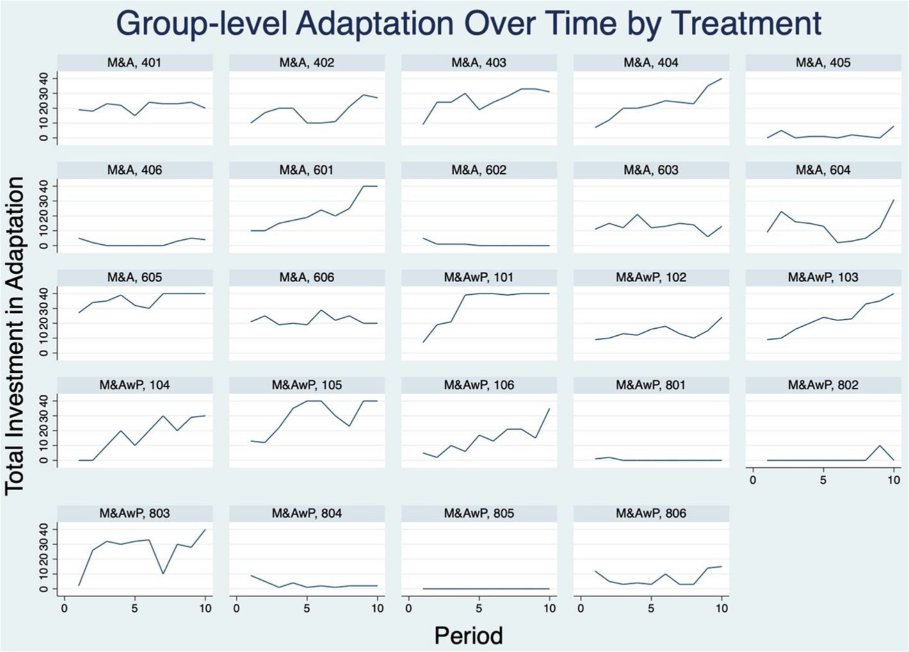
$$\frac{d\pi_i}{da_i} = 1.5(1 - 0.025M) - 1.$$

Investments in adaptation increase profits given low levels of aggregate mitigation (e.g., investing in a unit of adaptation increases profit by 0.5 when $M = 0$), but get smaller with positive group investments in mitigation. Indeed, the change in profit from adaptation investment turns negative once aggregate investment in mitigation exceeds 13 tokens, and in those circumstances it is more lucrative to hold onto tokens rather than invest them. While no self-interested payoff-maximizing player contributes to mitigation in a Nash equilibrium, we know that positive contributions to mitigation can be expected in social dilemma environments. If a player *expects* positive investments in mitigation then investing in adaptation is less lucrative than a marginal increase in profit of 0.5 tokens. In fact, given high enough investments in mitigation by other players the expected return from adaptation may be lower than the expected return from mitigation. To see this, consider the return from mitigation given that a player does not contribute anything to adaptation. This return is -0.50 tokens. The return from a token invested in adaptation drops below -0.50 once aggregate mitigation investments reach 27 tokens (i.e., at least 90 percent of others' total endowments). In summary, a player would need to expect very high investments in mitigation from others to find investing in mitigation more profitable than adaptation. This of course is an empirical question.

Group-Level Mitigation over Time by Treatment



Group-Level Adaptation over Time by Treatment



References

Barrett S, Dannenberg A (2012) Climate negotiations under scientific uncertainty. Proc Natl Acad Sci 109(43):17372–17376

Barrett S, Dannenberg A (2014) Sensitivity of collective action to uncertainty about climate tipping points. Nat Clim Chang 4(1):36–39

Barrett S, Dannenberg A (2016) An experimental investigation into pledge and review in climate negotiations. Clim Change 138(1–2):339–351

Bayramoglu B, Finus M, Jaques J (2018) Climate agreements in a mitigation-adaptation game. J Public Econ 165:101–113

Blanco E, Haller T, Walker JM (2017) Externalities in appropriation: Responses to probabilistic losses. Exp Econ 20:793–808

Blanco E, Dutcher E, Haller T (2020) Social dilemmas with public and private insurance against losses. J Econ Behav Organiz 180:924–937

Bochet O, Putterman L (2009) Not just babble: Opening the black box of communication in a voluntary contribution experiment. Eur Econ Rev 53:309–326

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

Borrero M, Rubio SJ (2021) An adaptation-mitigation game: does adaptation promote participation in international environmental agreements? Int Environ Agreem: Polit Law Econ (forthcoming)

Brandts L, Saijo T, Schram A (2004) How universal is behavior? A four country comparison of spite and cooperation in voluntary contribution mechanisms. Public Choice 119(3):381–424

Buob S, Stephen G (2011) To mitigate or to adapt: How to confront global climate change. Eur J Polit Econ 27:1–16

Camerer C (2003) Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton

Camerer C, Fehr E (2006) When does “economic man” dominate social behavior? Science 311(5757):47–52

Charness G, Dufwenberg M (2006) Promises and partnership. Econometrica 74(6):1579–1601

Chaudhuri A (2011) Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. Exp Econ 14(1):47–83

Chen X, Komorita SS (1994) The effects of communication and commitment in a public goods social dilemma. Organz Behav Human Decision Process 60:367–386

Crawford V (1998) A survey of experiments on communication via cheap talk. J Econ Theory 78(2):286–298

Croson R (2007) Theories of commitment, altruism and reciprocity: Evidence from linear public good games. Econ Inq 45(2):199–216

Ebert U, Welsch H (2012) Adaptation and mitigation in global pollution problems: Economic impacts of productivity, sensitivity, and adaptive capacity. Environ Resource Econ 52:49–64

Falk A, Heckman JJ (2009) Lab experiments are a major source of knowledge in the social sciences. Science 326:535–538

Fischbacher U (2007) z-tree: Zurich toolbox for ready-made economic experiments. Exp Econ 10(2):171–178

Fischbacher U, S. G¨achter, (2010) Social preferences, beliefs, and the dynamics of free riding in public goods experiments. Am Econ Rev 100(1):541–556

Fischbacher U, Gächter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. Econ Lett 71(3):397–404

Frey B, Meier S (2004) Social comparisons and pro-social behavior: Testing conditional cooperation in a field experiment. Am Econ Rev 94(5):1717–1722

Gächter S, Johnson EJ, Herrmann A (2007) Individual-level loss aversion in riskless and risky choices. IZA Discuss Paper Series 2961

Goeree JK, Holt CA, Laury SK (2002) Private costs and public benefits: unraveling the effects of altruism and noisy behavior. J Public Econ 83(2):255–276

Hasson R, Lofgren A, Visser M (2010) Climate change in a public goods game: Investment decision in mitigation versus adaptation. Ecol Econ 70(2):331–338

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) In search of homo economicus: behavioral experiments in 15 small-scale societies. Am Econ Rev 91(2):73–78

Ingham A, Ma J, Ulph A (2007) Climate change, mitigation and adaptation with uncertainty and learning. Energy Policy 35:5354–5369

Isaac RM, Walker J (1988) Communication and free-riding behavior: The voluntary contribution mechanism. Econ Inq 26:585–608

Jacquemet N, James A, Luchini S, Shogren JF (2017) Referenda under oath. Environ Res Econ 67:479–504

Jacquemet N, Luchini S, Rosaz J, Shogren JF (2019) Truth telling under oath. Manage Sci 65:426–438

Kane S, Shogren J (2000) Linking adaptation and mitigation in climate change policy. Clim Change 45:75–102

Koukoumelis A, Levati MV, Weisser J (2012) Leading by words: A voluntary contribution experiment with one-way communication. J Econ Behav Organ 8:379–390

Lazkano I, Marrouch W, Nkuiya B (2016) Adaptation to climate change: How does het-erogeneity in adaptation costs affect climate coalition? Environ Develop Econ 21(6):812–838

Ledyard (1995) Public goods: A survey of experimental research. In: Kagel J, Roth A (eds) Handbook of Experimental Economics. Princeton University Press, New Jersey

Marrouch W, Chaudhuri AR (2011) International environmental agreements in the presence of adaptation. FEEM Working Paper No. 35, 2011.

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

Neugebauer T, Perote J, Schmidt U, Loose M (2009) Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. J Econ Psychol 30(1):52–60

Nyborg K (2018) Reciprocal climate negotiators. J Environ Econ Manag 92:707–725

Oprea R, Charness G, Friedman D (2014) Continuous time and communication in a public goods experiment. J Econ Behav Organ 108:212–223

Ostrom E (2006) The value-added of laboratory experiments for the study of institutions and common-pool resources. J Econ Behav Organ 61(2):149–163

Ostrom E, Gardner R, Walker J (eds) (1994) Rules Games and Common-Pool Resources. University of Michigan Press, Michigan

Ostrom E, Walker J (2003) *Trust and Reciprocity: Interdisciplinary Lessons for Experimental Research.* Russell Sage Foundation

Palfrey TR, Rosenthal H (1991) Testing for effects of cheap talk in a public goods game with private information. Games Econom Behav 3:183–220

Pogrebna G, Krantz DH, Schade C, Keser C (2011) Words versus actions as a means to influence cooperation in social dilemma situations. Theor Decis 71:473–502

Rabin M (1993) Incorporating fairness into game theory and economics. Am Econ Rev 83:1281–1302

Sally D (1995) Conversation and cooperation in social dilemmas. Ration Soc 7(1):58–92

Shibata H, Winrich JS (1983) Control of pollution when the offended defend themselves. Economica 50(200):425–437

Smith V (1994) Economics in the laboratory. J Econ Perspect 8(1):113–131

Smith A (2013) Estimating the causal effect of beliefs on contributions in repeated public good games. Exp Econ 16:414–425

Snowberg E, Yariv L (2021) Testing the waters: Behavior across participant pools. Am Econ Rev 111(2):687–719

Tavoni A, Dannenberg A, Kallis G, Loschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. Proc Natl Acad Sci 108(29):11825–11829

Thöni C, Volk S (2018) Conditional Cooperation: Review and Refinement. Econ Lett 171:37–40

Wilson RK, Sell J (1997) Liar, liar: Cheap talk and reputation in repeated public goods settings. J Conf Resol 41(5):695–717

Zehaie F (2009) The timing and strategic role of self-protection. Environ Res Econ 44:337–350