Conditional Commitments and Cooperation in Public Goods: Theory and Experimental Evidence

Mehmet Y. Gürdal* Özgür Gürerk† Yeliz Kaçamak‡ Edip Kart§

April 5, 2022

Abstract

Conditional commitment devices such as price matching guarantees or legal contracts can be highly efficient and improve the outcomes for involved parties. In the context of public goods, however, empirical evidence on the cooperation-enhancing effect of conditional commitments is scarce. In this paper, we demonstrate that conditional commitments can indeed increase voluntary contributions to public goods. First, we theoretically analyze the effect of conditional commitments made through a mediator on contributions towards a public good. In this mediated game, the set of conditional commitments can be chosen such that a strong equilibrium exists, corresponding to the strong mediated equilibrium introduced in Monderer and Tennenholtz (2009). We run laboratory experiments and find, if a strong equilibrium exists, almost all interacting groups use conditional commitments and manage to coordinate on this cooperative equilibrium sustaining cooperation on a high level. When conditional commitments only allow for socially inefficient outcomes, we observe a substantial decline in their utilization, and the cooperation level drops significantly.

Keywords: prisoner's dilemma, cooperation, public goods, assortative matching.
JEL Classification Numbers: C72; C91; H41.

*Department of Economics, Boğaziçi University. E-mail: mehmet.gurdal@boun.edu.tr
†Corresponding author. Faculty of Law, Social Sciences & Economics, University of Erfurt. E-mail: guererk@uni-erfurt.de
‡Department of Economics, Boğaziçi University. E-mail: yeliz.kacamak@boun.edu.tr
§Department of Economics, Boğaziçi University.

1 Introduction

Cooperation and coordination are difficult to establish and sustain when decision makers have a conflict of interest, or a lack of mutual trust. In game theory, these types of situations are represented by games such as the prisoner's dilemma and the stag hunt game. In these interactions, also called social dilemmas, unsuccessful coordination and the lack of cooperation can lead to significant potential welfare losses for all actors involved.

One of the important economic social dilemma situations is the provision of public goods. The workhorse model to study behavior in public goods is the *voluntary contribution mechanism* (VCM), which involves a group of individuals contributing money or exerting effort in order to provide a public good. Experimental studies show that without any additional supporting mechanisms, in repeated interactions, contributions to public good typically decline over time until full collapse of cooperation (Ledyard 1995).

Although many different modifications are proposed for enhancing the efficiency of the VCM in public goods provision, arguably the most successful ones use a sanctioning device to punish free-riders. Allowing for costly punishment in these types of environments can sustain high contribution levels, regardless of whether the punishers are able to develop individual reputations (Ostrom et al. 1992) or not (Fehr and Gächter 2000, 2002). Costly punishment has also been shown to be highly effective across different societies (Herrmann et al. 2008). Additionally, when given a choice between environments with and without sanctioning institutions, the great majority of subjects choose the environment where free-riders can be punished (Gürerk et al. 2006, 2014). Besides sanctioning, some alternative interventions to sustain high contributions in the VCM involve the grouping of subjects based on previous cooperative tendencies (Gächter and Thöni 2005) or based on current contribution decisions (Gunnthorsdottir et al. 2010); voluntary leadership to choose contributions before other group members (Rivas and Sutter 2011), and allowing communication among group members (Bochet et al. 2006).

In this paper, we show that, besides the mechanisms described above, cooperation in the context of public good provision can also be sustained by conditional commitments as suggested by Monderer and Tennenholtz (2009). At the center of the mechanism that we theoretically analyze is a mediator who coordinates players' commitment decisions conditional on other players' commitments. The conditional commitments can be chosen such that a cooperative equilibrium exists.

In this context, a conditional commitment mechanism refers to simultaneously sent messages by the players and a centrally implemented outcome as a function of these messages. The messages are of the form "I agree to contribute K units out of my endowment ω if all other members send this same message, otherwise, I will contribute zero units". For example, if $K = \omega$, then each subject's choice of the above message would imply full contribution by each group member. If one or more subjects choose not to send this message (by choosing another message or by not sending a message and choosing their contributions independently), then each subject sending this message contributes zero. Consequently, we obtain a mechanism where players engage in commitments that are conditional on the commitments of other players. In this regard, the conditional commitment mechanism also differs from the conditional cooperation mechanisms studied earlier (Kocher et al. 2008).

Enlarging the strategy set in the VCM setting by the messages described above constitutes a *mediated game*. We find that subjects quickly converge to strong equilibrium when the mediated game has a strong equilibrium robust to coalitional deviations. On the other hand, when a strong equilibrium does not exist, the frequency of groups exhibiting successful coordination stays at relatively lower levels.

The study of conditional commitments is also essential since some recent technological innovations make the implementation of multi-lateral agreements easier. So called *smart contracts* based on the *blockchain technology* can be used in many different domains to execute transactions in a decentralized, enforceable, and automated manner. We will refer to some examples in section 6.

Our contribution to the literature is both theoretical and experimental. First, we formally document that experimental studies on conditional commitments such as the present paper and that by Oechssler et al. (2019) can be analyzed within the framework provided in Monderer and Tennenholtz (2009) and Kalai et al. (2010). Second, we experimentally test the effectiveness of conditional commitments in sustaining coordination and increasing cooperation in public goods with a VCM environment.

The paper is structured as follows: Section 2 introduces the related literature. In section 3, we describe the model by Monderer and Tennenholtz (2009), and outline how our conditional commitment mechanism implements a strong mediated equilibrium in the VCM environment. Section 4 is about the experimental procedures. In section 5, we present the results. Section 6 concludes.

2 Related Literature

As mentioned in the previous section, the workhorse model to study behavior in public goods is the voluntary contribution mechanism (VCM). The added value of the provision of the public good to the society (or for the involved parties) is reflected in the productivity parameter $R > 1$, i.e., each dollar contributed towards to public good generates a return higher than one dollar. Two important properties of the public good are (i) that its returns are shared equally among N group members, and (ii) no member can be excluded from receiving his part of the return. Every dollar contributed to the public good increases each member's earning by a certain amount, which is known as the marginal per capita return (MPCR). In essence, VCM settings with MPCR bigger than R/N but less than one constitute n-person prisoner's dilemma games, where the Nash equilibrium is zero contributions by everyone. The socially efficient outcome, however, involves everyone contributing all their endowments. The typical outcome for laboratory studies involving the VCM is, however, declining contributions towards the public good throughout time, regardless of the group size (Isaac and Walker 1988; Ledyard 1995).

Our study is related to two different strands of literature: (i) experimental studies of the voluntary contribution mechanism and (ii) conditional commitments and mediation in simultaneous games. Below we summarize the relevant literature from these two strands:

Among the studies in the first strand, our paper is closest to the experimental and theoretical study of the *binary conditional contribution mechanism* (BCCM) by Reischmann and Oechssler (2018). In their experiment, players may condition their binary contribution decisions (contribute or not) on the number of other players who agree to contribute ("I am willing to contribute to the public good if at least k agents (including myself) contribute in total."). They also allow heterogeneous returns from the public good for different players, and information about individual returns are allowed to be public or private across incomplete information and complete information treatments, respectively. Reischmann and Oechssler (2018) find that the BCCM leads to higher contribution rates compared to VCM. Using a behavioral model based on better response dynamics, they predict the stable long-run outcomes in the BCCM to be Pareto efficient, which is mostly confirmed by their experimental findings.

Another recent study by Oechssler et al. (2019) provides an extension of BCCM to a more general setting. While both BCCM and the setting in this paper allow commitments conditional on the commitments of other players, previous studies mostly allow choosing contributions conditional on the contributions of other players. As an example, Fischbacher et al. (2001) uses a variant of the strategy method and asks subjects to choose both an unconditional contribution and a conditional contribution based on the contributions of other players. They find that a substantial fraction of the subjects acts as conditional cooperators.

The second strand of studies mainly focus on mediation or commitments conditional on the commitments of other players in simultaneous games. Three important studies from this strand involve theoretical predictions relevant for the current study: Tennenholtz (2004), Kalai et al. (2010), and Monderer and Tennenholtz (2009). Among these, Tennenholtz (2004) studies a mechanism where the players in a prisoner's dilemma game can deploy their program, i.e., an algorithm of their strategy, on a central system simultaneously. For example, a mechanism in which players enter the following program into the system generates a cooperation equilibrium between players in the classic prisoner's dilemma game:

if P1 = P2 then DO (1, 0);
ELSE DO (0, 1);
STOP;

Here, P1 and P2 refer to the programs entered into the system, the DO $(a, 1-a)$ function refers to mixed strategy of playing "cooperate" with probability "a". Conditional cooperation, which is the efficient outcome, is ensured when both players enter the above program, and Tennenholtz (2004) refers to this as a "program equilibrium". Kalai et al. (2010) also show that endowing players with a large enough strategy set involving contingent commitments, allows players to sustain all kinds of cooperation equilibria. What is critical here is that, it is possible for players to choose commitment strategies conditional on other players' commitments. A real life example of this situation is the court conditions in which people (players) are represented by a lawyer, and the lawyer is instructed to choose her strategy conditional on the strategy of the other lawyer representing the opponent player.

The setup of the current study is mainly based on Monderer and Tennenholtz (2009). They introduce the *strong mediated equilibrium* which emerges when players are given the option to delegate their actions to a mediator who takes actions conditional on the delegation decisions of all players. The mediated game has a strong equilibrium, immune to coalitional deviations, and this equilibrium is called the strong mediated equilibrium. In Section 3, we demonstrate that the outcome of the conditional commitment mechanism in our setting constitutes a strong mediated equilibrium. In addition, in Section 3.3 we also discuss the relation between the BCCM mechanism by Reischmann and Oechssler (2018) and the strong mediated equilibrium concept by Monderer and Tennenholtz (2009). In a more recent theoretical study, Heitzig (2019) introduces the *conditional commitment function* (CCF) for simultaneous games. CCF's allow players to specify how to act conditionally on how other players commit to act. He shows that the modification of a certain class of games through the usage CCF's results in the equilibrium outcome to be in the core of the original game.

3 Model

In this section, we first describe the strong mediated equilibrium introduced by Monderer and Tennenholtz (2009) and then outline how our conditional commitment mechanism (CCM) implements a strong mediated equilibrium in the voluntary contribution mechanism (VCM) environment. As the binary conditional contribution mechanism (BCCM) by Reischmann and Oechssler (2018) is closely related to our study, we end the section with a description of the relationship between BCCM and the strong mediated equilibrium.

3.1 Strong Mediated Equilibrium

Let $\Gamma := (N, \{X_i\}_{i \in N}, \{u_i\}_{i \in N})$ be a finite strategic game, where $N := \{1, \ldots, n\}$ denotes the finite set of players, X_i is the nonempty strategy set of player i, and $u_i : \mathbf{X} \to \mathbb{R}$ is the payoff function of player i, where $\mathbf{X} = \bigtimes_{i \in N} X_i$.

A mixed strategy of a player i is a probability distribution θ_i over X_i, $\theta_i := \Delta(X_i)$. Similarly, the set of mixed strategy profiles of players is $\Theta = \bigtimes_{i \in N} \Delta(X_i)$.

A coalition is a nonempty subset of players. Let $S \subseteq N$ be a coalition and $\mathbf{X_S} = \bigtimes_{i \in S} X_i$ denote the set of strategy profiles for the coalition. Also, let S^c, the complement of S, denote the set of members outside the coalition. Then, for any coalition S, we can define a mixed strategy profile, $\Theta_S = \bigtimes_{i \in S} \Delta(X_i)$. Similarly, a correlated strategy for coalition S is defined as $c_S \in \Delta(X_S)$. The corresponding expected utility payoff of player is is denoted by U_i

In general, a strategy profile is a *strong equilibrium* if there is no coalition S for which all members can improve their utility by deviating from the selected strategy profile. Although, Monderer and Tennenholtz (2009) define three types of strong equilibrium, here we only focus on one type of these equilibria, which is sufficient for our analysis.

Definition 1. *Let $\theta \in \Theta$ be a mixed strategy profile where $\theta = (\theta_1, \ldots, \theta_n)$. Then, θ is a* ***strong equilibrium of type III*** *, if for every coalition S, and for every correlated strategy for S, $c_s \in \Delta(\mathbf{X_S})$, there exists a player $j \in S$ such that*

$$U_j(c_s \times (\bigtimes_{i \in S^c} \theta_i)) \leqslant U_j(\theta_1 \times \theta_2 \times \ldots \times \theta_n). \tag{1}$$

As it can be interpreted from Definition 1, strong equilibrium is immune to correlated deviations by coalitions. Hence, a successful deviation by a coalition S can occur only if all coalition members are strictly better off.

Before we define *strong mediated equilibrium*, we introduce the concept of *mediator*. A mediator is an entity that can act on behalf of a player when a player chooses to do so. In particular, players may send a message to the mediator, and the mediator decides on a pre-specified strategy to choose on behalf of the message sending players according to a prescribed action function. Note that the mediator does not enforce a behavior. Players may also choose not to send a message and play the game without the assistance of a mediator.

Definition 2. *A* ***mediator*** *for* Γ *is a tuple* $\mathcal{M} = (\{M_i\}_{i \in N}, \alpha = (\alpha_s)_{\varnothing \neq S \subseteq N})$*, where each* M_i *is a finite set of messages for every player* i*,* $M_i \cap X_i = \varnothing$*, and for every coalition S,* $\alpha_S : M_S \to \Delta(X_S)$*, where* $\alpha = (\alpha_s)_{\varnothing \neq S \subseteq N}$ *stands for the action function of the mediator.*

Let S be the set of players who choose to send a message to the mediator, then, the message profile for this group is $m_S = \{m_i\}_{i \in S}$, and the action function of the mediator is $\alpha_S(m_S) \in \Delta(X_S)$. In other words, the mediator determines the strategy profile $x_S \in \mathbf{X_S}$ for the members of S according to the probability distribution $\alpha_S(m_S)$. Any mediator $\mathcal{M}$ for Γ generates a new game called the mediated game, denoted by $\Gamma(\mathcal{M})$.

Definition 3. *Let* Γ *be a game in strategic form. A correlated strategy* $c \in \Delta(X)$ *is a* ***strong mediated equilibrium*** *if there exists a mediator for* Γ*,* $\mathcal{M}$*, and a vector of messages* $m \in M = \times_{i \in N} M_i$*, with* $\alpha_N(m) = c$*, such that* m *is a strong equilibrium of type III in* $\Gamma(\mathcal{M})$*. Such a mediator is said to strongly implement* c*.*

Definition 4. *A mediator is a* ***minimal mediator*** *if each message space is a singleton.*

Monderer and Tennenholtz (2009) prove that a minimal mediator can implement any strong mediated equilibrium in Γ.

3.2 The VCM Game and the Conditional Commitment Mechanism

We consider a set of players, $N = \{1, 2, , n\}$, where each player has an endowment of ω tokens. Players simultaneously choose their voluntary contribution, $x_i \in \{0, 1, \ldots, \omega\}$, to the public good and the rest of their endowment, $\omega - x_i$, is kept for themselves for private consumption. Players enjoy a utility of 1 for every token they keep for private consumption and a utility of γ_i for each token contributed by any player to the public good. The utility of an agent is as follows:

$$u_i = \omega - x_i + \gamma_i \sum_{j=1}^{n} x_j \tag{2}$$

In line with standard public good games, to make the problem interesting, we assume $\gamma_i \in [0, 1)$ for all $i \in N$, so that given the other players' contribution, the best response for player i is always contributing zero tokens. Further, we assume $\gamma_i n > 1$ for all $i \in N$, to ensure that full contribution by all players is a Pareto efficient outcome.

We now introduce a mediator, i.e., the conditional commitment mechanism, for this environment. CCM is defined as the mediator $\mathcal{M}^{\mathcal{CCM}} = (\{M_i\}_{i \in N}, \alpha = (\alpha_s)_{\varnothing \neq S \subseteq N})$ where the message space is a singleton, $M_i = \{r\}$ for all players. In words, players either choose to send the message r, hence partake in the mechanism, or choose a strategy x_i themselves. The game induced by the CCM is as follows. $\Gamma^{CMM} := (N, \{M_i\}_{i \in N}, \{U_i\}_{i \in N})$

The CCM we employ has a specific mediator that chooses to contribute the total endowment, ω, of players whenever all players choose to utilize the mediator, otherwise the mediator contributes 0 on behalf of the players who partake in the mechanism. In other words, the conditional commitment mechanism works as follows: "If all players choose message r, then all players contribute their endowment, ω, to the public good. Otherwise, those who choose message r contribute zero, and those who do not send a message choose their own contributions, $x_i \in \{0, 1, 2, ..\omega\}$".

Consider the following illustrative example. Let S be the set of players for whom $m_i = r$. Suppose all agents send the message r. Then, $S = N$, and $\alpha_S(m_S)$ prescribes the action $x_i = \omega$ for all $i \in N$. On the other hand, consider the case where at least one player chooses not to use the mediator. Then, $S \subset N$, and $x_i = 0$ for all $i \in S$.

The following proposition illustrates how CCM overcomes the free-riding problem for the public good game.

Proposition 1. *All players using CCM, $m_i = r \; \forall i$, constitutes a strong equilibrium if $\gamma_i < 1/n \; \forall i$.*

Proof. Definitions 1 and 3 imply that $m_i = r$ for all $i \in N$ is a *strong equilibrium* of the mediated game if there is no coalition that can strictly improve any of its members' payoff.

Consider the case where all players choose to partake in the conditional commitment mechanism. On the equilibrium path, everyone contributes the totality of their endowment to the public good, whereas, off the equilibrium path, everyone who partakes in the conditional commitment mechanism contributes 0.

Consider a deviation from this strategy by a coalition of players denoted by $S \subseteq N$, where $|S| = s$. Assume that each member of this coalition, while not sending a message, decides to contribute an amount of τ_i to the public good, where $0 \leqslant \tau_i \leqslant \omega \; \forall i$. Let τ denote the total contribution level of this coalition, $\tau = \Sigma_{i \in S} \tau_i$. Note that the players outside of the coalition, $j \in S^c$, will choose not to contribute to the public good, hence, $x_j = 0$ for all $j \in S^c$. [1]

Each member of the coalition S gets the following payoff.

$$u_{i \in S} = \omega - \tau_i + \gamma_i \tau \tag{3}$$

Then, $m_i = r$ for all $i \in N$ is a strong mediated equilibrium if

$$u_{i \in S}(\bigtimes_{i \in S} \tau_i \times \bigtimes_{j \in S^c} x_j) \leqslant u_{i \in S}(\omega_1 \times \omega_2 \times \ldots \times \omega_n) \tag{4}$$

[1]Note that because 0 is a dominant strategy for all players, there is no need to check whether players outside the coalition might deviate from the punishment strategy.

where $x_j = 0 \; \forall j \in S^c$. Thus, the following inequality must hold for all $i \in S$.

$$\begin{aligned} \omega - \tau_i + \gamma_i \tau &\leqslant \gamma_i n\omega \\ \omega - \tau_i &\leqslant \gamma_i (n\omega - \tau) \\ \frac{\omega - \tau_i}{n\omega - \tau} &\leqslant \gamma_i \end{aligned} \tag{5}$$

Since $|S| = s < n$, τ_i/τ is smaller than or equal to $1/n$ for at least one $i \in S$. Consequently, the left hand side is smaller than or equal to $1/n$ for at least one $i \in S$. As $\gamma_i > 1/n$ for all $i \in N$ by an earlier assumption, the inequality holds for at least one $i \in S$; hence, she does not benefit from deviating. Because at least one member of the coalition does not strictly benefit from the deviation, we conclude that $m_i = r$ for all $i \in N$ is a strong equilibrium of the mediated game. □

Note that full contributing by all players is Pareto efficient outcome of the original game whenever $\gamma > 1/n$. However, without the CCM full contribution is not an equilibrium of the original game. Therefore, by utilizing CCM players benefit from committing conditional on other players' commitments.

Note that the voluntary contribution mechanism with a mediator described as before can be extended to include semi-contributions in case of a commitment. However, whether the resulting mechanism constitutes a strong mediated equilibrium requires a stricter bound on the public good valuation, γ_i. Let $k \in (0, 1]$ define the endowment multiplier that characterizes the contribution in case of a commitment. Similar to before, the message signals, "If all other players choose message r, then, all players contribute a *pre-determined* portion k of their endowment, $k\omega$, to the public good. Otherwise, those who choose message r contribute zero, and those who do not send a message choose their own contributions, $x_i \in \{0, 1, 2, ..\omega\}$". The following corollary formalizes this observation.

Corollary 1. *All players using the conditional commitment mechanism is a strong equilibrium if* $\gamma_i \geqslant 1/nk \; \forall i$.

Proof. The only change from Proposition 1 is the level of contribution when an agreement occurs. Revising the proof by taking this change into account gives us the following inequality:

$$\frac{1}{nk} \leqslant \frac{\omega - \tau_i}{nk\omega - \tau} \leqslant \gamma_i \tag{6}$$

□

Note that the lower bound on γ_i decreases with k. This result implies that given the preferences captured by the specified utility, it is easier to reach an agreement if the contribution enforced by the mediator in case of commitment is higher. The earlier assumption we made, $\gamma_i < 1/n$, guarantees that full commitment mechanism is a strong mediated equilibrium. However, lower commitment inducing mechanisms are not guaranteed to constitute an equilibrium.

3.3 Strong Mediated Equilibrium and the BCCM

In this section, we will outline the relationship between the binary conditional contribution mechanism (BCCM) by Reischmann and Oechssler (2018) and the strong mediated equilibrium by Monderer and Tennenholtz (2009). While the original paper introducing BCCM focuses on the multi-period dynamics of the mechanism, here, we will consider a one-shot game.

BCCM is defined for an environment where $N = \{1, 2, ..., n\}$ is the set of players, and each player has an endowment of 1 token, which she can keep for private consumption or invest in the public good. Let $z_i \in Z_i = \{0, 1\}$ denote the strategy of a player where $Z_i = \{0, 1\} \forall i$. An outcome is defined as $z = \{z_1, z_2, ..., z_n\}$, with $z_i = 1$ indicating that player i invests her entire endowment to the public good and $z_i = 0$ indicating that she keeps all of her endowment for private consumption. Players have potentially different valuations (γ_i) for the public good and the utility of an agent can then be stated as:

$$u_i = 1 - z_i + \gamma_i \sum_{j=1}^{n} z_j \tag{7}$$

In the original setup, the outcome z defined above is the result of simultaneously chosen messages by each player, denoted by m_i. Here, $m_i \in M_i = \{1, 2, ...n + 1\}$ and choosing $m_i = k$ means *"I am willing to contribute to the public good if at least k agents (including myself) contribute in total."*. This message space implies that $z_i = 1$ whenever $m_i = 1$, and $z_i = 0$ whenever $m_i = n + 1$, regardless of the messages by other players. Hence, in order to allow players to choose their strategies by not sending a message, we need to slightly modify the original setup to have a mediated game as defined in section 3.

In this context, agents choose either to send a message $m_i \in M_i = \{2, 3, ..., n\}$ or not to send a message and play their own strategy $z_i \in Z_i = \{0, 1\}$. When the set of players sending a message is S, the message profile is $m_S = \{m_i\}_{i \in S}$ and , the outcome function of the mediator is $\alpha_S(m_S) \in \Delta(Z_S)$. For each $i \in S$, $\alpha_S(m_S)$ assigns probability 1 to $z_i = 1$ whenever $\sum_{j \in S} \mathbb{1}_{m_j \leqslant m_i} \geqslant m_i$ and probability 1 to $z_i = 0$, otherwise.

To illustrate BCCM further, consider example 2 from Reischmann and Oechssler (2018). $N = \{1, 2, 3, 4, 5\}$ and $\gamma_i = 0.4$ for $\forall i \in N$. In this setup there is one strong equilibrium where $z_i = 1, \forall i \in N$. In the mediated game where $m_i \in M_i = \{2, 3, 4, 5\}$, this outcome can be implemented when $m_i = 5$ for $\forall i \in N$. Alternatively, it could also be implemented with a singleton message space for players, where $M_i = \{5\}$. Note that this equilibrium is not unique and there are other Nash equilibria in the original game such as the case where any three player chooses $z_i = 1$ and the remaining two players choose $z_i = 0$. This type of an equilibrium could also be implemented in the mediated game, for example, by letting the message sending players be $S = \{1, 2, 3\}$, setting $m_i = 3, \forall i \in S$, and $z_i = 0$ for $i = 4$ and $i = 5$.

3.4 Extended Message Space

We previously described in section 3.2 the conditional commitment mechanism and the resulting mediated game using a minimal message space. In particular, each member could choose to send a message from a singleton set or choose her own contribution to the public good. However, in the experimental design, we use an extended message space $M_i = \{r_0, r_1, ..., r_\omega\}$, and a choice of the message r_k implies *"I am willing to contribute my endowment, ω, to the public good if all agents (including myself) choose to send a message; otherwise I will contribute k units out of ω"*. Notice that the message r_0 in this message space corresponds to the singleton message space defined in section 3.2.

Monderer and Tennenholtz (2009) shows that every strong mediated equilibrium in a finite strategic game can be implemented by a minimal mediator. The following Lemma 1 shows that the minimal mediator, CMM, introduced in section 3.2 can implement any strong mediated equilibrium in the mediated game with the extended message space.

Lemma 1. *Let $\mathcal{M}^{\mathcal{E}} = (\{M_i^E\}_{i \in N}, \alpha = (\alpha_s)_{\varnothing \neq S \subseteq N})$ where M_i^E denotes the extended message space, $M_i^E = \{r_o, r_1, ..., r_\omega\}$ $\forall i$ and r_i is described as above. Every strong equilibrium of type III in the mediated $\Gamma(\mathcal{M}^{\mathcal{E}})$ can be implemented by CMM.*

Proof. Let $\alpha' \in \Delta(\mathbf{X})$ be a strong mediated equilibrium implemented by $\mathcal{M}^{\mathcal{E}}$ by the profile $\mathbf{m} \in \mathbf{M^E}$. Hence, $\mathbf{m}$ is strong equilibrium of type III in the mediated game $\Gamma(\mathcal{M}^{\mathcal{E}})$ and $\alpha_{\mathbf{N}}(\mathbf{m}) = \alpha'$ [2]. Therefore, for any coalition S' which consists of members who do not send a message and contribute $x_i = \tau_i$ $\forall i \in S$

$$u_{i \in S'}(\bigtimes_{i \in S} \tau_i \times \bigtimes_{j \in S^c} x_j) \leqslant u_{i \in S}(\alpha'_1 \times \alpha'_2 \times \ldots \times \alpha'_n) \tag{8}$$

where $x_j = k$ such that $m_j = r_k$. Now consider CMM which restricts the set of messages for every i to $M_i = \{r_0\}$. Any member of a deviating coalition S' will get the payoff $u_{i \in S'}(\bigtimes_{i \in S} \tau_i \times \bigtimes_{j \in S^c} x_j)$ where $x_j = 0$ $\forall i \in S^c$.

$$u_{i \in S'}(\bigtimes_{i \in S} \tau_i \times 0 \times 0 \times 0) \leqslant u_{i \in S'}(\bigtimes_{i \in S} \tau_i \times \bigtimes_{j \in S^c} x_j) \leqslant u_{i \in S}(\alpha'_1 \times \alpha'_2 \times \ldots \times \alpha'_n) \tag{9}$$

Therefore, any strong mediated equilibrium implemented by the extended message space can also be implemented by CMM. □

The intuition of the proof is as follows. Note that on the equilibrium path all messages lead to the same outcome where all players contribute their endowment to the public good, hence, all messages r_k such that $k > 0$ is redundant, using the extended message space is equivalent to using the minimal mediator with the singleton message space. Whereas, off the equilibrium path, k such that $m_i = r_k$ works as a punishment strategy. As r_0 provides the harshest punishment any strategy profile that constitutes an equilibrium with a "nicer" ($k > 0$) punishment will readily constitute an equilibrium with the harshest ($k = 0$) punishment. Due to this equivalency, theoretical predictions regarding the strong equilibrium outcome in the mediated game remain unchanged. At the same time, the extended message space allows us to control observe the extent players send *unexploitable messages*, which we describe in more detail later in Section 5.

[2]Because we are implicitly considering full contribution when all players send a message, $\alpha' = \omega$

4 Experimental Design and Procedures

All experimental sessions were conducted at the economics laboratory of Boğaziçi University, İstanbul during May and November, 2018. Subjects were undergraduate students. The registration was organized with ORSEE (Greiner 2015) and the experiment was conducted using z-Tree (Fischbacher 2007). We conducted a total of 8 sessions each with 12 participants. Among these 96 subjects, 44 indicated their gender as female and 52 as male.

Each session began with the presentation of instructions on the computer screen and this was followed by questions for understanding. Subjects had to answer each question correctly in order to proceed. When this stage was completed, we proceeded with the experimental treatment. Among 8 sessions conducted, we had four sessions with full commitment we refer to as the FULL treatment, and four sessions with half commitment we refer to as the HALF treatment. 12 subjects in each of the sessions were randomly assigned to three groups of four members, and they played with the same group members for 10 consecutive periods. At each period, subjects were given an endowment of 20 points and decided how much of this endowment to invest in a group project and how much of it to keep for themselves. Every point kept by the subjects resulted in a payoff of one points for this subject, whereas each point invested in the group project resulted in a payoff of 0.4 points for each member of the group, i.e., the marginal per capita (MPCR) return equals to 0.4. The main task for the subject was making two decisions and we asked them to make these decisions on the same screen: (i) whether they consent for the *conditional contribution agreement*, and (ii) how much they would invest in the group project in case an agreement is not reached.

The conditional commitment agreement worked as follows: If all group members consented for this agreement, then each member automatically contributed a default amount set according to the treatment. In the FULL treatment, this amount was 100% of the endowment (20 points) and in the HALF treatment, this amount was 50% of the endowment (10 points). In case at least one group member did not consent to agree, each subject invested the amount he or she chose to be implemented in case an agreement could not be reached, as mentioned in decision (ii) above. Subjects made their decisions simultaneously and independently. Subject earnings were cumulative across periods and the period earnings for a subject was calculated as follows:

Payoff = (20 - individual investment to the group project) + 0.4 * (total investment to the group project)

At the end of each period, subjects could see how many of their group members consented for the agreement, the resulting total investment in the group project, their earnings for that period and their cumulative earnings. In a follow-up screen, they could see the same information for individual levels, yet the order of the members was randomized. Each session lasted nearly 40 minutes. Subjects were paid in cash and privately at the end of the each session. Points they earned in the game were converted into Turkish Liras (TL) as 10 points = 1 TL. In addition to their experimental earnings, Subjects were also paid 10 TL as the participation fee. On average, players in the FULL treatment earned 39.7 TL while in the HALF treatment, this average was 34.8 TL[3]

[3]At the time the experiment was conducted, 1 US Dollar was worth 5.40 TL.

5 Results

There are two key variables of interest in our analysis: the frequency of commitment agreement in groups, and the contribution levels chosen by the subjects to be implemented in case of the group not reaching a commitment agreement.

5.1 Commitment Agreements

In Figure 1, we present for both treatments, the number of groups reaching an agreement, as well as the overall frequency of consent for conditional commitment agreement across periods. Both treatments start at similar consent frequencies (around 80%), but the FULL treatment exhibits relatively higher consent frequencies starting from period 3 on and a vast majority of the groups engage in successful commitment in FULL. In the HALF treatment, agreement frequencies fluctuate at relatively lower levels and around 50% of the groups reach an agreement during different periods of this treatment. In Table 3 of Appendix section A, we present these numbers in detail for both treatments. Naturally, the observed differences between two treatments also exist at the group level. Figures 6 and 7 in Appendix section B present the frequency of agreement for distinct groups in each treatment. In the HALF treatment, the number of groups reaching agreement more than half of the time is 5 out of 12, whereas this number is 10 out of 12 for the FULL treatment. Nonparametric tests using matching groups as independent units of observations also confirm the general pattern we describe above. In particular, while there is no significant difference across treatments during period 1 in terms of consent frequency (Wilcoxon rank-sum (Mann-Whitney) test, $p = 1.000$) or reaching agreement (Wilcoxon rank-sum (Mann-Whitney) test, $p = 1.000$), we find a significant difference when all periods are considered, both in terms of consent frequency (Wilcoxon rank-sum (Mann-Whitney) test, $z = -2.79, p < 0.01$) and agreement (Wilcoxon rank-sum (Mann-Whitney) test, $z = -2.705, p < 0.01$).

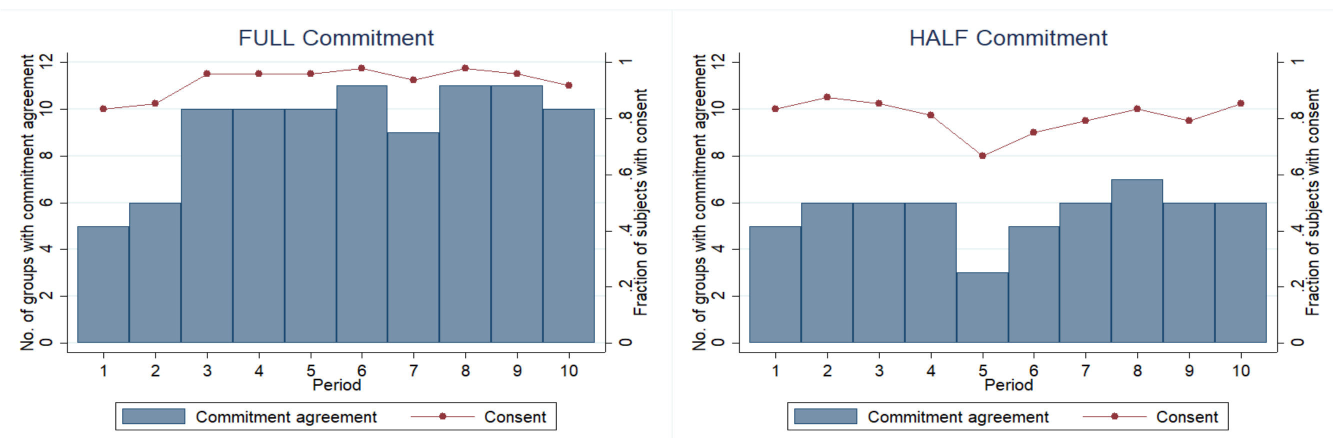

Figure 1: Evolution of consents and commitment agreements

Result 1. *While initial consent rates and agreements are similar in both treatments, over all periods they are significantly higher in the FULL treatment.*

5.2 Contributions

We continue with the contribution decisions of subjects that would be effective only if an agreement in their group is not reached. In Figure 2 we present the time pattern for this variable, conditional on whether the subject indicated consent for commitment or not.[4]

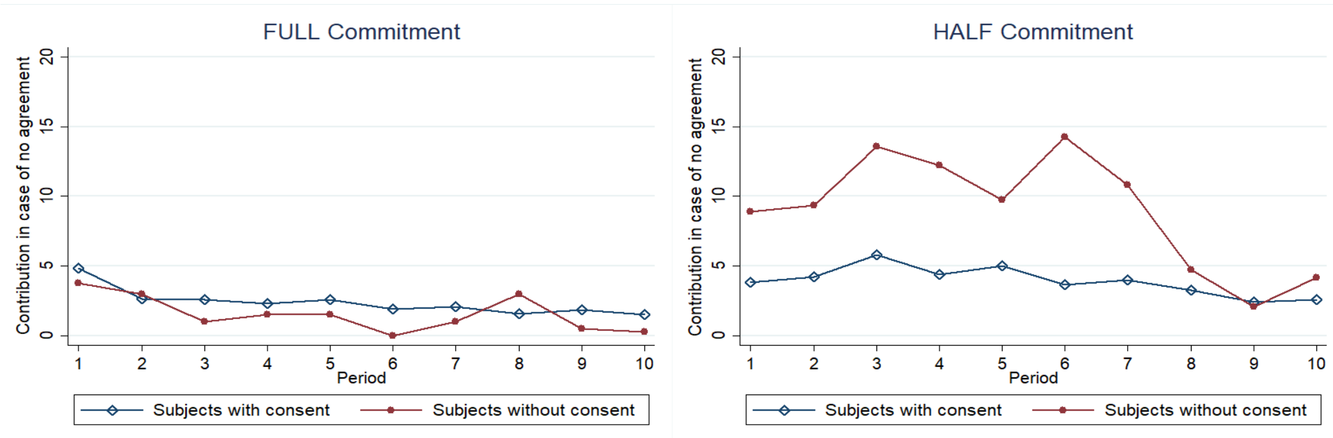

Figure 2: Evolution of contribution decisions, effective only in case of disagreement

In the FULL treatment, contributions start at low levels regardless of the consent decisions of subjects and generally decay for both groups of subjects. In the HALF treatment a more interesting pattern emerges. Contribution decisions of subjects who gave consent for commitment start at low levels but the extent of decay is slightly less pronounced compared to the FULL treatment. Subjects who did not give consent for commitment in HALF exhibit the highest contributions in the experiment. This group is relatively small in number since the majority of subjects give consent. However, a certain group of subjects among the non-consent-givers perhaps try to convey a particular message. They do not give consent for conditional commitment agreement but simultaneously choose to contribute more than the commitment amount in HALF (50% of the endowment). This behavior can be observed for the first 7 periods and it is also clearly visible in the separate histograms of contributions across two treatments and commitment decisions presented in Figure 8 of Appendix section B. Further support for this conjecture comes from the comparison of member contributions and the average contribution of the group in cases where groups couldn't reach an agreement.

At a given period, we refer to a subject as a net contributor (free-rider) if her contribution was more (less) than the group average. During periods without agreement in the HALF treatment, subjects who gave consent for agreement were net contributors 31% of the time, whereas those who didn't were net contributors 44% of the time. In the FULL treatment, the respective percentages were 29% and 19%. We

[4] The respective numbers are also presented in Table 4 in Appendix section A.

conclude that pro-social types were more likely to refrain from giving consent to conditional commitment agreement in the HALF treatment, probably to signal the potential of a Pareto improvement over the commitment default.

What is the effect of commitments on effective (realized) contributions after all, i.e., what are the actually realized contributions independent of whether a commitment agreement was reached or not. Figure 3 shows that in the FULL treatment effective contributions are clearly higher. In FULL from period 3 on, the average contributions are 75 percent or higher with respect to full contributions while in HALF they are always lower than 50 percent. This relatively large difference between treatments is significant (Wilcoxon rank-sum (Mann-Whitney) test, $z = -3.812, p < 0.01$).

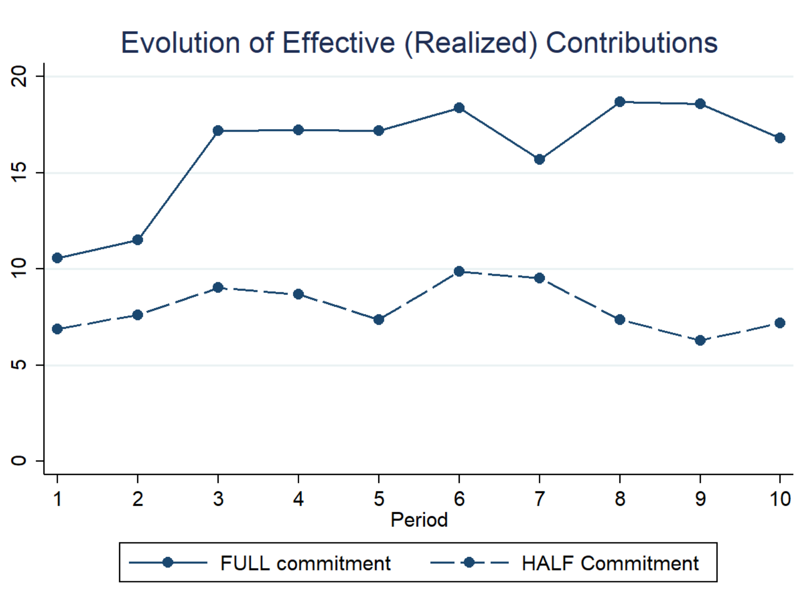

Figure 3: Effective Contributions

Result 2. *In FULL, contributions in case of disagreement for commitment do not differ among subjects with respect to given consent. In HALF, however, subjects who do not give consent make higher contribution decisions than subjects who give consent. Overall, effective contributions, i.e., contributions that are actually made are much higher in FULL.*

5.3 Payoffs

How do commitment and contribution decisions translate into payoffs across different treatments? Due to the structure of the payoffs and the commitment mechanism, the payoff for groups reaching an agreement in the FULL treatment is constant and equal to 32 $(= 20 * 1.6)$ while in the HALF treatment, it is also constant and equal to 26 $(= 10 + 10 * 1.6)$. For groups who could not reach an agreement, the mean payoff was 23.7 in the HALF treatment and 21.8 in the FULL treatment, both of which are only slightly higher than the initial endowment of the subjects (20 points) in every period. In Figure 4, we present the evolution of payoffs in both treatments.

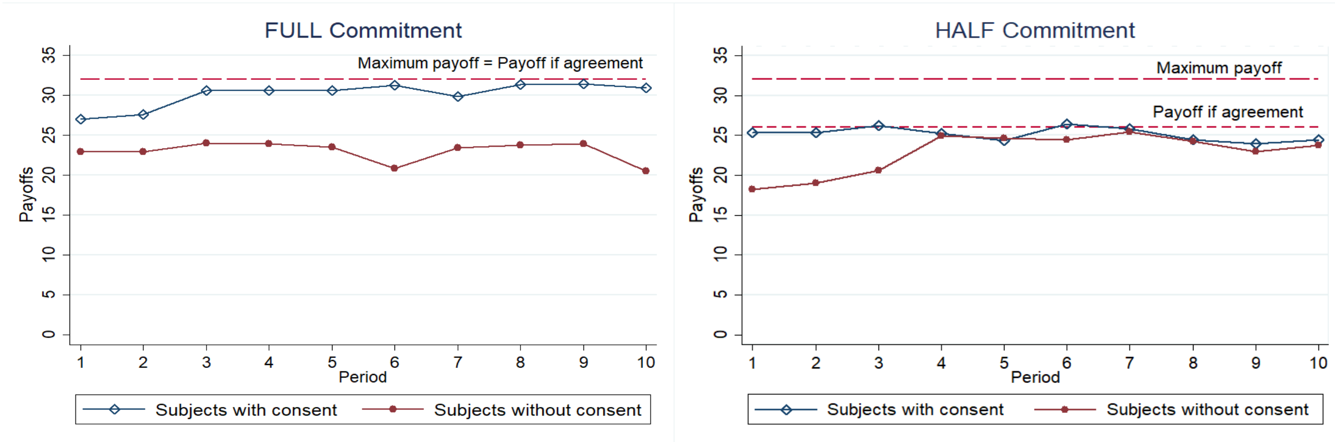

Figure 4: Evolution of Payoffs

In the FULL treatment, subjects with consent for commitment agreement start with slightly higher payoffs compared to those without consent. Later on, due to high frequency of successful agreements, this difference grows and subjects with consent get closer to the maximum possible payoff of 32 units for this treatment. In the HALF treatment, the initial differences show a similar pattern, yet this is potentially due to those subjects who do not give consent to the agreement and chose to contribute relatively high amounts instead. As the number of subjects with consent, and the contribution of these subjects decline, the differences in payoffs of two groups is mitigated and they both stabilize slightly below 25 units. Averaged over all subjects, payoffs are significantly higher in FULL than in HALF (Wilcoxon rank-sum (Mann-Whitney) test, $z = -3.812, p < 0.01$)

Result 3. *With respect to given consent, while in FULL payoffs differ, in HALF, they do not. Overall, payoffs in FULL are significantly higher than in HALF.*

As for the individual subjects, a key determinant of the payoffs was the extent that the subject chose the dominant strategy in the game: Giving consent to the conditional commitment agreement and setting the contribution to zero in case the group can not reach an agreement. As in Reischmann and Oechssler (2018), we term these choices as *unexploitable messages*. The choice of unexploitable messages start at relatively low frequencies, around 15% for the HALF treatment and slightly above 20% for the FULL treatment. As Figure 5 shows, the respective frequencies keep increasing throughout the game and reach above 50% for both treatments as of Period 10.

5.4 Individual Results

Next, in Table 1, we present the results from a series of probit regressions for the determinants of consent to conditional commitment agreement. Our independent variables are a dummy for the FULL treatment, a variable indicating the period in the game, a dummy for subject's consent for agreement in the previous period, a dummy for the groups' agreement decision from the previous period, subject's contribution

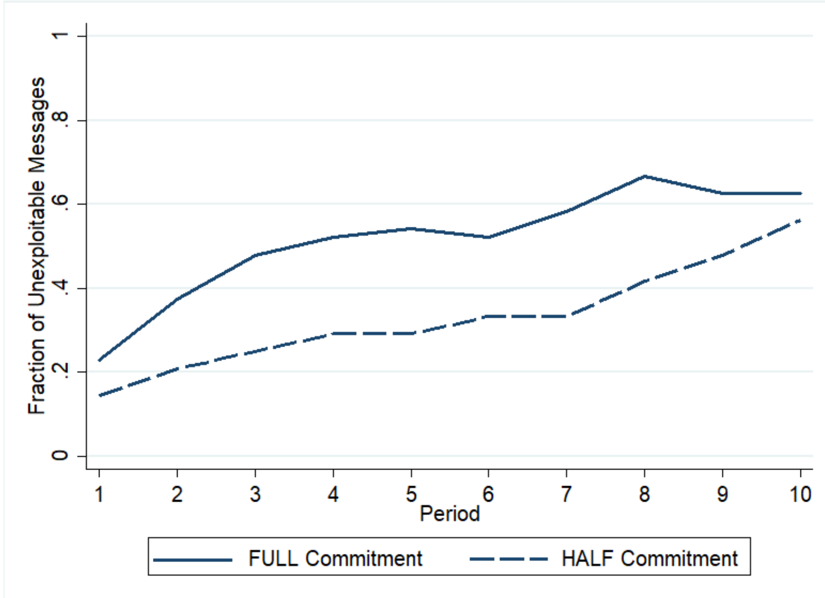

Figure 5: Unexploitable Messages

decision from the previous period to be implemented in case an agreement was not reached for that period, and finally our control variables for which detailed information is presented in Appendix section C.

Table 1: Determinants of Commitment Decisions

	(1)	(2)	(3)	(4)
Full Commitment	0.129***	0.114***	0.0862***	0.0817***
	(0.042)	(0.033)	(0.024)	(0.024)
Period	0.00382	0.00385	-0.00136	-0.000990
	(0.005)	(0.005)	(0.005)	(0.005)
Consent in $t-1$			0.172***	0.159***
			(0.037)	(0.033)
Group agreement in $t-1$			0.00161	0.00121
			(0.027)	(0.025)
Chosen contribution in $t-1$			-0.00545***	-0.00515***
			(0.002)	(0.002)
Controls	No	Yes	No	Yes
N	960	960	864	864

Marginal effects from probit regressions reported. Dependent variable: 1=Consent, 0=No consent.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In line with the aggregate statistics reported above, the FULL treatment significantly increases the likelihood of consent to agreement whereas the period in the experiment has no significant effect. As for the independent variables regarding subject's experience from the previous period, we see that the previous consent decision is a significant predictor of consent in the current period as well. While the group's agreement decision from the previous period has no significant effect on the consent in the current period, previous period's contribution decision (for the case of no agreement) has a small but significant negative effect. This is mostly due to subjects in HALF, who attempt to signal the possibility of a Pareto improvement by choosing a contribution level higher than the commitment default of 50% (10 out of 20 tokens).

We continue with a regression analysis on the determinants of contribution decisions of subjects. Note that subjects were asked to specify a contribution amount regardless of their consent to the commitment agreement. Our main objective here is to see how chosen contributions are jointly affected from the treatment (contribution level in case of agreement) and from subject's consent to commitment when we control for other relevant factors (specified in Appendix C). Results of this regression analysis is summarized in Table 2.

In line with the aggregate statistics we provided above, the significant coefficients of the FULL treatment and consent dummies as well as their interaction points out that chosen contributions are on average lower in the FULL treatment. In this treatment chosen contributions are higher for those who give consent to commitment agreement whereas the opposite is true for the HALF treatment. Consequently the lowest contributions are observed among those who do not give consent to commitment agreement in FULL, whereas the highest contributions are observed among those who do not give consent to commitment agreement in HALF. This, as we mentioned before, is due to a small group of subjects aiming for a Pareto improvement for their group above the level specified by the commitment default of 50% for this treatment.

Table 2: Determinants of Contribution Decisions

	(1)	(2)
Full Commitment	-7.481***	-7.381***
	(2.178)	(2.096)
Consent	-5.337***	-5.255***
	(1.737)	(1.767)
Full Commitment * Consent	5.965***	5.975***
	(1.998)	(2.034)
Period	-0.298***	-0.298***
	(0.074)	(0.075)
Constant	10.87***	8.990***
	(2.126)	(2.867)
Controls	No	Yes
N	960	960

Coefficients from OLS regressions reported. Dependent variable: Chosen Contribution.

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusion

How to increase and sustain cooperation in social dilemma situations? In this paper, we show that a mechanism involving conditional commitments handled by a mediator can help. We adapt theoretically the notion of strong mediated equilibrium proposed by Monderer and Tennenholtz (2009) to the case of public good provision. Using experiments, we demonstrate that this commitment mechanism can indeed increase cooperation. If a strong equilibrium exists, like in our FULL treatment, almost all groups use commitments and coordinate on this cooperative equilibrium. However, when conditional commitments only allow for socially inefficient outcomes like in our HALF treatment, we observe a decline in their utilization, and overall cooperation decreases.

When the outcome of the conditional commitment mechanism is a strong equilibrium, i.e., immune to coalitional deviations, we observe almost universal acceptance. Hence the overall efficiency level is very high. Contrary to this, when the outcome of this mechanism is not a strong equilibrium, a substantial fraction of players opt out of agreements and choose strategies that would, when followed by all group members, be Pareto improving over the mechanism's outcome. However, since the voluntary contribution mechanism suffers from free-riding, these individual choices fade out over time. Players also seem to learn to rationally use the conditional commitment device in a way to increase their personal benefit.

A functional conditional commitment mechanism might seem a bit far-fetched since it requires a collection of messages from multiple parties and implementing the resulting outcome in a binding fashion. However, as computational techniques are more frequently used for tackling social problems, as in the emergent new field of social computing, and new technologies such as smart contracts grow in use, conditional commitments will be easier to implement, through decentralized, automated and enforceable means. Roth and Shorrer (2021) demonstrate that introduction of a mediator, through which market participants can delegate their decision rights via individual messages, can improve the market outcome for these participants in environments like decentralized matching markets and ride-sharing platforms. Holden and Malani (2021) argue that when smart contracts become widely available, they can be utilized as effective remedies against the classical hold-up problem observed in many situations involving relation-specific investments.

Another interesting recent example that is similar in nature to conditional commitments is the National Popular Vote bill proposed for united States Presidential Elections (https://www.nationalpopularvote.com/). For states where this bill is enacted into law, electors will be the supporters of the candidate winning the national popular vote. When enough states commit to this, that is when committing states have at least 270 presidential electors, than this the outcome of these commitments practically implies the election of president by national popular vote. This example, along with other applications such as those using smart contracts, has recently been discussed in Dafoe et al. (2020).

To sum up, we contribute to the public goods literature by theoretically analyzing and experimentally demonstrating that conditional commitments handled by a mediator can be a promising mechanism to increase overall cooperation in social dilemma situations, particularly public goods.

References

Bochet, Olivier, Talbot Page, and Louis Putterman, "Communication and punishment in voluntary contribution experiments," *Journal of Economic Behavior & Organization*, 2006, *60* (1), 11–26.

Dafoe, Allan, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson, and Thore Graepel, "Open Problems in Cooperative AI," 2020.

Fehr, Ernst and Simon Gächter, "Cooperation and punishment in public goods experiments," *American Economic Review*, 2000, *90* (4), 980–994.

_ and _ , "Altruistic punishment in humans," *Nature*, 2002, *415* (6868), 137–140.

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

_ , Simon Gächter, and Ernst Fehr, "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics letters*, 2001, *71* (3), 397–404.

Gächter, Simon and Christian Thöni, "Social learning and voluntary cooperation among like-minded people," *Journal of the European Economic Association*, 2005, *3* (2-3), 303–314.

Greiner, Ben, "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.

Gunnthorsdottir, Anna, Roumen Vragov, Stefan Seifert, and Kevin McCabe, "Near-efficient equilibria in contribution-based competitive grouping," *Journal of Public Economics*, 2010, *94* (11-12), 987–994.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach, "The competitive advantage of sanctioning institutions," *Science*, 2006, *312* (5770), 108–111.

_ , _ , and _ , "On cooperation in open communities," *Journal of Public Economics*, 2014, *120*, 220–230.

Heitzig, Jobst, "Efficient Non-Cooperative Provision of Costly Positive Externalities via Conditional Commitments," *Available at SSRN*, 2019.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter, "Antisocial punishment across societies," *Science*, 2008, *319* (5868), 1362–1367.

Holden, Richard and Anup Malani, *Can Blockchain Solve the Hold-up Problem in Contracts?* Elements in Law, Economics and Politics, Cambridge University Press, 2021.

Isaac, R Mark and James M Walker, "Group size effects in public goods provision: The voluntary contributions mechanism," *The Quarterly Journal of Economics*, 1988, *103* (1), 179–199.

Kalai, Adam Tauman, Ehud Kalai, Ehud Lehrer, and Dov Samet, "A commitment folk theorem," *Games and Economic Behavior*, 2010, *69* (1), 127–137.

Kocher, Martin G, Todd Cherry, Stephan Kroll, Robert J Netzer, and Matthias Sutter, "Conditional cooperation on three continents," *Economics letters*, 2008, *101* (3), 175–178.

Ledyard, John O, "Public Goods: A Survey of Experimental Research," *The Handbook of Experimental Economics*, 1995, pp. 111–194.

Monderer, Dov and Moshe Tennenholtz, "Strong mediated equilibrium," *Artificial Intelligence*, 2009, *173* (1), 180–195.

Oechssler, Joerg, Andreas Reischmann, and Andis Sofianos, "The conditional contribution mechanism for repeated public goods: The general case," Technical Report, WZB Discussion Paper 2019.

Ostrom, Elinor, James Walker, and Roy Gardner, "Covenants with and without a sword: Self-governance is possible," *American political science Review*, 1992, *86* (2), 404–417.

Reischmann, Andreas and Joerg Oechssler, "The binary conditional contribution mechanism for public good provision in dynamic settings—Theory and experimental evidence," *Journal of Public Economics*, 2018, *159*, 104–115.

Rivas, M Fernanda and Matthias Sutter, "The benefits of voluntary leadership in experimental public goods games," *Economics Letters*, 2011, *112* (2), 176–178.

Roth, Benjamin N and Ran I Shorrer, "Making marketplaces safe: Dominant individual rationality and applications to market design," *Management Science*, 2021, *67* (6), 3694–3713.

Tennenholtz, Moshe, "Program equilibrium," *Games and Economic Behavior*, 2004, *49* (2), 363–373.

A Additional Tables

Table 3: Frequency of consent among subjects and realized number of agreements in groups

	Half Commitment			Full Commitment		
Period	Consent	No Agreement	Agreement	Consent	No Agreement	Agreement
1	83.33	7	5	83.33	7	5
2	87.50	6	6	85.42	6	6
3	85.42	6	6	95.83	2	10
4	81.25	6	6	95.83	2	10
5	66.67	9	3	95.83	2	10
6	75.00	7	5	97.92	1	11
7	79.17	6	6	93.75	3	9
8	83.33	5	7	97.92	1	11
9	79.17	6	6	95.83	1	11
10	85.42	6	6	91.67	2	10
All	80.62	64	57	93.33	27	93

Table 4: Contributions by subjects according to consent to agreement

	Half Commitment		Full Commitment	
Period	Consent	No Consent	Consent	No Consent
1	3.8	8.9	4.8	3.8
2	4.2	9.3	2.7	3.0
3	5.8	13.6	2.6	1.0
4	4.4	12.2	2.3	1.5
5	5.0	9.8	2.6	1.5
6	3.6	14.3	1.9	0.0
7	4.0	10.8	2.1	1.0
8	3.3	4.8	1.6	3.0
9	2.4	2.1	1.8	0.5
10	2.6	4.1	1.5	0.3
All	3.9	9.2	2.4	2.1

B Additional Figures

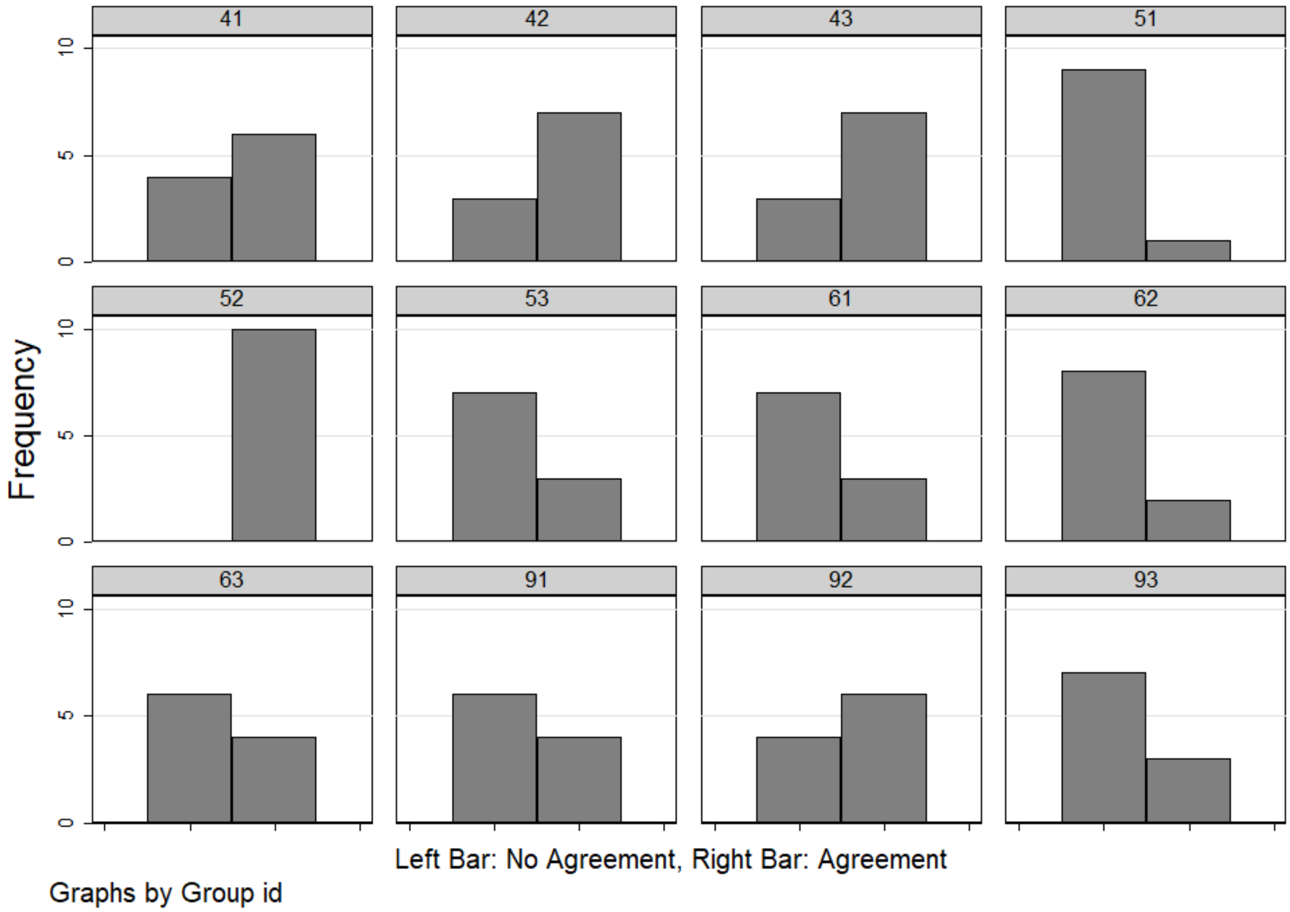

Figure 6: Agreement across groups in Half Contribution treatment

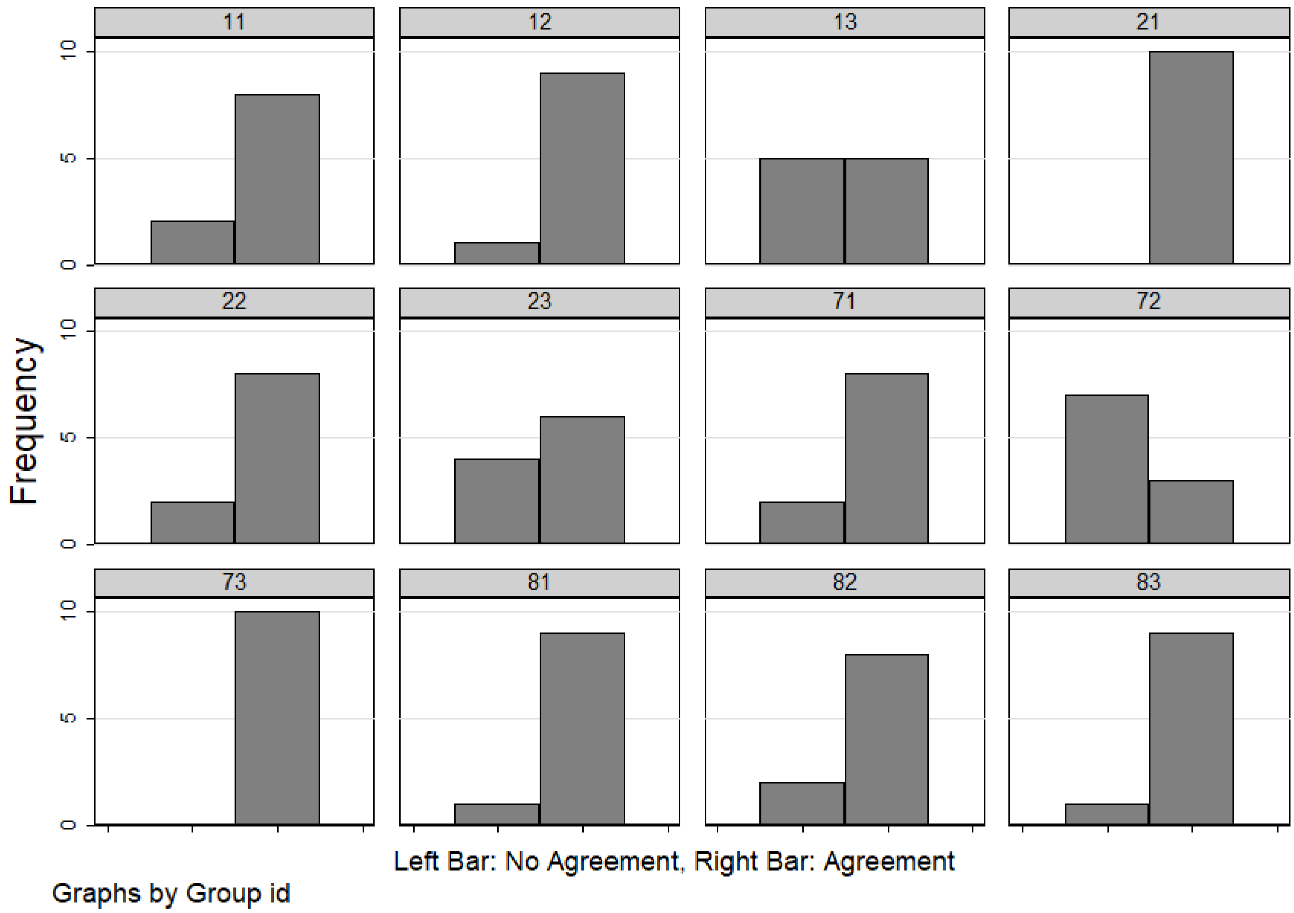

Figure 7: Agreement across groups in the FULL commitment treatment

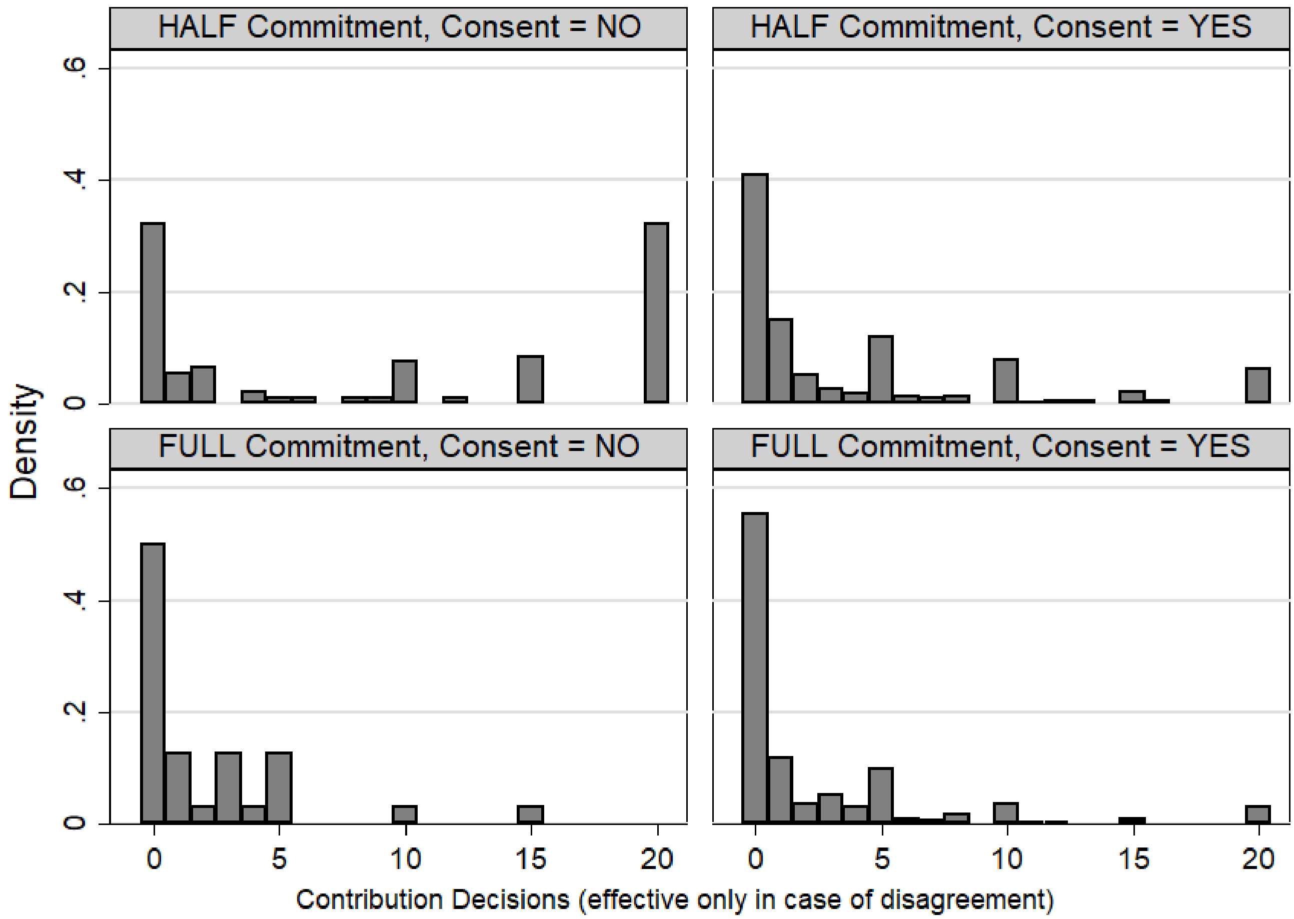

Figure 8: Contribution histogram

C Control Variables

Variable	Range	Definition
Male	$0/1$	1 if subject indicates gender as male.
Siblings	$[0, 5]$	number of siblings of the subject
Econ	$[0, 4]$	number of economics courses taken by the subject, censored at 4.
Friends	$[0, 4]$	number of people known in the session
Risk	$[1, 9]$	response to general risk question: "How willing are you to take risks in general?" (0 lowest -10 highest)
Trust	$0/1$	response to general trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"
Member	$0/1$	membership in organizations such as clubs, associations, foundations, political parties
Civic	$[5, 48]$	Sum of Civic 1 to Civic 5 indicated below, where subjects rate how justified they think the following behavior is
Civic 1	$[1, 10]$	claiming government benefits to which you are not entitled
Civic 2	$[1, 10]$	avoiding a fare on public transport
Civic 3	$[1, 10]$	cheating on taxes if you have the chance
Civic 4	$[1, 10]$	keeping money that you have found
Civic 5	$[1, 10]$	failing to report damage you have done accidentally to a parked vehicle

Table 5: Ranges and definitions for control variables used in the regressions.

D English translation of experimental instructions

The instructions below are translated from those used in **Full Commitment** treatment and they are a modified version of the instructions used in Herrmann et al. (2008). All instructions were presented on the computer screen. The instructions from the other treatments are similar and available upon request from the corresponding author.

General Information:

You are now taking part in an economic experiment studying how people make decisions under certain situations. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. This amount will be paid to you in cash at the end of the experiment.

It is prohibited to communicate with the other participants during the experiment. Should you have any questions, please ask us. If you violate this rule, you will be dismissed from the experiment.

During the experiment we will not speak in terms of TL, but in Points. During the experiment your entire earnings will be calculated in Points At the end of the experiment the total amount of Points you have earned will be converted to [national currency] at the following rate:

1 point = 10 Kuru(0.1 TL)

At the end of the experiment, you will also receive an extra 10 TL in cash as the participation fee in addition to your earnings in the experiment .

The experiment is divided into 10 separate rounds. In each round the participants are divided into groups of four. You will therefore be in a group with 3 other participants. The composition of the groups will stay the same for all ten rounds. In the following pages we describe the experiment in detail. Please click on OK when you are ready.

Decision Stage

At the beginning of each round each participant receives 20 tokens. We call this his or her endowment. You have to decide how many of the 20 tokens you want to contribute to the group's common project and how many of them to keep for yourself.

Every point you keep will be yours. On the other hand, each point invested in the group project will bring in 0.4 points to you and to each other member of the group.

We ask you to make two decisions in the same screen.

1) Your decision about whether you give consent to the conditional commitment agreement

2) Your decision about how much to invest in the group project in case conditional commitment agreement is not accepted.

If all members in your group give consent to the conditional commitment agreement, then you will all contribute 20 points each to the project. In other words, you will invest all your endowment in the group project.

If at least one member of the group does not give consent, then every member will contribute the amount they themselves choose.

Group members will make their decisions simultaneously and without seeing others' decisions.

Please click on OK when you are ready.

Earnings:

Your income in terms of points at each round is calculated as follows:

(20 - individual investment in the group project) + 0.4 * (total investment in the group project)

Each group member receives the same income from the project.. For example, assume that every member of the group gives consent to the conditional commitment agreement, then everyone will invest 20 points to the project and the total amount of investment will be 80. In this case, each member of the group will gain 0.4*80 = 32 points.

Now, assume that the conditional commitment agreement is not approved and group members make a total investment of 60 points. In that case, each member will earn 0.4*60 = 24 points from the project.

For each token, which you keep for yourself you earn an income of 1 point. Suppose you contributed this token to the project instead, then the total contribution to the project would rise by one token. Your income from the project would rise by 0.4*1=0.4 points. However the income of the other group members would also rise by 0.4 points each, so that the total income of the group from the project would rise by 1.6 points. Your contribution to the project therefore also raises the income of the other group members. On the other hand you earn an income for each token contributed by the other members to the project. For each point contributed by any member you earn 0.4*1=0.4 points.

Please click on OK when you are ready.

Information Screen:

After all group members make their decisions, in the following screen you can see the number of the group members who give consent to the conditional commitment agreement, the amount that you and the other group members contribute to the group project, your earnings in that round and your total earnings including the previous rounds.

On the next screen, similar information will be displayed at individual level for all members of the group. Your contribution will be displayed in blue in the first column while other members' contributions will be shown in the remaining three columns, in a random order. For instance, the contribution amount in the second column will generally show the choice of a different group member in each round. This setting also applies to amounts presented on other columns.

Control Questions:

Now, we will ask you some questions. You need to answer them correctly in order to proceed. We remind below how your earnings will be calculated, so that you can use this information while you answer the questions.

Your earnings: (20 - individual contribution to the group project) + 0.4 * (total contribution to the group project)

1) Every group member has an endowment of 20 points. If every member rejects the conditional commitment agreement, and no group member contributes to the group project, how much income will you earn in that round?

(Answer = 20)

2) Suppose that you give consent to the conditional commitment agreement and choose to contribute nothing to the project in case the agreement is not approved. Moreover, suppose that the other 3 group members also give consent to the agreement, thus the agreement is approved.

In the above mentioned case, how many points will you contribute to the project?

In the above mentioned case, how many points will the group members (including you) contribute to the project in total?

In the above mentioned case, what will your total earnings be?

(Answers = 20 & 80 & 32, respectively)

1) Assume that one or more members of the group do not give consent to the conditional contribution agreement and thus the agreement is not approved. You choose to contribute 1 point in case the agreement is not approved while the other members choose to contribute 2, 3 and 4 points, respectively.

In the above mentioned case, how many points will you contribute to the project?

In the above mentioned case, how many points will the group members (including you) contribute to the project in total?

In the above mentioned case, what will your total earnings be?
(Please also consider the amount you do not contribute and keep for yourself)?

(Answers = 1 & 10 & 23, respectively)

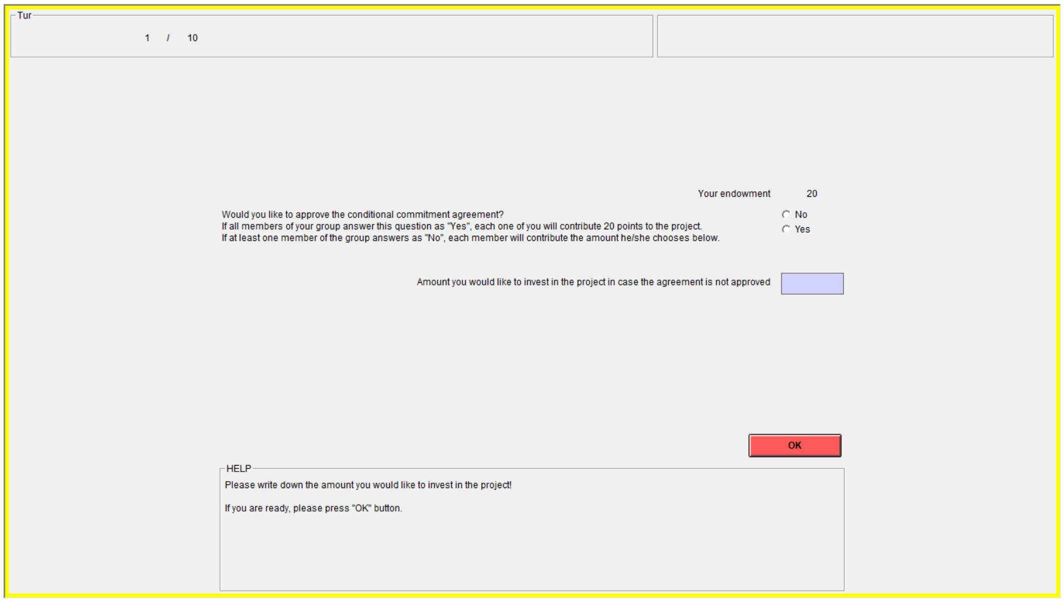

Figure 9: Screenshot of the experiment (decision screen)

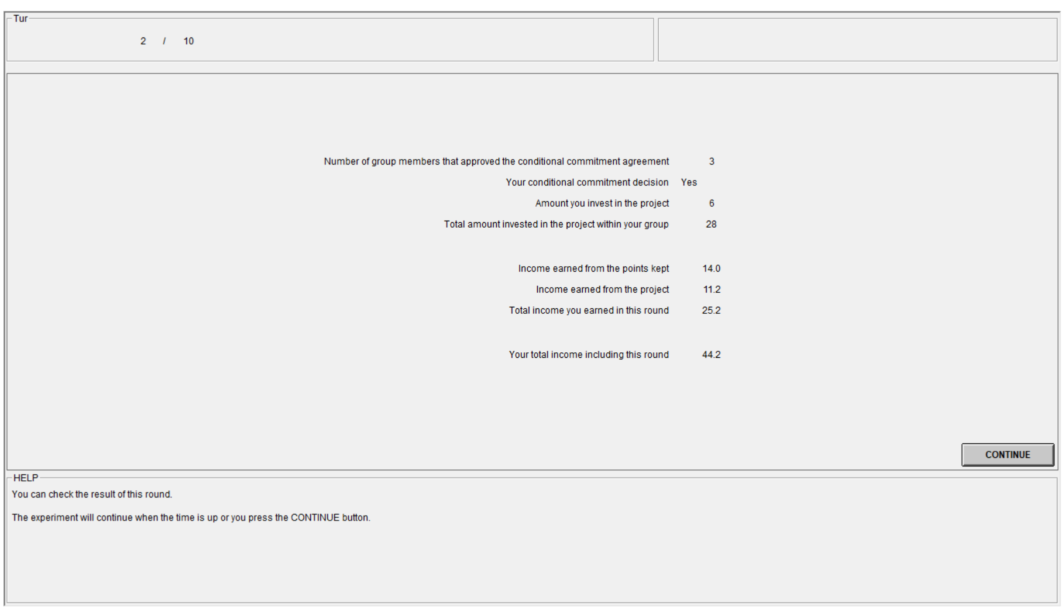

Figure 10: Screenshot of the experiment (check screen)

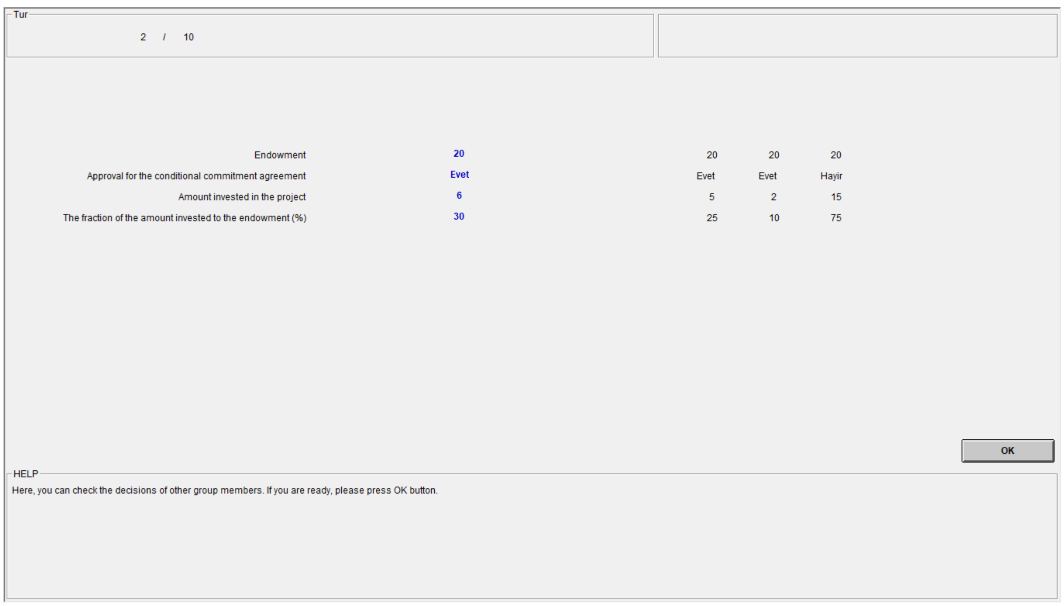

Figure 11: Screenshot of the experiment (individual level check screen)