Evolution of cooperation in the indefinitely repeated collective action with a contest for power

August 31, 2022

Yaroslav Rosokha† Xinxin Lyu† Denis Tverskoi‡ Sergey Gavrilets‡

Abstract

Social and political inequality among individuals is a common driving force behind the breakdown in cooperation. In this paper, we theoretically and experimentally study cooperation among individuals faced with a sequence of collective-action problems in which the benefits to cooperation are divided according to political power that is obtained through a contest. We have three main results. First, we find that cooperation predictably responds to the fundamental parameters of the collective-action problem. Specifically, it is increasing in the benefit to cooperation and how much benefit is gained from partial group cooperation, and decreasing in the number of players. Second, we find that when players are unrestricted in their expenditures in the contest, cooperation is much lower than when expenditures are set to a specific proportion of earnings. Finally, we find that individual norms and beliefs account for a substantial proportion of explained variance in individuals' decisions to cooperate.

Keywords: Cooperation, Contest, Dynamic Coordination Games, Indefinitely Repeated Games, Experimental Design, Beliefs, Individual and Social Norms

† Purdue University • yrosokha@purdue.edu • lyu75@purdue.edu
‡ University of Tennessee • dtversko@utk.edu • gavrila@utk.edu

1 Introduction

The 21st century has seen considerable social unrest across the developed and developing countries (e.g., Black Lives Matter Movement in the U.S., the Umbrella Movement in Hong Kong). Among the major driving forces behind these conflicts is political inequality between different ethnic, regional, or religious groups. The aim of our paper is to understand how human decision-makers cooperate when the benefits to cooperation are divided according to political power that is obtained through a contest. In particular, we develop a theoretical model that connects two famous, but largely disconnected problems. The first is a collective-action problem in which individuals face a decision on whether to undertake a risky collective action or take an individual action that guarantees a safe payoff. For example, consider a stag-hunt game (Rousseau, 1754) or a public goods game (Samuelson, 1954; Hirshleifer, 1983). The second is the contest for "political power" in which players face a decision on how much to spend to gain greater representation, which, in turn, translates into a more beneficial division of benefits from the collective action.[1] Building on earlier work (Houle, Ruck, Bentley, and Gavrilets, 2022; Tverskoi, Senthilnathan, and Gavrilets, 2021), we connect the two problems by assuming the benefits from the collective action are split based on the dynamically changing political power of the individuals. We further integrate the two problems by assuming interactions are indefinitely repeated, which creates opportunities for accumulation of power and cooperation breakdown over the long horizon that may not be present in the short run.

Our approach for deriving theoretical predictions for the decisions in the collective-action problem is twofold. First, we use a measure of strategic uncertainty developed for one-shot coordination games (Dal Bó, Fréchette, and Kim, 2021) to serve as a guideline for the choices during the initial interaction. Second, we use a model of myopic best-response to derive theoretical predictions for the long-run outcomes that incorporate the contest for power. Both approaches are consistent regarding the impact of the fundamental parameters of the decision to cooperate. Specifically, players are more likely to cooperate as the benefit to (partial) cooperation increases or the number of players decreases. The main theoretical results of the paper pertain to the long-term impact of the contest for power on the players' decision to cooperate. In particular, we show that when players do not have a choice regarding how much to spend in the contest, the cooperation is much higher than when they are unrestricted in their expenditure in contests for power.

To test our theoretical predictions, we design and run a controlled lab experiment. The experiment achieves three main objectives. First, we establish that human decision-makers respond to the fundamental parameters of the collective-action problem according to the theoretical predictions. Second, the results of our experiments also confirm that when human subjects are free to choose their expenditure in the contest, cooperation in the collective-action stage breaks down. Finally, as part of the experiment, we elicited individual beliefs as well as individual and social

[1] For example, in models of electoral competition (e.g., Baron, 1994; Grossman and Helpman, 1996), political parties use campaign spending to influence the voting behaviors to achieve more favorable outcomes. In the context of rent-seeking (e.g. Tullock, 1967; Krueger, 1974; Brock and Magee, 1978; Findlay and Wellisz, 1982), special interest lobbies compete for more favorable policies in areas with government restrictions such as taxes, subsidies, tariffs, and quotas.

norms. We then use the data on elicited beliefs and norms to estimate a behavioral model of choice (Gavrilets, 2021; Tverskoi, Xu, Nelson, Menassa, Gavrilets, and Chen, 2021; Tverskoi, Babu, and Gavrilets, 2022). Specifically, we show that in a dynamic setting in which individuals face an indefinite sequence of collective-action problems and contests for power, individual beliefs and norms play a prominent role in explaining individual behavior. Our experimental results on the effects of inequality in power, conformity and norms on cooperation complement an earlier test of the model predictions using country-level data linking economic inequality with social unrest in 75 countries between 1991 and 2016 (Houle, Ruck, Bentley, and Gavrilets, 2022).

Our paper contributes to several strands of literature. First, we contribute to the vast experimental literature on coordination games.[2] Early works in this stream include Van Huyck, Battalio, and Beil (1990) and Cooper, DeJong, Forsythe, and Ross (1992), who show that in a coordination game, human subjects tend to coordinate on the risk-dominant equilibrium.[3] More recently, Dal Bó, Fréchette, and Kim (2021) show that coordination in the experimental setting is better explained by a continuous measure of risk associated with choosing a cooperative action. We use this measure to make predictions in a much more complex dynamic setting. Specifically, in our experiments, subjects interact in a sequence of games whereby payoffs start with a stag-hunt coordination game but then evolve endogenously based on the resulting payoffs and the decision to invest in a contest for power. From this perspective, the most relevant papers are Cooper and Van Huyck (2018), who show that subjects are able transfer conventions between related coordination games presented in a sequence, Bornstein, Gneezy, and Nagel (2002), who show that in the presence of inter-group competition, more efficient outcomes can be achieved, and Cooper, Ioannou, and Qi (2018), who show that endogenous assignment to higher payoffs to coordinating on risky action leads to greater efficiency.

Second, we contribute to the experimental and theoretical literature that studies proportional-prize contest (Cason, Masters, and Sheremeta, 2020). Whereas the most famous theoretic and experimental analyses consider the winner-take-all lottery contests of Tullock (1980), a smaller stream considers proportional prize contests (Cason, Masters, and Sheremeta, 2010).[4] One of the most relevant papers is Savikhin and Sheremeta (2013), who study simultaneous decisions in a contest and public-goods game. The authors find that contributions to the public-goods game are not affected by the contest, whereas the (sub optimal) overbidding in the contest decreases, indicating a positive spill-over effect of the cooperative game on the competitive one. Our theory and experiment focus on a different combination of games integrated in a new, dynamic way. In particular, we consider the impact of the contest for power on the individual's decision to cooperate when the benefits to cooperation are split according to the power earned in the contest. Both theoretically and experimentally, we find that the an unrestricted contest for power leads to

[2] For a recent survey of experiments on coordination games, we refer the reader to Cooper and Weber (2020).

[3] In a two-player two-action coordination game, risk dominance is defined as a best response to the other player choosing 50-50 (Harsanyi and Selten, 1988).

[4] See Dechenaux, Kovenock, and Sheremeta (2015) for a review of experimental literature on winner-take-all Tullock contests.

significantly lower cooperation.

The third stream of literature that we contribute to studies dynamic repeated games. Work in this field has focused on behavior in common-pool resources games (Gardner, Ostrom, and Walker, 1990; Stoddard, Walker, and Williams, 2014; Vespa, 2020) and dynamic Prisoner's dilemma games (Vespa and Wilson, 2019; Rosokha and Wei, 2020). In addition, a growing literature exists that links behavior of individuals with changes in their personal norms and empirical and normative expectations (d'Adda, Dufwenberg, Passarelli, and Tabellin, 2020; Górges and Nosenzo, 2020; Andreozzi, Ploner, and Saral, 2020; Szekely, Lipari, Antonioni, Paolucci, Sánchez, Tummolini, and Andrighetto, 2021; Tverskoi, Guido, Andrighetto, Sánchez, and Gavrilets, 2022). Our work integrates these approaches by accounting for changes in individual beliefs and norms as the individual's power evolves during social interactions.

The rest of the paper is organized as follows: in section 2, we formalize the environment. In section 3, we develop three main hypotheses. Next, in section 4, we present details of the experimental design. We then present results of the experiment in section 5. Finally, we conclude in section 6.

2 Environment

We consider a society composed of $I = \{1, \ldots, n\}$ individual decision-makers interacting over an indefinite sequence of rounds. Each round, $t \in \{1, 2, 3, \ldots\}$, the decision-makers are engaged in a collective-action game (stage 1) and a contest for power (stage 2). Specifically, in stage 1 of period t, each player i chooses whether to participate ($a_{i,t} = 1$) or not ($a_{i,t} = 0$) in the production of a club good. The cost of participating in the production, $c > 0$, is the same across all players and is constant across time. Let $a_t = (a_{i,t}, a_{-i,t}) = (a_{1,t}, \ldots, a_{n,t})$ denote the action profile in period t, with $a_{-i,t}$ denoting an action profile of all players excluding i. The production amount $F(\bar{a}_t)$ is an S-shaped function of the proportion of players who decide to participate in the production, $\bar{a}_t = \frac{\sum_{i \in I} a_{i,t}}{n}$, as follows:

$$F(\bar{a}_t) = b \frac{(\bar{a}_t)^\kappa}{(\bar{a}_t)^\kappa + (a_0)^\kappa} \tag{1}$$

where $b > 0$ is the maximum benefit to cooperation, $a_0 \in (0, 1)$ is the "half-effort" parameter that determines the proportion of the group required to produce half of the maximum benefit, $(\frac{b}{2})$, and $\kappa \geq 1$ is the parameter that determines the steepness of the production function (Gavrilets, 2015).

Unlike the widely studied collective-action problems, such as stag-hunt or public-goods games, in our environment, the share of the production that player i gets in period t depends on how much effort, e_{t-1}, players spent in stage 2 of period $t-1$ on obtaining "political power" over the division. Specifically, the division in a round is determined according to the proportional-prize contest among all cooperators based on the exerted effort. Thus, player i's payoff in stage 1 is

$$\pi_i^1(a_t, e_{t-1}) = R_0 + a_{i,t} \Big(\frac{e_{i,t-1}}{a_t \cdot e_{t-1}} F(\bar{a}_t) - c \Big), \tag{2}$$

where $a_t \cdot e_{t-1} = \sum_{i \in I} a_{i,t} e_{i,t-1}$ is the dot product of the two vectors equal to the sum of all efforts by cooperating players, and $R_0 > c$ is an endowment.[5] Then, the payoff in round t is

$$\pi_i(a_t, e_{t-1}, e_t) = \pi_i^1(a_t, e_{t-1}) - e_{i,t}. \tag{3}$$

In this paper, we aim to achieve three main goals. First, we would like to establish that human decision-makers respond to the fundamental parameters of the collective-action problem (b, n, a_0). Second, we would like to understand how the contest for power interacts with the decision to cooperate in the collective production. Finally, we consider an individual's beliefs and norms about cooperation to provide insights into the forces that may drive decisions to cooperate or defect in this highly dynamic environment.

2.1 Parameters

Table 1: Summary of Treatment Parameters

Treatment		Parameters						Production function, $F(\bar{a})$				
		b	a_0	n	R_0	c	κ	$\bar{a} = 0$	.25	.5	.75	1
T1	EXO / END	109	.812	2	60	20.4	12	0		0		100
T2	EXO	109	.812	4	60	20.4	12	0	0	0	30	100
T3	END	218	.812	2	60	20.4	12	0		0		200
T4	EXO / END	218	.812	4	60	20.4	12	0	0	0	60	200
T5	END	109	.406	2	60	20.4	12	0		100		108
T6	EXO / END	109	.406	4	60	20.4	12	0	0	100	108	108

Notes: Production function, $F(\bar{a})$, is given by equation (1). b denotes the maximum benefit to cooperation; n denotes the number of players in the environment; a_0 denotes the "half-effort" parameter, which determines proportion of the group that is required to achieve half of b. EXO denotes a treatment with an exogenously specified proportion of earnings in stage 1 that are contributed in stage 2. END denotes a treatment in which players make decisions in stage 2.

As mentioned above, our first goal is to establish that decision-makers in this environment respond to the fundamental parameters in a predictable way. To this end, we vary three fundamental parameters: $b \in \{109, 218\}$, $n \in \{2, 4\}$, and $a_0 \in \{0.406, 0.819\}$ and fix $R_0 = 60$, $c = 20.4$, $\kappa = 12$, and $e_{i,0} = 0$, $\forall i \in I$ across all treatments. Summary of the resulting treatment parameters, including collective production, are presented in Table 1. In addition, Table 2 presents the focal player's stage-game payoffs in round 1 of a supergame. We choose the parameters so that payoffs in the first round of interaction are comparable to previously studied two-player stag-hunt games (Dal Bó, Fréchette, and Kim, 2021; Schmidt, Shupp, Walker, and Ostrom, 2003). For example, the payoffs in one of the games studied in Dal Bó, Fréchette, and Kim (2021) are the same as in round 1 of

[5] In the case of $a_t \cdot e_{t-1} = 0$ and $a_t \cdot \mathbb{1} \neq 0$ (where $\mathbb{1}$ is a vector of ones), we define $\pi(a_t, e_{t-1}) = R_0 + a_{i,t}\left(\frac{1}{a_t \cdot \mathbb{1}} F(\bar{a}_t) - c\right)$. In the case of $a_t \cdot \mathbb{1} = 0$, we define $\pi(a_t, e_{t-1}) = R_0$.

the T1 parameter combination with the exception that the payoff to (D,C) in T1 is 60, whereas the payoff to (D,C) in Dal Bó, Fréchette, and Kim (2021) is 65.

Table 2: Stage-Game Payoffs when All Players Have the Same Power

Parameters	n = *2*			n = *4*				
	T1	0	1	T2	0	1	2	3
b = *109*	C	40	90	C	40	40	50	65
a_0 = *0.812*	D	60	60	D	60	60	60	60
	T3	0	1	T4	0	1	2	3
b = *218*	C	40	140	C	40	40	60	90
a_0 = *0.812*	D	60	60	D	60	60	60	60
	T5	0	1	T6	0	1	2	3
b = *109*	C	140	94	C	40	90	76	67
a_0 = *0.406*	D	60	60	D	60	60	60	60

Notes: Payoff for choosing C(cooperate) and D(defect) when all players have equal power. Columns denote how many other players choose C (out of $n - 1$). Players always have equal power in Round 1 of a match, but may have equal power in other rounds depending on players' choices in prior rounds.

Our second goal is to understand how the contest for power influences the decisions to cooperate. To this end in some of the treatments, we restrict the investment in the contest to be a constant fraction of the earnings from the collective action. We use abbreviations EXO and END to differentiate between an exogenous and an endogenous contest treatment (see Table 1). Specifically, in the exogenous treatment, players are restricted to invest a fixed proportion (10%) of their stage 1 earnings in the contest for the next round's power. By contrast, in the endogenous treatment, the only restriction on players' spending is the intrarounds budget constraint (i.e., in stage 2 of a given round, subjects may not exceed what they earned in stage 1). Finally, our third goal is to understand how norms and beliefs impact behavior. To this end, we elicit subjects' round-by-round beliefs and norms. We then test whether the belief and norm data help better explain subjects' observed behaviors.

3 Hypotheses

Theoretical analysis of the indefinitely repeated coordination games does not provide a clear prediction regarding whether decision-makers will cooperate or defect. On the one hand, any sequence of stage-game Nash equilibria (NE) is supported as a subgame perfect equilibrium (SPE) and given that both cooperation and defection are stage-game NE, either could be played. On the other hand, infinitely many trigger strategies could be supported as an SPE as well. For example, consider a strategy that prescribes cooperating in stage 1 and a contribution of a fixed fraction of stage 1 earnings to stage 2 contest as long as the other cooperates in stage 1 and contributes the same fraction in stage 2. Any deviation, either by defecting in stage 1 or by changing the amount in stage 2, will trigger punishment of defections forever. Therefore, for theoretical guidance, we rely on two behaviorally grounded approaches. First, we consider the *size of the basin of attraction of cooperation* (henceforth, $SizeBC$) of the stage game as a predictor of the behavior in rounds 1 of a supergame. Focusing on the behavior in round 1 has several advantages: (i) It is an important determinant of how the interaction unfolds, because behavior in later rounds is not independent of previous rounds; (ii) in round 1 of each supergame, all players have the same power, and thus conditional on parameters of the collective-action problem, play the same game; and (iii) in round 1, subjects have not yet participated in the contest for power, which may add an additional layer of complexity to the analysis. The second approach we take focuses on the long-term outcomes. In particular, we use a model of *myopic best response* which has been widely used among economists (Kandori, Mailath, and Rob, 1993; Young, 1993; Kandori and Rob, 1995; Hopkins, 1999) and evolutionary game theorists (Smith, 1982; Matsui, 1992; Sandholm, 1998; Alós-Ferrer, 2003; Roca, Cuesta, and Sánchez, 2009; Szolnoki and Perc, 2014; Tverskoi, Senthilnathan, and Gavrilets, 2021; Houle, Ruck, Bentley, and Gavrilets, 2022). Notably, the approach has found recent empirical support in economics experiment on repeated coordination games (Mäs and Nax, 2016). In addition, Offerman, Sonnemans, and Schram (2001) note that subjects tend to be adaptive and less strategic in complicated experimental environments, as is the case in our experiment.

3.1 Size of Basin of Attraction of Cooperation in Round 1

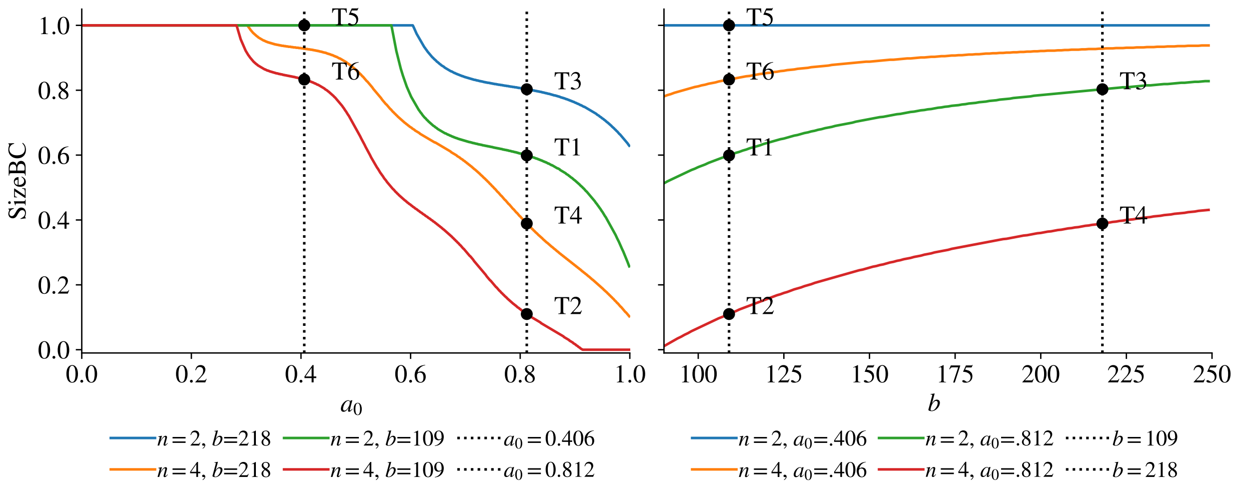
To predict behavior in Round 1, we focus on a measure of strategic uncertainty developed for one-shot games by Dal Bó, Fréchette, and Kim (2021). In particular, for the two-player version of the game, we define $SizeBC$ of the stage game as the maximum probability of the other subject playing defect that still makes cooperation a best response. Specifically, let θ_{-i} be the probability that the other player chooses to cooperate. Then, to calculate the $SizeBC$, we find the maximum value of $(1-\theta_{-i}) \in [0,1]$ such that

$$\theta_{-i}\pi_i^1((1,1),e_0) + (1-\theta_{-i})\pi_i^1((1,0),e_0) \geq R_0$$

$$\Rightarrow SizeBC = \begin{cases} 1, \text{ if } a_0 \leq 0.5\left(\frac{b}{c}-1\right)^{\frac{1}{\kappa}}, \\ \frac{(2c+2ca_0^\kappa - b)(1+2^\kappa a_0^\kappa)}{b(1+2a_0^\kappa - 2^\kappa a_0^\kappa)}, \text{ otherwise.} \end{cases}$$

In Appendix A.1, we show that for the parameters chosen for the experiment, $SizeBC$ is increasing in b and decreasing in a_0. Note that if $SizeBC$ is greater than one half, then cooperation is risk dominant (Harsanyi and Selten, 1988). Furthermore, the higher $SizeBC$, the more robust cooperation is to strategic uncertainty and the more cooperation we expect to see in the experiment. To adapt this measure to games with $n > 2$ players, we follow Kim (1996), Morris, Rob, and Shin (1995), and Peski (2010) in assuming that players other than the focal player, have the same probability of cooperation.[6] Figure 1 presents how *SizeBC* changes with the treatment parameters.

Figure 1: Basin of Attraction and Behavior in Round 1



Notes: The figure presents the size of basin of attraction of cooperation (*SizeBC*) assuming the power is equally distributed. The left panel shows how *SizeBC* changes with a_0. The right panel shows how *SizeBC* changes with b. • denotes treatment parameters chosen for the experiment.

The comparison between treatments T3 and T1 as well as between T4 and T2 shows that *SizeBC* increases with the maximum benefit to cooperation (b). The comparison between treatments T1 and T2, T3 and T4, as well as T5 and T6 shows that *SizeBC* is decreasing in the group size (n). Finally, the comparison between T1 and T5 as well as T2 and T6 shows that *SizeBC* is decreasing in the proportion (a_0) required to achieve half of the possible benefit to cooperation. We summarize the resulting predictions with Hypothesis 1:

Hypothesis 1 *Cooperation responds to the parameters of the collective-action problem:*

(a) *Cooperation is increasing in the maximum benefit to cooperation (b),*

(b) *Cooperation is decreasing in the group size (n).*

(c) *Cooperation is decreasing in the proportion of the group (a_0) required to achieve half of the maximum benefit to cooperation.*

3.2 Myopic Best-Response, Contest for Power, and the Long-term Outcomes

To understand how the contest for political power interplays with decisions to cooperate, we consider two versions of the environment. In particular, in addition to the environment in which players freely choose how much to spend on the contest for power in stage 2 (which we denote as END),

[6]Kim (1996) generalizes the risk-dominance concept of Harsanyi and Selten (1988) to an N-player coordination game using the same approach. A similar approach is adopted for the p-dominant equilibrium by Morris, Rob, and Shin (1995) and the GR-dominance by Peski (2010).

we also consider a baseline, denoted as EXO, in which we exogenously restrict expenditures on the contest for power to be a fixed proportion of the earnings in stage 1 (i.e., players do not have a choice over effort, Tverskoi, Senthilnathan, and Gavrilets 2021; Houle, Ruck, Bentley, and Gavrilets 2022). By comparing the two models (and resulting treatments), we have a better understanding of reasons cooperation may break down. Next, we introduce the best-response functions for both versions of the model and characterize the myopic best-response equilibria.

3.2.1 Exogenous Power Revision

For the model of exogenous power revision, we restrict effort in stage 2 to be a fixed proportion, $\gamma \in (0,1)$, of the payoff in stage 1:

$$e_{i,t} = \gamma \pi_i^1(a_t, e_{t-1}), \forall i \in I. \tag{4}$$

We assume that in stage 1 of period $t+1$, player i decides whether to cooperate, by best responding to the choices in period t. That is, in stage 1 of period $t+1$, player i chooses

$$a_{i,t+1} = BR_i^a(a_{-i,t}, e_t) = \underset{a_i \in \{0,1\}}{\operatorname{argmax}} \ \pi_i^1((a_i, a_{-i,t}), e_t). \tag{5}$$

Definition 1 *An action profile a^* is a myopic-best-response equilibrium in the exogenous version of the model if*

$$a_i^* = BR_i^a(a_{-i}^*, \hat{e}), \forall i \in I, \tag{6}$$

where

$$\hat{e}_i = \gamma \pi_i^1(a^*, \hat{e}), \forall i \in I. \tag{7}$$

In Appendix A.2, we provide further details. In particular, we show that all equilibria are symmetric in that all cooperators (if exist) exert the same effort and all defectors provide the same effort. As a result, no more than $n+1$ equilibria (with 0, 1 ,.., or n cooperators, respectively) can exist. Moreover, we provide conditions for the existence of these equilibria. Notably, because γ impacts all payoff combinations in the same way, the outcomes do not depend on the actual proportion.

3.2.2 Endogenous Power Revision

For the model of endogenous power revision, in addition to the decision to cooperate in stage 1, players must decide on the effort to spend in the contest for power in stage 2, $e_{i,t} \in [0, \pi_i^1(a_t, e_{t-1})]$. Note, however, the effort spent in stage 2 of period t directly affects not only the current payoff, but also the next-period payoff (which also depends on $a_{i,t+1}$). Therefore, to make the analysis manageable, we assume the individual simultaneously chooses the effort $e_{i,t}$ in stage 2 of period t and the action $a_{i,t+1}$ in stage 1 of period $t+1$ to maximize her expected total earnings by best responding to the previous choices (a_t, e_{t-1}). That is, if $a_{-i,t} \cdot e_{-i,t-1} \neq 0$ or $a_{-i,t} = 0$ in stage 2 of period t, player i chooses

$$(a_{i,t+1}, e_{i,t}) = BR_i^{a,e}(a_t, e_{t-1}) = \underset{a_i \in \{0,1\}, e_i \in [0, \pi_i^1(a_t, e_{t-1})]}{\operatorname{argmax}} \Big\{ -e_i + \delta \pi_i^1((a_i, a_{-i,t}), (e_i, e_{-i,t-1})) \Big\}, \quad (8)$$

where $a_{-i,t} \cdot e_{-i,t-1} = \sum_{j \in I \setminus \{i\}} a_{j,t} e_{j,t-1}$ is the total expenditure of all cooperating players except i, and $\delta \in (0,1)$ is the probability of continuing the game to the next round (for more details, see Appendix A.2).

Definition 2 *A strategy profile (a^*, e^*) is a myopic-best-response equilibrium in the endogenous version of the model if*

$$(a_i^*, e_i^*) = BR_i^{a,e}(a^*, e^*), \forall i \in I. \quad (9)$$

Proposition 1 *All equilibria in the endogenous version of the model are symmetric, in that all $n_C^* \in \{0, 1, .., n\} \setminus \{1\}$ cooperators (if they exist) exert the same effort $e_C^* = \delta\Big(1 - \frac{1}{n_C^*}\Big)\frac{F(n_C^*/n)}{n_C^*}$, and all $n - n_C^*$ defectors (if they exist) exert the same effort $e_D^* = 0$.*

The conditions for equilibrium existence as well as the proof of Proposition 1 can be found in Appendix A.2. In addition, as a corollary, we show that no more than n equilibria (with $n_C^* \in \{0, 1, ..., n\} \setminus \{1\}$ cooperators, respectively) can exist.

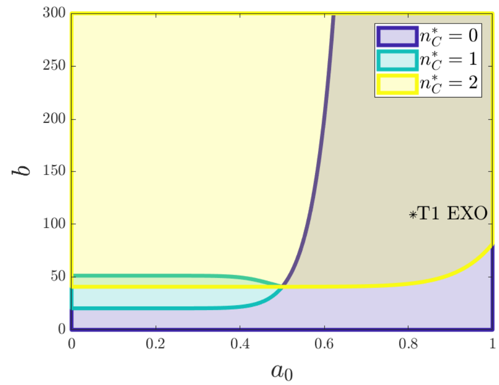
3.2.3 Endogenous versus Exogenous Comparison

Figure 2 presents the summary of the theoretical results. The figure shows parameter regions for which a particular symmetric equilibrium (denoted by the number of cooperators) exists. In the figure, we also mark the treatments of the experiment that we run. The main takeaway from the theoretical results is that allowing players to compete for power leads to lower cooperation. The most stark example is that the T4 parameter combination with the endogenous scenario is predicted to have no cooperation, whereas for the same parameter combination in the exogenous scenario, full cooperation (all four players) can be supported in equilibrium.

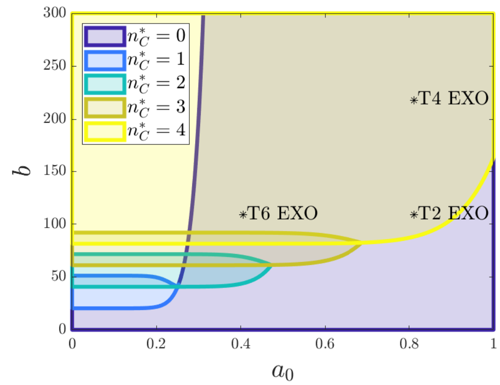
The intuition behind the above result is as follows. First, under the myopic best-response framework, the defectors are not motivated to invest in the competition if they have a choice. However, if they are forced to do so exogenously, they have an extra incentive to switch to cooperation. Second, if the power is revised endogenously, cooperators are motivated to cooperate if their share of the jointly produced resource exceeds individual costs plus individual investments in competition as compared with just their share of the jointly produced resource if power is revised exogenously. We summarize the above results with the following hypothesis:

Hypothesis 2 *Cooperation is lower in endogenous-power-revision treatments than in exogenous-power-revision treatments.*

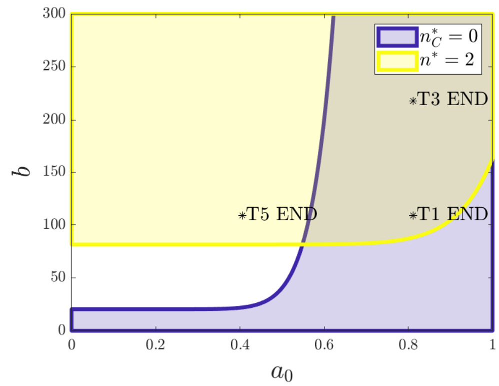
Figure 2: Myopic-Best-Response Equilibria



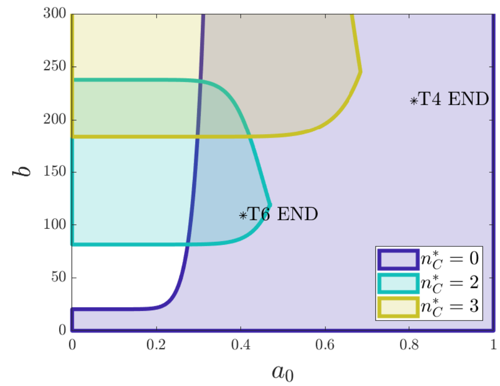
(a) EXO, $n = 2$



(b) EXO, $n = 4$



(c) END, $n = 2$

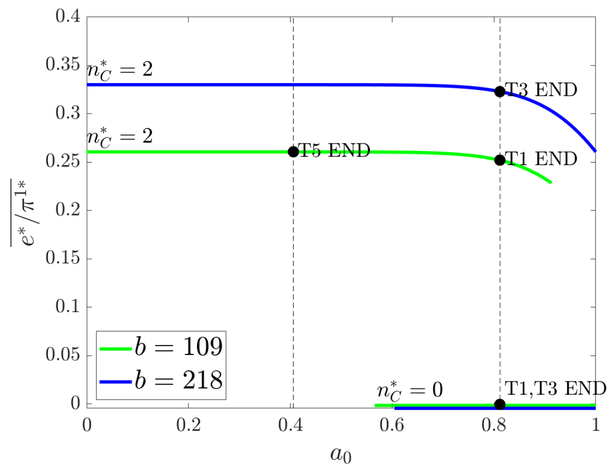


(d) END, $n = 4$

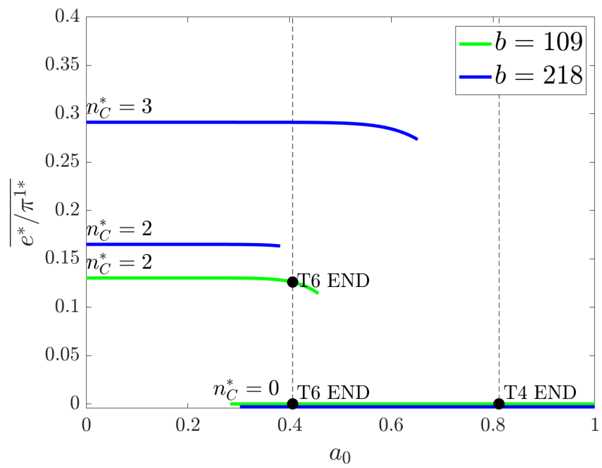
Notes: Shaded regions correspond to symmetric equilibria with $n_C^* \in \{0, .., n\}$ cooperators and $n - n_C^*$ defectors. b denotes the maximum benefit to cooperation. a_0 denotes the proportion of the group that is required to produce $\frac{b}{2}$. * denotes treatment parameters chosen for the experiment.

In addition to the results on cooperation in the collective-action stage, Figure 3 presents a summary of the theoretical predictions regarding the average equilibrium effort in the endogenous version of the model. The figure shows the equilibrium effort in the contest for power as the average proportion of the payoff from stage 1, $\overline{e^*/\pi^{1*}} = \frac{1}{n}\sum_{i=1}^{n} \frac{e_i^*}{\pi_i^{1*}}$. The main takeaway is that for the treatments of the experiment that we run, the average proportion of the payoff an individual spends in the context responds to the fundamental parameters of the collective-action problem similarly to the cooperation described in Hypothesis 1.

Figure 3: Contest Expenditures in Equilibrium



(a) END, $n = 2$



(b) END, $n = 4$

Notes: The figure presents an average proportion of the payoff earned at stage 1 that an individual spends in the contest at stage 2, $\overline{e^*/\pi^{1*}} = \frac{1}{n}\sum_{i=1}^{n}\frac{e_i^*}{\pi_i^{1*}}$. When multiple equilibria exist, all are shown using the same color. For example, three equilibria for the case of $n = 4$ and $b = 218$ are in blue (top blue, middle blue, and bottom blue). • denotes the experimental treatments.

3.3 Beliefs, Norms, and Within-Supergame Interactions

A distinct feature of our environment is that subjects face payoffs that depend on the political power over the division obtained through a contest. That is, a contest for power introduces additional considerations, such as unequal payoffs and subjective evaluations of what others will or should do given a particular power distribution. To help sort through the myriad of outcomes, we consider beliefs and norms that subjects hold. In particular, we follow Gavrilets (2021) in assuming that the behavioral utility function has four components: expected payoffs given beliefs, conformity with the behavior of others, social norms about appropriateness of behavior, and personal norms about appropriateness of behavior. Next, we elaborate on each component.

A number of experimental studies have found evidence of best responding to beliefs in one-shot coordination games (Harsanyi and Selten, 1988; Cooper, DeJong, Forsythe, and Ross, 1990; Heinemann, Nagel, and Ockenfels, 2009; Bosworth, 2017) as well as in more complicated repeated games (Nyarko and Schotter, 2002, Davis, Ivanov, and Korenok, 2016, Gill and Rosokha, 2020, Aoyagi, Fréchette, and Yuksel, 2020). To capture an individual's tendency to best respond to beliefs, the behavioral utility function will include the expected payoff given the belief about the behavior of others in the group: $\pi_i^1(a_{i,t}, e_{t-1}, \theta_{-i,t}) = \mathbb{E}[\pi_i^1((a_{i,t}, a_{-i,t}), e_{t-1})|\theta_{-i,t}]$.

Although many subjects tend to best respond to the beliefs, previous studies have also found that a substantial fraction fail to do so (Nyarko and Schotter, 2002, Costa-Gomes and Weizsäcker, 2008, Heinemann, Nagel, and Ockenfels, 2009). To help explain why subjects may not best-respond, we consider three types of norms: (1) descriptive social norms; (2) injunctive social norms; and (3) personal norms.

Following Bicchieri (2005, 2016) we define a descriptive social norm as a behavioral rule that individuals are willing to comply with, provided that most people conform to it. That is, descriptive norms are based on the first-order beliefs of what others will do. To operationalize how descriptive social norms enter the utility function, we define $C(a_{i,t}, \theta_{-i,t}) = -\mathbb{E}\left[(a_{i,t} - \bar{a}_{-i,t})^2 \middle| \theta_{-i,t}\right]$ as the expected disutility associated with not conforming with the expected actions of others. That is, we need to compare each subject's choice with what they expect others will do, and say that subjects conforms with others if their own actions match their expectations about others.

Following Krupka and Weber (2013), we define injunctive social norms, $SN(a_{i,t}, e_{t-1})$, as collective perceptions regarding the appropriateness of action $a_{i,t}$ given a particular power distribution (determined by e_{t-1}). Finally, following Burks and Krupka (2012), we define the personal norm, $PN(a_{i,t}, e_{t-1})$, as a individual's own perception of the appropriateness of an action $a_{i,t}$ given a particular power distribution (determined by e_{t-1}). Both social norms and personal norms have been found to be important drivers of individual behaviors and decision-making, including cooperation (Camerer and Fehr, 2004; Fehr and Fischbacher, 2004a; Fehr and Schurtenberger, 2018), prosocial behavior (Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009; Bénabou, Falk, and Tirole, 2020), and punishment (Fehr and Gächter, 2000; Fehr and Fischbacher, 2004b).[7]

To summarize, we propose that an individual i makes her decision regarding cooperation in the collective action in round t based on the utility function (for more details, see Appendix A.3):

$$u_i(a_{i,t}, e_{t-1}, \theta_{-i,t}) = \beta_{1,i}\pi_i^1(a_{i,t}, e_{t-1}, \theta_{-i,t}) + \beta_{2,i}C(a_{i,t}, \theta_{-i,t}) + \beta_{3,i}SN(a_{i,t}, e_{t-1}) + \beta_{4,i}PN(a_{i,t}, e_{t-1}), \tag{10}$$

and we put forward the following hypothesis:

Hypothesis 3 *Beliefs and norms explain cooperative behavior in the collective-action stage.*

4 Experimental Design and Administration

To establish that individuals' decision to cooperate and compete responds to the main parameters of the environment, we designed a between-subjects experiment that systematically varies (i) the benefit to full cooperation, b, (ii) the number of subjects in each group, n, and (iii) the proportion of subjects that is required to achieve half of maximum payoff to cooperation, a_0. To show that the nature of the contest over political power – exogenous versus endogenous – has a substantial impact on cooperation in the collective action, we included treatment pairs for the same parameter combinations. Finally, to understand whether behavioral factors may influence individuals to co-

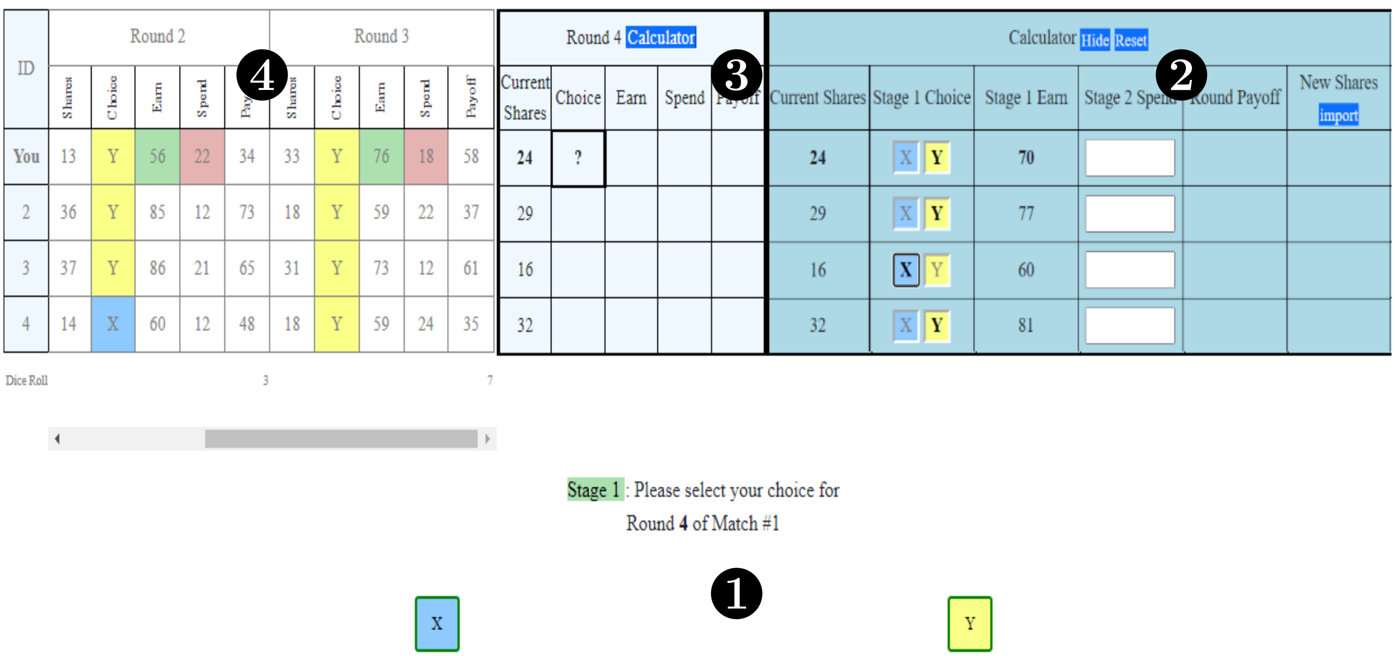
[7]The literature on the effect of personal and social norms is vast and includes the public-goods game (Fischbacher and Gächter, 2010; Kölle and Quercia, 2021; Reuben and Riedl, 2013), the collective-risk social dilemma (Szekely, Lipari, Antonioni, Paolucci, Sánchez, Tummolini, and Andrighetto, 2021), the dictator game (d'Adda, Dufwenberg, Passarelli, and Tabellini, 2020), the common-pool resource game (Tverskoi, Guido, Andrighetto, Sánchez, and Gavrilets, 2022), Bertrand games (Krupka, Leider, and Jiang, 2017), trusting games (Krupka, Leider, and Jiang, 2020), and a set of different games (dictator game, dictator game with tax, ultimatum game, and third-party punishment game) (Bašić and Verrina, 2021).

operate in our environment, we elicited beliefs about other group members' choices, personal and social norms, and measures of risk aversion, loss aversion, social preference, and cognitive ability.

4.1 Indefinitely Repeated Collective Action with Contest for Power

To implement the infinitely repeated interactions in the lab, we follow Roth and Murnighan (1978) with subjects interacting in fixed groups for a random number of decision rounds. In particular, at the end of each decision round, there the supergame ends with a 0.1 probability and continues with 0.9 probability. Thus, on average, each supergame lasts 10 rounds; however the actual realizations vary.[8] At the end of each supergame, subjects are randomly rematched to avoid a long-term reputation effect. Each decision round contains two stages: collective action and the contest for power. Next, we describe each stage in more detail.

Figure 4: Stage 1 Interface Screenshot



Notes: The screenshot shows the decision screen in the T6 END treatment. The neutral action names X and Y correspond to D (defect) and C (cooperate). The screenshot shows (1) decision entry, (2) hypothetical payoff calculator, (3) current round summary with power distribution in the first column (neutral "current shares" was used instead of "power"), and a question mark denoting current decision, (4) scrollable history.

4.1.1 Stage 1: Collective-Action Decision

In stage 1, subjects simultaneously decide whether to participate in the production of a collective good (i.e., cooperate). Figure 4 presents the decision screen for stage 1 of the T6 END treatment.

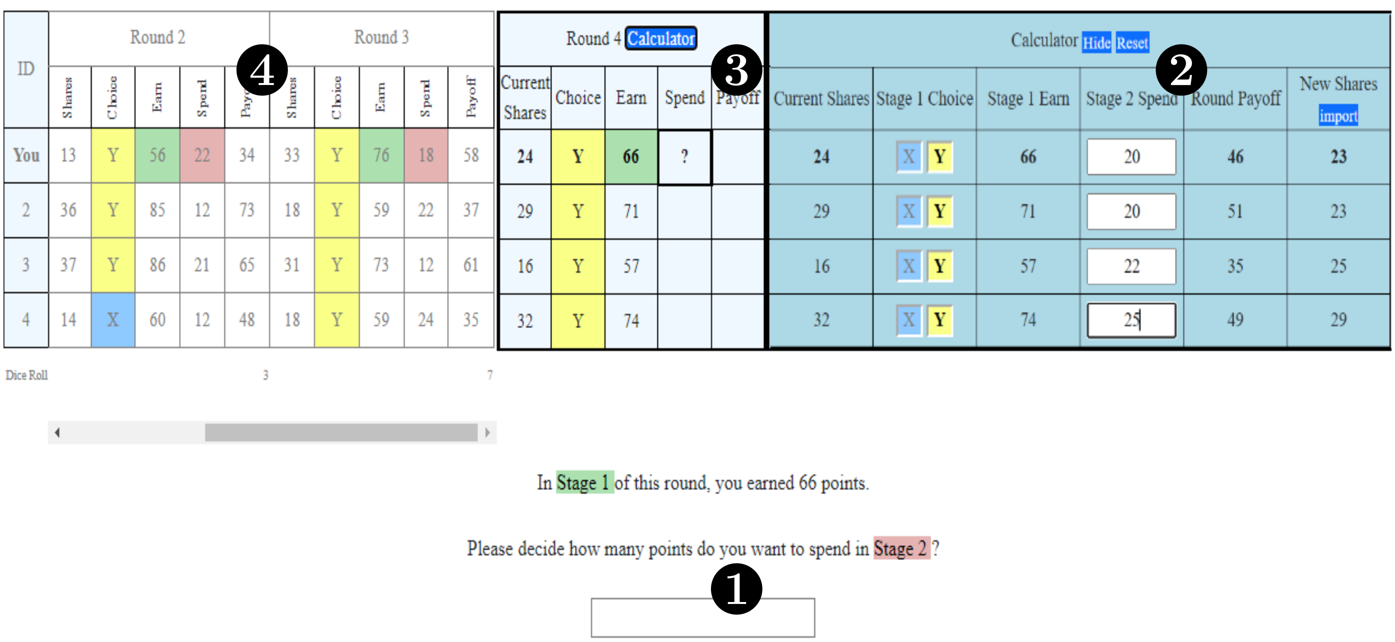
[8]Table D-4 in the Appendix presents supergame length sequences used in our experiment.

Given the complexity of the environment and the dynamic consequences of decisions, we provide a hypothetical calculator (2 in Figure 4). Using the calculator, subjects could enter a hypothetical scenario and see the resulting payoffs for the round as well as a consequence on the power in the following round.

4.1.2 Stage 2: Contest for Power

After all subjects make their stage 1 decisions, the experiment proceeds to stage 2. Figure 5 presents the screenshot of the stage 2 interface for the T6 END treatment. In the END treatment, subjects need to decide how many points to spend in the contest for power. In particular, we use neutral phrases such as ”shares” when referring to power (see 2 in Figure 5). The points they spend in stage 2 cannot exceed their earnings in stage 1. In the EXO treatment, subjects don’t have the option to specify how many points to spend. Instead, the screen notifies them that 10% of their stage 1 earnings (rounded to the nearest integer) are spent in stage 2.

Figure 5: Stage 2 Interface Screenshot



Notes: The screenshot shows the decision screen in the T6 END treatment. The neutral action names X and Y correspond to D (defect) and C (cooperate). The screenshot shows (1) stage 2 decision entry (for endogenous case), (2) hypothetical payoff calculator, (3) updated current-round summary with power distribution in the first column (neutral “current shares” was used instead of ”power”), stage 1 decisions in the second column, stage 1 earnings in the third column (self stage 1 earn is highlighted with green cell, and a question mark denoting current decision, (4) scrollable history.

4.2 Elicitation of Beliefs, Norms, and Individual Characteristics

In the first and 10th match of the END treatment and in the first, 10th, and 20th match of the EXO treatment, we elicit subjects' beliefs and norms.[9] The belief and norm elicitation is done in every round of a supergame immediately following the stage 1 decision. Specifically, we ask subjects three elicitation questions. The first question uses binarized scoring rule (Hossain and Okui, 2013, Erkal, Gangadharan, and Koh, 2020) to elicit subjects' beliefs about other subjects' choices.[10] With the second question, we elicit how appropriate their two actions are on a 4-point Likert scale (1 = inappropriate, 2 = somewhat inappropriate, 3 = somewhat appropriate, and 4 = appropriate). In particular, the aim is to elicit subjects' personal ethical norms, which can not be financially incentivized (as discussed in Young, 1998, Bicchieri and Chavez, 2010, Burks and Krupka, 2012). With the third question, we elicit subjects' social norms by describing the task as a coordination game. We follow Krupka and Weber (2013) in the elicitation structure, except we decided to not incentivize the answers given the time constraints, the complexity of the compensation procedure, and the complexity of the environment.[11]

We were concerned that round-by-round belief and norms elicitation may influence the behavior in the experiment, therefore we ran 14 pilot sessions (7 without elicitation and 7 with elicitation) for three parameter combinations T1, T2, and T6. Comparing subjects' behaviors across these three parameter combinations in Appendix B, we find no impact of elicitation on subjects' decision makings. Therefore, we summarize this design check with Remark 1:

Remark 1 *Belief and norm elicitation did not impact subjects decisions to cooperate and compete.*

4.3 Elicitation of Individual Characteristics and Demographic Variables

Before the main experiment, we ask subjects to complete five individual tasks: (i) risk-aversion elicitation, (ii) loss-aversion elicitation, (iii) elicitation of social preferences for advantageous inequality, (iv) elicitation of social preferences for disadvantageous inequality, and (v) cognitive ability. The first four tasks are organized as multiple price lists following Holt and Laury (2002), Rubin, Samek, and Sheremeta (2018), and Kerschbamer (2015). The fifth task is composed of 11 matrix-reasoning questions (Condon and Revelle, 2014). We incentivized subjects' decisions by randomly picking one of the four tasks to pay. If the picked task was a multiple-price-list task, we randomly pick one of the decisions and paid subjects based on their choice. If the cognitive ability task was picked, we paid subjects a flat rate of $4. In Appendix C we provide screenshots with more details for each task.

[9] As part of the main dataset, we include data from the pilot experiment, which had some variation in the timing and number of elicitations. See Appendix B for details.

[10] Following the suggestions from Danz, Vesterlund, and Wilson (2021), we provide the full details of the incentive mechanism upon request. Subjects needed to actively click a button to go over the mathematical details.

[11] For example, subjects could face different power distributions after round 1, making having enough people to evaluate the same scenario for each answer infeasible.

4.4 Experimental Protocol and Administration

For the experiment, we recruited 388 subjects and ran 26 sessions at the Vernon Smith Experimental Economics Laboratory at Purdue University between February and April 2022. Table D-5 in the Appendix presents a summary of the nine treatments. Each treatment contained at least two sessions and at least 40 participants across sessions. On average, subjects earned \$22.16 (including the \$5 show-up fee) in our experiment.

Given the complexity of the environment, we took extra steps to ensure subjects understood the interface and the consequences of the cooperation and competition decisions. First, we developed an interactive interface to engage subjects throughout the instructions (see Appendix C). Second, to facilitate better understanding of how earnings and new shares are determined in stages 1 and 2, subjects had to go through five examples with step-by-step calculations. To eliminate any bias, we generated the power distribution and the choices at random.[12] Third, subjects had to answer seven comprehension questions. Although the questions were not incentivized, participants could only proceed if the answer was correct. Lastly, throughout the experiment, including the waiting pages, they had access to the payoff calculator.

5 Experimental Results

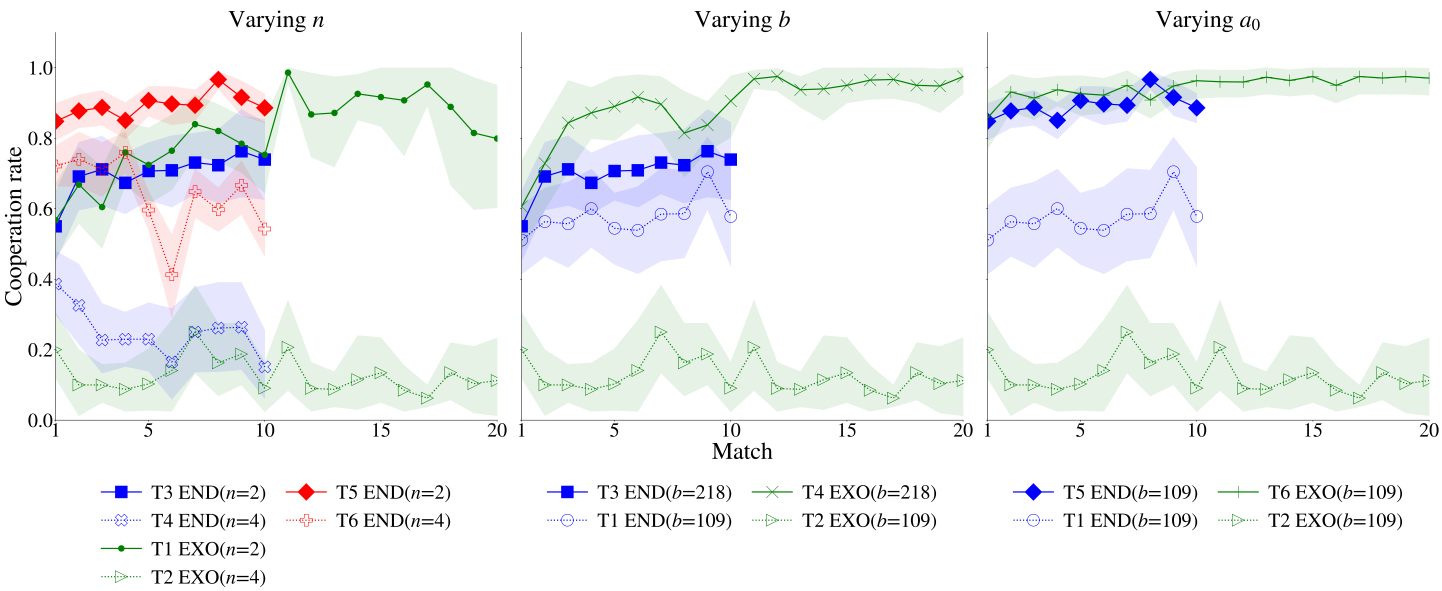
The results section is organized as follows. First, in section 5.1, we focus on the impact of fundamental parameters of the collective-action problem on the decisions of human subjects to cooperate. Next, in section 5.2, we explore the endogenous power revision and how it affects the proclivity to cooperate. Finally, in section 5.3, we estimate a behavioral model that takes into account an individual's beliefs and norms.

5.1 Effect of the Parameters of the Collective-Action Problem

Figure 6 presents the average cooperation rate across matches observed in our experiment. The three panels in the figure present the comparison of treatments based on n, b, and a_0, respectively. In particular, to make the comparison easier, we use the same color for a pair of treatments that have the same parameters other than the varied parameters. For example, treatments T1 EXO and T2 EXO in the left panel are presented in the same color (green) to indicate that all parameters with the exception of the number of participants are the same. The solid line with solid circles corresponding to T1 EXO is clearly higher than the dashed line with empty triangles corresponding to T2 EXO, indicating the strong negative impact of increasing the number of players in the group.

[12] For stage 1, subjects see five randomly generated power distributions and random choices made by each subject. They then see how their earnings in stage 1 are calculated step by step. For stage 2, in the END treatment, they see randomly generated spending, whereas in the EXO treatment they, see how the randomly generated choices from stage 1 determine the spending in stage 2.

Figure 6: Impact of Fundamental Parameters on Cooperation



Notes: The cooperation rate is the fraction of rounds in which an individual cooperated in a match. From left to right, the three panels show the impact of varying n, b, and a_0. In each panel, colors indicate pairs of treatments to be compared. For each pair, a solid line with filled markers indicates the treatment with the greater cooperation. The shaded areas show 90% bootstrapped confidence interval, treating a group in a match as one observation unit.

The raw data in Figure 6 suggest subjects respond to the game parameters as theory predicts. These results are confirmed by random-effects regressions presented in Table 3. In particular, the regressions show that the effects are highly significant whether we focus on Round 1 or all rounds, and whether we control for preferences and demographics.[13] We summarize results on the role of parameters with Result 1.

Result 1 *Hypothesis 1 is supported: the decision to cooperate in the collective-action stage responds to the fundamental parameters:*

(a) *Cooperation is increasing in the maximum benefit to cooperation (b),*

(b) *Cooperation is decreasing in the group size (n),*

(c) *Cooperation is decreasing in the proportion of the group (a_0) required to achieve half of the maximum benefit to cooperation.*

A notable observation is that round 1 cooperation rates in T1 and T3 treatments are comparable to previous one-shot stag-hunt experiments that employed similar stage-game payoffs. For example, in a game with the same payoffs as T1 for three out of four action profiles, Dal Bó, Fréchette, and Kim (2021) reports an average cooperation rate of 78.57%, whereas the average cooperation rate is

[13] In Table D-7 of the Appendix D, we provide full set of estimates including preferences and demographics.

79.2% in T1 EXO and 67.6% in T1 END treatments, respectively. In addition to the similar levels of cooperation, the upward trend across matches is present in both instances. Moreover, Dal Bó, Fréchette, and Kim (2021) find that increasing the size of the basin of attraction of stag (which is equivalent to the $SizeBC$) increases the prevalence of cooperation. In our experiment, such an increase corresponds to the comparison of T1 to T3. Our data are consistent with their finding, because the average cooperation rate increases from 67.6 % in T1 END to 72% in T3 END (p-value < 0.01).

Table 3: Cooperation in Stage 1

	All rounds			Round 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Greater n ($n = 4$)	-0.47***	-0.45***	-0.46***	-0.38***	-0.30***	-0.32***
	(0.07)	(0.06)	(0.06)	(0.08)	(0.07)	(0.06)
Greater b ($b = 218$)	0.45***	0.42***	0.43***	0.36***	0.26***	0.27***
	(0.11)	(0.10)	(0.09)	(0.11)	(0.08)	(0.08)
Greater a_0 ($a_0 = 0.812$)	-0.65***	-0.57***	-0.58***	-0.61***	-0.40***	-0.41***
	(0.09)	(0.09)	(0.08)	(0.11)	(0.08)	(0.08)
Choose effort (END)	-0.44***	-0.44***	-0.45***	-0.22***	-0.20***	-0.21***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.06)
Own R1 coop in Match 1		0.13***	0.13***		0.28***	0.28***
		(0.03)	(0.03)		(0.03)	(0.03)
Others' R1 coop in Match $t-1$		0.04**	0.04**		0.14***	0.14***
		(0.01)	(0.01)		(0.02)	(0.02)
(Length of match $t-1$) / 100		0.10	0.10		0.07	0.07
		(0.06)	(0.06)		(0.07)	(0.07)
Constant	1.41***	1.25***	1.30***	1.32***	0.90***	0.96***
	(0.08)	(0.09)	(0.12)	(0.09)	(0.09)	(0.14)
Observations	42,392	39,912	39,912	5,088	4,700	4,700
Number of Subjects	388	388	388	388	388	388
Preferences	No	No	Yes	No	No	Yes
Demographics	No	No	Yes	No	No	Yes

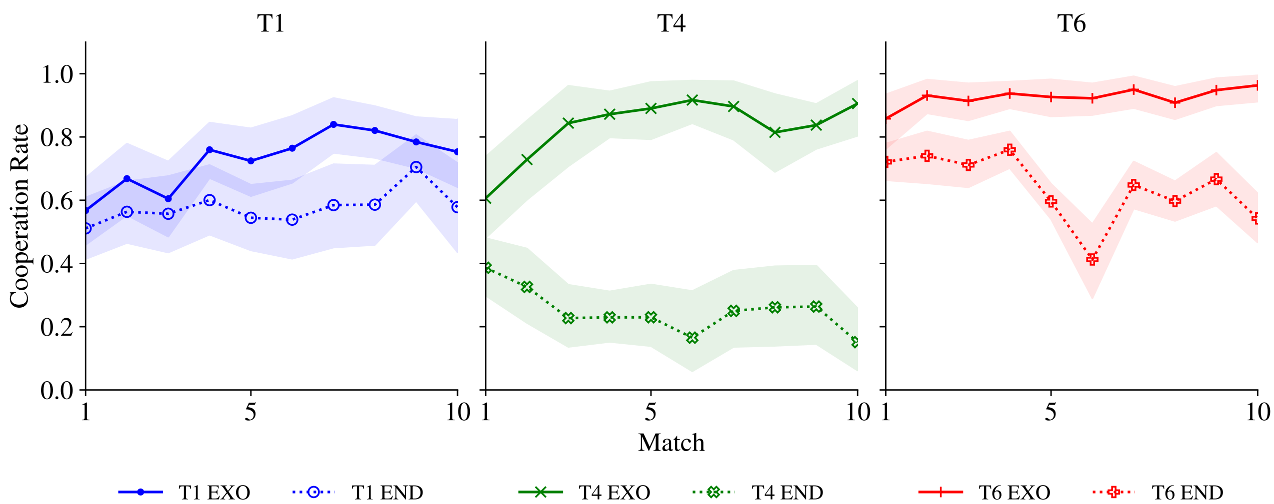
Notes: The table reports results from random-effects regressions using data across all nine treatments. The dependent variable is 1 if subjects chose "Y"(cooperation) in stage 1, and 0 otherwise. Preference measures include risk aversion, loss aversion, other-regarding preference in disadvantageous and advantageous inequality, and cognitive ability. Demographics include age, gender, major, and subjects' high school location (US or not). Standard errors are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.2 Endogenous Power Revision

In this section, we focus on the impact of contests for power on individuals' decisions to cooperate in the collective-action problem. In particular, we compare the END treatments with the EXO treatments and show that cooperation indeed decreases in the endogenous-power-revision treatment as the theory in section 3.2 predicts. We then take a closer look at the END treatment to see how well the theory predicts the competition in the contest for power. In addition, we note several observations regarding the interplay between cooperation in the collective-action stage and competition in the contest for power.

Figure 7 presents the average cooperation rate for the three pairs of treatments that isolate the impact of the contest for power in stage 2. In particular, across the three pairs, consistent with the theoretical predictions derived in section 3.2, EXO treatments have significantly higher cooperation rates. The regressions presented in Table 3 confirm the strong significance of these results.

Figure 7: Impact of Endogenous Power Revision on Cooperation

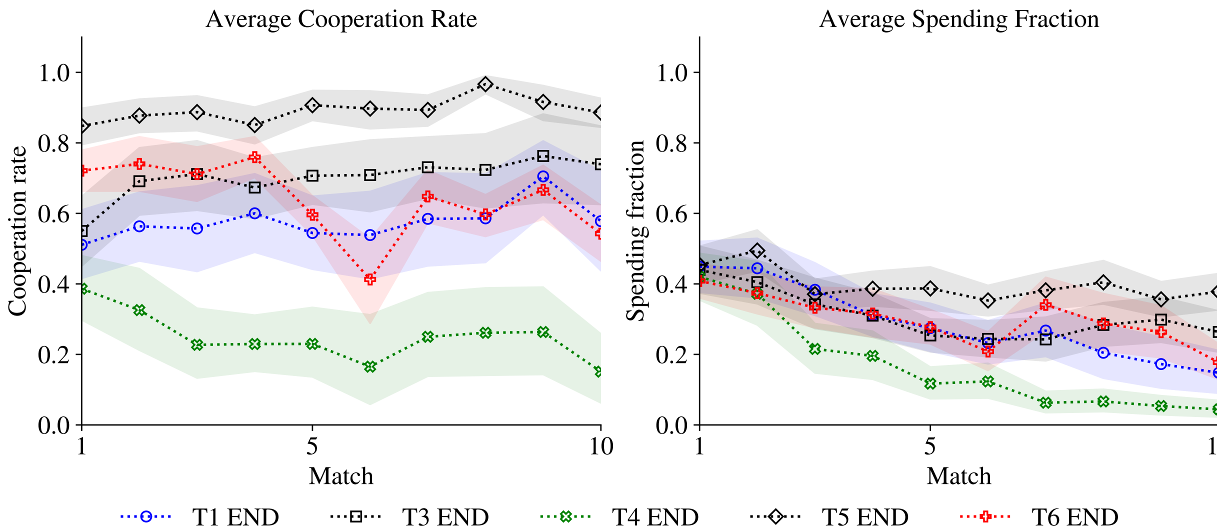


Notes: The figure presents the average cooperation rate for all rounds over 10 matches. Each panel contains one treatment pair of EXO and END treatments. For each pair, a solid (dotted) line indicates the EXO (END) treatment. The shaded areas show 90% bootstrapping confidence intervals, treating a group in a match as one observation unit.

Result 2 *Hypothesis 2 is supported: cooperation in the collective-action stage is significantly lower when subjects compete in the contest for power.*

Figure 8 presents the average cooperation rate in the stage 1 collective-action problem (left panel) and the average spending rate in the stage 2 contest for power (right panel) across the five END treatments of our experiment. The ranking of cooperation rates among the two-player settings (T1, T3, and T5) and four-player settings (T4 and T6) are as the theory in section 3.1 predicted. Regarding the spending in the contest for power, the highest proportion spent is in the T5 treatment, followed by T3, T6, T1, and lastly T4. Generally, these results are consistent with the theoretical predictions discussed in section 3.2. In particular, Figure 3 shows that T5 was unambiguously predicted to have higher proportions of spending than T1, T6 and T4, all of which held. The theoretical comparison of T5 and T3 is less clear because of the multiplicity of equilibria in the T3 case (with one equilibrium higher and one lower than T5).

Figure 8: Cooperation and Spending across END Treatments



Notes: The figure presents the average cooperation rate and spending fraction for all rounds over 10 matches for the five END treatments. In each figure, different marker symbol indicate different treatments. The shaded areas show 90% bootstrapping confidence interval, treating a group in a match as one observation unit.

Although the theory had accurate comparative static predictions, the actual level and the symmetry rarely hit the mark. In particular, our theoretical predictions based on myopic best-response generated symmetric equilibria with all cooperators spending the same amount in the contest and all defectors spending zero. In the experiment, we see a considerable degree of heterogeneity within cooperators as well as expenditures by the defectors. For example, Table 4 shows a regression of subjects' spending in stage 2 on some metrics capturing the state of the game in a round. Some of the results (e.g., negative trend across matches and increased expenditures based on the payoff from the stage 1) confirm observations from Figure 8. More interestingly, however, are results that are not directly observable from the raw data. In particular, the strong negative impact of inequality (measured as the standardized power variance) and significantly positive baseline expenditure (captured with the constant term) indicate other factors may be at play.

Table 4: Spending in Stage 2

	All		Defectors Only		Cooperators Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Pay from Cooperation	0.18***	0.18***			0.16***	0.18***
	(0.04)	(0.04)			(0.03)	(0.03)
Power Inequality		-0.06***		-0.03***		-0.09***
		(0.01)		(0.01)		(0.02)
My Power (%)		-0.05*		0.03*		-0.10***
		(0.03)		(0.01)		(0.03)
Match Number		-1.53***		-1.25***		-1.58***
		(0.36)		(0.31)		(0.50)
Length of Match $t-1$		-0.13*		-0.09		-0.13**
		(0.08)		(0.08)		(0.07)
Constant	0.30***	0.52***	0.24***	0.40***	0.36***	0.58***
	(0.04)	(0.12)	(0.04)	(0.07)	(0.03)	(0.15)
Observations	16,104	14,664	6,590	6,005	9,514	8,659
Number of Subjects	216	216	212	207	212	210
Preferences	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes

Notes: The table reports results from random-effects regressions using data from the five END treatments. The dependent variable is the stage 2 spending. Columns (1)-(2) show estimates based on all individuals. Columns (3)-(4) show individuals who choose to defect in the current round. Columns (5)-(6) show individuals who choose to cooperate in the current round. Power inequality is calculated as the group variance over the maximal variance a group can obtain (when $n = 2$, the maximal variance is 0.5; when $n = 4$, the maximal variance is 0.75. In both cases, the maximal variance happens when one person has 100% power and the rest has 0%). Preferences include risk aversion, loss aversion, other-regarding preference in disadvantageous and advantageous inequality, and cognitive ability. Demographics include age, gender, major, and the subjects' high school location. Standard errors in parentheses are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.3 Beliefs and Norms

In this section, we focus on the individual's beliefs and norms that were elicited as part of our experiment to help explain cooperative (and non-cooperative) behavior by the human participants. In particular, we find beliefs and norms respond to the environmental parameters (b, n, a_0) as well as to the nature of the power-revision contests (endogenous vs. exogenous) similarly to the cooperation decision (see Table D-16 in the Appendix). In addition, we find that the round-by-round beliefs are relatively accurate, with the average accuracy rate of 78.2% across the whole experiment and a minimum accuracy rate of 49.4% that was observed in the T3 END treatment. Finally, we find an average best-response rate of 77.8% in match 1 and an average best response rate of 88.0% in match 10.

Despite the fact that approximately 40% of subjects always best responded to their beliefs in every round of match 1 and approximately 50% of subjects always best responded to their beliefs in match 10, a significant proportion of subjects best responded at a much lower rate. For example, approximately 50% of participants in Match 1 and 25% of participants in Match 10 best responded less than 80% of the rounds they faced (see Figure D-2 in the Appendix for the full distribution). To help explain why human subjects may not best respond all the time, we estimate a behavioral model that incorporates normative factors into the random utility framework. In particular, we estimate model 10 using a logistic mixed-effect regression.[14] The results are presented in Table 5.

Results of the regression analysis show that the expected payoffs ($p < 0.01$), personal norms ($p < 0.01$), and conformity with perceived actions of others ($p < 0.01$) are associated with individual decisions to cooperate. The effect of injunctive norms ($p = 0.06$) is less salient. The less important effect of injunctive norms is well in line with the previous research (Tverskoi, Guido, Andrighetto, Sánchez, and Gavrilets, 2022). A possible explanation is that individuals did not know each other, and were randomly reshuffled every match. In addition, we found that the expected payoffs and personal norms have the highest contributions to the marginal R-squared among all the predictors, whereas the contribution of conformity is higher than that of the injunctive norms.

We perform several diagnostics of our model and robustness checks of the results. In particular, the share of the variance explained by fixed effects is 0.64 (marginal R-squared), while the share of the variance explained by both, fixed and random effects is 0.88 (conditional R-squared) indicating a good overall fit. The variance inflation scores range from 1.16 to 1.56, indicating that we did not detect multicollinearity. In addition, the Kolmogorov-Smirnov test ($p = 0.39$) and bootstrap outlier test ($p = 0.17$) indicate that no evidence of an incorrect specification of the model. Regarding the robustness of results, we check various specifications of the link function and various assumptions on the correlation structure (see Tables D-17-D-18 in the Appendix). We find that our main conclusions on the strong significant effects of the three variables (expected payoffs, conformity, and personal norms) and their contributions towards the R-squared hold. We also check results when splitting the endogenous and exogenous treatments. The results support our conclusions on the significance of expected payoffs, conformity, and personal norms. The difference between the treatments is that expected payoffs contribute more, while conformity and injunctive norms contribute less to the marginal R-squared in the endogenous treatments compared to the exogenous treatments (see Table D-19 in Appendix D).

[14] Mixed-effects regression analysis is performed using R 3.6.6. The “performance” package is used to compute pseudo R-squared metrics (Lüdecke, Ben-Shachar, Patil, Waggoner, and Makowski, 2021), the “lme4” package is used for the mixed-model estimation (Bates, Mächler, Bolker, and Walker, 2015), the ”DHARMa” package - for the residuals diagnostics (Hartig and Hartig, 2017), and the “glmm.hp” package - for hierarchical partitioning to calculate the individual contributions of each predictor to marginal Nakagawa R-squared for generalized mixed-effect models (Lai, Zou, Zhang, and Peres-Neto, 2022).

Table 5: Effects of Beliefs and Norms on Cooperation

	(1)	(2)	(3)	(4)	(5)	(6)	R^2-dec
Intercept	0.54**	0.54*	0.51	0.58***	0.10	1.51**	-
	(0.23)	(0.31)	(0.34)	(0.23)	(0.13)	(0.64)	
Expected payoffs	0.13***	-	-	-	0.07***	0.06***	0.34
	(0.01)				(0.01)	(0.01)	
Conformity	-	2.79***	-	-	1.35***	1.49***	0.21
		(0.15)			(0.29)	(0.19)	
Injunctive norms	-	-	1.37***	-	0.19*	0.21**	0.15
			(0.09)		(0.10)	(0.10)	
Personal norms	-	-	-	1.67***	1.24***	1.25***	0.30
				(0.10)	(0.11)	(0.11)	
Preferences	No	No	No	No	No	Yes	-
Demographics	No	No	No	No	No	Yes	-
Observations	8100	8100	8100	8100	8100	8100	-
AIC	4883	5434	5697	5126	4097	4122	-
BIC	4918	5469	5732	5161	4216	4297	-
marginal R^2_{Nak}	0.50	0.36	0.29	0.45	0.64	0.64	-
conditional R^2_{Nak}	0.86	0.73	0.79	0.82	0.88	0.88	-

Notes: The table reports results from the mixed-effects logistic regression using data from matches 1, 10, and 20 (if available) across the nine treatments. The dependent variable is a dummy variable $a_{i,t}$ indicating whether a subject i in round t chooses to cooperate. To capture heterogeneity among individuals, we assume random intercepts and random slopes (slopes vary among individuals). To capture the session-level effects, we assume that an intercept varies among sessions and among participants of the sessions. The marginal Nacagawa's R-squared shows a proportion of the variance explained by fixed effects, whereas the conditional Nacagawa's R-squared shows a proportion of the variance explained by both, fixed and random effects. The last column shows the results of the hierarchical partitioning of the marginal Nacagawa's R-squared. Preferences include risk aversion, loss aversion, other-regarding preference in disadvantageous and advantageous inequality, and cognitive ability. Demographics include age, gender, major, and the subjects' high school location. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6 Conclusion

In this paper, we study a model of cooperation and competition in which players split the benefits of cooperation according to the political power obtained in a contest. Our main contributions are threefold. First, we provide a theoretical foundation based on the framework of myopic best-response to show that the contest for power introduces additional considerations that decrease cooperation of the players in the cooperation stage. Second, we design and conduct an experiment to test our theoretical predictions. Finally, we estimate a behavioral model of cooperation in which a decision is based on subjective beliefs and norms regarding appropriateness of behavior in a particular situation.

Our experimental results show that human subjects predictably respond to the main parameters of the collective-action problem. For example, an increase in the benefit to cooperation results in a greater frequency of subjects cooperating, as well as greater expenditures in the contest for power. The most novel result of the paper, however, is the comparison of the endogenous and exogenous contest for power. Specifically, in the exogenous contest, we restrict players to contribute a fixed proportion of earnings from the collective-action problem, whereas in the exogenous case, they are free to choose the amount of their contribution. We find both the theory and experiments are consistent in that players significantly reduce cooperation in the collective action when the contest is not restricted. These results provide insight into the design of institutions in which cooperation is desired, but which also include a competitive stage (e.g., elections).

From the data obtained in the experiment, we estimate a behavioral model based on beliefs and norms elicited from human subjects (Gavrilets, 2021; Tverskoi, Xu, Nelson, Menassa, Gavrilets, and Chen, 2021; Houle, Ruck, Bentley, and Gavrilets, 2022; Tverskoi, Babu, and Gavrilets, 2022). We find that beliefs matter in two ways. First, they matter as a determinant of the expected payoffs of available actions. Thus, choosing actions with higher expected payoffs captures best-responding behaviors. Second, they matter as a measure of descriptive social norms. These norms allow us to specify a measure of conformity. In our estimation exercise, we find that expected payoffs account for 34% of explained variance in cooperation. The remaining variance is due to personal norms (30%), conformity (21%), and injunctive social norms (15%). Overall, our results show that understanding human cooperation is hardly possible without accounting for the effects of inequality in power, conformity and norms (Gavrilets, 2015; Houle, Ruck, Bentley, and Gavrilets, 2022).

Our study opens a number of interesting avenues for future research. First, we focused on societies composed of individual decision-makers. In the real world, political power is often held by groups or factions. Thus, studying whether groups would act differently would be interesting (e.g., Cooper and Kagel, 2005). Second, given the complexity of the experiment and some of the elicitation procedures, we did not elicit beliefs in every interaction. Understanding the evolution of beliefs both within and across interactions (e.g., Szekely, Lipari, Antonioni, Paolucci, Sánchez, Tummolini, and Andrighetto 2021) would be important. To this end, adding elicitations throughout the experiment and in the contest stage would be interesting. Finally, many real-world collective-action problems are subject to unexpected shocks (e.g., flood impact on the public infrastructure); therefore, establishing the degree to which such uncertainties affect the outcomes of collective-action problems would be interesting.

Acknowledgments

Supported by the U. S. Army Research Office grants W911NF-14-1-0637 and W911NF-18-1-0138, the Office of Naval Research grant W911NF-17-1-0150, and the Air Force Office of Scientific Research grant FA9550-21-1-0217.

References

Alós-Ferrer, C. (2003): "Finite population dynamics and mixed equilibria," *International Game Theory Review*, 5(03), 263–290.

Andreoni, J., and B. D. Bernheim (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 77(5), 1607–1636.

Andreozzi, L., M. Ploner, and A. S. Saral (2020): "The stability of conditional cooperation: beliefs alone cannot explain the decline of cooperation in social dilemmas," *Scientific Reports*, 10, 13610.

Aoyagi, M., G. R. Fréchette, and S. Yuksel (2020): "Beliefs in repeated games," *Working Paper*.

Baron, D. P. (1994): "Electoral competition with informed and uninformed voters," *American Political Science Review*, 88(1), 33–47.

Bašić, Z., and E. Verrina (2021): "Personal norms—and not only social norms—shape economic behavior," *MPI Collective Goods Discussion Paper*, 2020/25(2020/25).

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015): "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 67(1), 1–48.

Bénabou, R., A. Falk, and J. Tirole (2020): "Narratives, imperatives, and moral persuasion," *University of Bonn, mimeo*.

Bénabou, R., and J. Tirole (2006): "Incentives and prosocial behavior," *American economic review*, 96(5), 1652–1678.

Bicchieri, C. (2005): *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

——— (2016): *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C., and A. Chavez (2010): "Behaving as expected: Public information and fairness norms," *Journal of Behavioral Decision Making*, 23(2), 161–178.

Bornstein, G., U. Gneezy, and R. Nagel (2002): "The effect of intergroup competition on group coordination: An experimental study," *Games and Economic Behavior*, 41(1), 1–25.

Bosworth, S. J. (2017): "The importance of higher-order beliefs to successful coordination," *Experimental Economics*, 20(1), 237–258.

Brock, W. A., and S. P. Magee (1978): "The economics of special interest politics: The case of the tariff," *The American Economic Review*, 68(2), 246–250.

Burks, S. V., and E. L. Krupka (2012): "A multimethod approach to identifying norms and normative expectations within a corporate hierarchy: Evidence from the financial services industry," *Management Science*, 58(1), 203–217.

Camerer, C. F., and E. Fehr (2004): "Measuring social norms and preferences using experimental games: A guide for social scientists," *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 97, 55–95.

Cason, T. N., W. A. Masters, and R. M. Sheremeta (2010): "Entry into winner-take-all and proportional-prize contests: An experimental study," *Journal of Public Economics*, 94(9-10), 604–611.

——— (2020): "Winner-take-all and proportional-prize contests: theory and experimental results," *Journal of Economic Behavior & Organization*, 175, 314–327.

Condon, D. M., and W. Revelle (2014): "The international cognitive ability resource: Development and initial validation of a public-domain measure," *Intelligence*, 43, 52–64.

Cooper, D. J., C. A. Ioannou, and S. Qi (2018): "Endogenous incentive contracts and efficient coordination," *Games and Economic Behavior*, 112, 78–97.

Cooper, D. J., and J. H. Kagel (2005): "Are two heads better than one? Team versus individual play in signaling games," *American Economic Review*, 95(3), 477–509.

Cooper, D. J., and J. Van Huyck (2018): "Coordination and transfer," *Experimental Economics*, 21(3), 487–512.

Cooper, D. J., and R. A. Weber (2020): "Recent advances in experimental coordination games," *Handbook of Experimental Game Theory*, pp. 149–183.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1992): "Communication in Coordination Games*," *The Quarterly Journal of Economics*, 107(2), 739–771.

Cooper, R. W., D. V. DeJong, R. Forsythe, and T. W. Ross (1990): "Selection criteria in coordination games: Some experimental results," *The American Economic Review*, 80(1), 218–233.

Costa-Gomes, M. A., and G. Weizsäcker (2008): "Stated beliefs and play in normal-form games," *The Review of Economic Studies*, 75(3), 729–762.

d'Adda, G., M. Dufwenberg, F. Passarelli, and G. Tabellin (2020): "Social norms with private values: Theory and experiments," *Games and Economic Behavior*, 124, 288–304.

d'Adda, G., M. Dufwenberg, F. Passarelli, and G. Tabellini (2020): "Social norms with private values: Theory and experiments," *Games and Economic Behavior*, 124, 288–304.

Dal Bó, P., G. R. Fréchette, and J. Kim (2021): "The determinants of efficient behavior in coordination games," *Games and Economic Behavior*.

Danz, D., L. Vesterlund, and A. J. Wilson (2021): "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*.

Davis, D., A. Ivanov, and O. Korenok (2016): "Individual characteristics and behavior in repeated games: An experimental study," *Experimental Economics*, 19(1), 67–99.

Dechenaux, E., D. Kovenock, and R. M. Sheremeta (2015): "A survey of experimental research on contests, all-pay auctions and tournaments," *Experimental Economics*, 18(4), 609–669.

Erkal, N., L. Gangadharan, and B. H. Koh (2020): "Replication: Belief elicitation with quadratic and binarized scoring rules," *Journal of Economic Psychology*, 81, 102315.

Fehr, E., and U. Fischbacher (2004a): "Social norms and human cooperation," *Trends in cognitive sciences*, 8(4), 185–190.

——— (2004b): "Third-party punishment and social norms," *Evolution and human behavior*, 25(2), 63–87.

Fehr, E., and S. Gächter (2000): "Fairness and retaliation: The economics of reciprocity," *Journal of economic perspectives*, 14(3), 159–181.

Fehr, E., and I. Schurtenberger (2018): "Normative foundations of human cooperation," *Nature Human Behaviour*, 2(7), 458–468.

Findlay, R., and S. Wellisz (1982): "Endogenous tariffs, the political economy of trade restrictions, and welfare," in *Import competition and response*, pp. 223–244. University of Chicago Press.

Fischbacher, U., and S. Gächter (2010): "Social preferences, beliefs, and the dynamics of free riding in public goods experiments," *American economic review*, 100(1), 541–56.

Gardner, R., E. Ostrom, and J. M. Walker (1990): "The nature of common-pool resource problems," *Rationality and society*, 2(3), 335–358.

Gavrilets, S. (2015): "Collective action problem in heterogeneous groups," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683), 20150016, Publisher: The Royal Society.

——— (2021): "Coevolution of actions, personal norms and beliefs about others in social dilemmas," *Evolutionary Human Sciences*, 3, e44.

Gill, D., and Y. Rosokha (2020): "Beliefs, Learning, and Personality in the Indefinitely Repeated Prisoner's Dilemma," *Available at SSRN 3652318*.

Górges, L., and D. Nosenzo (2020): "Measuring Social Norms in Economics: Why It Is Important and How It Is Done," *Analyse & Kritik*, 42, 285–311.

Grossman, G. M., and E. Helpman (1996): "Electoral competition and special interest politics," *The Review of Economic Studies*, 63(2), 265–286.

Harsanyi, J. C., and R. Selten (1988): "A general theory of equilibrium selection in games," *MIT Press Books*, 1.

Hartig, F., and M. F. Hartig (2017): "Package 'DHARMa'," *Vienna, Austria: R Development Core Team*.

Heinemann, F., R. Nagel, and P. Ockenfels (2009): "Measuring strategic uncertainty in coordination games," *The review of economic studies*, 76(1), 181–221.

Hirshleifer, J. (1983): "From weakest-link to best-shot: The voluntary provision of public goods," *Public choice*, 41(3), 371–386.

Holt, C. A., and S. K. Laury (2002): "Risk aversion and incentive effects," *American economic review*, 92(5), 1644–1655.

Hopkins, E. (1999): "A note on best response dynamics," *Games and Economic Behavior*, 29(1-2), 138–150.

Hossain, T., and R. Okui (2013): "The binarized scoring rule," *Review of Economic Studies*, 80(3), 984–1001, Publisher: Oxford University Press.

Houle, C., D. J. Ruck, R. A. Bentley, and S. Gavrilets (2022): "Inequality between identity groups and social unrest," *Journal of the Royal Society Interface*, 19(188), 20210725.

Kandori, M., G. J. Mailath, and R. Rob (1993): "Learning, mutation, and long run equilibria in games," *Econometrica: Journal of the Econometric Society*, pp. 29–56.

Kandori, M., and R. Rob (1995): "Evolution of equilibria in the long run: A general theory and applications," *Journal of Economic Theory*, 65(2), 383–414.

Kerschbamer, R. (2015): "The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test," *European Economic Review*, 76, 85–103.

Kim, Y. (1996): "Equilibrium selection inn-person coordination games," *Games and Economic Behavior*, 15(2), 203–227.

Kölle, F., and S. Quercia (2021): "The influence of empirical and normative expectations on cooperation," *Journal of Economic Behavior & Organization*, 190, 691–703.

Krueger, A. O. (1974): "The political economy of the rent-seeking society," *The American economic review*, 64(3), 291–303.

Krupka, E., S. Leider, and M. Jiang (2020): "Renegotiation Behavior and Promise-Keeping Norms," in *Working Paper*. Working Paper.

Krupka, E. L., S. Leider, and M. Jiang (2017): "A meeting of the minds: informal agreements and social norms," *Management Science*, 63(6), 1708–1729.

Krupka, E. L., and R. A. Weber (2013): "Identifying social norms using coordination games: Why does dictator game sharing vary?," *Journal of the European Economic Association*, 11(3), 495–524.

Lai, J., Y. Zou, J. Zhang, and P. R. Peres-Neto (2022): "Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca. hp R package," *Methods in Ecology and Evolution*, 13(4), 782–788.

Lüdecke, D., M. S. Ben-Shachar, I. Patil, P. Waggoner, and D. Makowski (2021): "performance: An R package for assessment, comparison and testing of statistical models," *Journal of Open Source Software*, 6(60).

Mäs, M., and H. H. Nax (2016): "A behavioral study of "noise" in coordination games," *Journal of Economic Theory*, 162, 195–208.

Matsui, A. (1992): "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, 57(2), 343–362.

Morris, S., R. Rob, and H. S. Shin (1995): "p-Dominance and belief potential," *Econometrica: Journal of the Econometric Society*, pp. 145–157.

Nyarko, Y., and A. Schotter (2002): "An experimental study of belief learning using elicited beliefs," *Econometrica*, 70(3), 971–1005.

Offerman, T., J. Sonnemans, and A. Schram (2001): "Expectation formation in step-level public good games," *Economic Inquiry*, 39(2), 250–269.

Peski, M. (2010): "Generalized risk-dominance and asymmetric dynamics," *Journal of Economic Theory*, 145(1), 216–248.

Reuben, E., and A. Riedl (2013): "Enforcement of contribution norms in public good games with heterogeneous populations," *Games and Economic Behavior*, 77(1), 122–137.

Roca, C. P., J. A. Cuesta, and A. Sánchez (2009): "Promotion of cooperation on networks? The myopic best response case," *The European Physical Journal B*, 71(4), 587–595.

Rosokha, Y., and C. Wei (2020): "Cooperation in queueing systems," *Available at SSRN 3526505.*

Roth, A. E., and J. K. Murnighan (1978): "Equilibrium behavior and repeated play of the prisoner's dilemma," *Journal of Mathematical psychology*, 17(2), 189–198.

Rousseau, J.-J. (1754): "Discourse on the origin and basis of inequality among men," *Rousseau J.-J. Traktaty*, pp. 31–109.

Rubin, J., A. Samek, and R. M. Sheremeta (2018): "Loss aversion and the quantity–quality tradeoff," *Experimental Economics*, 21(2), 292–315.

Samuelson, P. A. (1954): "The pure theory of public expenditure," *The review of economics and statistics*, pp. 387–389.

Sandholm, W. H. (1998): "Simple and clever decision rules for a model of evolution," *Economics Letters*, 61(2), 165–170.

Savikhin, A. C., and R. M. Sheremeta (2013): "Simultaneous decision-making in competitive and cooperative environments," *Economic Inquiry*, 51(2), 1311–1323.

Schmidt, D., R. Shupp, J. M. Walker, and E. Ostrom (2003): "Playing safe in coordination games:: the roles of risk dominance, payoff dominance, and history of play," *Games and Economic Behavior*, 42(2), 281–299.

Smith, J. M. (1982): *Evolution and the Theory of Games*. Cambridge university press.

Stoddard, B., J. M. Walker, and A. Williams (2014): "Allocating a voluntarily provided common-property resource: An experimental examination," *Journal of Economic Behavior & Organization*, 101, 141–155.

Szekely, A., F. Lipari, A. Antonioni, M. Paolucci, A. Sánchez, L. Tummolini, and G. Andrighetto (2021): "Evidence from a long-term experiment that collective risks change social norms and promote cooperation," *Nature communications*, 12(1), 1–7.

Szolnoki, A., and M. Perc (2014): "Evolution of extortion in structured populations," *Physical Review E*, 89(2), 022804.

Tullock, G. (1967): "The welfare costs of tariffs, monopolies, and theft," *Economic inquiry*, 5(3), 224–232.

——— (1980): "Efficient rent seeking," *Toward a Theory of the Rent-seeking Society*, pp. 97–112.

Tverskoi, D., S. Babu, and S. Gavrilets (2022): "The spread of technological innovations: effects of psychology, culture, and policy interventions," *J. R. Soc. Open Science*, 9, 211833.

Tverskoi, D., A. Guido, G. Andrighetto, A. Sánchez, and S. Gavrilets (2022): "Disentangling material, social, and cognitive determinants of human behavior and beliefs," *SocArXiv*.

Tverskoi, D., A. Senthilnathan, and S. Gavrilets (2021): "The dynamics of cooperation, power, and inequality in a group-structured society," *Scientific reports*, 11(1), 1–16.

Tverskoi, D., X. Xu, H. Nelson, C. Menassa, S. Gavrilets, and C.-F. Chen (2021): "Energy saving at work: Understanding the roles of normative values and perceived benefits and costs in single-person and shared offices in the United States," *Energy Research & Social Science*, 79, 102173.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990): "Tacit coordination games, strategic uncertainty, and coordination failure," *The American Economic Review*, 80(1), 234–248.

Vespa, E. (2020): "An experimental investigation of cooperation in the dynamic common pool game," *International Economic Review*, 61(1), 417–440.

Vespa, E., and A. J. Wilson (2019): "Experimenting with the transition rule in dynamic games," *Quantitative Economics*, 10(4), 1825–1849, Publisher: Wiley Online Library.

Young, H. P. (1993): "The evolution of conventions," *Econometrica: Journal of the Econometric Society*, pp. 57–84.

——— (1998): "Social norms and economic welfare," *European Economic Review*, 42(3-5), 821–830.

Appendix A Details of Theoretical Predictions

Appendix A.1 The Size of Basin of Attraction of Cooperation

In this section, we present the calculation of the size of the basin of attraction of cooperation in our experiment for both $n = 2$ and $n = 4$ cases. The size of the basin of attraction $SizeBC$ provides a continuous measure of the strategic uncertainty. The higher $SizeBC$ is, the more robust cooperation is to strategic uncertainty. In a two-player game, it is defined as the maximum probability of the other player choosing defection such that choosing cooperation is still a best response to the row (focal) player. In the four-player game, when every player has the same power (such as in round 1), we impose one additional assumption that all other players have the same probability of playing defection[15]. Thus, the definition of $SizeBC$ maintains to be the maximum probability of other players choosing defection that makes cooperation remain a best response to the row (focal) player. Note that we compare the round 1 game only and use it as an equilibrium selection criteria to explain the cooperation rate difference among treatments with different game parameters.

Appendix A.1.1 $n = 2$

When there are two decision-makers in the game (such as the case for the T1, T3, T5 treatments), the payoffs of cooperation(C) and defection(D) in round 1 are:

	0	1
C	$\pi_i^1((1,0),0)$	$\pi_i^1((1,1),0)$
D	R_0	R_0

where

$$\pi_i^1((1,0),0) = R_0 + F(0.5) - c = R_0 + b\frac{0.5^\kappa}{0.5^\kappa + a_0^\kappa} - c$$

$$\pi_i^1((1,1),0) = R_0 + \frac{F(1)}{2} - c = R_0 + b\frac{1^\kappa}{2(1^\kappa + a_0^\kappa)} - c$$

Consider a focal (row) individual. Let pr be the probability of the other player choosing defection. By the definition, $SizeBC$ is the maximum value of $pr \in [0,1]$ such that:

$$pr\pi_i^1((1,0),0) + (1-pr)\pi_i^1((1,1),0) \geq R_0$$

$$pr\left(\frac{0.5^\kappa}{0.5^\kappa + a_0^\kappa} - \frac{1}{2(1+a_0^\kappa)}\right) \geq \frac{c}{b} - \frac{1}{2(1+a_0^\kappa)}$$

$$pr(1 + 2a_0^\kappa - 2^\kappa a_0^\kappa) \geq \frac{(2c + 2ca_0^\kappa - b)(1 + 2^\kappa a_0^\kappa)}{b}$$

The term $(2c + 2ca_0^\kappa - b)$ is negative for the chosen parameters b, c and $a_0 \in (0,1)$. The term $(1 + 2a_0^\kappa - 2^\kappa a_0^\kappa)$ is positive if $a_0 \leq \frac{1}{(2^\kappa - 2)^\kappa}$. As a result (and keeping in mind that $pr \in [0,1]$),

$$SizeBC = \begin{cases} 1, \text{ if } a_0 \leq 0.5\left(\frac{b}{c} - 1\right)^{\frac{1}{\kappa}}, \\ \frac{(2c+2ca_0^\kappa - b)(1+2^\kappa a_0^\kappa)}{b(1+2a_0^\kappa - 2^\kappa a_0^\kappa)}, \text{ otherwise.} \end{cases}$$

[15] We also assume that there exists at most one mixed NE in the game

These results are presented in Figure 1. In the case of $a_0 > 0.5(b/c-1)^{\frac{1}{\kappa}}$, we can calculate the derivatives:

$$\frac{\partial SizeBC}{\partial b} = \frac{2c\left(a_0^\kappa + 1\right)\left(2^\kappa a_0^\kappa + 1\right)}{b^2\left(\left(2^\kappa - 2\right) a_0^\kappa - 1\right)}$$

$$\frac{\partial SizeBC}{\partial a_0} = -\frac{2\kappa a_0^{\kappa-1}\left(b\left(2^\kappa - 1\right) + c\left(2^\kappa\left(2^\kappa - 2\right) a_0^{2\kappa} - 2^{\kappa+1} a_0^\kappa - 2^{\kappa+1} + 1\right)\right)}{b\left(\left(2^\kappa - 2\right) a_0^\kappa - 1\right)^2}$$

It is straightforward to show that $\frac{\partial SizeBC}{\partial b} > 0$ and $\frac{\partial SizeBC}{\partial a_0} < 0$ for the parameters chosen in the experiment.

Appendix A.1.2 $\quad n = 4$

When there are four decision makers in the game (T2, T4, T6 treatments), the payoffs of cooperation and defection in round 1 can be presented as:

	0	1	2	3
C	$\Pi(0)$	$\Pi(1)$	$\Pi(2)$	$\Pi(3)$
D	R_0	R_0	R_0	R_0

where $\Pi(x) = \pi_i^1((1, a_{-i}), 0)$ and $a_{-i} \cdot \mathbb{1} = x, \forall x \in I \backslash \{n\}$. Consequently,

$$\begin{aligned}
\Pi(0) &= R_0 + F(0.25) - c = R_0 + b\frac{0.25^\kappa}{0.25^\kappa + a_0^\kappa} - c \\
\Pi(1) &= R_0 + \frac{F(0.5)}{2} - c = R_0 + b\frac{0.5^\kappa}{2(0.5^\kappa + a_0^\kappa)} - c \\
\Pi(2) &= R_0 + \frac{F(0.75)}{3} - c = R_0 + b\frac{0.75^\kappa}{3(0.75^\kappa + a_0^\kappa)} - c \\
\Pi(3) &= R_0 + \frac{F(1)}{4} - c = R_0 + b\frac{1^\kappa}{4(1^\kappa + a_0^\kappa)} - c
\end{aligned}$$

Consider a focal (row) individual. To calculate $SizeBC$, we assume the probability of any other player choosing defection is the same and denote it as pr. By the definition, $SizeBC$ is the maximum value of $pr \in [0,1]$ such that:

$$L(pr) = pr^3 F(0.25) + \frac{3}{2} pr^2(1-pr)F(0.5) + pr(1-pr)^2 F(0.75) + \frac{1}{4}(1-pr)^3 F(1) - c \geq 0.$$

Our results on $SizeBC$ are based on the following lemma.

Lemma. There exists a solution $pr \in [0,1]$ to the equation $L(pr) = 0$ if

$$\frac{1}{4}\sqrt[\kappa]{\frac{b}{c} - 1} < a_0 < \sqrt[\kappa]{\frac{b}{4c} - 1},$$

In addition, there exists a unique solution $pr^* \in [0,1]$ to the equation $L(pr) = 0$, and $SizeBC = pr^*$ if

- (a) $a_0 < \frac{1}{\sqrt[\kappa]{3\left(\frac{4}{3}\right)^\kappa - 4}}$, or
- (b) $a_0^1 < a_0 < a_0^2$, where $a_0^{1,2} = \frac{-(-3\cdot 6^\kappa + 2\cdot 3^\kappa + 7\cdot 4^\kappa) \pm \sqrt{(-3\cdot 6^\kappa + 2\cdot 3^\kappa + 7\cdot 4^\kappa)^2 - 12\cdot 3^\kappa\cdot(8^\kappa + 6\cdot 4^\kappa - 4\cdot 6^\kappa)}}{2\cdot(8^\kappa + 6\cdot 4^\kappa - 4\cdot 6^\kappa)}$,

Proof. It is straightforward to show that $L(0) > 0$ and $L(1) < 0$ if $\frac{1}{4}\sqrt[\kappa]{\frac{b}{c} - 1} < a_0 < \sqrt[\kappa]{\frac{b}{4c} - 1}$. Then, according to the mean value theorem, there exists a solution $pr \in (0, 1)$ to the equation $L(pr) = 0$. The existence is proved.

In addition, if $a_0 < \frac{1}{\sqrt[\kappa]{3\left(\frac{4}{3}\right)^{\kappa} - 4}}$, we obtain that $L'(0) > 0$, which implies the uniqueness of the solution to the equation $L(pr) = 0$ on $[0, 1]$ since $L(pr)$ is a cubic function.

If $a_0^1 < a_0 < a_0^2$, it follows that $F(0.75) > \frac{3}{2}F(0.5) + \frac{1}{4}F(1)$. As a result, the coefficient of $L(pr)$ in front of pr^3 is positive. Consequently, $L(pr) < 0$ if pr is very small, and $L(pr) > 0$ if pr is very large. As a result, according to the mean value theorem and keeping in mind that $L(pr)$ is a cubic function, we prove the uniqueness of the solution to the equation $L(pr) = 0$ on $[0, 1]$. Lemma is proved.

As a result, for the chosen values of parameters b, c, κ, and $a_0 \in (0.30, 0.83)$, there is an unique solution $pr^* \in [0, 1]$ to the equation $L(pr) = 0$, and $SizeBC = pr^*$.[16] We found the corresponding values of the $SizeBC$ numerically. The results are summarized in figure 1.

Appendix A.2 Myopic best response

Here we present detailed theoretical results on the exogenous and endogenous versions of the model assuming that individuals use myopic best response to make the decision $a_{i,t} \in \{0, 1\}$ to cooperate in stage 1, and their effort $e_{i,t} \in [0, \pi_i^1(a_t, e_{t-1})]$ to spend in the contest for power in stage 2 (in the endogenous version of the model). For both versions of the model, we derive the best response function, show that all existing equilibria are symmetric, and analyze them providing conditions for their existence.

Appendix A.2.1 Exogenous power revision.

Under the assumption of myopic best response, each player best responds against the profile of the other players' choices in the previous round. That is, in stage 1 of period $t + 1$, player i chooses

$$a_{i,t+1} = BR_i^a(a_{-i,t}, e_t) = \underset{a_i \in \{0,1\}}{\operatorname{argmax}} \, \pi_i^1((a_i, a_{-i,t}), e_t)$$

which transforms to

$$a_{i,t+1} = BR_i^a(a_{-i,t}, e_t) = \begin{cases} 1, & \text{if } \frac{e_{i,t}}{e_{i,t} + a_{-i,t} \cdot e_{-i,t}} F_C(\bar{a}_{-i,t}) \geq c, \\ 0, & \text{otherwise,} \end{cases} \tag{11}$$

where $F_C(\bar{a}_{-i,t}) = F\left(\frac{1+(n-1)\bar{a}_{-i,t}}{n}\right)$ is the total production when player i cooperates and $\bar{a}_{-i,t}$ is the proportion of cooperators among other players. Intuitively, this means that a player cooperates if her share of the jointly produced resource covers the costs of cooperation.

Proposition 2 *All myopic best response equilibria (see Definition 1 in the main text) existing in the model are symmetric in that all cooperators (if exist) exert the same effort $\hat{e}_C$, and all defectors (if exist) provide the same effort $\hat{e}_D$.*

Proof. First, according to Definition 1, all defectors exert the same effort

$$\hat{e}_D = \gamma R_0. \tag{12}$$

[16] Precisely, a_0 should be in $(0.28, 0.83)$ if $b = 109$; and a_0 should be in $(0.30, 0.91)$ if $b = 218$.

Second, assume there are two cooperators, $i, j \in I$. Then, according to Definition 1,

$$\frac{\hat{e}_i}{\hat{e}_j} = \frac{\pi_i^1(a^*, \hat{e})}{\pi_j^1(a^*, \hat{e})} = \frac{R_0 - c + \frac{\hat{e}_i}{a^* \cdot \hat{e}} \cdot F(\bar{a}^*)}{R_0 - c + \frac{\hat{e}_j}{a^* \cdot \hat{e}} \cdot F(\bar{a}^*)},$$

which implies that

$$(R_0 - c)(\hat{e}_i - \hat{e}_j) = 0, \text{ so that } \hat{e}_i = \hat{e}_j = \hat{e}_C,$$

where

$$\hat{e}_C = \gamma\Big(R_0 - c + \frac{1}{n_C^*} F(n_C^*/n)\Big). \tag{13}$$

The proposition is proved. As a corollary, there can be no more than $n+1$ equilibria in the model (with $n_C^* \in \{0, 1, .., n\}$ cooperators, respectively; and $\hat{e}_C$, $\hat{e}_D$ defined by equations 13 and 12, respectively). By Definition 1, to obtain conditions for the existence of the symmetric equilibrium with n_C^* cooperators, one should check that (1) for each cooperator, cooperation is the best response to actions of others (given $\hat{e}_D$ and $\hat{e}_C$) if $n_C^* - 1$ of them cooperate; and (2) for each defector, defection is the best response to actions of others (given $\hat{e}_D$ and $\hat{e}_C$) if n_C^* of them cooperate.

Employing Equation 11 we obtain that the first condition is equivalent to

$$c \leq \frac{F(n_C^*/n)}{n_C^*} \tag{14}$$

(meaning a cooperator will not be interested in withdrawing from cooperation if her share in the jointly produced resource exceeds her cost of cooperation), while the second one is equivalent to

$$F\left(\frac{n_C^* + 1}{n}\right) < \frac{c}{R_0}\left((n_C^* + 1)R_0 - n_C^* c + F(n_C^*/n)\right) \tag{15}$$

(meaning a defector will not be interested in cooperating if the resource $F\left(\frac{n_C^*+1}{n}\right)$ that can be jointly produced with the cooperators does not exceed a special threshold which increases with increasing costs of cooperation c and resource $F(n_C^*/n)$ produced by the cooperators if the focal individual keeps defecting).

Finally, we show that the above equilibria are stable to small perturbations in individual effort.

Definition 3 *Let a^* is a myopic best response equilibrium in the exogenous version of the model with the corresponding effort $\hat{e}$ defined by equation 7. Then, a^* is stable to small perturbations in effort if $\hat{e}$ is a locally stable equilibrium of the system*

$$e_{i,t} = \gamma \pi_i^1(a^*, e_{t-1}), i \in I. \tag{16}$$

Proposition 3 *All symmetric equilibria existing in the exogenous version of the model are stable to small perturbations in power.*

Proof. It is straightforward that we can consider System 16 only for cooperators. Consider a symmetric equilibrium with $n_C^* > 0$ cooperators. Then, the Hessian matrix H for the corresponding System 16 is an $n_C^* \times n_C^*$ matrix with elements $H_{k,k} = \frac{\gamma F(n_C^*/n)}{\hat{e}_C (n_C^*)^2}(n_C^* - 1), \forall k \in \{1, .., m\}$, and $H_{l,k} = -\frac{\gamma F(n_C^*/n)}{\hat{e}_C (n_C^*)^2}, \forall k \neq l \in \{1, .., m\}$. The Hessian matrix has eigenvalues $\frac{\gamma F(n_C^*/n)}{\hat{e}_C n_C^*}$ and 0 with multiplicities $n_C^* - 1$ and 1, respectively. $\hat{e}_C$ is a locally stable equilibrium of the System 16 if the absolute values of all eigenvalues are less than 1. This holds if $R_0 > c$, which is an assumption of our model. The proposition is proved.

Appendix A.2.2 Endogenous power revision

In the endogenous power revision model, an individual makes decision to cooperate in stage 1, and decides on the effort to spend in the contest for power in stage 2. Since the effort $e_{i,t}$ individual i spends in stage 2 of period t directly affects not only the current payoff, but also the next period payoff (which also depends on the individual decision on cooperation $a_{i,t+1}$ in stage 1 of period $t+1$), we assume that the individual simultaneously chooses the effort $e_{i,t}$ in stage 2 of period t and action $a_{i,t+1}$ in stage 1 of period $t+1$ to maximize her expected total earnings in stage 2 of period t and stage 1 of period $t+1$ using myopic best response. That is, if $a_{-i,t} \cdot e_{-i,t-1} \neq 0$ or $a_{-i,t} = 0$, in stage 2 of period t, player i chooses

$$(a_{i,t+1}, e_{i,t}) = BR_i^{a,e}(a_t, e_{t-1}) = \underset{a_i \in \{0,1\}, e_i \in [0, \pi_i^1(a_t, e_{t-1})]}{\operatorname{argmax}} \Big\{ -e_i + \delta \pi_i^1((a_i, a_{-i,t}), (e_i, e_{-i,t-1})) \Big\},$$

Proposition 4 *Assume that $a_{-i,t} \cdot e_{-i,t-1} = 0$ and $a_{-i,t} \neq 0$. Then player i to maximize her expected earnings $E_i(a_i, e_i) = -e_i + \delta \pi_i^1((a_i, a_{-i,t}), (e_i, e_{-i,t-1}))$ on the set $\{0,1\} \times [0, \pi_i^1(a_t, e_{t-1})]$,*

- *is motivated to cooperate (i.e., $a_i = 1$) and make effort $e_i = \varepsilon$, where $\varepsilon > 0$ and $\varepsilon \to 0$ if $c < F_C(\bar{a}_{-i,t})$,*
- *chooses to defect with a zero effort, otherwise.*

Proof. Note that $\operatorname{argmax}_{e_i \in [0, \pi_i^1(a_t, e_{t-1})]} E_i(0, e_i) = 0$. Also note that $E_i(1,0) = \delta\Big(R_0 - c + \frac{F_C(\bar{a}_{-i,t})}{1 + a_{-i,t} \cdot \mathbb{1}}\Big) < -\varepsilon + \delta(R_0 - c + F_C(\bar{a}_{-i,t})) = E_i(1, \varepsilon)$ if $\varepsilon \in \Big(0, \delta F_C(\bar{a}_{-i,t}) \frac{a_{-i,t} \cdot \mathbb{1}}{1 + a_{-i,t} \cdot \mathbb{1}}\Big)$.

First, consider a case with $F_C(\bar{a}_{-i,t}) \leq c$. Then, $E_i(0,0) > E_i(1, \varepsilon), \forall \varepsilon > 0$. As a result,

$$\underset{a_i \in \{0,1\}, e_i \in [0, \pi_i^1(a_t, e_{t-1})]}{\operatorname{argmax}} E_i(a_i, e_i) = (0,0).$$

Second, consider a case with $F_C(\bar{a}_{-i,t}) > c$. Then, $E_i(0,0) < E_i(1, \varepsilon)$ if $\varepsilon \in \Big(0, \delta(F_C(\bar{a}_{-i,t}) - c)\Big)$. As a result, an individual maximizing her expected earnings should choose among strategies $(1, \varepsilon)$, where $\varepsilon \in \Bigg(0, min\Big\{\delta F_C(\bar{a}_{-i,t}) \frac{a_{-i,t} \cdot \mathbb{1}}{1 + a_{-i,t} \cdot \mathbb{1}}, \delta(F_C(\bar{a}_{-i,t}) - c)\Big\}\Bigg)$. Since $E_i(1, \varepsilon) = -\varepsilon + \delta(R_0 - c + F_C(\bar{a}_{-i,t}))$, an individual should choose $\varepsilon > 0$ such that $\varepsilon \to 0$. The proposition is proved.

Proposition 5 *If $a_{-i,t} = 0$,*

$$(a_{i,t+1}, e_{i,t}) = BR_i^{a,e}(a_t, e_{t-1}) = \begin{cases} (1,0), & \text{if } c \leq F_C(0), \\ (0,0), & \text{otherwise.} \end{cases} \tag{17}$$

Proof. Since $E_i(1, e_i) = -e_i + \delta(R_0 - c + F_C(0))$, one concludes that $\operatorname{argmax}_{e_i \in [0, \pi_i^1(a_t, e_{t-1})]} E_i(1, e_i) = 0$. Moreover, $\operatorname{argmax}_{e_i \in [0, \pi_i^1(a_t, e_{t-1})]} E_i(0, e_i) = 0$. Then, $BR^{a,e}(a_t, e_{t-1}) = 1$ if $E_i(1,0) \geq E_i(0,0)$ and $BR^{a,e}(a_t, e_{t-1}) = 0$, otherwise, which is equivalent to the statement of the proposition. The proposition is proved.

Proposition 6 *Let $\sqrt{\delta F_C(\bar{a}_{-i,t}) a_{-i,t} \cdot e_{-i,t-1}} - a_{-i,t} \cdot e_{-i,t-1} \leq \pi_i^1(a_t, e_{t-1})$, and $a_{-i,t} \cdot e_{-i,t-1} > 0$. Then,*

$$(a_{i,t+1}, e_{i,t}) = BR_i^{a,e}(a_t, e_{t-1}) =$$

$$= \begin{cases} (1, \sqrt{\delta F_C(\bar{a}_{-i,t}) a_{-i,t} \cdot e_{-i,t-1}} - a_{-i,t} \cdot e_{-i,t-1}), & \text{if } \sqrt{a_{-i,t} \cdot e_{-i,t-1}} < \sqrt{\delta}(\sqrt{F_C(\bar{a}_{-i,t})} - \sqrt{c}), \\ (0,0), \text{otherwise.} \end{cases} \tag{18}$$

Proof. First, note that $\operatorname{argmax}_{e_i \in [0, \pi_i^1(a_t, e_{t-1})]} E_i(0, e_i) = 0$, and

$$\underset{e_i \in [0, \pi_i^1(a_t, e_{t-1})]}{\operatorname{argmax}} E_i(1, e_i) = \begin{cases} \sqrt{\delta F_C(\bar{a}_{-i,t}) a_{-i,t} \cdot e_{-i,t-1}} - a_{-i,t} \cdot e_{-i,t-1}, \text{ if } a_{-i,t} \cdot e_{-i,t-1} < \delta F_C(\bar{a}_{-i,t}), \\ 0, \text{otherwise.} \end{cases}$$

Since $E_i(0,0) > E_i(1,0)$, one concludes that $BR^{a,e}(a_t, e_{t-1}) = (1, \sqrt{\delta F_C(\bar{a}_{-i,t}) a_{-i,t} \cdot e_{-i,t-1}} - a_{-i,t} \cdot e_{-i,t-1})$ if $a_{-i,t} \cdot e_{-i,t-1} \leq \delta F_C(\bar{a}_{-i,t})$ and $E_i(0,0) \leq E_i(1, \sqrt{\delta F_C(\bar{a}_{-i,t}) a_{-i,t} \cdot e_{-i,t-1}} - a_{-i,t} \cdot e_{-i,t-1})$. Otherwise, $BR^{a,e}(a_t, e_{t-1}) = (0,0)$. A straightforward algebraic manipulations show that this is equivalent to the statement of the proposition. The proposition is proved.

Proof of Proposition 1 in the main text. Consider an equilibrium with $n_C^* \in \{0, 1, .., n\}$ cooperators, and $n - n_C^*$ defectors. Assume that $n_C^* < n$, then, as follows from Propositions 4-6, each defector has an effort $e_D^* = 0$.

Assume that $n_C^* = 1$. Let player i is the cooperator. Then, $a_{-i}^* = 0$, and according to Proposition 5, $e_i^* = 0$, and $F_C(\bar{a}_{-i,t}) = F_C(0) = F(1/n) \geq c$. Then, consider a defector $j \in I \backslash \{i\}$. Note, that $a_{-j}^* \cdot e_{-j}^* = 0$ and $a_{-j}^* \neq 0$. As a result, according to Proposition 4, j is motivated to defect if $c \geq F_C(\bar{a}_{-j,t}) = F_C(1/n) = F(2/n)$, which leads to a contradiction $c \leq F(1/n) < F(2/n) \leq c$ since F is a monotonically increasing function on $(0, +\infty)$. Consequently, a state with $n_C^* = 1$ cooperator is not an equilibrium in the model.

Assume that $n_C^* > 1$. Let i is a cooperator. Then, (1) $a_{-i}^* e_{-i}^* \neq 0$, and (2) $\sqrt{\delta F_C(\bar{a}_{-i}^*) a_{-i}^* e_{-i}^*} - a_{-i}^* e_{-i}^* \leq \pi_i^1(a^*, e^*)$.

First, we prove statement (1). Indeed, assume that $a_{-i}^* e_{-i}^* = 0$. First, note that $a_{-i}^* \neq 0$. Let j be another cooperator. Then, $e_j^* = 0$. Second, according to Proposition 4, i is motivated to make an infinitely small but non-zero effort. As a result, $E_j(1,0) = \delta(R_0 - c) < \delta R_0 = E_j(0,0)$, which is a contradiction.

Second, we prove statement (2). Indeed, assume that $\sqrt{\delta F_C(\bar{a}_{-i}^*) a_{-i}^* e_{-i}^*} - a_{-i}^* e_{-i}^* > \pi_i^1(a^*, e^*)$. Then, $e_i^* = \pi_i^1(a^*, e^*)$. As a result, $E(1, e_i^*) = (\delta - 1)\pi_i^1(a^*, e^*) < 0 < \delta R_0 = E(0,0)$, which is a contradiction.

As a result of (1) and (2), we can apply Proposition 6 to two cooperators, i and j. Specifically, $e_i^* = \sqrt{\delta F(n_C^*/n) a_{-i}^* \cdot e_{-i}^*} - a_{-i}^* \cdot e_{-i}^*$ and $e_j^* = \sqrt{\delta F(n_C^*/n) a_{-j}^* \cdot e_{-j}^*} - a_{-j}^* \cdot e_{-j}^*$ which implies

$$a^* \cdot e^* = \sqrt{\delta F(n_C^*/n)(a^* \cdot e^* - e_i^*)} = \sqrt{\delta F(n_C^*/n)(a^* \cdot e^* - e_j^*)}$$

which, in turn, implies

$$e_i^* = e_j^* = e_C^*,$$

where

$$e_C^* = \delta\Big(1 - \frac{1}{n_C^*}\Big)\frac{F(n_C^*/n)}{n_C^*}.$$

The proposition is proved. As a corollary, there can be no more than n equilibria with $n_C^* \in \{0, 1, ..., n\} \backslash \{1\}$ cooperators, respectively. Below, we provide conditions for their existence.

First, consider the symmetric equilibrium with 0 cooperators. To obtain conditions for the existence, one should check that for each defector, to defect and exert a zero effort is the best response to strategies of others if all of them defect and exert a zero effort. Employing Proposition 5, one concludes that the above condition is equivalent to

$$F(1/n) < c. \tag{19}$$

Second, consider the symmetric equilibrium with $n_C^* > 1$ cooperators exerting the same effort e_C^*, and all defectors (if exist) exerting the same effort e_D^*. To obtain conditions for the existence of this equilibrium, one should check that (1) for each cooperator, to cooperate and exert effort e^*_C is the best response to actions of others if $n^*_C - 1$ of them cooperate and exert effort e^*_C, while $n - n^*_C$ of them defect and exert effort e^*_D; and (2) for each defector, to defect and exert a zero effort is the best response to actions of others if n^*_C of them cooperate and exert effort e^*_C, while $max\{0, n - n^*_C - 1\}$ of them defect and exert effort e^*_D.

According to Proposition 6, the first condition is equivalent to

$$cn^*_C \leq \frac{F(n^*_C/n)}{n^*_C} \tag{20}$$

(meaning a cooperator will not be interested in withdrawing from cooperation if her share of the jointly produced resource exceeds total costs of all cooperators), while the second one is equivalent to

$$F\left(\frac{n^*_C + 1}{n}\right) < \left(\sqrt{c} + \sqrt{\left(1 - \frac{1}{n^*_C}\right)F(n^*_C/n)}\right)^2 \tag{21}$$

(meaning a defector will not be interested in cooperation if the resource $F\left(\frac{n^*_C+1}{n}\right)$ she can produce together with the cooperators does not exceed a special threshold that increases with increasing cost of cooperation c and the resource $F(n^*_C/n)$ produced by the cooperators).

Appendix A.3 Additional details on the behavioral utility function (Equation 10)

Here we provide some additional details on the behavioral utility function 10. Since an individual choice on cooperation is a binary variable $a_{i,t} \in \{0, 1\}$, it is based on the difference in the utilities associated with cooperation and defection, respectively:

$$\Delta u_i(e_{t-1}, \theta_{-i,t}) = u_i(1, e_{t-1}, \theta_{-i,t}) - u_i(0, e_{t-1}, \theta_{-i,t}) =$$

$$= \beta_{1,i}\Delta\pi^1_i(e_{t-1}, \theta_{-i,t}) + \beta_{2,i}\Delta C(\theta_{-i,t}) + \beta_{3,i}\Delta SN(e_{t-1}) + \beta_{4,i}\Delta PN(e_{t-1}), \tag{22}$$

where $\theta_{-i,t} \in [0, 1]^{n-1}$ captures beliefs of player i about choices of others. Specifically, $\theta_{-i,t} = (\theta^1_{-i,t}, .., \theta^{n-1}_{-i,t})$, where each $\theta^k_{-i,t}$, $k \in I\backslash\{i\}$ is the beliefs of the i-th player about the probability that the k-th player will choose to cooperate. Let $e_{-i,t-1} = (e^1_{-i,t-1}, .., e^{n-1}_{-i,t-1})$ be the effort of other players in the group. Below we provide additional details on the components of the above utility function that are relevant to our experimental setups (where there are either $n = 2$ or $n = 4$ players in each group).

First, consider the term related to the expected material payoff, $\Delta\pi^1_i(e_{t-1}, \theta_{-i,t}) = \pi^1_i(1, e_{t-1}, \theta_{-i,t}) - \pi^1_i(0, e_{t-1}, \theta_{-i,t}) = \mathbb{E}[\Delta\pi^1_i(a_{-i,t}, e_{t-1})|\theta_{-i,t}]$. If $n = 2$,

$$\Delta\pi^1_i(e_{t-1}, \theta_{-i,t}) = -c + \left[\theta^1_{-i,t}\nu(e_{i,t-1}, e_{-i,t-1})F(1) + (1 - \theta^1_{-i,t})F(0.5)\right],$$

where

$$\forall x \in R, \forall x_{-i} \in R^m : \nu(x_i, x_{-i}) = \begin{cases} \frac{x_i}{x_i + \mathbb{1} \cdot x_{-i}}, & \text{if } x_i + \mathbb{1} \cdot x_{-i} \neq 0, \\ \frac{1}{m+1}, & \text{otherwise.} \end{cases}$$

If $n = 4$,

$$\Delta\pi^1_i(e_{t-1}, \theta_{-i,t}) = -c + \theta^1_{-i,t}\theta^2_{-i,t}\theta^3_{-i,t}\nu(e_{i,t-1}, e_{-i,t-1})F(1) + \Big(\theta^1_{-i,t}\theta^2_{-i,t}(1 - \theta^3_{-i,t})\nu(e_{i,t-1}, (e^1_{-i,t-1}, e^2_{-i,t-1})) +$$

$$+\theta^1_{-i,t}\theta^3_{-i,t}(1-\theta^2_{-i,t})\nu(e_{i,t-1},(e^1_{-i,t-1},e^3_{-i,t-1}))+\theta^2_{-i,t}\theta^3_{-i,t}(1-\theta^1_{-i,t})\nu(e_{i,t-1},(e^2_{-i,t-1},e^3_{-i,t-1}))\Bigg)F(0.75)+$$

$$\Bigg(\theta^1_{-i,t}(1-\theta^2_{-i,t})(1-\theta^3_{-i,t})\nu(e_{i,t-1},e^1_{-i,t-1})+\theta^2_{-i,t}(1-\theta^1_{-i,t})(1-\theta^3_{-i,t})\nu(e_{i,t-1},e^2_{-i,t-1})+$$

$$+\theta^3_{-i,t}(1-\theta^1_{-i,t})(1-\theta^2_{-i,t})\nu(e_{i,t-1},e^3_{-i,t-1})\Bigg)F(0.5)+(1-\theta^1_{-i,t})(1-\theta^2_{-i,t})(1-\theta^3_{-i,t})F(0.25).$$

Second, consider the term related to conformity, $\Delta C(\theta_{-i,t}) = C(1,\theta_{-i,t}) - C(0,\theta_{-i,t}) = -\mathbb{E}\big[(1-\bar{a}_{-i,t})^2 - (\bar{a}_{-i,t})^2\big|\theta_{-i,t}\big]$. If $n = 2$,

$$\Delta C(\theta_{-i,t}) = 2\theta^1_{-i,t} - 1.$$

If $n = 4$,

$$\Delta C(\theta_{-i,t}) = \theta^1_{-i,t}\theta^2_{-i,t}\theta^3_{-i,t} + \frac{1}{3}\Big[\theta^1_{-i,t}\theta^2_{-i,t}(1-\theta^3_{-i,t}) + \theta^1_{-i,t}\theta^3_{-i,t}(1-\theta^2_{-i,t}) + \theta^2_{-i,t}\theta^3_{-i,t}(1-\theta^1_{-i,t})\Big]-$$

$$-\frac{1}{3}\Big[\theta^1_{-i,t}(1-\theta^2_{-i,t})(1-\theta^3_{-i,t})+\theta^2_{-i,t}(1-\theta^1_{-i,t})(1-\theta^3_{-i,t})+\theta^3_{-i,t}(1-\theta^1_{-i,t})(1-\theta^2_{-i,t})\Big]-(1-\theta^1_{-i,t})(1-\theta^2_{-i,t})(1-\theta^3_{-i,t}).$$

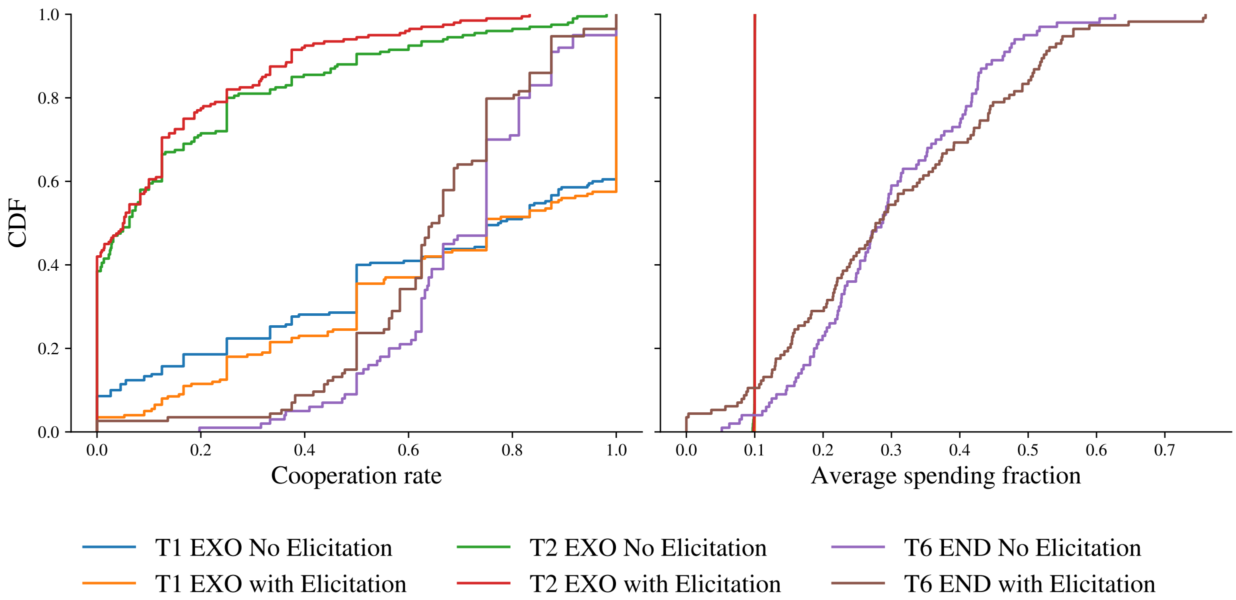
Appendix B Pilot Experiment

Table B-1: Pilot Sessions Summary

Pilot	Treatment	Matches	Elicitation in Match	Subjects	Main Analysis
P1	T6 END	6	1	16	Yes
P2	T6 END	10	1, 10	16	Yes
P3	T6 END	10	None	24	No
P4	T6 END	10	None	16	No
P5	T6 END	10	1,10	20	Yes
P6	T1 EXO	10	1,10	24	Yes
P7	T1 EXO	10	1,10	16	Yes
P8	T1 EXO	10	None	24	No
P9	T1 EXO	10	None	18	No
P10	T2 EXO	20	None	16	No
P11	T2 EXO	20	None	24	No
P12	T2 EXO	20	1,20	16	Yes
P13	T2 EXO	20	1,20	24	Yes

Notes: As part of the main data set analyzed in the paper, we only include sessions with elicitation (P1, P2, P5, P6, P7, P12, and P13).

Figure B-1: Cooperation and Spending With and Without Elicitation



Notes: The figure presents the cumulative distribution of the average cooperation rate and spending fraction for pilot sessions with and without the elicitation tasks. Each panel contains all 13 pilot sessions. Different colors indicate different treatments with or without elicitation, treating each group in each match as one observation unit. The permutation test for the average cooperation rate between sessions with and without elicitation shows a minimal difference : T1 EXO (p-value=0.29); T2 EXO (p-value=0.09); T6 END (p-value = 0.15).

Table B-2: Regressions Comparing Cooperation With and Without Elicitation

	All rounds			Round 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Elicitation	**-0.01**	**0.01**	**0.01**	**-0.02**	**0.02**	**0.02**
	(0.14)	**(0.10)**	**(0.10)**	**(0.18)**	**(0.09)**	**(0.09)**
Own R1 coop in Match 1		0.31***	0.32***		0.46***	0.45***
		(0.06)	(0.06)		(0.05)	(0.06)
Others' R1 coop in match t-1		0.02	0.02		0.16***	0.16***
		(0.02)	(0.03)		(0.03)	(0.03)
(Length of match t-1) / 100		0.16	0.16		-0.07	-0.06
		(0.14)	(0.14)		(0.10)	(0.10)
Risk aversion			-0.12			-0.20
			(0.10)			(0.14)
Loss aversion			-0.03			0.03
			(0.06)			(0.07)
Disadvantageous inequality aversion			0.18*			-0.04
			(0.11)			(0.13)
Advantageous inequality aversion			-0.13*			-0.10
			(0.07)			(0.10)
Cognitive ability			0.02			0.11
			(0.09)			(0.10)
Age under 20			-0.01			-0.03
			(0.05)			(0.05)
Female			-0.02			0.01
			(0.05)			(0.05)
Major in STEM			0.03			0.05
			(0.05)			(0.04)
High school in US			0.07			0.05
			(0.05)			(0.03)
Constant	0.48***	0.28***	0.27*	0.59***	0.24***	0.28
	(0.10)	(0.08)	(0.16)	(0.13)	(0.08)	(0.19)
Observations	23,418	22,402	22,402	3,276	3,022	3,022
Number of Subjects	254	254	254	254	254	254

Notes: The table reports results from random-effects regressions using all pilot data. The dependent variable is 1 if subjects chose "Y"(cooperation) in stage 1, and 0 otherwise. Elicitation is a dummy variable indicating whether the session has elicitation tasks for beliefs and norms. Standard errors are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B-3: Regressions Comparing Spending With and Without Elicitation

	All		Defectors Only		Cooperators Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Elicitation	**0.10**	**-0.30**	**-2.14**	**-2.27**	**2.86**	**3.03**
	(3.34)	**(3.79)**	**(2.85)**	**(3.03)**	**(3.14)**	**(3.47)**
Pay from Cooperation	0.28***	0.23***			0.20***	0.21***
	(0.07)	(0.08)			(0.06)	(0.08)
Power inequality		-0.09***		-0.04**		-0.11***
		(0.02)		(0.02)		(0.02)
My power (%) in current round		0.01		0.01		-0.05
		(0.02)		(0.02)		(0.04)
Match number		-1.30***		-0.86***		-1.30***
		(0.26)		(0.18)		(0.26)
Length of match t-1		-0.09		-0.15		0.25
		(0.25)		(0.13)		(0.25)
Risk aversion		9.64		8.40**		10.43
		(8.51)		(3.36)		(10.35)
Loss aversion		-4.88		2.42		-7.37
		(5.32)		(2.87)		(5.32)
Disadvantageous inequality aversion		16.15***		13.54***		15.32**
		(3.65)		(3.51)		(6.56)
Advantageous inequality aversion		10.11		7.78		11.29
		(6.38)		(5.94)		(8.51)
Cognitive ability		-11.70***		-3.06		-18.67***
		(2.40)		(4.97)		(3.58)
Age under 20		-2.84		-0.98		-2.92
		(2.09)		(1.77)		(2.56)
Female		0.70		0.58		-0.04
		(3.67)		(2.60)		(4.26)
Major in STEM		-2.02		-4.01		-0.90
		(3.71)		(3.64)		(3.73)
High school in US		3.64*		1.48		3.85
		(1.87)		(2.00)		(3.06)
Constant	15.22***	16.95	12.07***	5.13	18.38***	23.07
	(1.65)	(12.54)	(0.99)	(9.03)	(1.67)	(14.48)
Observations	4,980	4,612	2,013	1,924	2,967	2,688

Notes: The table reports results from random-effects regressions using pilot data in the T6 END treatments. The dependent variable is the stage-2 spending. Elicitation is a dummy variable indicating whether the session has elicitation tasks for beliefs and norms. Columns (1)-(2) show estimates based on all individuals. Columns (3)-(4) show individuals who choose to defect in the current round. Columns (5)-(6) show individuals who choose to cooperate in the current round. Power Inequality: group variance over the maximal variance a group can obtain (when $n = 2$, the maximal variance is 0.5; when $n = 4$, the maximal variance is 0.75. In both cases, the maximal variance happens when one person has 100 % power and the rest has 0%). Standard errors in parentheses are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix C Experimental Instructions (T6 END)

Experiment Overview

Today's experiment will last about 90 minutes.

You will be paid a show-up fee of $5 together with any money you accumulate during this experiment. The amount of money you accumulate will depend partly on your actions, partly on the actions of other participants, and partly on chance. This money will be paid at the end of the experiment in private and in cash.

It is important that during the experiment you remain **silent**. If you have a question or need assistance of any kind, please **raise your hand, but do not speak** - and an experiment administrator will come to you, and you may then whisper your question.

In addition, please **turn off your cell phones and put them away now**.

Anybody who breaks these rules will be asked to leave.

Agenda:

- Part 1
- Part 2
- Questionnaire

Part 1

Part 1 of the experiment contains 5 tasks.

At the end of the experiment, one of the 5 tasks will be chosen at random.

The chosen task will decide your compensation for part 1.

Tasks #1-4: Instructions

In Tasks #1-4, you will be asked to make a series of decisions. Each decision is a choice between two options -- Option A and Option B.

For example, in Task #1, the outcome of **Option A** is uncertain. The outcome of **Option B** is a sure amount.

The decisions for each task will be organized into a table (see below). Notice that for the practice task there are a total of 5 lines in the table. You should think of each line as a separate decision you need to make. At this time, please make your choice for each decision line. Note that your compensation will not depend on the practice task.

Decision	Option A		Option B	Your Choice
#1	$8 with 50% chance $0 with 50% chance		**$0.5**	**B**
#2	$8 with 50% chance $0 with 50% chance		**$2**	**B**
#3	$8 with 50% chance $0 with 50% chance		**$4**	**B**
#4	$8 with 50% chance $0 with 50% chance		**$6**	**B**
#5	$8 with 50% chance $0 with 50% chance		**$8**	**B**

Next, we will explain the compensation procedure for Tasks #1-4.

Tasks #1-4: Compensations

In each of the Tasks #1-4, you will make a series of decisions. At the end of the experiment, if one of these tasks is chosen for compensation, then the computer will randomly select one of the decisions to be carried out. As an example, we will demonstrate how compensation works with the practice tasks that you just did.

The procedure involves two steps:

- First, one of the decisions from the task will be chosen at random. For example, in the practice task you had 5 decisions, so one will be chosen randomly by the computer (with equal probability).
- Second, depending on the chosen decision in step one, a random draw might be necessary. For example, if you chose Option A over Option B, then a random draw will be made to determine the outcome of Option A. For example, in the practice task, Option A has 50% to pay $8 and 50% to pay $0.

To demonstrate the compensation procedure, we will perform the two steps using the practice task.

Decision	Option A	Option B	Your Choice
#1	**$8** with 50% chance; **$0** with 50% chance;	**$0.5**	B
#2	**$8** with 50% chance; **$0** with 50% chance;	**$2**	**B**
#3	**$8** with 50% chance; **$0** with 50% chance;	**$4**	B
#4	**$8** with 50% chance; **$0** with 50% chance;	**$6**	B
#5	**$8** with 50% chance; **$0** with 50% chance;	**$8**	B

1. The randomly chosen decision is Decision #2

2. For the selected decision, you have chosen Option B.

-> Therefore, your payoff is $2.

To gain further understanding of the compensation procedure, you may try again. You may also raise your hand and and an experiment administrator will come up to you. At that time you can quietly ask a question.

Try Again

In the actual experiment, this procedure will be carried out **only once**. If you are ready to begin, please click **Next**.

Task #1

Please make a choice for **each of the 16 decisions** in this task.

Decision	Option A		Option B	Your Choice
#1	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$0.5**	__
#2	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$1**	__
#3	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$1.5**	__
#4	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$2**	__
#5	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$2.5**	__
#6	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$3**	__
#7	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$3.5**	__
#8	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$4**	__
#9	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$4.5**	__
#10	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$5**	__
#11	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$5.5**	__
#12	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$6**	__
#13	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$6.5**	__
#14	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$7**	__
#15	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$7.5**	__
#16	**$8** with 50% chance; **$0** with 50% chance	○ ○	**$8**	__

Task # 2

Please make a choice for **each of the 16 decisions** in this task. Notice that the amount in red will be subtracted from your earnings in case that outcome happens.

Decision	Option A		Option B	Your Choice
#1	-$0.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#2	-$1 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#3	-$1.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#4	-$2 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#5	-$2.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#6	-$3 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#7	-$3.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#8	-$4 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#9	-$4.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#10	-$5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#11	-$5.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#12	-$6 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#13	-$6.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#14	-$7 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#15	-$7.5 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__
#16	-$8 with 50% chance; **$4.00** with 50% chance	○ ○	**$0.00**	__

Task #3-4 Descriptions

For Task #3-4, you are matched with another, randomly selected participant in this room. Your decision is to determine the payoff for you and the participant you are matched with. Your decision and the decisions of all other participants are independent and made anonymously. **You will never learn the identity of the other participant you are paired with.**

You will make 9 decisions in a decision tables similar to the previous tasks. Each decision is a choice between two options -- Option A and Option B. Each option specifies a payment for you and the participant you are matched with.

For example, in Task #3, **option A** specifies a sure and equal payment for you and the participant you are matched with. **Option B** varies your payment and specifies a sure payment for the participant you are matched with.

Decision	Option A		Option B	Your Choice
#1	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3**; the participant you are matched with receives $2.5.	__
#2	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4**; the participant you are matched with receives $2.5.	__
#3	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$5**; the participant you are matched with receives $2.5.	__

If this task is selected for your payment at the end of the experiment, one of the decisions from the task will be chosen at random. You and the participant you are matched with will receive the money you selected in the chosen decision row.

Notice that if the participant you are matched with has this task selected for the payment, you will receive the money determined by them.

Task #3

Please make a choice for **each of the 9 decisions** in this task.

Decision	Option A		Option B	Your Choice
#1	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.0**; the participant you are matched with receives $2.5.	__
#2	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.25**; the participant you are matched with receives $2.5.	__
#3	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.5**; the participant you are matched with receives $2.5.	__
#4	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.75**; the participant you are matched with receives $2.5.	__
#5	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.0**; the participant you are matched with receives $2.5.	__
#6	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.25**; the participant you are matched with receives $2.5.	__
#7	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.5**; the participant you are matched with receives $2.5.	__
#8	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.75**; the participant you are matched with receives $2.5.	__
#9	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$5.0**; the participant you are matched with receives $2.5.	__

Task #4

Please make a choice for **each of the 9 decisions** in this task.

Decision	Option A		Option B	Your Choice
#1	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.0**; the participant you are matched with receives $5.5.	__
#2	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.25**; the participant you are matched with receives $5.5.	__
#3	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.5**; the participant you are matched with receives $5.5.	__
#4	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$3.75**; the participant you are matched with receives $5.5.	__
#5	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.0**; the participant you are matched with receives $5.5.	__
#6	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.25**; the participant you are matched with receives $5.5.	__
#7	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.5**; the participant you are matched with receives $5.5.	__
#8	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$4.75**; the participant you are matched with receives $5.5.	__
#9	You earn **$4.0**; the participant you are matched with receives $4.0.	○ ○	You earn **$5.0**; the participant you are matched with receives $5.5.	__

Task #5

This task is made up of 11 questions. You will have 7 minutes to complete the task.

The top right-hand corner of the screen will display the time remaining.

If this task is selected at the end of the experiment, you will get a flat payment of $4.

The answers you give in the task will not affect part 2 of the experiment in any way.

Begin Test

Part 2

Part 2 of the experiment is made up of **10 matches**.

At the start of each match you will be randomly paired with another participant in this room.

You will then play a number of rounds with that participant (this is what we call a "match").

How Matches Work

Each match will last for a random number of **rounds**:

- At the end of each round the computer will roll a ten-sided fair dice.
- If the computer rolls a number less than 10, then the **match continues** for at least one more round .
- If the computer rolls a 10 , then the **match ends** after the current round.

To test this procedure, click the 'Test' button below. You will need to test this procedure 5 times.

Round

Dice Roll

Remember that at the end of each round the computer rolls a ten-sided fair dice. The match ends when the computer rolls a 10.

Test #1

Round Overview

Each round has two decision stages.

Stage 1

- You and the other participant need to choose either action X or action Y.
  - If you choose action X , you will earn 60 points.
  - If you choose action Y , you will earn 40 points plus a **proportion** of either 0 , 100 , 108 , or 108 points depending on how many participants choose Y, how many shares you own, and how many shares other participants who choose Y own.
    - In total, there are 100 shares.
    - The number of shares that you and other participants own in each round will be known before the decisions in Stage 1.
    - The number of shares will be revised at the end of each round based on decisions in Stage 2.
    - We will provide more details in the next page.

Stage 2

- You and the other participant can purchase shares by spending points earned in Stage 1.
- The number of shares you get in the next round will be equal to the percentage of total points spent in Stage 2 that were spent by you.

Your payoff of each round = points you earn in Stage 1 - points you spend in Stage 2

At the end of the experiment, your total payoff (accumulated across all rounds and matches) will be converted into cash at the exchange rate of 500 points = $1.

Next, we will provide more details about each stage, including examples.

Stage 1 Details

ID	Example Round				
	Current Shares	Choice	Earn	Spend	Payoff
You	**25**	?			
2	25				
3	25				
4	25				

Dice Roll

At the beginning of each round, you will see shares of all participants in a match presented in a table like the one above.

In Stage 1, you and the other participant will choose either action X or action Y. Currently, your choice is marked with a '?' denoting that it has not yet been made.

If you choose action X, you will earn **60 points** regardless of what everyone else chooses and regardless of how many shares you and the other participant own.

If you choose action Y, you will earn **40 points** + amount * **your proportion**, where

- the amount is either 0, 100, 108, or 108 points depending on whether 1, 2, 3, or 4 participants choose Y,
- **your proportion of shares** is your proportion of shares among those who chose Y.

Throughout the experiment, you will be provided with a calculator to check different scenarios.

ID	Example Round				
	Current Shares	Choice	Earn	Spend	Payoff
You	**40**	**Y**	**83**		
2	30	Y	72		
3	20	Y	61		
4	10	Y	50		

Dice Roll

To see an example, click the 'Example' button below. You will need to see 5 examples.

Example #1

Suppose your own 40 shares, and participants 2, 3, 4 own 30, 20, 10 shares, respectively.

Suppose that all participants choose Y.

You will earn **83 points** $= 40 + \frac{40}{40+30+20+10} \times 108$.

To elaborate, because four participants choose Y (you and all other participants) , the amount to be divided is 108 points.

Your proportion of that amount is **0.4** because you own 40 shares and all participants choose Y.

Recap

*If you choose action X , you will earn **60 points** regardless of what everyone else chooses and regardless of how many shares you and other participants own.*

*If you choose action Y , you will earn **40 points** + amount * **your proportion** , where*

- *the amount is either 0, 100, 108, or 108 points depending on whether 1, 2, 3, or 4 participants choose Y,*
- ***your proportion of shares** is your proportion of shares among those who chose Y.*

Stage 2 Details

ID	Example Round				
	Current Shares	Choice	Earn	Spend	Payoff
You	**25**	**Y**	**90**	**?**	
2	25	X	60		
3	25	X	60		
4	25	Y	90		

Dice Roll

Once all participants in the match have chosen their actions in Stage 1, the summary table will be updated to reflect that choice and the experiment will proceed to Stage 2.

In Stage 2, everyone will choose how many points earned in Stage 1 to spend on shares for the next Round. Currently, your choice is marked with a '?' denoting that it has not yet been made.

The number of shares you get in the next round will be determined by what percentage of total points spent in Stage 2 was spent by you. (If no one spent any point, each of you will get 25 shares in the next round.)

Again, you will be provided with a calculator to check different scenarios.

Stage 2 Examples

ID	Example Round					Next Round				
	Current Shares	Choice	Earn	Spend	Payoff	New Shares	Choice	Earn	Spend	Payoff
You	**25**	**Y**	**90**	**76**	**14**	67				
2	25	X	60	10	50	9				
3	25	X	60	19	41	17				
4	25	Y	90	9	81	8				

Dice Roll

To see an example, click the 'Example' button below. You will need to see 5 examples.

Example #1

Suppose you spend 76 points in stage 2, and participants 2, 3, 4 spend 10, 19, 9 points, respectively.

Your payoff in this round is 14 points.

If the match continues to a new round,the number of shares you will own at the beginning of next round will be $\frac{76}{76+10+19+9} \times 100 = 67$

.

Recap

- ***Your payoff of each round*** *= points you earn in Stage 1 - points you spend in Stage 2*
- *The number of shares you get in the next round will be determined by what percentage of total points spent in Stage 2 was spent by you. (If no one spent any point, each of you will get 25 shares in the next round.)*
- *At the end of each round the computer will roll a ten-sided fair dice. If the computer rolls a number less than 10, then the match continues for at least one more round*

How to use the Calculator

ID	Example Round Calculator					Calculator Hide Reset					
	Current Shares	Choice	Earn	Spend	Payoff	Current Shares	Stage 1 Choice	Stage 1 Earn	Stage 2 Spend	Round Payoff	New Shares import
You	**25**	?				**25**	X **Y**	**90**	2	**88**	**12**
2	25					25	**X** Y	60	4	56	24
3	25					25	X **Y**	90	5	85	29
4	25					25	**X** Y	60	6	54	35

Dice Roll

Click Calculator. This will open the calculator.

Specify the choices in the "Stage 1 Choice" column. Once all of the choices are filled in you will see how much you earn in Stage 1.

Specify how much is spent in Stage 2 by typing the spends for each participant. Once all choices and spendings are specified, the rest of the table will be automatically filled. You will see your payoff for this round and shares in the next round.

Click import. This will import the "New Shares" column into the "Current Shares" column.

Click Reset This will reset the table.

Click Hide. This will hide the table.

How History Will be Recorded

ID	Round 2					Round 3					Round 4				
	Shares	Choice	Earn	Spend	Payoff	Shares	Choice	Earn	Spend	Payoff	Current Shares	Choice	Earn	Spend	Payoff
You	30	Y	81	8	73	28	X	60	6	54	**20**	?			
2	20	X	60	6	54	21	Y	83	8	75	27				
3	20	Y	67	7	60	24	X	60	6	54	20				
4	30	Y	81	8	73	28	Y	97	10	87	33				

The history of all variables will be recorded as presented in the example table above.

You will be able to see the full history of your current match by scrolling to the left of the history.

The history table will be cleared at the beginning of each new match.

Additional Questions

In addition to the decisions at stage 1 and stage 2, you will be asked three questions in each round of the first and last match:

- **Q1** What do you think the chances are that the other participant will choose X or Y?
- **Q2** How appropriate do YOU think your choices in this round are?
- **Q3** How socially appropriate will MOST PEOPLE agree your actions are?

Question 1 Instructions

Question 1 and the response table are presented below. During the experiment, you will need to enter the percent for the X column only. The Y column will be updated automatically.

Question 1: What do you think the chances are that the other participant will choose X or Y?

ID	X	Y
Participant 2:	%	%
Participant 3:	%	%
Participant 4:	%	%

You can earn up to 10 points based on the accuracy of your answer. <u>You can secure the largest chance of earning points by reporting your most-accurate guess.</u>

You don't need to understand the details of how the payment works for this question. If you are interested, please click the "More Details". If you are not interested, you can stop reading this page and click 'Next' now.

More Details

Your chance of receiving 10 points is determined by the following formulas:

- $[1-[\frac{100-\text{your guess}}{100})^2]\times 100$ *if the other participant chooses X*
- $[1-(\frac{\text{your guess}}{100})^2]\times 100$ *if the other participant chooses Y*

To illustrate, suppose you guess that the chance of the participant 2 chooses X is 70.

- *If s/he chooses X, your chance of receiving 10 points is* $[1-(\frac{100-70}{100})^2]\times 100 = 91$
- *If s/he chooses X, your chance of receiving 10 points is* $[1-(\frac{70}{100})^2]\times 100 = 51$

To determine whether you receive 10 points, the computer will randomly draw a number between 0 and 100. Each number between 0 and 100 is equally likely to be picked. If the number drawn by the computer is less than or equal to your chance of receiving 10 points as determined by the formulas above, then you will receive the points. Otherwise, you will receive 0 point.

Question 2 Instructions

Question 2 and the response table are presented below. During the experiment, you will need to enter your evaluation for your action X and your action Y.

Question 2: How appropriate do YOU think your actions in this round are?

	Inappropriate	Somewhat Inappropriate	Somewhat Appropriate	Appropriate
X	○	○	○	○
Y	○	○	○	○

We ask you to truthfully report your evaluation. You will get 10 points for any answer you provide.

By appropriate, we mean behavior that you think is the "correct" or "ethical" thing to do. You can think both actions are appropriate or inappropriate.

Question 3 Instructions

Question 3 and the response table are presented below. During the experiment, you will need to evaluate your action X and Y.

Question 3: How socially appropriate will MOST PEOPLE agree your actions are?

	Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
X	○	○	○	○
Y	○	○	○	○

We ask you to truthfully report your evaluation. You will get 10 points for any answer you provide.

By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do. Another way to think about what we mean is that if a person were to choose a socially inappropriate action, then someone else might be angry at the person for doing so. Show me an example

Consider the following situation: Individual A is at a local coffee shop near campus. While there, individual A notices that someone has left a wallet at one of the tables. We list the 4 possible actions individual A can choose in the table below. You will rate each action on social appropriateness by clicking on the radial for that action.

Individual A's Actions	*Very socially inappropriate*	*Somewhat socially inappropriate*	*Somewhat socially appropriate*	*Very socially appropriate*
Take the wallet	○	○	○	○
Ask others nearby if the wallet belongs to them	○	○	○	○
Leave the wallet where it is	○	○	○	○
Give the wallet to the shop manager	○	○	○	○

Suppose you believe that most people think that taking the wallet is "very socially inappropriate", asking others nearby if the wallet belongs to them is "somewhat socially appropriate", leaving the wallet is "somewhat socially inappropriate" and giving the wallet to the shop manager is "very socially appropriate". Then you would have rated the actions in the following way: Click me to show how

In the experiment, we ask you to evaluate how socially appropriate you actions are. That is, suppose we present the history and the current shares in your match to everyone in the room, what will be the most common response regarding how appropriate your action X and Y are?

Match #1

ID	Round 2					Round 3					Round 4 Calculator					Calculator Hide Reset					
	Shares	Choice	Earn	Spend	Payoff	Shares	Choice	Earn	Spend	Payoff	Current Shares	Choice	Earn	Spend	Payoff	Current Shares	Stage 1 Choice	Stage 1 Earn	Stage 2 Spend	Round Payoff	New Shares import
You	13	Y	56	22	34	33	Y	76	18	58	24	?				24	X Y	70			
2	36	Y	85	12	73	18	Y	59	22	37	29					29	X Y	77			
3	37	Y	86	21	65	31	Y	73	12	61	16					16	X Y	60			
4	14	X	60	12	48	18	Y	59	24	35	32					32	X Y	81			
Dice Roll					3					7											

Stage 1 : Please select your choice for
Round **4** of Match #1

X

Y

Match #1

ID	Round 2					Round 3					Round 4 Calculator				
	Shares	Choice	Earn	Spend	Payoff	Shares	Choice	Earn	Spend	Payoff	Current Shares	Choice	Earn	Spend	Payoff
You	13	Y	56	22	34	33	Y	76	18	58	24	Y			
2	36	Y	85	12	73	18	Y	59	22	37	29				
3	37	Y	86	21	65	31	Y	73	12	61	16				
4	14	X	60	12	48	18	Y	59	24	35	32				
Dice Roll					3					7					

Question 1: What do you think the chances are that the other participant will choose X or Y?

ID	X	Y
Participant 2:	%	%
Participant 3:	%	%
Participant 4:	%	%

Match #1

ID	Round 2					Round 3					Round 4 Calculator				
	Shares	Choice	Earn	Spend	Payoff	Shares	Choice	Earn	Spend	Payoff	Current Shares	Choice	Earn	Spend	Payoff
You	13	Y	56	22	34	33	Y	76	18	58	24	Y			
2	36	Y	85	12	73	18	Y	59	22	37	29				
3	37	Y	86	21	65	31	Y	73	12	61	16				
4	14	X	60	12	48	18	Y	59	24	35	32				
Dice Roll					3					7					

Question 2: How appropriate do YOU think your actions in this round are?

	Inappropriate	Somewhat Inappropriate	Somewhat Appropriate	Appropriate
X				
Y				

Question 3: How socially appropriate will MOST PEOPLE agree your actions are?

	Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate
X				
Y				

Match #1

ID	Round 2					Round 3					Round 4 Calculator				
	Shares	Choice	Earn	Spend	Payoff	Shares	Choice	Earn	Spend	Payoff	Current Shares	Choice	Earn	Spend	Payoff
You	13	Y	56	22	34	33	Y	76	18	58	24	Y	66	?	
2	36	Y	85	12	73	18	Y	59	22	37	29	Y	71		
3	37	Y	86	21	65	31	Y	73	12	61	16	Y	57		
4	14	X	60	12	48	18	Y	59	24	35	32	Y	74		
Dice Roll					3					7					

Calculator Hide Reset					
Current Shares	Stage 1 Choice	Stage 1 Earn	Stage 2 Spend	Round Payoff	New Shares import
24	X Y	66	20	46	23
29	X Y	71	20	51	23
16	X Y	57	22	35	25
32	X Y	74	25	49	29

In Stage 1 of this round, you earned 66 points.

Please decide how many points do you want to spend in Stage 2 ?

Appendix D Additional Tables and Figures

Table D-4: Supergame Lengths

Supergame Number:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Sequence #1:	4	2	3	2	9	11	3	4	2	19	11	35	2	5	8	5	2	10	27	6
Sequence #2:	9	6	21	5	3	7	12	5	5	12	18	7	7	17	5	2	48	4	2	9
Sequence #3:	16	4	9	11	18	21	2	3	14	2	6	22	13	9	3	9	7	6	18	17
Sequence #4:	2	7	36	41	2	5	7	12	3	18	2	21	2	1	1	5	6	2	9	1

Table D-5: Summary of Experiment Administration

Treatment				Administration				Demographics		
Name	b	n	a_0	Sessions	Subjects	Sequence	Earnings	% Male	% STEM	% US HS
T1 END	109	2	0.812	4	40	#1,#2,#3,#4	20.5	35.0	60.0	77.5
							(1.0)	(7.6)	(7.8)	(6.7)
T1 EXO	109	2	0.812	3	52	#1*, #1*, #3	20.8	51.9	63.5	75.0
							(0.6)	(7.0)	(6.7)	(6.1)
T2 EXO	109	4	0.812	2	40	#1*, #1*	18.7	50.0	65.0	75.0
							(0.7)	(8.0)	(7.6)	(6.9)
T3 END	218	2	0.812	4	40	#1,#2, #3,#4	24.3	57.5	67.5	75.0
							(1.1)	(7.9)	(7.5)	(6.9)
T4 END	218	4	0.812	2	44	#1, #2	22.3	47.7	59.1	70.5
							(1.0)	(7.6)	(7.5)	(7.0)
T4 EXO	218	4	0.812	2	40	#1, #4	26.0	52.5	62.5	85.0
							(0.5)	(8.0)	(7.8)	(5.7)
T5 END	109	2	0.406	4	40	#1, #2, #3, #4	25.7	55.0	65.0	77.5
							(0.9)	(8.0)	(7.6)	(6.7)
T6 END	109	4	0.406	3	52	#1*, #1*,#1*	19.6	69.2	57.7	67.3
							(0.6)	(6.5)	(6.9)	(6.6)
T6 EXO	109	4	0.406	2	40	#1, #3	22.8	47.5	67.5	87.5
							(0.7)	(8.0)	(7.5)	(5.3)
Total				26	388		22.2	52.3	62.9	76.3
							(0.3)	(2.5)	(2.5)	(2.2)

Notes: Standard errors are in parentheses. % STEM denotes proportion of participants that are in STEM majors. % US HS denotes the proportion of participants that completed high-school in the US. #1* indicates a session from the pilot *Aside:* Only one sequence was used in the pilot. We ran an additional sessions for T1 EXO because the number of subjects from the pilot was below 40.

Table D-6: Average Cooperation Rate across Treatments

	All Rounds			Round 1		
	all Match	Match 1-5	Match 6-10	all Match	Match 1-5	Match 6-10
T1 END	57.4	55.5	59.5	67.6	63.0	72.7
	(2.26)	(2.92)	(3.51)	(2.48)	(3.38)	(3.61)
T1 EXO	75.9	66.5	79.2	79.2	69.6	80.4
	(1.72)	(2.97)	(2.54)	(1.68)	(2.97)	(2.47)
T2 EXO	12.8	11.8	16.6	12.5	11.5	15.5
	(1.28)	(2.28)	(2.98)	(1.25)	(2.39)	(2.56)
T3 END	69.7	66.7	73.2	76.3	72.0	81.2
	(1.94)	(2.59)	(2.87)	(2.18)	(3.2)	(2.84)
T4 END	24.9	28.0	21.8	40.2	44.5	35.9
	(2.2)	(2.82)	(3.36)	(2.84)	(3.65)	(4.29)
T4 EXO	89.4	78.8	87.4	85.1	72.0	80.0
	(1.26)	(3.4)	(2.55)	(1.32)	(3.0)	(2.58)
T5 END	89.3	87.4	91.2	96.2	95.5	97.0
	(0.92)	(1.35)	(1.24)	(0.93)	(1.44)	(1.19)
T6 END	64.3	70.6	56.0	85.1	85.0	85.2
	(1.8)	(1.96)	(2.9)	(1.94)	(2.11)	(3.57)
T6 EXO	94.7	91.3	93.8	94.5	89.5	93.5
	(0.69)	(1.71)	(1.34)	(0.92)	(2.68)	(1.86)

Notes: "all Match" means match 1-20 for EXO treatments and match 1-10 for END treatments. Standard errors (in parentheses) are calculated by taking one group (with either 2 or 4 subjects) in one match as a unit of observation.

Table D-7: Cooperation in Stage 1 (Full Table 3)

	All rounds			Round 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Larger n ($n = 4$)	-0.47***	-0.45***	-0.46***	-0.38***	-0.30***	-0.31***
	(0.07)	(0.06)	(0.06)	(0.08)	(0.07)	(0.06)
Greater b ($b = 218$)	0.45***	0.42***	0.43***	0.36***	0.26***	0.27***
	(0.11)	(0.10)	(0.09)	(0.11)	(0.08)	(0.08)
Greater $a0$ ($a0 = 0.812$)	-0.65***	-0.57***	-0.58***	-0.61***	-0.40***	-0.40***
	(0.09)	(0.09)	(0.08)	(0.11)	(0.08)	(0.08)
Choose effort (END)	-0.44***	-0.44***	-0.44***	-0.22***	-0.20***	-0.20***
	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.06)
Own R1 coop in Match 1		0.13***	0.12***		0.28***	0.27***
		(0.03)	(0.03)		(0.03)	(0.03)
Others' R1 coop in Match $t-1$		0.04**	0.04**		0.14***	0.14***
		(0.01)	(0.01)		(0.02)	(0.02)
(Length of match $t-1$) / 100		0.10	0.10		0.07	0.07
		(0.06)	(0.06)		(0.07)	(0.07)
Risk aversion			-0.05			-0.09
			(0.08)			(0.09)
Loss aversion			-0.07*			-0.06
			(0.04)			(0.05)
Disadvantageous inequality aversion			-0.06			-0.14**
			(0.06)			(0.07)
Advantageous inequality aversion			-0.02			-0.01
			(0.05)			(0.06)
Cognitive ability			-0.01			-0.03
			(0.06)			(0.06)
Age under 20			-0.05*			-0.07*
			(0.03)			(0.04)
Female			0.04			0.00
			(0.03)			(0.03)
Major in STEM			-0.01			0.03
			(0.04)			(0.04)
High school in US			0.02			0.01
			(0.03)			(0.03)
Constant	1.41***	1.25***	1.38***	1.32***	0.90***	1.09***
	(0.08)	(0.09)	(0.13)	(0.09)	(0.09)	(0.14)
Observations	42,392	39,912	39,912	5,088	4,700	4,700
Number of Subjects	388	388	388	388	388	388

Notes: The table reports results from random-effects regressions using data across all nine treatments. The dependent variable is 1 if subjects chose "Y"(cooperation) in stage 1 and 0 otherwise. Standard errors are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D-8: Spending in Stage 2 (Full Table 4)

	All		Defectors Only		Cooperators Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Pay from Cooperation	0.18***	0.18***			0.16***	0.18***
	(0.04)	(0.04)			(0.03)	(0.03)
Power Inequality		-0.06***		-0.03***		-0.09***
		(0.01)		(0.01)		(0.02)
My power (%) in current round		-0.05*		0.03*		-0.10***
		(0.03)		(0.01)		(0.03)
Match number		-1.53***		-1.25***		-1.58***
		(0.36)		(0.31)		(0.50)
Length of match t-1		-0.13*		-0.09		-0.13**
		(0.08)		(0.08)		(0.07)
Risk aversion		-8.25		-2.15		-7.41
		(6.10)		(6.70)		(5.99)
Loss aversion		1.24		0.23		4.12
		(4.65)		(3.45)		(4.72)
Disadvantageous inequality aversion		3.10		1.75		4.00
		(5.58)		(4.17)		(6.16)
Advantageous inequality aversion		4.35		-0.38		5.09
		(5.18)		(4.74)		(6.05)
Cognitive ability		-9.68		-2.26		-8.96
		(6.48)		(4.39)		(7.90)
Age under 20		-0.86		1.39		-0.56
		(2.06)		(2.01)		(2.22)
Female		2.49		0.01		2.71
		(2.39)		(2.17)		(2.76)
Major in STEM		0.65		1.94		-1.14
		(2.94)		(2.38)		(3.03)
High school in US		2.93		2.83		2.75
		(2.57)		(1.92)		(2.92)
Constant	0.30***	0.55***	0.24***	0.32**	0.36***	0.59***
	(0.04)	(0.19)	(0.04)	(0.14)	(0.03)	(0.23)
Observations	16,104	14,664	6,590	6,005	9,514	8,659
Number of Subjects	216	216	212	207	212	210

Notes: The table reports results from random-effects regressions using data from the five END treatments. The dependent variable is the stage-2 spending. Columns (1)-(2) show estimates based on all individuals. Columns (3)-(4) show individuals who choose to defect in the current round. Columns (5)-(6) show individuals who choose to cooperate in the current round. Power Inequality: group variance over the maximal variance a group can obtain (when $n = 2$, the maximal variance is 0.5; when $n = 4$, the maximal variance is 0.75. In both cases, the maximal variance happens when one person has 100 % power and the rest has 0%). Standard errors in parentheses are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D-9: Spending in Stage 2 in All Rounds All Matches

	All		Defectors Only		Cooperators Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Pay from Cooperation	0.17***	0.20***			0.15***	0.21***
	(0.04)	(0.04)			(0.03)	(0.03)
Power Inequality	-0.07***		-0.03***		-0.12***	
	(0.01)		(0.01)		(0.02)	
My power (%) in current round		-0.06*		0.03*		-0.15***
		(0.03)		(0.01)		(0.03)
Match number	-1.53***	-1.57***	-1.26***	-1.27***	-1.58***	-1.60***
	(0.36)	(0.37)	(0.32)	(0.33)	(0.50)	(0.50)
Length of match t-1	-0.13*	-0.14*	-0.09	-0.10	-0.13**	-0.14**
	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.06)
Constant	0.49***	0.51***	0.42***	0.38***	0.52***	0.61***
	(0.11)	(0.12)	(0.07)	(0.07)	(0.15)	(0.15)
Observations	14,664	14,664	6,005	6,005	8,659	8,659
Number of Subjects	216	216	207	207	210	210
Preferences	Yes	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports results from random-effects regressions using data across the five END treatments. The dependent variable is the stage-2 spending. Variables of interest is the stage-1 earning, separate by pay from cooperation and the pay of defection (60 tokens). The first two columns show all individual data. Columns (3)-(4) show individuals who choose to defect in the current round. Columns (5)-(6) show individuals who choose to cooperate in the current round. Note defectors will get 60 tokens as stage 1 earning, thus the constant is omitted in columns (3) and (4). Power Inequality: group variance over the maximal variance a group can obtain (when $n = 2$, the maximal variance is 0.5; when $n = 4$, the maximal variance is 0.75. In both cases, the maximal variance happens when one person has 100 % power and the rest has 0%). Preferences include risk aversion, loss aversion, other-regarding preference in disadvantageous and advantageous inequality, cognitive ability. Demographics include age, gender, major, and subjects high school location. Standard errors in parentheses are clustered at the session level. Columns (4)-(6) shows the regression results only on round 1 data. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D-10: Average Belief about Others' Cooperation (Y)

	All Rounds			Round 1		
	Match 1	Match 10	Match 20	Match 1	Match 10	Match 20
T1 END	43.2	55.3	-	50.6	32.2	-
	(3.46)	(4.76)		(3.79)	(4.36)	
T1 EXO	52.4	72.7	80.8	48.0	27.4	10.4
	(3.11)	(3.98)	(7.13)	(2.63)	(3.18)	(3.65)
T2 EXO	28.9	-	12.1	63.0	-	87.1
	(3.02)		(3.02)	(2.93)		(3.09)
T3 END	51.4	69.2	-	45.1	25.8	-
	(3.42)	(3.94)		(3.02)	(4.51)	
T4 END	42.7	14.7	-	47.0	71.5	-
	(2.54)	(2.5)		(4.02)	(4.7)	
T4 EXO	52.4	89.3	89.4	48.7	17.8	15.4
	(3.59)	(2.54)	(2.63)	(4.13)	(3.07)	(3.07)
T5 END	75.0	89.7	-	26.3	9.1	-
	(3.0)	(2.42)		(3.9)	(2.78)	
T6 END	61.3	56.1	-	37.7	12.1	-
	(2.33)	(2.97)		(2.87)	(2.99)	
T6 EXO	76.3	93.3	94.9	39.4	10.0	5.4
	(2.96)	(2.08)	(1.87)	(3.78)	(3.15)	(1.96)

Notes: The two sessions T2 EXO were conducted in the pilot phase. Later, we decided to collect the data in match 1, 10, 20 for EXO treatments. Standard errors (in parentheses) are calculated by taking each individual as a unit of observation.

Table D-11: Average Personal Norm

	All Rounds			Round 1		
	Match 1	Match 10	Match 20	Match 1	Match 10	Match 20
			About Defection (X)			
T1 END	3.2	2.9	-	3.1	2.7	-
	(0.11)	(0.15)		(0.16)	(0.2)	
T1 EXO	2.9	2.5	2.2	3.2	2.5	1.9
	(0.11)	(0.14)	(0.33)	(0.1)	(0.14)	(0.31)
T2 EXO	3.4	-	3.6	3.3	-	3.6
	(0.08)		(0.11)	(0.12)		(0.11)
T3 END	2.9	2.6	-	3.0	2.5	-
	(0.12)	(0.17)		(0.16)	(0.21)	
T4 END	3.2	3.5	-	3.2	3.2	-
	(0.12)	(0.11)		(0.14)	(0.18)	
T4 EXO	2.9	1.9	1.9	3.0	2.3	2.0
	(0.17)	(0.18)	(0.19)	(0.18)	(0.18)	(0.19)
T5 END	2.9	2.8	-	2.8	2.7	-
	(0.13)	(0.15)		(0.18)	(0.19)	
T6 END	3.1	3.2	-	3.1	2.6	-
	(0.08)	(0.12)		(0.12)	(0.19)	
T6 EXO	2.7	2.6	2.5	2.9	2.7	2.6
	(0.16)	(0.18)	(0.18)	(0.15)	(0.19)	(0.21)
			About Cooperation (Y)			
T1 END	3.3	3.2	-	3.3	3.5	-
	(0.09)	(0.11)		(0.13)	(0.1)	
T1 EXO	3.3	3.5	3.6	3.4	3.6	3.8
	(0.09)	(0.09)	(0.19)	(0.1)	(0.1)	(0.11)
T2 EXO	2.8	-	2.7	3.0	-	2.8
	(0.15)		(0.19)	(0.17)		(0.2)
T3 END	3.1	3.3	-	3.2	3.6	-
	(0.1)	(0.1)		(0.14)	(0.09)	
T4 END	3.1	2.7	-	3.5	3.2	-
	(0.1)	(0.16)		(0.11)	(0.17)	
T4 EXO	3.5	3.8	4.0	3.5	3.8	3.9
	(0.12)	(0.07)	(0.03)	(0.12)	(0.06)	(0.04)
T5 END	3.4	3.5	-	3.6	3.7	-
	(0.1)	(0.09)		(0.11)	(0.08)	
T6 END	3.3	3.1	-	3.4	3.7	-
	(0.08)	(0.12)		(0.1)	(0.09)	
T6 EXO	3.7	3.7	3.7	3.7	3.8	3.8
	(0.07)	(0.08)	(0.08)	(0.09)	(0.07)	(0.07)

Notes: Responses were coded as 1 = "Inappropriate", 2 = "Somewhat inappropriate", 3="Somewhat appropriate", 4="Appropriate". The two sessions T2 EXO were conducted in the pilot phase. Later, we decided to collect the data in match 1, 10, 20 for EXO treatments. Standard errors (in parentheses) are calculated by taking each individual as a unit of observation.

Table D-12: Average Social Norm

	All Rounds			Round 1		
	Match 1	Match 10	Match 20	Match 1	Match 10	Match 20
			About Defection (X)			
T1 END	2.9	2.9	-	2.9	2.8	-
	(0.11)	(0.14)		(0.14)	(0.17)	
T1 EXO	2.8	2.5	2.2	3.1	2.5	2.2
	(0.11)	(0.13)	(0.3)	(0.1)	(0.13)	(0.27)
T2 EXO	3.2	-	3.6	3.3	-	3.6
	(0.1)		(0.11)	(0.11)		(0.12)
T3 END	3.0	2.6	-	3.0	2.5	-
	(0.1)	(0.15)		(0.12)	(0.19)	
T4 END	3.0	3.3	-	3.0	2.9	-
	(0.1)	(0.11)		(0.12)	(0.16)	
T4 EXO	2.8	1.9	1.9	2.8	2.4	2.0
	(0.14)	(0.17)	(0.18)	(0.14)	(0.17)	(0.18)
T5 END	3.1	3.1	-	3.2	3.1	-
	(0.11)	(0.13)		(0.13)	(0.16)	
T6 END	3.0	3.2	-	3.0	2.9	-
	(0.07)	(0.1)		(0.11)	(0.16)	
T6 EXO	2.8	2.8	2.7	3.0	2.8	2.9
	(0.14)	(0.18)	(0.18)	(0.14)	(0.19)	(0.19)
			About Cooperation (Y)			
T1 END	3.3	3.2	-	3.2	3.3	-
	(0.09)	(0.11)		(0.12)	(0.12)	
T1 EXO	3.3	3.5	3.6	3.2	3.6	3.7
	(0.09)	(0.1)	(0.19)	(0.1)	(0.1)	(0.19)
T2 EXO	2.9	-	2.9	3.0	-	3.0
	(0.12)		(0.17)	(0.12)		(0.17)
T3 END	3.2	3.3	-	3.2	3.5	-
	(0.07)	(0.1)		(0.1)	(0.11)	
T4 END	3.2	2.8	-	3.4	3.3	-
	(0.09)	(0.13)		(0.1)	(0.13)	
T4 EXO	3.4	3.8	3.9	3.4	3.8	3.9
	(0.11)	(0.06)	(0.04)	(0.12)	(0.09)	(0.06)
T5 END	3.2	3.4	-	3.4	3.5	-
	(0.1)	(0.1)		(0.11)	(0.13)	
T6 END	3.3	3.1	-	3.4	3.6	-
	(0.07)	(0.11)		(0.08)	(0.1)	
T6 EXO	3.5	3.7	3.6	3.4	3.7	3.7
	(0.09)	(0.08)	(0.08)	(0.11)	(0.07)	(0.08)

Notes: Responses were coded as 1 = "Very socially inappropriate", 2 = "Somewhat socially inappropriate", 3="Somewhat socially appropriate", 4="very socially appropriate". The two sessions T2 EXO were conducted in the pilot phase. Later, we decided to collect the data in match 1, 10, 20 for EXO treatments. Standard errors (in parentheses) are calculated by taking each individual as a unit of observation.

Table D-13: Average Belief Deviation across Treatments

	All Rounds			Round 1		
	Match 1	Match 10	Match 20	Match 1	Match 10	Match 20
T1 END	44.0	31.0	-	49.4	46.0	-
	(2.48)	(4.0)		(3.79)	(5.36)	
T1 EXO	39.1	19.5	14.3	50.9	35.2	10.4
	(2.32)	(2.5)	(3.8)	(2.64)	(3.99)	(3.65)
T2 EXO	36.1	-	14.2	41.6	-	20.9
	(2.28)		(2.97)	(2.76)		(4.07)
T3 END	42.5	27.5	-	49.6	31.4	-
	(2.5)	(3.36)		(3.12)	(5.3)	
T4 END	41.4	16.5	-	47.3	34.1	-
	(1.24)	(2.04)		(2.41)	(4.03)	
T4 EXO	42.3	10.2	12.2	48.5	30.9	17.1
	(2.03)	(1.93)	(2.8)	(2.03)	(3.49)	(3.15)
T5 END	28.3	15.2	-	31.8	11.6	-
	(3.12)	(3.41)		(4.59)	(3.58)	
T6 END	40.5	24.5	-	44.5	13.9	-
	(1.68)	(2.39)		(2.02)	(3.19)	
T6 EXO	29.0	5.7	4.0	47.3	14.9	7.5
	(2.88)	(1.71)	(1.56)	(3.25)	(3.61)	(2.32)

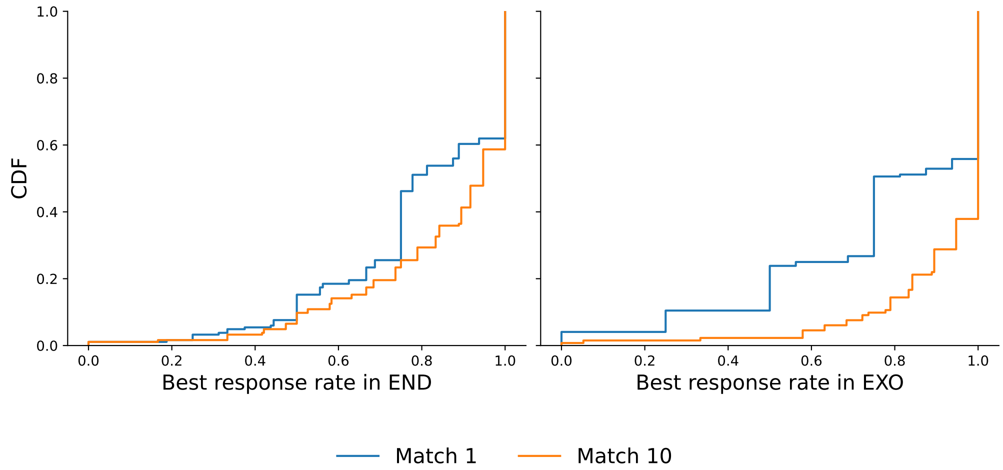
Notes: Belief deviations are measured as the absolute values of the difference between the reported belief and the actual choice of the other subject. Standard errors (in parentheses) are calculated by taking each individual in each match as a unit of observation

Table D-14: Average Best Response Rate across Treatments

	All Rounds			Round 1		
	Match 1	Match 10	Match 20	Match 1	Match 10	Match 20
T1 END	79.2	77.3	-	87.5	78.1	-
	(3.33)	(4.77)		(5.3)	(7.42)	
T1 EXO	76.0	89.2	87.7	73.1	80.8	100.0
	(3.73)	(2.49)	(3.65)	(6.21)	(5.52)	(0.0)
T2 EXO	80.6	-	90.8	82.5	-	87.5
	(3.74)		(2.98)	(6.08)		(5.3)
T3 END	70.6	83.3	-	57.5	84.4	-
	(4.26)	(4.62)		(7.92)	(6.52)	
T4 END	77.2	89.8	-	86.4	77.3	-
	(3.2)	(2.17)		(5.23)	(6.39)	
T4 EXO	65.6	91.8	87.9	67.5	80.0	82.5
	(5.27)	(2.02)	(4.88)	(7.5)	(6.41)	(6.08)
T5 END	84.8	88.6	-	92.5	97.5	-
	(3.27)	(2.87)		(4.22)	(2.5)	
T6 END	80.3	84.2	-	75.0	97.2	-
	(3.02)	(2.26)		(6.06)	(2.78)	
T6 EXO	85.8	96.1	97.1	77.5	90.0	97.5
	(3.61)	(2.53)	(2.5)	(6.69)	(4.8)	(2.5)

Notes: Best response to belief is a binary measurement that equates to 1 if a subject's choice is the best response to the expected payoff calculated based on her reported beliefs. The unit of observation is each individual in each match. Standard errors (in parentheses) are calculated by taking each individual in each match as a unit of observation.

Figure D-2: Best-Response Rates



Notes: The figure presents the cumulative distribution of subjects with different best response rate in match 1 and 10. The left panel shows all subjects in the END treatment. The right panel shows all subjects in EXO treatment. The best response rate for each subject is calculated as the average frequency of making the best responding choice within each match.

Table D-15: How Belief Related Variables Respond to Game Parameters

	Average belief		Belief accuracy		Best responding	
	(1)	(2)	(3)	(4)	(5)	(6)
Greater n (n=4)	-37.15***	-38.97***	-0.16	-1.59	0.03	0.02
	(5.18)	(4.92)	(4.37)	(4.34)	(0.03)	(0.03)
Greater b (b=218)	36.49***	37.04***	9.96***	10.06***	0.04	0.04
	(8.86)	(8.20)	(3.65)	(3.57)	(0.03)	(0.03)
Greater x0 (x0 = 0.812)	-59.68***	-60.56***	-13.59***	-13.26***	-0.07**	-0.07**
	(6.93)	(6.40)	(4.08)	(3.80)	(0.04)	(0.03)
Choose effort (endogenous)	-38.59***	-39.60***	-13.74***	-14.12***	-0.08***	-0.08***
	(6.37)	(5.90)	(3.41)	(3.30)	(0.03)	(0.02)
Constant	128.57***	134.18***	-10.31*	-10.40	0.91***	0.97***
	(6.81)	(9.52)	(5.36)	(7.71)	(0.04)	(0.08)
Observations	8,100	8,100	8,100	8,100	8,100	8,100
Number of Subjects	388	388	388	388	388	388
Preferences	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes

Notes: Belief deviations are measured as the absolute values of the difference between the reported belief and the actual choice of the other subject. Conformity measures how close my choice is to the expected choice of others. Best response to belief is a binary measurement that equates to 1 if a subject's choice is the best response to the expected payoff calculated based on her reported beliefs. Standard errors (in parentheses) are calculated by taking each individual in each match as a unit of observation

Table D-16: How Norms Respond to Game Parameters

	Descriptive social norm		Injunctive social norm		personal norm	
	(1)	(2)	(3)	(4)	(5)	(6)
Greater n (n=4)	-37.15***	-38.97***	-1.08***	-1.19***	-0.75***	-0.83***
	(5.18)	(4.92)	(0.21)	(0.21)	(0.22)	(0.21)
Greater b (b=218)	36.49***	37.04***	1.29***	1.32***	1.08***	1.10***
	(8.86)	(8.20)	(0.39)	(0.36)	(0.36)	(0.35)
Greater x0 (x0 = 0.812)	-59.68***	-60.56***	-1.38***	-1.34***	-0.81**	-0.78**
	(6.93)	(6.40)	(0.32)	(0.32)	(0.32)	(0.33)
Choose effort (endogenous)	-38.59***	-39.60***	-1.37***	-1.39***	-1.12***	-1.13***
	(6.37)	(5.90)	(0.29)	(0.25)	(0.27)	(0.24)
Constant	128.57***	134.18***	2.24***	2.45***	1.63***	1.90***
	(6.81)	(9.52)	(0.32)	(0.58)	(0.30)	(0.51)
Observations	8,100	8,100	8,100	8,100	8,100	8,100
Number of Subjects	388	388	388	388	388	388
Preferences	No	Yes	No	Yes	No	Yes
Demographics	No	Yes	No	Yes	No	Yes

Notes: The table reports results from Random Effects GLS Regressions pulling all individual data across the nine treatments. Columns (1) and (2) show how the average beliefs of others choosing to cooperate respond to game parameters. Columns (3)-(4) and (5)-(6) separately show how the personal norms and social norms of choosing to cooperate respond to game parameters. We assign values of 1,2,3,4 to the appropriateness evaluations of (very socially) inappropriate, somewhat (socially) inappropriate, somewhat (socially) appropriate, and (very socially) appropriate. The personal norm and social norm are measured as the appropriateness value difference between own cooperation and defection choice. For example, if a person assigns 'inappropriate' to the defection choice and 'appropriate' to the cooperation choice, her personal norm is calculated as 4-1 = 3. Variables of interest are dummy variables indicating whether the corresponding game parameter equal to the specified value. Preferences include risk aversion, loss aversion, other-regarding preference in disadvantageous and advantageous inequality, cognitive ability. Demographics include age, gender, major, and subjects high school location. Standard errors in parentheses are clustered at the session level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D-17: Robustness Check on Table 5: Linear Model, Fixed Effects

	(1)	(2)	(3)	(4)	(5)	R^2-dec
Intercept	0.56***	0.57***	0.58***	0.57***	0.51***	-
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	
Expected payoffs	0.01***	-	-	-	0.01***	0.30
	(0.00)				(0.00)	
Conformity	-	0.29***	-	-	0.12***	0.21
		(0.03)			(0.02)	
Injunctive norms	-	-	0.13***	-	0.01*	0.14
			(0.01)		(0.00)	
Personal norms	-	-	-	0.15***	0.10***	0.35
				(0.01)	(0.01)	
Observations	8100	8100	8100	8100	8100	-
within R^2	0.19	0.16	0.13	0.22	0.32	-
between R^2	0.48	0.74	0.28	0.42	0.71	-
overall R^2	0.31	0.43	0.20	0.32	0.52	-

Notes: The table reports results of the fixed effect estimator for the linear regression model with clustered standard errors pulling individual data at matches 1, 10, and 20 (in the exogenous treatments) across the nine treatments. The dependent variable is a dummy variable $A_{i,t}$ indicating whether a subject i in round t chooses to cooperate or not. The results on different alternative models are shown in columns(1)-(5). Standard errors in parentheses are clustered at the session level. The quality of the models is controlled by the R-squared metrics. The last column shows the results of the within R-squared decomposition based on the dominance analysis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D-18: Robustness Check on Table 5: Linear Model, Random Effects

	(1)	(2)	(3)	(4)	(5)	wR^2-dec	bR^2-dec
Intercept	0.52***	0.55***	0.55***	0.55***	0.49***	-	
	(0.04)	(0.04)	(0.05)	(0.03)	(0.00)		
Expected payoffs	0.01***	-	-	-	0.01***	0.30	0.26
	(0.00)				(0.00)		
Conformity	-	0.33***	-	-	0.15***	0.21	0.46
		(0.03)			(0.02)		
Injunctive norms	-	-	0.13***	-	0.00	0.14	0.10
			(0.01)		(0.00)		
Personal norms	-	-	-	0.15***	0.09***	0.35	0.18
				(0.01)	(0.01)		
Observations	8100	8100	8100	8100	8100	-	-
within R^2	0.19	0.16	0.13	0.22	0.32	-	-
between R^2	0.48	0.74	0.28	0.42	0.73	-	-
overall R^2	0.31	0.43	0.20	0.32	0.52	-	-

Notes: The table reports results of the random effect estimator for the linear regression model with clustered standard errors pulling individual data at matches 1, 10, and 20 (in the exogenous treatments) across the nine treatments. The dependent variable is a dummy variable $a_{i,t}$ indicating whether a subject i in round t chooses to cooperate or not. The results on different alternative models are shown in columns(1)-(5). Standard errors in parentheses are clustered at the session level. The quality of the models is controlled by the R-squared metrics. The last two columns show the results of the within and between R-squared decomposition based on the dominance analysis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D-19: Robustness Check on Table 5: Split by Treatment

	All	R^2-dec	EXO	R^2-dec	END	R^2-dec
Intercept	0.10	-	0.19	-	0.05	-
	(0.13)		(0.25)		(0.12)	
Expected payoffs	0.07***	0.34	0.03**	0.24	0.08***	0.53
	(0.01)		(0.01)		(0.01)	
Conformity	1.35***	0.21	2.39***	0.32	1.03***	0.11
	(0.29)		(0.35)		(0.21)	
Injunctive norms	0.19*	0.15	0.23	0.18	0.21*	0.07
	(0.10)		(0.17)		(0.12)	
Personal norms	1.24***	0.30	0.97***	0.26	1.27***	0.30
	(0.11)		(0.17)		(0.13)	
Observations	8100	-	4052	-	4048	-
AIC	4097	-	1707	-	2393	-
BIC	4216	-	1814	-	2500	-
marginal R^2_{Nak}	0.64	-	0.64	-	0.63	-
conditional R^2_{Nak}	0.88	-	0.84	-	0.90	-

Notes: The table reports results from the mixed-effects logistic regression using data from matches 1, 10, and 20 (when available) from all treatments. The results on the separate regressions for EXO and END treatments are also shown. The dependent variable is a dummy variable $a_{i,t}$ indicating whether a subject i in round t chooses to cooperate. To capture heterogeneity among individuals, we assume random intercept and random slopes (slopes vary among individuals). To capture the session-level effects, we assume that an intercept varies among sessions and among participants of the sessions. Note that the marginal Nacagawa's R-squared shows a proportion of the variance explained by fixed effects, while the conditional Nacagawa's R-squared shows a proportion of the variance explained by both, fixed and random effects. The results of the hierarchical partitioning of the marginal Nacagawa's R-squared for all models are also presented. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.