TI 2013-176/VII
Tinbergen Institute Discussion Paper

Performance and Relative Incentive Pay: The Role of Social Preferences

Pablo Hernandez[1]
Dylan Minor[2]
Dana Sisak[3]

[1] *New York University Abu Dhabi ;*
[2] *Northwestern University, United States of America;*
[3] *Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute, The Netherlands.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Performance and Relative Incentive Pay: The Role of Social Preferences*

Pablo Hernandez
New York University Abu Dhabi

Dylan B. Minor
Northwestern University

Dana Sisak
Erasmus University Rotterdam
& Tinbergen Institute

This version: October 2013

Abstract

Under relative performance pay, other-regarding workers internalize the negative externality they impose on other workers. In one form—increased own effort reduces others' payoffs—this results in other-regarding individuals depressing efforts. In another form—punishment reduces the payoff of other workers—groups with other-regarding individuals feature higher efforts because it is more difficult for these individuals to sustain low-effort (collusive) outcomes. We explore these effects experimentally and find other-regarding workers tend to depress efforts by 15% on average. However, selfish workers are nearly three times more likely to lead workers to coordinate on minimal efforts when communication is possible. Hence, the social preferences composition of a team of workers has nuanced consequences on efforts.

Keywords: *Social Preferences, Relative Performance, Collusion, Leadership*

*We would like to thank participants of seminars and conferences in Norwich, Rotterdam, Mannheim, Munich, Trier, Fresno, Budapest, Chicago, Zurich as well as Juan Atal, Ernesto Dal Bo, Josse Delfgaauw, Robert Dur, Dirk Engelmann, Sacha Kapoor, Martin Kolmar, John Morgan, Felix Vardy and Bauke Visser.

1 Introduction

Relative performance incentives are a common feature of the workplace environment. They appear in many different forms: be it monthly or yearly bonuses or promotions within an organization. An interesting feature of relative incentive pay is that a worker's performance also affects his or her co-workers compensation; in particular, it imposes a negative externality. An increase in one's own performance will not only increase one's own compensation, but inevitably also decrease a co-workers pay, at least in expectation. How this externality affects the incentives of a worker will crucially depend on whether a worker incorporates this reduction in his or her own effort decision. A worker who incorporates a co-worker's payoffs can also be referred to as "other-regarding" while a worker basing decisions solely on own payoffs can be referred to as "selfish;" these different types of workers should then respond differently to relative incentives. A robust finding in the literature is that individuals have heterogeneous degrees of other-regardingness (e.g., see Andreoni and Miller (2002) and Fisman, Kariv and Markovitz (2007)), which suggests that the effectiveness of relative incentives will depend on particular worker social preferences.

In this paper, we explore how social preferences of this form affect effort provision in a relative incentive framework using a controlled laboratory environment. In particular, we measure subjects' social preferences using dictator menus and then relate these to their effort decisions under relative incentives. Thus, we contribute to the understanding of the efficacy of relative incentives.

A first-order question for firms is what types of workers to hire. If a firm uses relative incentives, it is especially important not to treat each hiring decision in isolation; since individual effort decisions are strategically linked to others' effort decisions, group composition becomes quite important. Our experimental design allows us to randomly assign subjects with different levels of other-regardingness into groups and thus identify the effect of group composition on effort. In addition, we employ an indefinitely repeated game setting to capture the feature that workplace interactions are usually not one shot but instead have some probability of continuing in the future. In this setting, we find that groups with more other-regarding workers tend to depress efforts. When communication is not part of the work environment, each other-regarding group member depresses overall effort by 15%. At the individual level we find that, absent communication, own social preferences matter for effort decisions, while individuals are not significantly affected by the other group members' social preferences.

Communication is, of course, an important feature of many workplace settings. In an indefinitely repeated relative performance setting, communication can help workers coordinate their effort choices to their mutual benefit. To facilitate such coordination it is expected a leader will emerge. Here we use the term leader in the sense of leader as a coordinator, as argued by Kreps (1986) and Hermalin (2012). To shed additional light on the effect of social preferences, we further analyze whether

social preferences relate to the emergence of a coordinating leader—an individual suggesting to the group to coordinate on minimal effort, which is the Pareto optimal outcome from the workers' viewpoint. We find that selfish individuals are 2.7 times more likely than other-regarding individuals to successfully lead their groups to such a "collusive" outcome. Controlling for this sort of leadership, we still find that even with communication, other-regarding subjects depress their effort relative to selfish ones by 50%. This implies that the effect of social preferences on work performance under relative incentives is a nuanced one. On the one hand, other-regarding workers have a tendency to depress effort, apparently through their internalizing of their efforts' negative externality. On the other hand, with the availability of communication, selfish workers seem more likely to help direct the group to the lowest of efforts. Thus, heterogenous groups may actually be the worst for a principal interested in maximizing workforce effort.

In order to eliminate possible confounds such as differences in beliefs or degrees of patience, we have subjects in a different treatment face computerized simulated subjects exhibiting choice behavior similar to that of past human subjects. Thus, while strategic incentives are left intact, social preferences are "turned off" in this treatment. In this setting, we find that by the end of the relative performance stage, other-regarding and selfish subjects are indistinguishable, lending support to our hypothesis that differences in social preferences are driving our results.

The structure of the paper is as follows. In the next section we review the relevant literature. In Section 3 we describe our experimental design and offer a simple theoretical framework. Section 4 provides the results of the laboratory experiment. In Section 5 we conclude and discuss organizational design implications.

2 Literature

The significant body of literature that documents different degrees of social preferences (e.g. Andreoni and Miller (2002); Fisman, Kariv and Markovitz (2007); DellaVigna (2009)) has led researchers to investigate their effects on public good contributions and other pro-social behaviors under different incentive schemes (e.g. see Bowles and Polania-Reyes (2012) for a survey). Fehr and Fischbacher (2002) also point at that when scholars disregard social preferences, they fail to understand the determinants and consequences of incentives. In our paper, we explore the effects of social preferences on productivity in the setting of relative performance incentives. To our knowledge, the only past study that explores this issue is Bandiera, Rasul and Barankay (2005). They find that fruit pickers in the UK cooperate on lower levels of effort under relative performance pay only when monitoring is possible. However, they also find that workers with social ties even more strongly depress effort when monitoring is available. Social ties could capture social preferences; however, they could also capture the salience of punishment should one "defect" from low efforts. As a result, it is unclear whether social preferences induce lower efforts in this setting.

Our paper complements this work by directly measuring participants' social preferences (à la Andreoni and Miller (2002)) and randomly combining groups to identify the link between social preferences and the responses to relative performance incentives. In order to do so, we employ an indefinitely repeated game setting to study the effect of social preferences even when collusion is possible.

In indefinitely repeated games, Pareto improvements over the one-shot Nash equilibrium can be obtained as equilibrium outcomes if the value of the future is high enough. Versions of this "folk theorem" can be found in Friedman (1971), Fudenberg and Maskin (1986). Indeed, it has been documented that individuals are able to achieve the Pareto-optimal outcomes quite often. For example, Palfrey and Rosenthal (1994) found cooperation rates from 29% to 40% and Dal Bo (2005) found cooperation rates of 38%. There is, however, a great variety of outcomes in this literature, which may or may not correspond to equilibrium outcomes derived from standard economic models, as Dreber, Fudenberg and Rand (2011) argue. Hence, while one purpose of our paper is to study the effect of social preferences on effort in relative performance schemes, a second purpose is to explain the different sources of variation in group productivity by controlling for groups' social preference composition.

This variation comes from the fact that indefinitely repeated games may have multiple equilibria. A usual criticism of the theory is that it does not provide sharp predictions about equilibrium selection (e.g. Dal Bo and Frechette (2011)). One method of dealing with equilibrium selection in games of coordination is analyzing the behavior of a leader, as argued by Kreps (1986) and Hermalin (2012). The emergence of such a leader may be related to social preferences. Indeed, a leadership-social preference link is reported in recent work by Gaechter, Nosenzo, Renner and Sefton (2012) and Kocher, Pogrebna and Sutter (2013). Our work complements theirs in that we explore the endogenous emergence of leaders, whereas their leaders are determined exogenously—leaders are assigned and then behavior is explored. In addition, whereas we study leadership through communication, the other papers study leadership by example and by asserting authority, respectively. Finally, our work also contributes to the literature on communication in games with multiple equilibria (e.g. Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross, 1992, Ledyard 1995; Seely, Van Huyck and Battalio 2007); while the extant literature is concerned about the effect of communication on the frequency of Pareto optimal outcomes, we instead explore how a group's social preference composition leads to patterns of communication (e.g., leadership emergence) that result in players coordinating on their Pareto optimal outcome.

3 Experimental Design

In total, we conducted 8 experimental sessions with 168 subjects. Participants were students from UC Berkeley, enrolled in the X-lab subject pool. Sessions lasted approximately 60 minutes from reading instructions to subject payment, which averaged

approximately $16 per subject. Participants were not allowed to take part in more than one session. The treatments were programmed and conducted using *z-Tree* developed by Fischbacher (2007).

We had the dual purpose of identifying subjects' social preferences and measuring their choices when facing a relative performance incentive scheme. In order to achieve this, the experiment was divided into three stages. In the first stage, we randomly matched subjects into anonymous groups of three individuals. Participants were then given 100 tokens for each of 9 periods and played a dictator game with their group members (including themselves). In each period participants faced different "prices" or token exchange rates of giving to each group member. Prices varied such that we could both identify individuals' willingness to give to others and individuals' willingness to give between others when facing different prices of giving.[1] We use these 9 periods to classify our subjects in terms of social preferences. In periods 10 and 11 we conducted allocation decisions with positive sloped budget sets as in Andreoni and Miller (2002) where subjects are given an allocation and decide on the overall exchange rate. We will use these decisions to test whether aversion to disadvantageous inequality matters in addition to other-regardingness in responding to relative incentives. These results are reported in the Appendix. Since we follow the categorization of Andreoni and Miller (2002), we are thus considering unconditional rather than conditional social preferences.

Subjects did not learn their other group members' choices to avoid uncontrolled learning. Participants were told that for 5 out of a total of 11 allocation decisions one of the group members' choices would be randomly selected to compute payoffs.

We use this first stage, in particular decisions in rounds 1 to 9, to classify participants as "Selfish" or "Other-Regarding."[2] An archetypal Selfish type, is only interested in his own monetary payoff and thus should never allocate any tokens to his or her group members. Thus we classify as Selfish all subjects that throughout rounds 1-9 do not allocate any tokens to another group member. The remainder of subjects are classified as Other-Regarding. We consider various other possible classifications in the analysis found in our online appendix; however, they provide little additional insight to this simple classification.

For the second stage, participants were again randomly matched with two other players for the remainder of the experiment. They participate in a relative performance game modeled after Bandiera, Barankay and Rasul (2005). The purpose of this stage was to give players the possibility to collude by jointly providing low levels

[1] Fisman et al. (2007) uses a slightly different nomenclature to describe distributional preferences. They call *preferences for giving* the fundamentals that rule the trade-off between individiual and others' payoffs and *social preferences* the ones that govern the allocation between others. Our study does not focus on that distinction, therefore we employ the following terminology: We use "social preferences" or "other regarding concerns" indistinctly to represent non-selfish behavior.

[2] From now on we use the capitalized form of selfish and other-regarding to refer to our categorization. Thus we do not imply that a subject we categorize as selfish necessarily always acts in a selfish manner, but only that given our categorization, he or she most closely resembles this type.

Treatment	Subjects
Chat & Observability	63
No Chat & Observability	63
No Chat & No Observability	21
Robot	21
Total	168

Table 1: Summary of Treatments

of effort. Thus, we implemented an indefinitely repeated game with continuation probability of $\delta = 95\%$. In order to gain consistency across treatments, we randomly drew the number of periods before running the sessions as in Fudenberg, Rand, and Dreber (2012).

We also varied factors considered important for creating and sustaining collusion. In particular, in the first treatment ("Chat/Observability") we allowed chat via computer terminals *during* each period and observability of choices and payoffs *after* every period. In the second treatment ("No Chat/Observability") we did not allow for chat but continued with observability after each period. In the third treatment ("No Chat/No Observability") neither chat nor observability was allowed. In this treatment, subjects only learned their own payoff after each period. Thus, this treatment allows us to identify effort levels when a player's group members' efforts are not directly observable. Since *a priori* we did not know how easily subjects would collude, we wanted to provide different degrees of difficulty of coordinating.

If we were able to mechanically switch on and off subject's social preferences, we could directly identify the effect of social preferences on effort. Unfortunately, this is not generally possible. However, we conducted a final treatment where we approximate this idea. Instead of facing human subjects, a subject played against their computer, which simulated the play of past subjects' decisions ("Robot" treatment). This treatment attempted to "switch off" social preferences by making it clear to subjects that even though they faced the same consequences for their choices as if playing human subjects, their effort decisions no longer affected any person's payoffs.

Table 1 provides a summary of these treatments.

A subject's payoff was calculated as follows. Note these figures are in Berkeley Bucks \$, converted at \$66.6 Berkeley Bucks to 1 US\$, which is how it was presented to subjects.[3] Each participant received an endowment of \$12 (Berkeley Bucks \$) each period from which they could choose costly effort. Effort costs \$1 for each unit of effort. Total payoff was then

$$\pi_i = 12 + \frac{x_i}{\overline{x}} 15 - x_i$$

[3]A copy of the instructions given to subjects is available in the appendix.

where $\overline{x} = \frac{\Sigma x_j}{3}$ is the average effort across $i's$ group and i chooses effort $x_i \in [1, 12]$.[4] Hence, each participant's effort is discounted by the average effort, so a higher average effort will reduce payoffs, *ceteris paribus*. This is the relative performance evaluation similar to the contract used by Bandiera, Barankay, and Rasul (2005).

The stage game (or one-shot) Nash equilibrium for homogeneous and Selfish players is to play $x_i = 10$ for all i, which is below 12 (the upper bound of the action space). Coordinating on $x_i = 1$ is sustained by a continuation probability $\overline{\delta} > 60\%$ (optimal one-shot deviation from Pareto Dominant outcome is to play $x_i \simeq 7.5$). Therefore, our $\delta = 95\%$ should guarantee the feasibility of collusion for utility maximizing rational Selfish agents.

For the final stage, subjects were again given the same allocation price menus as in the first stage. Critically, subjects did not know they were going to have this final allocation opportunity. Instead, they were told at the beginning of the experiment they would have a final stage with some additional opportunities to increase their payoffs. We do not consider this data in this study.

After the allocation decisions, subjects completed a risk aversion test as in Holt and Laury (2002), and a basic demographic questionnaire. We now turn to our theoretical analysis.

3.1 Theoretical Considerations

In this section, we explore the various incentives when group composition varies in terms of social preferences. Consider first the incentives in the stage game of an indefinitely repeated game without communication—as in our No Chat/Observability treatment. Assume that a player believes her group members' efforts are x_{1o} and x_{2o} and her utility is given by

$$\pi_i = 12 + \frac{x_i}{\bar{x}} \times 15 - x_i + \rho_i \left[\left(12 + \frac{x_{1o}}{\bar{x}} \times 15 - x_{1o} \right) + \left(12 + \frac{x_{2o}}{\bar{x}} \times 15 - x_{2o} \right) \right], \quad (1)$$

where $\rho_i \in [0, 1]$ represents the degree of Other-Regardingness.[5] If $\rho_i = 0$, the player is a Selfish individual and only values his own payoff. If instead $\rho_i > 0$, the player is Other-Regarding (i.e., she at least partially internalizes the negative externality that she imposes on others), and increasingly so as ρ_i increases. For $\rho_i = 1$ the individual places equal weight on all group members' payoffs.

As can be shown, the (one-shot) best response function is

[4] Although subjects were not told to do so, almost all entered effort choices as an integer. We had an effort lower bound of 1 to create an upper bound for payoffs. The effort upper bound of 12 came from the periodic endowment of \$12.

[5] Note that we have assumed an Other-Regarding individual cares about each of her other members' payoffs equally. Although this specification is a stylized version of Fehr and Schmidt (1999), it allows us to illustrate the role of social preferences on collusive outcomes.

$$x_i^*(\rho_i) = \max\left\{\sqrt{45(1-\rho_i)(x_{1o}+x_{2o})} - (x_{1o}+x_{2o}), 1\right\} \tag{2}$$

Best response effort is strictly decreasing in the degree of Other-Regardingness ρ_i. Thus, subjects who put higher weight on others' payoffs choose lower levels of effort. Note that in the extreme case of $\rho_i = 1$ (i.e., an individual maximizing a utilitarian welfare function), the best-response is always to choose minimal effort. Solving for the stage game Nash equilibrium yields

$$x_i = \max\left\{\frac{90(1-\rho_i)(1-\rho_j)(1-\rho_k)(1-2\rho_i+\rho_i\rho_j+\rho_i\rho_k-\rho_j\rho_k)}{(3-2(\rho_i+\rho_k+\rho_j)+\rho_i\rho_j+\rho_i\rho_k+\rho_j\rho_k)^2}, 1\right\}$$

In an interior solution for all three group members ($x_1, x_2, x_3 > 1$), aggregate effort equals

$$X = x_1 + x_2 + x_3 = \frac{90(1-\rho_i)(1-\rho_j)(1-\rho_k)}{(3-2(\rho_i+\rho_k+\rho_j)+\rho_i\rho_j+\rho_i\rho_k+\rho_j\rho_k)} \tag{3}$$

It is easily verified that aggregate effort is strictly decreasing in the degree of other-regardingness of each single group member. Note that if $\rho_i = \rho_j = \rho_k \geq \frac{9}{10}$ the unique stage game equilibrium will be exactly the "collusive outcome": all play minimal effort of 1 even in a shot game. Of course, playing the one-shot best response in each period is an equilibrium in the dynamic game. In this case, Other-Regarding players unambiguously depress overall group effort.

In dynamic games a collusive equilibrium outcome is another possibility. For this possibility, punishment is essential. Other-Regarding individual, however, have less incentives to exert punishment upon observing a defection. For example, consider the canonical grim-trigger strategy. Group members begin by coordinating on minimum efforts and continue each period until at least one group member deviates from minimum effort; in response to such a deviation, a player punishes by choosing the stage game Nash equilibrium effort forever. Note that the aggregate effort of the two other group members in the stage game Nash equilibrium (for an interior solution) from the viewpoint of i is equal to

$$X_{-i} = \frac{180(1-\rho_i)(1-\rho_j)^2(1-\rho_k)^2}{(3-2(\rho_i+\rho_k+\rho_j)+\rho_i\rho_j+\rho_i\rho_k+\rho_j\rho_k)^2} \tag{4}$$

It can be verified that X_{-i} is decreasing in ρ_j and ρ_k. Consequently, the punishment for a grim trigger strategy is less for group member i as her other group members are increasingly Other-Regarding. Hence, this suggests that it could be *more* difficult for groups with more Other-Regarding players to sustain the collusive outcome of extreme effort depression. This is especially the case for groups with only one Selfish individual, as he has the strongest incentives to take advantage of his

group members—his cost of deviating goes down as others' Other-Regarding concerns go up. In Table 2, we illustrate how the costs and benefits of deviating relate to effort decisions for a group of 3 individuals, with varying numbers of Selfish ($\rho_i = 0$) and Other-Regarding ($\rho_i = .5$) players. For an example of Other-Regardingness hindering collusive outcomes, we calculate the minimal continuation probability necessary to sustain a collusive outcome of $(1, 1, 1)$ under grim-trigger strategies. A higher minimal continuation probability is associated with lower costs of deviation for at least one group member, so players need to care more about the future to sustain collusion. In the extreme, a continuation probability of one means that only a group of individuals who care equally about the present and the future are able to collude. We find that the continuation probability necessary for sustaining collusion is non-monotonic in the number of Other-Regarding group members. In particular, the minimum continuation probability is lowest in a group with only Other-Regarding individuals, but highest in a group with one and only one Selfish player. In fact, given our parameters and grim-trigger strategies, a group with only one Selfish individual will not be able to collude. Thus, while a group of only Other-Regarding members should be worst for a principal, as collusion is most likely and periodic effort is lowest, it may be the case that the principal only needs to replace one Other-Regarding worker with a Selfish one to maximize total effort.

Since we randomly determine group composition, our No Chat/ Observability treatment allows us to test if Other-Regarding subjects exert lower effort on average. If players' decisions converge on Nash stage-game outcomes, we expect Other-Regarding players to unambiguously decrease group efforts. However, Other-Regarding subjects may make it more difficult to collude and sustain collusion due to their incentives to punish less severely, which could increase average group effort.

	Total effort (one-shot Nash)	Lowest continuation probability to sustain collusion
3 Other-Regarding	15	0.364
2 Other-Regarding & 1 Selfish	18	> 1
1 Other-Regarding & 2 Selfish	23.5	0.791
3 Selfish	30	0.609

Table 2: Group composition and effort provision (Selfish: $\rho_i = 0$, Other-Regarding: $\rho_i = .5$). The second column shows aggregate equilibrium effort in the one-shot game. The third column shows the lowest continuation probability necessary to sustain collusion under grim-trigger strategies in the indefinitely repeated game.

In this analysis, we have illustrated the potentially opposing effects of social preferences on effort in a dynamic setting using one-shot Nash equilibria and grim-trigger strategies. Indefinitely repeated games, however, typically admit a myriad of equilibria. Nonetheless, communication may help players converge on a particular equilibrium through the emergence of a coordinating leader (e.g. Hermalin 2012). Thus,

it may also be the case that introducing communication creates another channel for social preferences to influence collusive outcomes, through the likelihood of a subject emerging as a leader suggesting coordination on minimal effort. We explore this possibility via our Chat/Observability treatment. To determine the net effect of all of these channels, we now turn to our empirical analysis.

4 Empirical Analysis

4.1 Examples of Decisions

We begin with some examples of actual giving and effort rates of particular groups to illustrate subjects' behavior. We analyze the effect of social preferences on effort in Section 4.3 and coordination leadership in Section 4.4. Figure 1 illustrates the patterns of decisions across time. In the first stage (periods 1 to 9), we can observe the number of tokens each player in the group keeps for him or herself. In the second stage, (periods 12 to 40) we observe the choice of effort ranging from 1 to 12.[6]

Starting with Panel 1 we observe a heterogeneous pattern of keeping in the first stage: One subject keeps everything to himself, while the others share almost equally. Thus, this group consists of one Selfish and two Other-Regarding subjects. Furthermore, it provides an example of "perfect collusion" in the Chat/Observability treatment: Subjects coordinate on minimal effort during almost the entire second stage.

Coordination on minimum effort $(1, 1, 1)$ also occurs absent communication. Panel 2 provides an example in the No Chat/Observability treatment on how subjects slowly manage to coordinate on lower efforts.

Panel 3 shows a group from the Chat/Observability treatment. In this case, behavior in the second stage is surprising: Participants play a strategy that is not the Pareto-dominant one. Subjects alternate between providing maximal and minimal effort. In each period a different subject reaps the rents of outperforming the other subjects With the help of the chat, they perfectly coordinate on this synchronized play. Although this does not allow the subjects to reach the maximal group payoff, this form of collusion still leads to high payoffs relative to the one-shot Nash outcome.

Finally, communication does not guarantee successful coordination. Our last example, Panel 4 provides a case in point. In this group from the Chat/Observability treatment, subjects choose maximal efforts in almost every round.

[6]We omitted periods 10 and 11 from the graphs. They are used for an extended categorization of subjects in the Appendix.

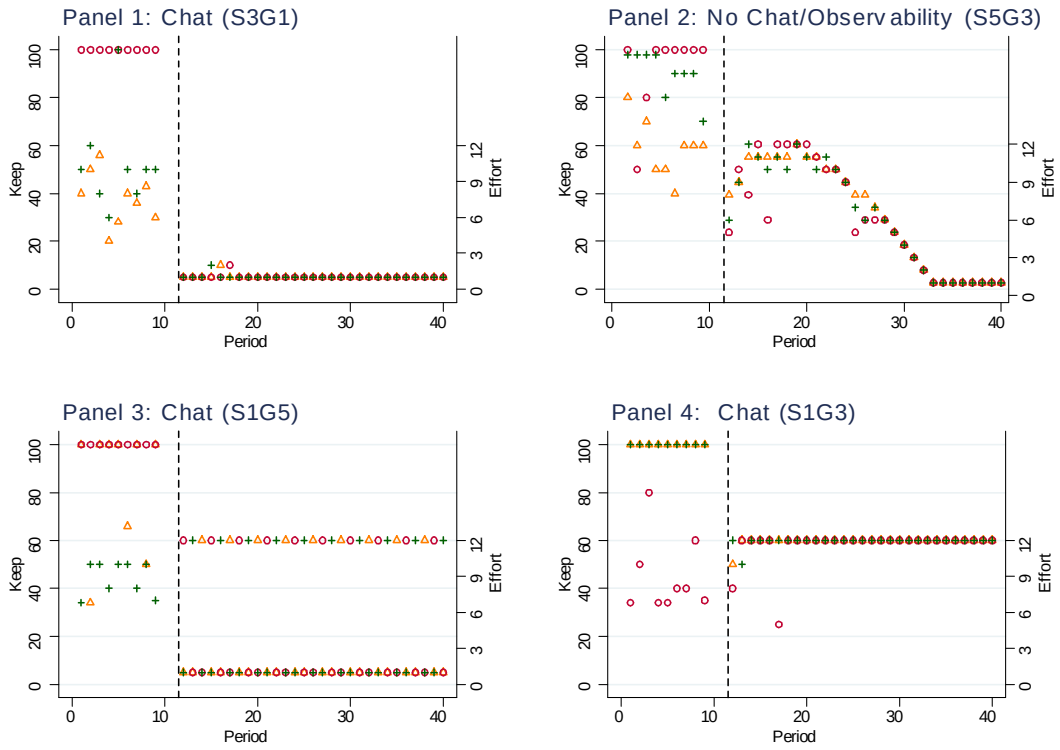


Figure 1: Examples of group giving and investment decisions (S denotes session number and G group number).

4.2 Categorizing Social Preference Types from Giving Menus

Table 3 summarizes the mean choices of our subjects under all 9 price vectors in treatments: 1) Chat/Observability, 2) No Chat/Observability and 3) No Chat/No Observability.[7] We will analyze the Robot treatment in section 4.6.

We see that regardless of the price of giving, subjects keep on average just above 70% of their endowment. Using these choices, we sort our subjects into Selfish and Other-Regarding. A subject is categorized as Selfish if he or she does not allocate any tokens to the other group members in any of the nine periods. All subjects who at some point allocated tokens to their group members are categorized as Other-Regarding. We explore other categorizations in the Appendix. Taking together the three treatments (Chat/Observability, No Chat/Observability and No Chat/No Observability) most of the participants (79.59%) are categorized as Other-Regarding. The balance of 20.41% of subjects are categorized as Selfish.

As described in Section 3 subjects were randomly allocated into groups without

[7]These vectors (a, b, c) represent the price a of giving to one's self, the price b of giving to player 1, and the price c of giving to player 2.

Period	Price vector	Keep (min, max)	Give to 1	Give to 2
1.	$(1,1,1)$	70.66 (33,100)	15.21	14.13
2.	$\left(1,\frac{1}{2},\frac{1}{2}\right)$	73.39 (0,100)	13.24	13.37
3.	$\left(1,\frac{3}{4},\frac{3}{4}\right)$	71.82 (0,100)	13.98	14.20
4.	$\left(1,\frac{5}{4},\frac{5}{4}\right)$	72.29 (20,100)	14.13	13.59
5.	$\left(1,\frac{3}{2},\frac{3}{2}\right)$	71.03 (20,100)	14.67	14.30
6.	$\left(1,1,\frac{2}{3}\right)$	71.80 (0,100)	15.90	12.30
7.	$\left(1,1,\frac{3}{4}\right)$	73.46 (0,100)	15.03	11.51
8.	$\left(1,\frac{3}{4},\frac{1}{2}\right)$	77.09 (0,100)	12.33	10.58
9.	$\left(1,\frac{5}{4},\frac{3}{4}\right)$	72.72 (0,100)	16.18	11.10

Table 3: Giving Rates.

regard to their social preference type. Figure 2 shows the distribution of Selfish subjects across groups. Since subjects were allocated randomly and Selfish subjects are relatively rare we do not observe groups with only Selfish group members in the Chat/Observability and No Chat/Observability treatments. Otherwise, we do observe random variations across groups in the number of Selfish subjects which we will use to identify the effect of group composition in the next sections.

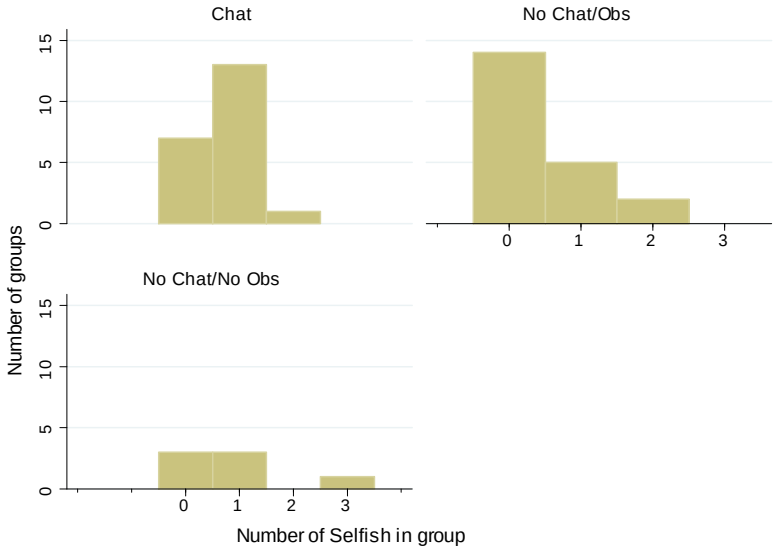


Figure 2: Allocation of Selfish across groups.

4.3 Social Preferences and Effort

Figure 3 provides a summary of effort choices over time by treatment. In all three treatments we observe average effort of around 8 units at the beginning of the relative incentives stage. As expected, there is a strong tendency to coordinate on lower efforts over time when subjects are able to communicate and observe past behavior in the Chat/Observability treatment (dashed line). When communication is not possible, observability by itself does little in sustaining lower levels of effort overall. In both such treatments, average effort stays close to the one-shot Nash equilibrium level for the Selfish type (dotted and solid line).

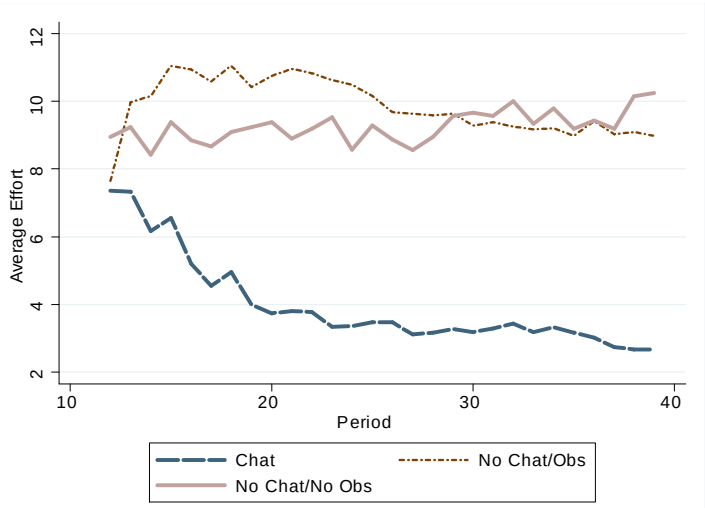


Figure 3: Average effort by treatment over time.

How do individual social preferences and group composition relate to efforts? To give an answer to this question we exploit the random allocation of subjects into groups. We compare behavior of groups with different numbers of Selfish and Other-Regarding individuals in each of the three treatments.

Figure 4 gives a first overview of our findings. Consider first panel a) in the upper left corner. We compare the average effort of subjects categorized as Selfish with the average effort of subjects categorized as Other-Regarding. We see that for all three treatments, average effort is higher for subjects categorized as Selfish, although a t-test rejects equality only for the No Chat/Observability treatment (p-values: $p<0.60$ in Chat/Observability; $p<0.01$ in No Chat/Observability and $p<0.35$ in No Chat/No Observability). Comparing the No Chat/Observability and No Chat/No Observability treatments, average efforts are similar to the one-shot Nash equilibrium efforts (i.e., efforts of 10 with $\rho = 0$) rather than a collusive outcome. Accordingly, as predicted in Section 3.1 for "non-collusive" outcomes, Other-Regarding subjects provide lower efforts on average.

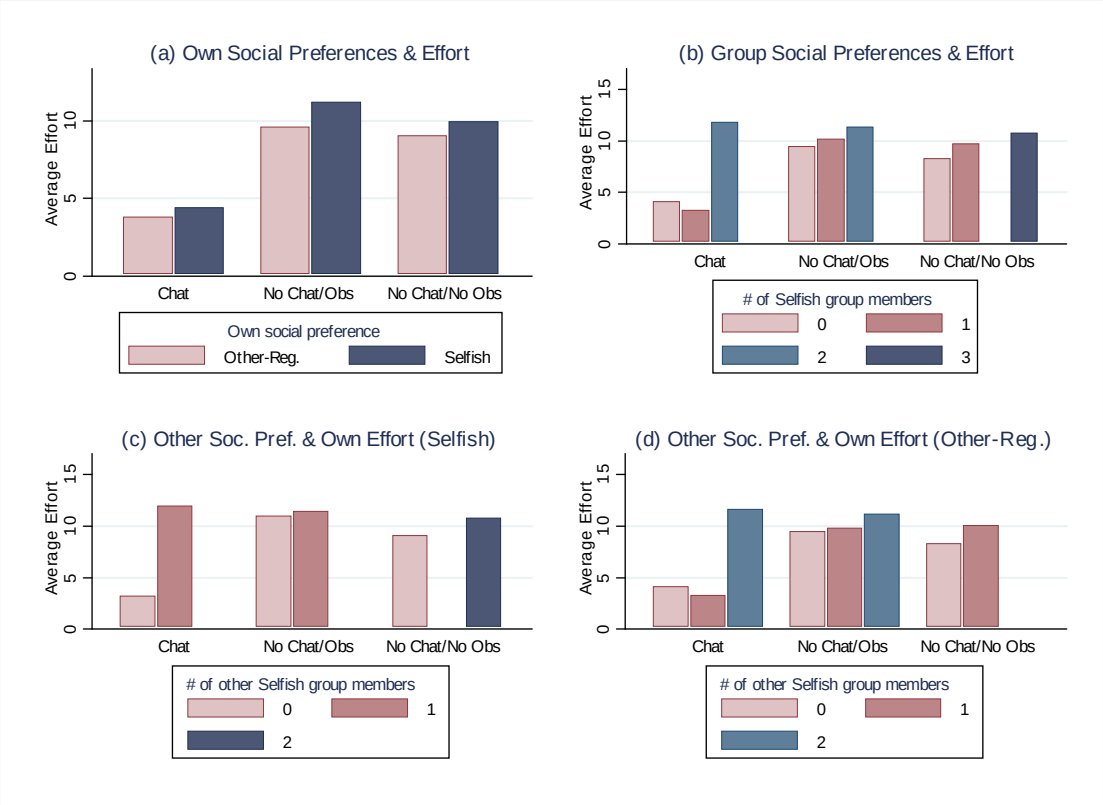


Figure 4: Overview of Effects of Social Preferences on Effort.

Panel b) in the upper right corner takes the view of the principal, who is interested in group effort. In the graph, we consider average group effort as a function of the number of Selfish players within a group. In the two treatments where communication was not possible, we observe that each additional Selfish group member increases average group effort. In particular, a group with only Other-Regarding members is worst for the principal. The monotonicity in social preferences and the relatively high levels of effort again suggest that collusion does not play a big role and the results are driven by the incentives of the stage game. In contrast, when communication is possible, the effect of Other-Regarding group members is non-monotonic. Average group effort is lowest when there is only one Selfish individual in the group. The latter result is especially interesting in light of our discussion in Section 3.1: we argued that a group with only Other-Regarding members will have lower stage-game aggregate effort and require a lower minimal discount factor than a group with one or more Selfish individuals. Furthermore, a Selfish individual in a group of Other-Regarding has the strongest incentive to take advantage of her fellow group members and defect. However, we observe a totally different pattern. Groups with only one Selfish member show the lowest levels of effort. While these charts are suggestive, not all differences in group average effort are statistically significant. Thus, we explore differences in

group effort choices as a function of the number of Selfish subjects through regression analysis as reported in Table 4.

Panel c) in the bottom left corner (respectively, panel d) in the bottom right corner) shows how individual effort of a Selfish (respectively, Other-Regarding) subject varies with the social preferences of the *other* group members. For a subject categorized as Selfish we find that effort is increasing in the number of other Selfish group members for all treatments. In contrast, for an Other-Regarding subject we find monotonicity of efforts only in the non-communication treatments. This suggests that, absent communication, both Selfish and Other-Regarding subjects respond more aggressively to fellow group members if they are Selfish. In the Chat/Observability treatment, however, Selfish respond more aggressively to other Selfish group members, while Other-Regarding respond less aggressively to the addition of the first Selfish individual to the group.

Overall, these results suggest that, absent communication, average efforts are consistent with one-shot Nash equilibrium strategies. When communication is introduced, however, efforts resemble the collusive outcome and results are somewhat surprising: The presence of one Selfish individual leads to lowest aggregate efforts. We return to analyze this treatment in more detail in Section 4.4.

To further explore our results, we construct a dependent variable of group effort averaged over all rounds of play (at stage 2, our relative performance stage). Groups are randomly assigned, so these averages are independent of individual assignment to groups. Table 4 reports the results of regressing average group effort on the number of Selfish individuals in a group. Column 1 shows the results for the Chat/ Observability treatment, column 2 and column 3 report the results for the No Chat/Observability treatment and No Chat/No Observability treatment, respectively. In the Chat/Observability treatment, we do not find a significant effect of Selfish group members. This is to be expected given that we identified a non-monotonic relationship from Figure 4. In contrast, when communication is not possible (No Chat/Observability treatment), and when neither communication nor observability are allowed (No Chat/No Observability treatment), each Selfish group member increases average group effort by approximately .9 units on average, which equals a 12% increase over our baseline mean effort of roughly 7.5 per period.

In the remainder of this section, we explore further the results of our No Chat/ Observability treatment. To disentangle the effect of one's own social preference from group composition effects we estimate a random effects model for the No Chat/ Observability treatment, clustering standard errors on the group level.[8] We exclude the Chat/Observability treatment in which composition effects are also feasible as we will devote the next section to this treatment. Table 5 reports our results. We find further evidence that Other-Regarding subjects choose significantly less effort. Now controlling for group composition, these subjects choose 1.5 fewer units of effort.

[8]Throughout the paper when using a random effects regression we cluster at the group level. Results are qualitatively unchanged when clustering at the individual level.

	Chat	No Chat / Obs	No Chat / No Obs
# Selfish	1.063	0.872**	0.860***
	(1.626)	(0.379)	(0.163)
Constant	3.180***	9.453***	8.540***
	(1.022)	(0.440)	(0.266)
Observations	21	21	7
Adjusted R^2	-0.012	0.081	0.656

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 4: Effect of a groups' social preference composition on group effort.

	Effort	
Period	-0.0538*	(0.0294)
Selfish	1.478***	(0.401)
# Other Selfish	0.569	(0.412)
Constant	10.85***	(0.502)
Observations	1827	
R^2 within/between	0.0322/0.0954	

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 5: Effect of own and others social preferences on own effort (treatment 2).

The group composition effect on the other hand, is positive but insignificant. Thus it seems that it is mainly the own social preference type that determines effort choices throughout the relative performance stage in this treatment.[9]

Although not included in this table, we have explored whether one's reaction to the social preferences of one's group members depends on one's own social preference type. We did not find any significant interaction effect between own social preference type and group members' social preference type. We do not include lagged effort choices due to the issue of inconsistent estimates from such a specification. Nonetheless, when doing so, our results are qualitatively the same. In addition, since effort choices are constrained to be between 1 and 12, we re-run our analysis using a Tobit

[9]In principle, one could conduct this analysis also for the No Chat/No Observability treatment. In this treatment feedback during the game is minimal and thus it is unclear how to interpret any interaction effects. If we run the regression, we do find a positive and significant interaction effect (more other Selfish correlates with higher effort). Due to the small sample size of this supporting treatment, however, we do not want to overly emphasize this result.

panel model. We find these results are qualitatively the same. We also conducted our individual level analysis controlling for gender, education major, and risk preferences, and find the results qualitatively unchanged. Finally, rather than using social preference types as regressors, we conduct individual-level regressions using instead the average amount of endowment kept by a subject to examine if subjects' intensity of social preferences matters. We find consistent results with this measure of selfishness. We also consider an alternative classification of social types: we classify Selfish subjects as those that keep on average at least 90% of their endowment (as opposed to 100%). Using this less stringent definition of Selfish subjects we find that the magnitude of the coefficient estimates on Selfish types decreases, but is still significant. However, for the group level regressions although the sign is still correct, the coefficient estimates are no longer significant.

Overall, the similarity of effort levels across the "No Chat" treatments and the positive difference between Selfish and Other-Regarding individuals' average efforts within each treatment, suggest that collusion does not play a major role. More importantly, these results seem to be driven by Other-Regarding individuals internalizing the negative externality of their effort choice by exerting one-shot competitive efforts.

When communication is possible, social preferences do not seem to affect efforts linearly in the number of Selfish group members. We found that groups with only one Selfish member seem to be most successful at choosing low efforts. In the next section we investigate deeper into the reason behind this non-monotonicity. In order to do this we differentiate between 1) encouraging low efforts through chat and 2) effectively choosing low efforts. Social preferences might relate differently to these two aspects of non-competitive efforts. To disentangle these effects we analyze the chat messages of each group and identify leaders that initiate collusive-like behavior and their social preferences.

4.4 Leadership

In the Chat/Observability treatment, a subject can take the initiative through chat, asking the group members to jointly exert low effort. This way the problem of equilibrium selection can be overcome. This channel was absent in all other treatments. We elicit this measure of "leadership" from the chat messages. We differentiate between two kinds of leaders: "First Leader" and "Right Leader."[10] A First Leader is the first subject to propose coordination on low efforts, without consideration for the actual effort level proposed. Thus, this is a relatively broad category. A Right Leader, on the other hand, is the first to propose coordinating on the minimum effort case (i.e., all providing effort of 1).[11] We identify 18 First Leaders (29%) and 13 Right Leaders

[10] We initially collected a third category: "Failed Leader" for a subject who called on his group members to decrease efforts but was not listened to/followed. This is a rare event in our study and thus we do not include this variable in our analysis.

[11] We also had both a research assistant from Erasmus University Rotterdam and from Northwestern University independently code the leadership variables. Using these classifications, we find

(21%) among the 63 subjects (21 groups) in the Chat/Observability treatment (11 subjects are both a Right Leader and a First Leader).

We start by providing a breakdown of the social preferences of the subjects we identified as leaders. Panel 1 of Figure 5 shows the distribution of social preference types in the sample of First Leaders and in the sample of Non-First Leaders. Panel 2 shows the distribution for Right Leaders and Non-Right Leaders.

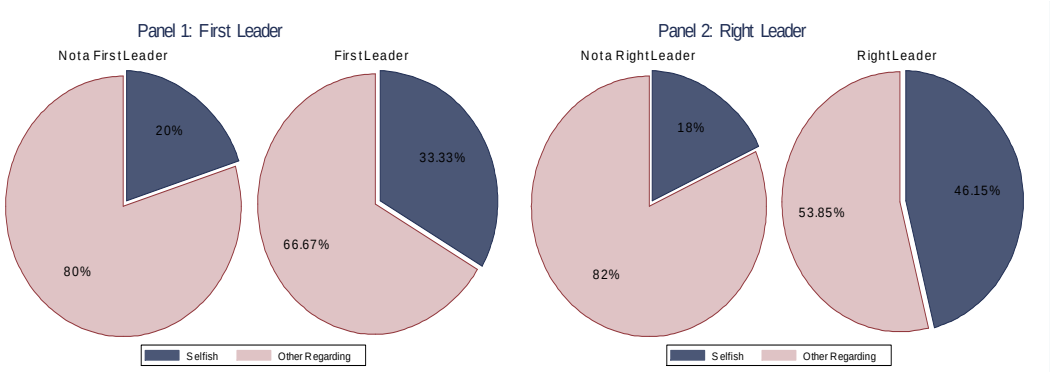


Figure 5: Social preferences of leaders.

For First Leaders and Right Leaders, we observe that Selfish individuals are more likely to be both of these types of leaders. Using a chi–squared test we do not find the difference to be significant for First Leaders (p=0.26) while it is significant at the 5% level for Right Leaders (p=0.03). Thus we find that, indeed, social preferences are linked to the emergence of a coordination leader.

Given that we control for the effect of social preferences that runs through leadership in suggesting low efforts, which is specific to the Chat/Observability treatment, it is still true that low efforts are related to group members' social preferences similarly as in the other treatments. Table 6 reports the results of a random effects model for the Chat/Observability treatment. Column 1 shows a regression without considering leader emergence, analogous to the one in Table 5 for the No Chat/Observability treatment. In column 2 we add as a control whether a Right Leader has emerged (i.e., a dummy variable that takes on a value of one once a Right Leader emerged in the given group) and whether the subject herself is a Right Leader (i.e., a time-invariant dummy variable that takes on value of one for all subjects who are classified as Right Leader). We only consider the variable Right Leader because First Leader is not found to be related to social preferences.[12] Notice that the coefficients of own social preference as well as group members' social preferences are highly significant and larger in magnitude once controlling for leadership in this way. This means that

similar results in our following analysis. The instructions given to the RAs are provided in the appendix.

[12] Our results are robust to including controls for First Leader emergence and type as well, though coefficients on social preferences become smaller in magnitude.

after controlling for the effect of social preferences running through leadership emergence, social preferences lead to significantly lower group efforts. This is consistent with our hypothesis of Other-Regarding individuals reducing efforts. The effect is slightly larger in magnitude than in the No Chat/Observability treatment. We find that a Selfish subject puts in 2 units effort more per period than an Other-Regarding subject. Furthermore, the presence of an additional Selfish group member increases a subject's own effort by 2 units per period.

	(1) Effort	(2) Effort	(3) Effort
Period	-0.133***	-0.0725***	-0.0728***
	(0.0276)	(0.0250)	(0.0245)
Selfish	1.069	2.054***	2.797***
	(1.596)	(0.737)	(0.687)
# Other Selfish	1.060	2.067***	2.864***
	(1.581)	(0.694)	(0.600)
Right Leader Exists		-5.709***	-3.661***
		(0.637)	(0.423)
Right Leader		0.0784	0.107
		(0.350)	(0.338)
RLeader*Selfish			-2.729***
			(0.678)
RLeader*#OthSelf			-2.800***
			(0.562)
Constant	6.628***	7.353***	6.911***
	(1.471)	(0.741)	(0.789)
Observations	1827	1827	1827
R^2-within/between	0.1000/0.0379	0.2117/0.7465	0.2184/0.7848

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6: Effect of social preferences on individual effort controling for leadership (Chat/Observability treatment).

Column 3 investigates further into the timing of the effect of social preferences. We include interactions of social preference measures and the emergence of a leader. We

find that social preferences depress efforts before a Right Leader emerges in a group. Once a leader emerges there is no difference between Selfish and Other-Regarding choices. Selfish are thus no more likely to deviate from a collusive outcome. Finally, note that the coefficient of Right Leader is insignificant. Thus, Right Leaders do not lead also by good example, i.e. putting in lower effort but only through cheap talk.

We conclude that social preferences are an important determinant of group effort also in the Chat/Observability treatment, though in a more nuanced way. On the one hand, subjects can use communication to coordinate the group on a collusive outcome. Such a "leader" tends to be a Selfish individual. This explains why the presence of one Selfish individual reduces efforts in the Chat/Observability treatment. On the other hand, controlling for the relation of leadership and social preferences, Other-Regarding subjects have a tendency to put in lower effort than their Selfish counterparts, exactly as in the non-communication treatments, suggesting these individuals internalize the externality their effort inflicts on their group members. Finally, from a principal's perspective our results suggest that in a work environment where communication is possible a heterogeneous social-preference group leads to the lowest work effort.

4.5 Propensity to "Collude"

Thus, far we have been exploring the relationship between social preferences and depressed efforts. In this section, we explicitly consider the most extreme version of depressed efforts: "collusion." While we are naturally not able to observe our subjects' strategies directly, we take an indirect approach and measure the frequency of "collusive outcomes," outcomes that are consistent with coordination on minimum efforts: all three players coordinate on efforts of 1 (i.e., efforts of $(1, 1, 1)$) for the last 3 periods of play. We additionally include as "collusive outcome" the setting where all three players coordinate on the outcome of two players choosing effort of 1 while a third player chooses maximal (payoff) effort of 12, and then the players alternate the player that gets the maximal payoff. As it might be expected, this latter form of collusion is only witnessed in the Chat/Observability treatment where subjects were allowed to coordinate via chat.

Table 7 reports the proportion of groups achieving the "collusive outcome" in the Chat/Observability treatment. Here, we partition groups by the number of Selfish members, which we have observations for groups with $0, 1$, or 2 Selfish members. Similar to our results on efforts in Section 4.4, when chat is available, groups with 1 Selfish member are more likely to exhibit the collusive outcome than groups with no Selfish members. When we expand the definition of "collusion" to include the case of the group cycling efforts of $(1, 1, 12)$ across players, we again find groups with 1 Selfish member are more successful at achieving the collusive outcome than groups with no Selfish members. This further corroborates our result that social preferences seem to matter in nuanced ways when communication is possible: Selfish individuals

play an important role in facilitating coordination on the collusive outcome.

# Selfish group members	Propensity to "collude" on $(1,1,1)$	Propensity to "collude" on $(1,1,1)$ or alternating $(1,1,12)$
0 (7 groups)	43%	57%
1 (13 groups)	77%	92%
2 (1 group)	0%	0%

Table 7: Propensity to "collude" by # of Selfish in the Chat/Observability treatment.

For the No Chat/Observability treatment, coordinating on a "collusive outcome" was more difficult, since subjects were not able to chat. As shown in Table 8, we find for this setting that 1 out of 21 (a mere 5%) ends up with minimum efforts and only if the group has no Selfish members. If we expand the definition of "collusive outcome" to include two subjective cases of collusion (we report their behavior in the appendix), then we find one additional group with no Selfish members and one additional group with 1 Selfish member successfully "collude." These results lend some support to our prediction that groups with only Other-Regarding are most likely to successfully collude, though because of the relatively rare occurrence of collusive outcomes, these have to be taken with some caution. Generally, it seems that collusion is not a main driver of behavior in this treatment and results seem more consistent with the predictions of the one-shot game.

# Selfish group members	Propensity to "collude" on $(1,1,1)$	Propensity to "collude" (self-classification)
0 (14 groups)	7%	14%
1 (5 groups)	0%	20%
2 (2 group)	0%	0%

Table 8: Propensity to "collude" by # of Selfish in the No Chat/Observability treatment.

One might object at this point that individuals we categorize as Selfish are the ones that understand the game and optimal strategy better than individuals we categorize as Other-Regarding. Thus, naturally they will be the ones suggesting non-competitive efforts, not because of their social preferences, but because of their better understanding of the game. While this is indeed a possibility, this rationale alone does not explain that, when controlling for leadership, Other-Regarding subjects put in lower efforts on average than Selfish ones. This is especially true for the No Chat/Observability treatment: without communication, we might expect that a subject understanding the game better would try to lead by example in order to

induce the other group members to follow his or her lead. In order to investigate further, in the No Chat/Observability treatment, we categorize subjects as attempting to be leaders when expending an effort less than four given that in the round before his or her group members expended efforts larger than 9. We do not systematically observe Selfish subjects leading "by example" with reduced efforts to induce the optimal strategy to their group members. If anything, we observe Other-Regarding types "trying out" low efforts; however, we do not find a statistical difference in the two distributions (Fisher's exact test, p-value = 0.67). Furthermore, we do not find that subjects with a background in Economics or Business are more likely to be Right Leaders (Fisher's exact test, p-value = 0.67). To more rigorously alleviate this and similar concerns, we conduct one more treatment, which is designed to "turn off" social preferences. Of course, this is very hard with human subject interaction. Nonetheless, our final treatment, which we present in the next section, attempts to approximate just such a procedure, by matching humans with computer simulated subjects.

4.6 Robot Treatment

This treatment is similar to the No Chat/Observability treatment in the sense that subjects cannot communicate but get to observe the efforts and payoffs of their group members after each period. The crucial difference is that in stage 2, instead of randomly pairing subjects to other subjects we paired them to two simulated subjects we call robots.[13] In particular, we programmed 42 "robot" subjects who react to past effort decisions by approximating what real subjects did in the No Chat/Observability treatment. Specifically, each "robot" chooses current period effort based on last period's own effort and effort choices of the other two subjects in the same way the real subject did on which it is based on. Critical in this treatment is that it is no longer the case a subject's effort choices impose a negative externality on other players, as the robots receive no payoffs. Thus the fundamental difference between the No Chat/Observability and the Robot treatment is that the latter attempts to "turn off" subjects' social preferences since their actions no longer affect any other player. Note, however, that social preferences are not completely absent, as the *robots'* choices simulate decisions by participants whose social preferences did matter. Thus, subject's decisions can reflect beliefs about the past subjects' social preferences. This in fact is helpful for us, as it allows us to distinguish an alternative hypothesis: "Selfish" subjects differ in their beliefs about their group members' (re-)actions from "Other-regarding" subjects. If this were the case, we should still see a difference between Selfish and Other-regarding effort choices in this treatment. Differences in effort should vanish in this treatment, however, if beliefs about other players' social preferences do not play a role in depressing own effort choices. Furthermore,

[13] We provide additional description of this treatment, as well as analysis on the efficacy of the robots in our appendix.

other potential confounds such as skill differences or differences in patience between "Selfish" and "Other-Regarding" are also not "turned off" by this treatment, allowing us further to test the appropriateness of our initial categorization.

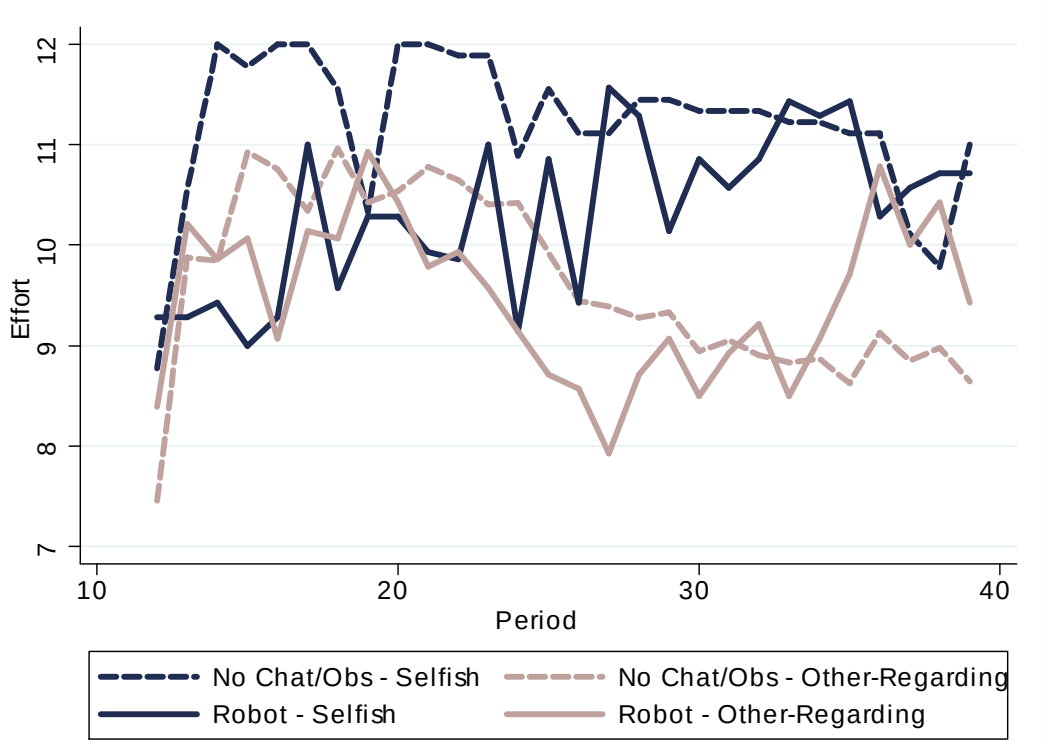


Figure 6: Comparing efforts between Selfish and Other-Regarding types over time.

We first compare subject behavior for the No Chat/Observability treatment and the Robot treatment graphically. Figure 6 depicts the effort profiles over the 29 periods of play by treatment for Selfish and Other-Regarding individuals. We find that in the first half of the relative performance stage (16 periods from periods 12 to 27) the effort of Selfish and Other-Regarding subjects in the Robot treatment is not statistically different (t-test, p-value 0.2122), supporting the validity of our categorization. There is some effort divergence in the intermediate term though—however, by the end of the relative performance stage, efforts of different social types converge back to similar effort levels. In fact, in the last 5 rounds a t-test cannot reject equality of efforts (p-value 0.1578). Interestingly, efforts of all social preference types in the Robot treatment converge towards the efforts of Selfish subjects in the No Chat/Observability treatment. For the last 5 periods a t-test cannot reject equality of efforts of any social preference type in the Robot Treatment compared to Selfish in the No Chat/Observability treatment (i.e., Selfish in the No Chat/Observability treatment vs. Selfish in the Robot treatment, p-value .7315; Selfish in the No Chat/Observability treatment vs. Other-Regarding in the Robot treatment, p-value .1578). When we compare Other-

Regarding individuals' efforts across treatments we do find a significant difference (Other-Regarding in the No Chat/Observability treatment vs. Other-Regarding in the Robot treatment, p-value .0016). If we include the final ten periods of effort though, there is a statistical difference in effort between Other-Regarding and Selfish players (p-value .002) in the Robot treatment, as suggested by the chart.

Thus, while predictions are borne out in the first half, we find only partial evidence of equal behavior between Selfish and Other-Regarding players for the entire last half of the relative performance game in the Robot treatment. Perhaps, subjects forget that they are playing "robot" subjects and began behaving as if they are playing "real" subjects. We did attempt to minimize this possibility by reminding subjects on each effort-entry screen that their effort choice will not affect the payoffs of any participants. Unfortunately, we cannot rule out that subjects disregarded this message after 15 periods. It nonetheless does seem these results suggest beliefs are not driving the difference in choices for different types of players: beliefs should loom largest in creating differences at the beginning of the relative-performance game before they converge based on experience. However, we observe just the opposite pattern.

If instead analyzing individual rather then average aggregate effort choices, which may mask individual behavior, we find a similar pattern of similar effort choices across social preference types. Table 9 reports the results of regressing individual effort on own and group members' social preference types for the No Chat/Observability and the Robot treatment. The coefficient estimate for Selfish is half the value as in the No Chat/Observability treatment and is no longer significant, though we do note the sample size is smaller.

	No Chat/Observability Effort		Robot Effort	
Period	-0.0538*	(0.0294)	0.0168	(0.0285)
Selfish	1.478***	(0.401)	0.824	(0.813)
# Other Selfish	0.569	(0.412)	-0.280	(0.996)
Constant	10.85***	(0.502)	9.152***	(0.685)
Observations	1827		609	
R^2 - within/between	0.032/0.095		0.003/0.049	

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 9: Effect of social preferences on individual effort treatment 2 vs. treatment 4.

Overall, we believe the Robot treatment provides further evidence that social preferences (and not beliefs) matter in creating and sustaining non-competitive efforts.

5 Conclusion

We explored how a relatively new dimension of worker heterogeneity affects the performance of workers subject to relative performance pay. In particular, we found that a basic form of social preferences, the degree of other-regardingness, is substantially linked to reduced effort choices, but in a nuanced manner. First, players categorized as Selfish are more likely to coordinate their group members to minimal efforts, when communication is available. Second, controlling for the existence and emergence of such leaders, players categorized as other-regarding exert lower levels of effort—an average of over 50% lower effort. Thus, when communication is available, a group that is heterogenous in social preferences can most successfully create and sustain very low efforts over those groups with no Selfish members. However, when communication is not available, groups of Other-Regarding players produce the lowest levels of effort. Since we find little evidence of collusive outcomes, this is again consistent with the theory that Other-Regarding players internalize their efforts' negative externality imposed on other players' payoffs.

To further validate our findings, we also attempted to "switch off" subjects' social preferences through our Robot treatment. For this experiment, we simulated the responses of human subjects via machine, thus removing a player's negative externality. By the end of the treatment, Other-Regarding subjects acted like Selfish subjects. This provided further evidence that other-regarding individuals are indeed depressing their efforts as they internalize the negative externality of higher effort.

Our findings have important policy implications for personnel policy. In organizations with more other-regarding workers (e.g., non profit firms or firms engaged in corporate social responsibility), relative performance incentives are likely to not be as effective as in other organizations. For firms already using relative incentive pay, screening workers according to their social preferences could improve performance. Human resource departments often provide potential workers with psychological based exams. These could readily incorporate explicit measures of other-regardingness. Similarly, information obtained from resumes, such as a potential worker's involvement in philanthropic activities, could shed light on a worker's degree of other-regardingness.

When workers are closely engaged so that communication flows freely and output is easily observed, relative performance schemes are also more likely to encourage noncompetitive behavior. In this setting, it is the Selfish worker that is likely to instigate a particularly low effort regime with other-regarding workers.

We note that we did not consider the case where workers might value their firm's payoff. Thus, our results can be seen as applying to settings where ownership is dispersed or the worker is removed from the top of the hierarchy. Finally, our measure of leadership is endogenous to the effort exerted in each group. It is an interesting challenge to design an experiment in which leadership varies with incentives and analyze how it relates to social preferences.

Although our setting only allows for the possibility of valuing *negative* externalities, to the extent workers also value their *positive* externalities, other-regarding preferences could mitigate the free rider problem amongst teams. That is, a team of workers with Other-Regarding preferences that receive a share of the common output are more likely to provide higher outputs, as they further value their effort's positive effects on their team members. We leave these topics for future research.

References

[1] Andreoni, J. and J. Miller (2002) "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica*, 70 (2), 737-753.

[2] Bandiera, O., Barankay, I and I. Rasul (2005) "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *The Quarterly Journal of Economics*, 120 (3), 917-962.

[3] Bowles, S., and S. Polania-Reyes (2012) "Economic Incentives and Social Preferences: Substitutes or Complements?" *Journal of Economic Literature*, 50 (2), 368-425.

[4] 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.

[5] Dal Bo, P., (2005) "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games," *The American Economic Review*, 5, 1591-1604.

[6] Dal Bo, P., and G. R. Fréchette (2011) "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence," *The American Economic Review* 101 (1), 411-429.

[7] DellaVigna, S. (2009) "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47 (2), 315-72.

[8] Dreber, A., D. Fudenberg and D. G. Rand (2011) "Who Cooperates in Repeated Games?" Working Paper.

[9] Erkal, N., L. Gangadharan, and N. Nikiforakis (2011) "Relative Earnings and Giving in a Real-Effort Experiment," *American Economic Review*, 101 (3), 3330-3348.

[10] Fehr, E. and U. Fischbacher (2002) "Why Social Preferences Matter—The Impact of Non-Selfish Motives on Competition, Cooperation and Incentives," *The Economic Journal,* 112, C1-C33.

[11] Fehr, E., and Schmidt, K. M. (1999). "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics*, 114(3), 817-868.

[12] Fischbacher, U. (2007) "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics,* 10 (2), 171-178.

[13] Fisman, R., S. Kariv, and D. Markovits (2007) "Individual Preferences for Giving," *American Economic Review*, 97 (5), 1858–1876.

[14] Friedman, J. (1971): "A Noncooperative Equilibrium for Supergames," *Review of Economic Studies*, 38, 1-12.

[15] Fudenberg, D., and E. Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or Incomplete Information," *Econometrica*, 54, 533-554.

[16] Fudenberg, D., Rand, D. G., and A. Dreber (2012). "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World," *American Economic Review*, 102 (2), 720-49.

[17] Gaechter, S, Nosenzo, D, Renner, E. and M. Sefton (2012). "Who Makes a Good Leader? Cooperativeness, Optimism and Leading-by-Example," *Economic Inquiry*, 50(4), 867–879.

[18] Hermalin, B. (2012). "Leadership and Corporate Culture," *Handbook of Organizational Economic*s (R. Gibbons and J. Roberts, eds.), Princeton University Press.

[19] Kocher, M. G., G. Pogrebna and M. Sutter, (2013) "Other-Regarding Preferences and Management Styles," *Journal of Economic Behavior & Organization*, 88, 109-132.

[20] Kreps, D. M. (1986), "Corporate Culture and Economic Theory," in M. Tsuchiya, ed., Technology, Innovation,and Business Strategy, Tokyo: Nippon Keizai Shumbunsha Press.

[21] Ledyard, J. (1994) "Public Goods: a Survey of Experimental Research," J. Kagel, A. Roth (Eds.), Handbook of Experimental Economics, Princeton University Press, Princeton.

[22] Palfrey, T. and H. Rosenthal (1994) "Repeated Play, Cooperation and Coordination: An Experimental Study," *Review of Economic Studies*, 61 (3), 545-565.

[23] Seelya, B., J. Van Huyck and R. Battalio (2007) "Credible Assignments can Improve Efficiency in Laboratory Public Goods Games," *Journal of Public Economics*, 89 (8), 1437–1455.

6 Appendix

6.1 Broader Social Preference Classifications

In this section we explore two alternative social preference categorizations. In particular we will use dictator menus 1-11 to classify subjects into different types depending on their choices. First we follow Andreoni and Miller (2002) and use menus 1-9 to broaden the category of Other-Regarding into subjects who tend to give more when the prize of giving increases (we call them Complements) and subjects which tend to react by giving less (we call these individuals Substitutes). The idea is that the former represents the motive of fairness, while the latter represents the motive of efficiency. Thus, menus 1-9 measure whether a subject values fairness or efficiency under favorable inequality. In a second analysis, we use dictator menus 10-11 to see whether subjects have an aversion to unfavorable inequality (i.e., unfavorable in terms of their own payoff relative to others). In the following, we provide more detail on the these categorization procedures, as well as some additional analysis using these expanded categories.

Complements vs. Substitutes

We use decision menus 1 to 9 (see Table 3 for an overview) to classify participants as "Selfish", "Complement" (Rawlsian) or "Substitute" (Utilitarian). To do so, we first compute the relative giving rates of an archetypal Selfish, Utilitarian and Rawlsian individual according to the preferences in Table 10. We denote player i's monetary payoff as π_i and the total number of players n. Thus, an archetypal Selfish type, is only interested in her own monetary payoff. In contrast, an archetypal Rawlsian player only values the minimal monetary payoff of all of her group member's payoffs. Finally, an archetypical Substitute simply maximizes her group's total monetary payoff.

Social Preference Types	**Utility**
Selfish	π_i
Complement (Rawlsian)	$\min\{\pi_i, \pi_j\}$
Substitute (Utilitarian)	$\pi_i + \Sigma_{j \neq i} \pi_j$

Table 10: Overview of social preference types.

To categorize subjects, we then measure the Euclidian distance from each of the participants' decisions to each of these archetypes' decisions. We compute such distance for each choice and then we compare the average distance across periods to each archetype's decision. We classify subjects as the archetype whose decision is closest to the subject's decision.[14] For treatments 1-3 we find that, for our subject

[14]Since we only use relative giving rates between the other two group members, our classification

population, 20% are Selfish, 63% are Complements and 17% are Substitutes. Consistent with Andreoni and Miller (2002), hereafter AM, we find that 20% of subjects are (perfectly) Selfish, whereas AM find that 23% of subjects are perfectly Selfish. 6.5% of our subject are classified as perfect Substitutes, while AM find 6.2%. In contrast to AM we only classify one subject as a perfect Complement, while they find 14.2% are perfect Complements. Different from AM, we do not have any "weak" Selfish types, as we categorize all Other-Regarding subjects (i.e., subjects that give to others) as either Complement or Substitute types.

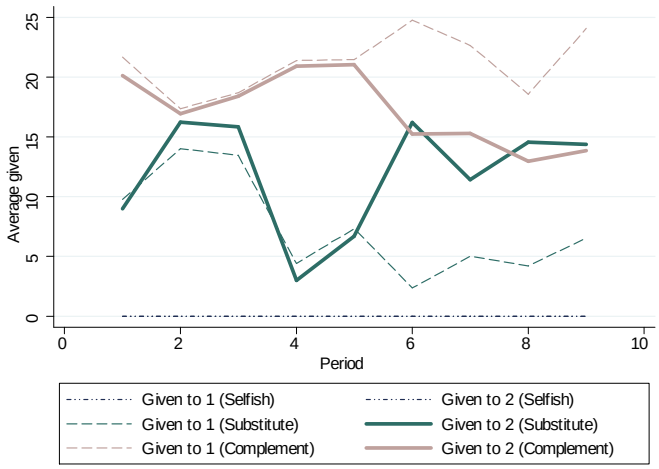


Figure 7: Giving rates by social preference types.

Figure 7 illustrates giving behavior under our broader categorization of social preferences types. We see that Selfish types, by definition, never give anything to their group members. In contrast, Other-Regarding types give positive amounts, on average, for every price vector. When the price of giving increases, Substitutes typically react by decreasing their giving rate, while Complements do the opposite. This is most easily seen for periods 6 to 9 where the price of giving to individual 2 is always lower than the price of giving to individual 1 as can be seen in Table 3 . Thus, as archetypal types would do, Complements react by allocating more to individual 1 while Substitutes react by allocating more to individual 2.

Table 11 is analogous to Table 4 and shows the results of a regression of average group effort on the number of Complements and Substitutes in a group. Both Complement and Substitute group members reduce group effort relative to Selfish group members in the No Chat/Observability treatment and No Chat/No Observability treatment by approximately .8 units. In the Chat/Observability treatment, a linear regression again does not yield significant results; this is to be expected given the

does not account for the intensity of social preferences. We can control for intensity separately by including the overall giving rate of a subject.

discussion in the main text of the confound of leadership. We will again consider the effect social preferences on leadership and explore whether it differs by Complements and Substitutes.

Table 12 is analogous to Table 5. Here, we present the results of a random effect panel regression model for the No Chat/Observability treatment that considers the effect of own and others' social preference type on individual effort. The results from our main analysis suggesting that Other-Regarding members exhibit lower efforts relative to more Selfish group members holds also when we consider our subcategories of Other-Regarding: Complements and Substitutes. Complements as well as Substitutes exhibit lower effort than their Selfish counterparts. In fact, we cannot reject the null hypothesis that Complements and Substitutes depress effort by the same magnitude (p-value 0.7102). Furthermore, we see that most of the effort reduction is driven by their own preference type (i.e., around 1.5 units) while the coefficients on the other group members' social preference types are of the same sign, but much smaller in magnitude and insignificant.

Finally, we turn to disentangling the effect of social preferences on leadership and individual effort provision in the Chat/Observability treatment. Figure 8 reports the distribution of social preferences among Non-Right Leaders and Right Leaders as defined in Section 4.4. As before, Selfish are significantly more likely to become Right Leaders (chi-squared test, p-value=0.034). The opposite is true for Complements (p-value=0.031). Finally, for Substitutes we do not find a significant effect on leadership propensity (p-value=0.678).

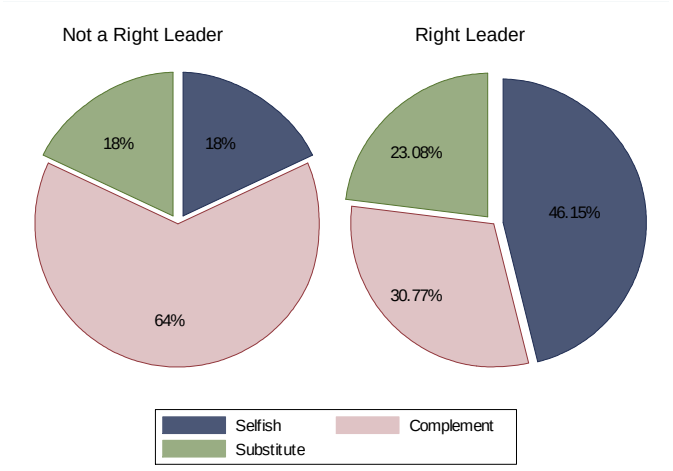


Figure 8: The distribution of social preferences among Right Leaders and non-Right Leaders.

In order to disentangle the effect of social preferences on the propensity to initiate coordination from the effect on effort choice, we run a random effect panel regression analogous to Table 6 for the Chat/Observability treatment.

	Chat/Obs Avg Effort (Grp/Sess)	No Chat/Obs Avg Effort (Grp/Sess)	No Chat/No Obs Avg Effort (Grp/Sess)
# Complements	-0.593	-0.873**	-0.919***
	(1.582)	(0.389)	(0.154)
# Substitutes	-1.742	-0.856	-0.604*
	(2.009)	(0.685)	(0.265)
Constant	5.952	12.06***	11.02***
	(4.017)	(0.942)	(0.388)
Observations	21	21	7
Adjusted R^2	-0.036	0.030	0.637

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 11: Group composition and average group effort.

	(1) Effort		(2) Effort	
Period	-0.0538*	(0.0294)	-0.0538*	(0.0294)
Selfish	1.478***	(0.401)		
# Other Selfish	0.569	(0.412)		
Complement			-1.410***	(0.386)
Substitute			-1.714**	(0.854)
# Other Substitutes			-0.427	(0.669)
# Other Complements			-0.604	(0.411)
Constant	10.85***	(0.502)	13.46***	(1.188)
Observations	1827		1827	
R^2 within/between	0.0322/0.0954		0.0322/0.0994	

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 12: Effect of own and others social preferences on own effort (No Chat/Observability).

	(1) Effort		(2) Effort	
Period	-0.133***	(0.0276)	-0.0727***	(0.0249)
Complement	-0.458	(0.901)	-1.884**	(0.760)
Substitute	-0.997	(1.301)	-2.245**	(0.891)
# Other Complements			-1.880***	(0.723)
# Other Substitutes			-2.348***	(0.847)
Right Leader Exists			-5.690***	(0.636)
Right Leader			0.0990	(0.353)
Constant	7.839***	(1.265)	13.36***	(1.844)
Observations	1827		1827	
R^2-within/between	.100/.012		.212/.751	

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 13: Leadership and Social Preferences

We report these results in Table 13. The first column does not control for the emergence of a Right Leader and whether or not an individual turns out to be a Right Leader. The coefficients on the social preferences are insignificant, though they do indicate an effort reduction by Complements and Substitutes. Controlling for the emergence of a Right Leader and controlling for being a Right Leader increases the magnitude of both coefficients by approximately 1 unit, both statistically significant at the 1% level. Also, the social preference types of the other group members matter. Having Complement or Substitute group members decreases own effort by about 2 units as well. Overall we conclude that there is a differences in the propensity to initiate coordination by Substitutes and Complements; however, effort choice is relatively similar.

Unfavorable Inequality

In a second classification, we use dictator menus 10-11 to differentiate subjects by their propensity to reduce their own payoff in order to reduce unfavorable inequality. Subjects were given an allocation vector and were able to choose an exchange rate between zero and two which translated tokens into payoffs for all group members. Thus, an exchange rate of 2 maximizes aggregate output, while an exchange rate of zero minimizes inequality. Table 14 summarizes the two menus and the decisions of subjects in Treatments 1-3. Overall, many subjects were willing to reduce their own payoff at least once to reduce inequality. Furthermore, the fraction of subjects who destroy some of their payoff goes up and the average exchange rate goes down when the allocation becomes more unfavorable. For our analysis, we denote a subject as

Jealous when he or she chose an exchange rate of less than two in any of the two menus. In treatments 1-3, 64% of subjects are classified as Jealous.

Menu (Allocation)	Mean	Percent where rate=2
10 (20,40,40)	1.796	77%
11 (2,49,49)	1.322	58%

Table 14: Average exchange rate chosen in menu 10 and 11.

Using the category of Selfish/Other-Regarding as well as Jealous/Non-Jealous we construct 4 new social preference categories:[15]

- Disinterested: not Jealous and Selfish (10%)
- Benevolent: not Jealous and Other-Regarding (26%)
- Spiteful: Jealous and Selfish (10%)
- Inequity Averse: Jealous and Other-Regarding (54%)

Table 15 reports the results of an OLS regression of average group effort on the number of Benevolent, Spiteful and Inequity Averse with Disinterested as the omitted category analogous to Table 4. In the Chat/Observability treatment, we do not find any significant effect of these social preferences types. In the No Chat/Observability treatment we find that Spiteful group members are responsible for highest group effort. On average, an additional Spiteful subject increases group effort by 1.5 units. We do not find significant differences for all of other social preference types. In contrast, in the No Chat/No Observability treatment, Disinterested group members lead to highest group efforts. Given the low number of observations for this treatment, however, this finding needs to be treated with caution.

Finally, we explore whether this extended categorization yields new insights on the propensity to initiate coordination when communication is possible. Figure 9 reports the distribution of social preferences for Non-Right Leaders (left panel) and Right Leaders (right panel) for the Chat/Observability treatment. As can be seen, Spiteful individuals have the highest propensity of becoming a Right Leader. While there are not enough observations for the Disinterested to make any meaningful statement—only 2 out of the 63 subjects in this treatment are Disinterested—we see that both types of Other-Regarding subjects have a lower propensity of becoming a Right Leader. This is especially so for Inequity Averse subjects. Thus, relative to an Inequity Averse, a Spiteful subject is 3.3 times more likely to emerge as a Right Leader.

[15] Population proportions are for Treatments 1-3

	(1) Avg Effort (Grp/Sess)	(2) Avg Effort (Grp/Sess)	(3) Avg Effort (Grp/Sess)
# Spiteful	-4.822	1.488***	-3.199**
	(3.274)	(0.421)	(0.764)
# Inequity Averse	-4.766	-0.807	-1.777**
	(2.945)	(0.489)	(0.324)
# Benevolent	-4.614	-0.761	-2.276***
	(3.101)	(0.512)	(0.337)
Constant	17.73*	11.79***	13.97***
	(8.834)	(0.934)	(0.764)
Observations	21	21	7
Adjusted R^2	0.087	0.017	0.844

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 15: Group Effort and Inequality Aversion (omitted category: Desinterested)

Finally, controlling for the emergence of a leader, as in Table 6, we can separate the relation of social preferences and leadership emergence from general effort choices. Table 16 summarizes the results. Note that we pooled Disinterested with Spiteful subjects due to the lack of observations for Disinterested in this treatment (i.e., only 2 subjects out of 63). Overall the results mirror our results from the main section. Inequity Averse subjects behave similar to Benevolent ones, though we only get significance for the Inequity Averse. This could be driven by the lower numbers of Benevolent subjects.

Conclusion

To summarize, the main results of our two alternative categorizations are:

- Both Substitutes and Complements reduce effort relative to Selfish types. We do not find significant differences in Substitutes' and Complements' effort choices.
- When communication is possible, Complements are less likely to initiate cooperation through chat, while this is not the case for Substitutes.
- There is (weak) evidence that especially Spiteful subjects lead to high group effort provision. There is not much difference between Benevolent and Inequity Averse subjects in terms of their effort choices.

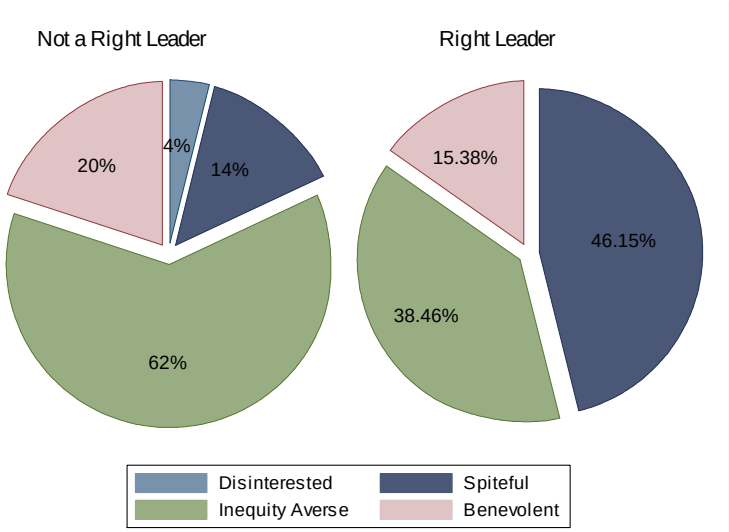


Figure 9: Distribution of social preferences among non-Right Leaders and Right Leaders under extended categorization two.

- Spiteful subjects are most likely to become leaders, while Inequity Averse subjects are least likely.
- Overall, a simple categorization into Selfish and Other-Regarding explains most of the variation in the data.

6.2 Appendix B - Subjectively Categorized Collusion

Figure 10 shows the effort choices of groups S4G1, S5G3 and S5G5 that we categorize as ultimately "colluding." Group S5G3 achieves the collusive outcome in the strictest sense—all group members choose minimal effort of 1 in the final periods. The other two groups we subjectively categorize as coordinating on low efforts.

6.3 Robot Details

For this treatment, we needed to develop a program that would create a similar experience for a subject playing a computer to if she was instead playing actual subjects. By experience we mean if the human subject played certain strategies, she would obtain similar results whether she played actual subjects or the computer. To accomplish this, we used actual subject behavior from the No Chat/Observability treatment to determine how the computer would respond to a subject's effort choices in the Robot treatment. In particular, we had the computer choose effort each period based on the composition of efforts of players in the last period. Although in practice subjects could use an entire history of play to determine their action for the current

	(1) Effort		(2) Effort	
Period	-0.133***	(0.0276)	-0.0766***	(0.0255)
Inequity Averse	-0.698	(0.910)	-0.682**	(0.333)
Benevolent	-0.276	(1.523)	-0.698	(0.601)
# other Inequity Averse			-2.831*	(1.679)
# other Benevolent			-2.223	(1.721)
Right Leader Exists			-5.316***	(0.633)
Right Leader			0.149	(0.403)
Constant	7.839***	(1.265)	14.61***	(3.338)
Observations	1827		1827	
R^2- within/between	.1/.01		.212/.719	

Standard errors in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 16: Leadership

period, regression analysis shows virtually all of history's effect on current choices is captured in just the last period of play.

Recall each subject can choose efforts between 1 and 12. This provides 12^3, or 1,728 possible effort outcomes for any given period. However, most subjects only faced a small fraction of all these possible outcomes, or what we refer to as "states." Thus, we collapse the 1,728 to 27 possible states by creating a coarse partition of efforts. In particular, we bucket effort into low (1-4 units), medium (5-8 units), or high (9-12 units). In addition, we assume a player does not care about the identity of which player provides a higher effort, should they be different efforts. This reduces the possible "states" to 18. With this coarser partition, at least one player faced each of these possible 18 states in the No Chat/Observability treatment. Our next step is to then build a set of strategies for 63 simulated players, which are based on each of the 63 actual subjects' actions in the No Chat/Observability treatment. For each of the possible "states," we create a transition matrix for each simulated player. The transition matrix contains the simulated player's action for each of the possible 18 "states" they might face. Often a given subject had historically chosen a different action when facing the same "state." In this case, we assign a probability for taking each action based on the historical likelihood of the human subject choosing each action. In the event a subject did not face a given "state" in the No Chat/Observability treatment, we impute the simulated subject's action as the average action of all players that faced such a "state." The 13 (of 63) subjects who faced the smallest number of "states" responded to just 3 "states" and the subject who faced the most "states," reacted to 11 "states" (out of 18). The mean of different "states" faced by a given

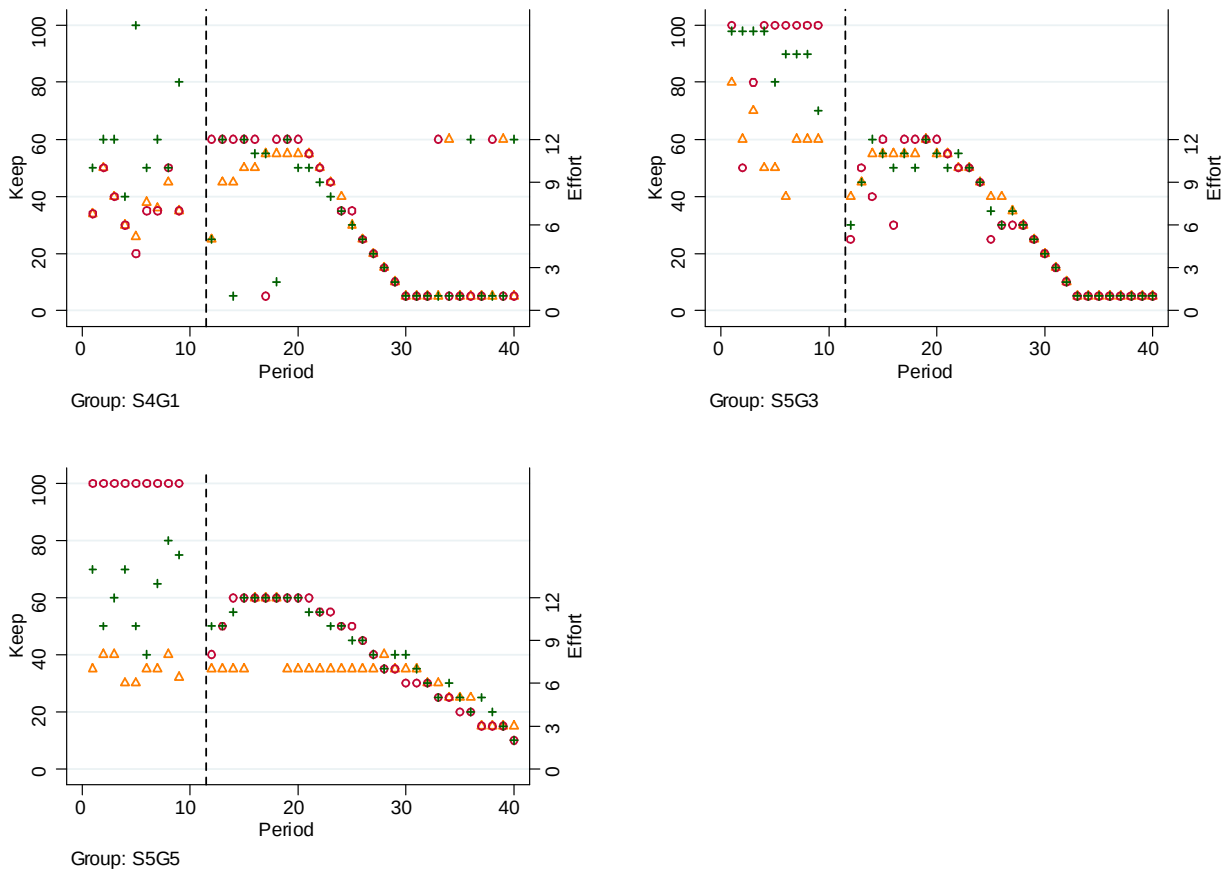


Figure 10: Choices of groups classified as "colluding."

subject was 5.2 and the median was 4. In the end, after imputation, we had created a complete transition matrix that assigned likelihood of each action for each of the 18 "states" for all 63 simulated subjects.

For the robot treatment, when subjects reached the relative performance stage, they were randomly assigned to two simulated subjects (out of the possible 63) that would react to the past period's efforts based on the transition matrix. For the first period, however, the selected simulated subject simply chose the same effort as the corresponding human subject did in the No Chat/Observability treatment for the first period of the relative performance stage.

Before running our experiment, we wanted to make sure the simulated subjects' behavior resembled real subjects. Again, for this treatment, we were attempting to "turn off" social preferences by presenting subjects with the same play experience as when facing real subjects but without generating any negative externality against the payoffs of their opponents. We performed two tests to check for the validity of our

simulated subjects (i.e., robots). First, we matched the simulated subjects into the same group pairings the human subjects experienced. For each of these 21 groups, we then ran 1000 repetitions of each group interacting over 29 periods. Table 17 reports the result of this simulation. A very common outcome for the human subjects was for groups to end with all players choosing high efforts. In fact, four groups all chose maximal effort of 12 in the final period. When these four group pairings are instead played by simulated players, they end up with this maximal outcome 95%, 91%, 71%, and 23% of the time. They all end up in the "state" of (high, high, high) effort (i.e., all players choosing effort above 8), 60-97% of the time. In terms of the extreme outcome of effort depression, colluding on effort choices of (1,1,1), there is only one group of human subjects that achieved this. This one group represents 5% of all human subject groups. The simulated group of these same members ends with (1,1,1) 7% of the time and the "state" (low,low,low) effort roughly 13% of the time. In contrast, this same group ends at highest efforts of (12,12,12) just .6% of the time.

		% of the time in which the robots' finished in:					
Group	Final effort	all 12	all < 4	2:< 4, 1:12	all > 8	all 1	2:> 8 1:≤ 4
S4G1	12,1,1	0.002	0.235	0.181	0.245	0.126	0.124
S4G2	6,12,12	0.083	0.002	0	0.57	0	0.033
S4G3	9,9,12	0.251	0	0	0.871	0	0
S4G4	12,5,12	0.464	0.003	0.002	0.636	0.001	0.029
S4G5	12,12,10	0.751	0	0	0.838	0	0.117
S4G6	12,10,12	0.028	0	0	0.966	0	0
S4G7	12,4,11	0.173	0.004	0.014	0.211	0	0.099
S5G1	10,9,11	0.007	0.005	0.002	0.574	0	0.004
S5G2	12,12,8	0.03	0.044	0.021	0.07	0.013	0.084
S5G3	1,1,1	0.006	0.129	0	0.472	0.071	0.016
S5G4	12,4,12	0	0	0	0	0	0.168
S5G5	2,3,2	0.091	0.25	0.002	0.124	0	0.008
S5G6	12,12,12	0.231	0.001	0.036	0.604	0.001	0.219
S5G7	11,12,5	0.037	0.003	0.003	0.084	0.001	0.088
S6G1	12,12,12	0.952	0	0	0.973	0	0.027
S6G2	7,8,12	0.313	0	0	0.683	0	0
S6G3	12,5,4	0.035	0.009	0.002	0.125	0.003	0.037
S6G4	12,12,1	0.015	0	0.062	0.098	0	0.833
S6G5	12,12,12	0.707	0	0	0.722	0	0.032
S6G6	12,12,12	0.907	0	0	0.971	0	0.029
S6G7	9,9,9	0.013	0	0	0.913	0	0.044

Table 17: Simulations (1000 repetitions of each group)

A second test we conducted was to simply randomly match all simulated subjects into groups of three and then compare the distribution of these group out-

comes to the distribution of actual group outcomes of human subjects in the No Chat/Observability treatment. Table 18 reports these findings. We did this in a series of 100, 1,000, and 10,000 repetitions of group pairings. While again just one group, or 5%, of human subject groups colluded, in our largest samples, we found 1% of simulated groups perfectly colluded (i.e. ended up in (1,1,1) efforts). In terms of maximal effort, whereas 19% of human subject groups ended with choosing (12,12,12), 17% of randomly matched robot groups experienced the same ending. For the common outcome of human subjects finishing in groups with effort choices of (high,high,high) (i.e., effort all higher than 8), human subjects achieved this 43% of the time versus the robot groups did so 49% of the time. Although, frequencies are not identical to the realized draw of 21 human subject groups, we were comforted by these simulations that these robots reasonably resemble human subject behavior.

		Simulations		
Last round effort	% Human	% Robot (100)	% Robot (1000)	% Robot (10000)
all 12	0.19	0.17	0.19	0.19
all ≤ 4	0.10	0.02	0.02	0.02
2: ≤ 4, 1: 12	0.05	0.00	0.01	0.01
all > 8	0.43	0.49	0.53	0.53
all 1	0.05	0.00	0.00	0.01
2: > 8, 1: ≤ 4	0.13	0.10	0.07	0.06

Table 18: Randomly matched groups (simulations)

6.4 Leader Classification Details

Attached file

6.5 Instructions for Subjects

Attached file

Instructions for RA

The Excel sheet has 21 tabs, each one provides data of chat messages for a group of three players. The variables are:

Session: identifies which experimental session the individual participated in, numbering 1 to 3
Group: identifies the group number that the participant was assigned to, numbering 1 to 7 in a given Session
Subject: an indicator for a particular participant number, numbering 1 through 21 for a particular Session
Period: records which period the chat or effort choice took place, ranging from 12 to 41
Effort: effort choice of participants for a given period, ranging from 1 to 12
Chat: records any chat message a player sends for other group members to be read for a given period. The period for Chat is recorded in chronological order. That is, a message coded in period 12.16 was sent before a message coded in period 12.25. However, note, any message recorded as period 13.XX was made after the effort choice for period 12 but BEFORE the effort choice for period 13.

We need you to classify any subjects that behave according to any of the following definitions of leaders. In particular, record in a new Excel sheet, the Session number, Group number and Individual ID of the respective leader (as defined below) and the period that the leadership chat takes place.

First leader is defined as:
"The first person in a group to suggest coordinating and his/her other group members follow the suggestion"

Right leader is defined as:
"The first person to suggest coordinating on efforts of (1,1,1) and his/her other group members follow the suggestion"

Failed leader is defined as:
"The first person to suggest coordinating and his/her other group members do NOT follow his/her suggestion"

What follows is an example. Please note to enter the period as simply the chat period without the decimal places. For example, if the Right Leader suggested to coordinate on effort (1,1,1) in period 12.16, then simply enter period 12.

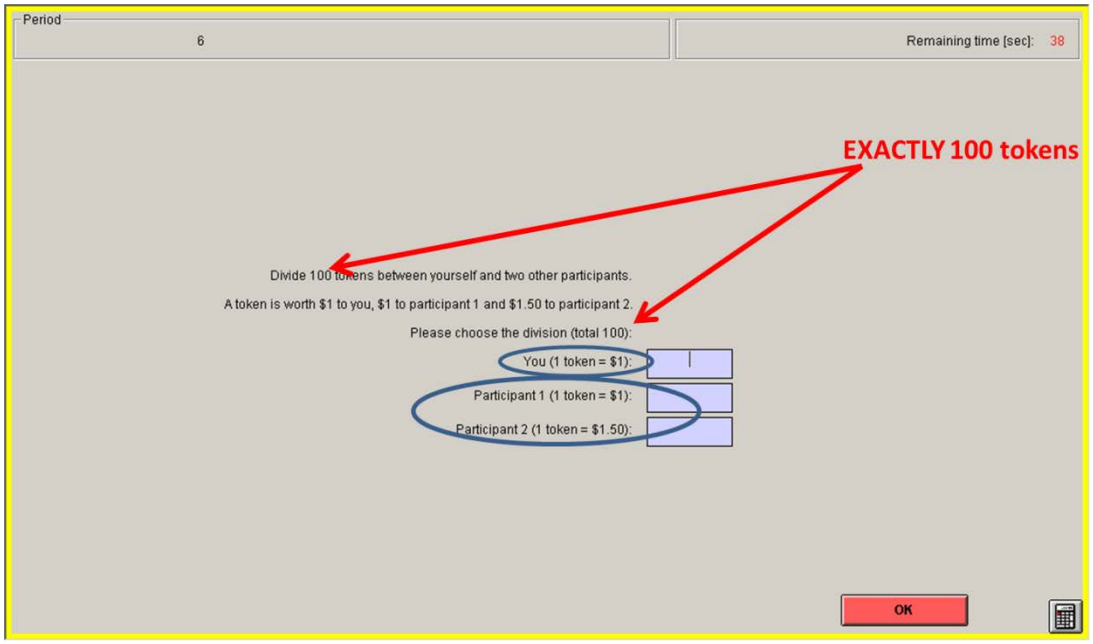
Session	Group	Subject	Period	First Leader	Right Leader	Failed Leader
1	1	2	17	X	X	
2	2	6	23	X		
3	3	16	33			X

Instructions

This is an experiment in the economics of decision-making. If you **follow these instructions carefully** and make good decisions, you might earn a considerable amount of money. The currency we will use throughout the instructions and the experiment is the Berkeley Buck. We will denote it as "$" and the exchange rate is $ 66.6 Berkeley Bucks per US$ dollar.

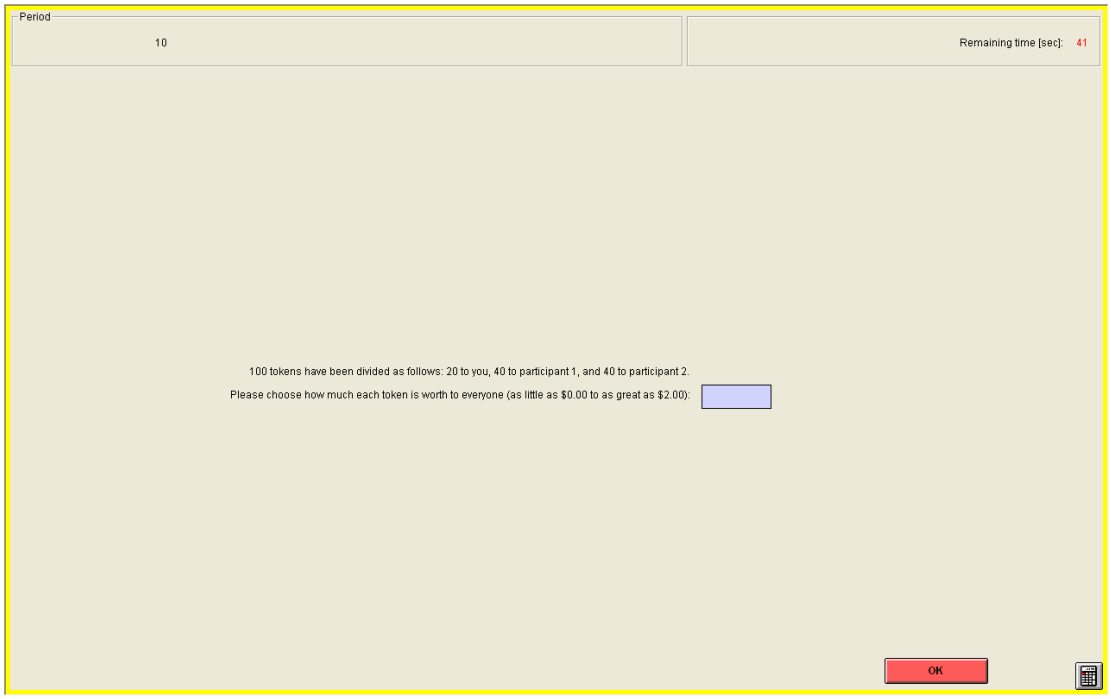
This experiment will occur in three stages today. The **first stage** will consist of dividing sums of money between yourself and two other randomly matched and anonymous participants. On each screen (there will be 11 screens in total for this stage), you will have to divide **exactly** 100 tokens between yourself and the two other participants in your group. The value of each token can vary for group members and for different screens.

Screens 1-9 will be similar to the screen shown in the following Figure:



The value of the token for each person (including you) is displayed just to the left of the input box where you will enter how many tokens each person will receive. For example, for the above screen, 1 token allocated to yourself yields you $1, 1 token designated to your other group member (labeled Participant 1) will yield him/her $1, and 1 token designated to the final group member (labeled Participant 2) will yield him/ her $1.50. You will need to allocate an amount of tokens to each person (including you) so that in **total 100** tokens are allocated. Thus, any one input box could have the number 0 to 100 entered, but all three boxes together must sum to 100.
<STOP READING HERE>

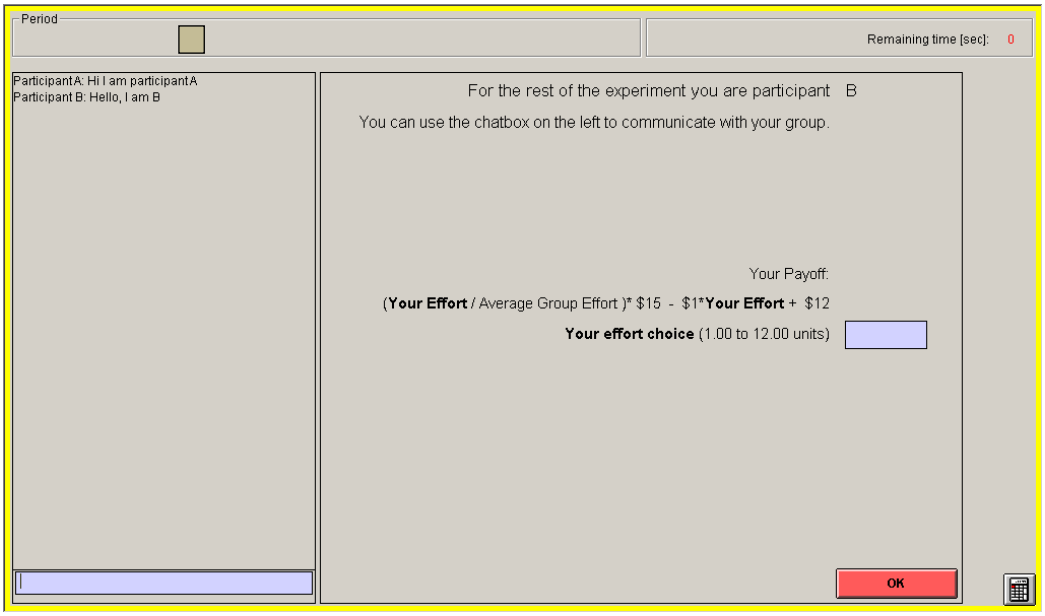
Screens 10-11 will be similar to the screen shown on the following Figure:



Here you are given an allocation of tokens to you and your group members and you have to determine the value of the tokens to each of you. You can choose a value as little as $0.00 to as great as $2.00 per token.

To determine your final payoff, one of your group members' decisions (including yours) will be randomly selected with equal chance in each screen (period). 5 of these 11 selected allocations will be randomly chosen with equal chance. These 5 selected allocations will be used to compute the final payoff for ALL members in your group (including you).
<STOP READING HERE>

In the **second stage** you will be grouped again randomly and anonymously with two other participants. You will remain matched with the same group members for the balance of the experiment. You will be making effort choices over a number of periods, as the following Figure shows:

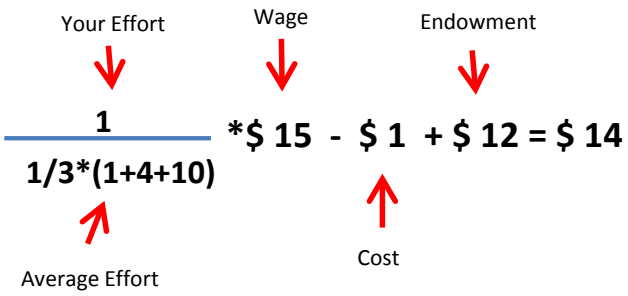


The total number of periods for this stage is unknown to all group members. Instead, there will be a 95% chance you will continue for another period.

In the Figure above, the first sentence in the center of the screen indicates your name (Participant A, B or C) for the rest of the experiment. The second sentence points out that you can chat with the other participants in your group, using the chat box on the left. The remaining information on the screen reminds you how your payoff for each screen will be calculated. Your payoff is calculated as follows:

For each period, each participant begins with a sum of $12 (Berkeley Bucks). You will choose effort between 1 and 12 units, where **each unit of effort costs $1**. After each period, you will be paid a wage of $15 TIMES your chosen effort DIVIDED by the average of your group of 3 participants' effort choices. This means that your effort will be evaluated relative to the average effort of all the participants in your group (yourself included). If your effort is higher than the average, the wage $15 will be multiplied by a number higher than one, and if it is less than average, it will be multiplied by a number lower than one.

For example, if you choose 1 unit of effort and another group member chooses 4 units of effort and the other chooses 10 units of effort, your TOTAL payoff for the period is $ 14, and it is computed as follows:

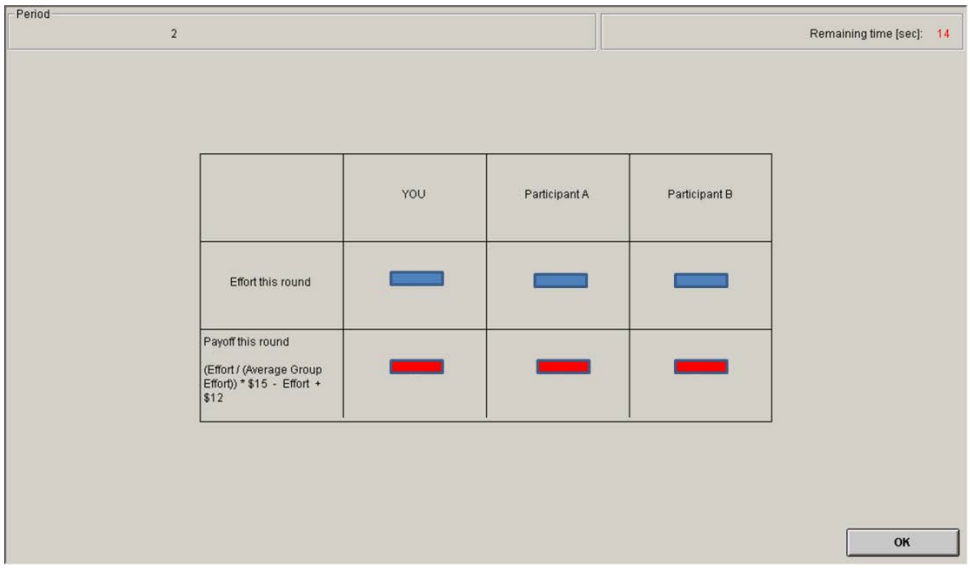


Notice that your effort DIVIDED by the average effort is equal to 1/5, so your relative compensation in this example would be 1/5 of $15 = $ 3. Hence, your total payoff would be $ 3 - $ 1 + $ 12 = $ 14.

As another example, if you choose 4 units of effort and the other two members each choose 3 units of effort, you will earn $26 (i.e., **(4/3.33)*$ 15-$ 4+$ 12=$ 26**), where 3.33=(4+3+3)*(1/3) is the average effort. The minimum you can make in each period is $ 12.8 and the maximum is $ 40.4.

You will have 45 seconds to enter your effort choice and to chat. Feel free to take the allocated time to choose and to chat with your group members. Your time remaining will appear on the upper right hand side of your screen. However, if a participant does not make his/her choice by the 45 seconds, the experimenter will prompt him/her to input his/her choice.
<STOP READING HERE>

After each period, you will see reported each of your group members' chosen efforts and the calculation of your payoff for that period. For instance, participant C will see the following (note the blue and red boxes below will have numbers in them during the experiment):



Your payoff for this second stage will be the sum of all payoffs over all periods of play for this stage.

<STOP READING HERE>

In the **final stage**, you will be given a questionnaire that can yield some additional payoffs. After all questionnaires are complete, final payments will be made to each of you.

Each screen you see throughout the experiment has all the instructions necessary for the decision in that screen.

Recall, during the session, all payoffs are expressed in terms of Berkeley Bucks. However, at the end of the session, all of your Berkeley Bucks will be converted at $66.6 Berkeley Bucks to $1 US. Thus, in US dollars, your final payment will be between $5 and more than $25, depending on how you do.

Thanks!