Do People Who Care About Others Cooperate More? Experimental Evidence from Relative Incentive Pay*

Pablo Hernandez
New York University AD

Dylan Minor
Northwestern University

Dana Sisak
Erasmus University Rotterdam
& Tinbergen Institute

This version: February 2015

Abstract

We propose an experiment to study the effect of group composition, in terms of other-regarding concerns, on effort under relative performance pay. We consider an indefinitely repeated interaction and we vary whether group members can communicate with each other. Although relative performance pay may lead other-regarding workers to reduce efforts as they internalize the negative externality they impose on other workers, indefinitely repeated interaction may incite selfish workers to take advantage of other-regarding ones, increasing overall efforts. We find other-regarding workers indeed 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 other-regarding composition of a team of workers has potentially nuanced consequences for firm performance.

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, Amsterdam 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 her own effort decision.

We propose a controlled laboratory environment which examines three channels through which agents may reduce effort under relative performance incentives. The first one is "other-regarding" concerns as agents may incorporate other agents' payoffs into their own effort choice. "Other-regarding" agents should respond differently to relative performance incentives compared to "selfish" agents. Thus, group composition in terms of other-regarding concerns should determine individual effort in strategic interactions such as relative performance pay. Even though individuals having heterogeneous degrees of other-regardingness (e.g., see Andreoni and Miller, 2002; Fisman, Kariv and Markovitz, 2007) is a robust finding in the literature, we know little about the effect of other-regarding concerns on the effectiveness of relative performance incentives. The second channel is indefinitely repeated interaction. Workplace interaction usually takes place for an indefinite period of time, so the "shadow of the future" may also affect agents behavior (e.g., see Dal Bó 2005). In particular, agents may reduce competition by colluding in low effort outcomes. At least to our knowledge, this is the only paper that explores the interplay between relative incentives, other-regarding concerns and indefinite horizon. The third channel is leadership. The potential for coordination in indefinitely repeated settings may stimulate leadership emergence (e.g., see Hermalin 2012). Although leaders make coordination focal in coordination games (e.g., see Cooper et al. 1992), we know little about their effect on agents' behavior in indefinitely repeated interactions under relative performance incentives. Leaders in this setting are important as they potentially direct individuals towards low effort outcomes, which in principle undermine productivity in real world settings.

In our experiment, we measure a proxy for subjects' other-regardingness using dictator games. We relate this proxy to their effort and leadership decisions in groups interacting repeatedly and indefinitely under relative incentives. We randomly assign subjects with different levels of other-regardingness into groups and thus identify the effect of group composition on effort. We also consider interactions without communication (the baseline) and with communication, in order to explore the role of verbal leadership on effort outcomes.

Regardless of communication, we find that groups with more other-regarding workers tend to depress total efforts. At the individual level we find that when communication is not part of the work environment, each other-regarding group member depresses overall effort by 15%. Collusive outcomes, where all group members depress efforts, rarely occur in this case. Thus, our results are consistent with other-regarding individuals internalizing the externality they impose without engaging in long-term strategic behavior.

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 as a coordinator, as argued by Kreps (1986) and Hermalin (2012). In our particular setting, we label as a "leader" an individual who suggest to the group to coordinate on minimal effort—which is the Pareto optimal outcome from the agents' viewpoint. Controlling for the emergence of this sort of leadership, we find that with communication, other-regarding subjects depress their effort relative to selfish ones by 50%. We also find that selfish individuals are 2.7 times more likely than other-regarding individuals to successfully lead their groups to such a "collusive" outcome.

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. This is in contrast to the idea that a mixed group will find it hardest to sustain low levels of effort due to the combination of other-regarding individuals having weaker punishment incentives while a selfish individual has a greater incentive to deviate. In fact, our results imply that heterogenous groups may actually be the worst for a principal interested in maximizing workforce effort using relative performance schemes.

In order to eliminate possible confounds such as differences in beliefs or degrees of patience, we have subjects in a final 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 the notion 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. 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 (for example 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. Loch and Wu, 2008; Dreber, Fudenberg and Rand, 2011; Bowles and Polania-Reyes, 2012; Kőszegi, 2013). Moreover, Fehr and Fischbacher (2002) point out 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 (e.g. see Kidd, Nicholas and Rai, 2013; Erkal, Gangadharan, and Nikiforakis, 2011; Rey-Biel, Sheremeta, and Uler, 2012; and Riyanto and Zhang, 2013). Similar to Gächter and Thöni (2005) and Fischbacher and Gächter (2010) we use one game (a dictator game as in Andreoni and Miller, 2002) to predict other-regarding concerns and relate those predictions to behavior in the relative performance game. Although our relative performance game is similar to the dilemmas used in those papers (players are better-off if they "cooperate" in low efforts), an important difference is that the interactions in our game are indefinitely repeated—which is a common feature of the workplace. Assessing the effect of social preferences on indefinitely repeated interaction is important because it is not clear (ex-ante) whether other-regarding concerns would depress efforts due to "internalizing" the negative externality imposed on others or would increase efforts due to lenient punishment in case of a deviation.

The importance of group composition on dimensions other than the degree of other-regardingness has been explored. Casas-Arce and Martinez-Jerez (2009) for example, find that relative performance incentives (tournaments in their setting) are less effective than piece rates when participants have heterogeneous abilities. The reason, the authors argue, is the discouragement effect of ability gaps. A similar result is found by Backes-Gellner and Pull (2013) in a sales contest within a German insurance firm. To our knowledge, the effect of group composition in terms of other-regardingness on efforts has not been explored, and yet there have been studies that show that individual other-regardingness is important. For example, Bandiera, Rasul and Barankay (2005) explore the role of social preferences in indefinitely repeated (or at least long-term) interactions. They find that workers with social ties depress effort. 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 forming groups whose members have varying degrees of social preferences. This allows us to identify the link between social preferences and the responses to relative performance incentives, both as a function of own preferences and group composition.

Usually workplace interactions are repeated indefinitely. This feature introduces the possibility of collusive outcomes: Pareto improvements over the one-shot Nash equilibrium can be obtained as equilibrium outcomes if the value of the future is high enough.[1] Although it has been documented that individuals are able to achieve the Pareto-optimal outcomes quite often—Palfrey and Rosenthal (1994) found cooperation rates from 29% to 40% in public goods games, and Dal Bo (2005) found cooperation rates of 38% in indefinitely repeated prisoner's dilemmas—there has been a great variety of outcomes in this literature. These 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 often feature 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 Gächter, 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, DeJong, Forsythe, and 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 7 experimental sessions with 147 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

[1] Versions of this "folk theorem" can be found in Friedman (1971) or Fudenberg and Maskin (1986).

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 beginning of the first stage, we randomly matched subjects into anonymous groups of three individuals and they remained in the same group for the remainder of this stage. 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.[2] We use these 9 periods to classify our subjects in terms of social preferences. In periods 10 and 11 we conducted allocation decisions with upwards-sloping budget sets as in Andreoni and Miller (2002) where subjects are given an allocation and decide on the overall exchange rate. In contrast to the previous dictator menus, here there is no possibility to distribute value between oneself and the other group members. The only choice a subject has is on the overall value of the endowment, not on how it is split up. 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. Finally, 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."[3] 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

[2]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.

[3]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	63
No Chat	63
Robot	21
Total	147

Table 1: *Summary of treatments*

players for the remainder of the experiment. The purpose of this stage was to give players the possibility to collude by jointly providing low levels 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 low levels of effort. In particular, in the first treatment ("Chat") we allowed chat via computer terminals *during* each period and observability of choices and payoffs *after* every period. We recorded the chat messages in order to identify coordination leaders and their social preferences. In the second treatment ("No Chat") we did not allow for chat but continued with observability after each period.

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:

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

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).[5] Note these figures are in Berkeley Bucks \$, converted at \$66.6 Berkeley Bucks to 1 US\$, which

[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 this is mathematically the same as a Tullock contest played by risk-averse individuals. That is, the principal has a total pool of 45 Berkeley Bucks to distribute across workers based on

is how it was presented to subjects.[6] 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.

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 $x_i = 1$ under grim-trigger strategies 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 coordinating on low efforts for utility maximizing rational Selfish agents.

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 experimental results.

4 Experimental Results

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 (measured on the left y-axis). In the second stage, (periods 12 to 40) we observe the choice of effort ranging from 1 to 12 (measured on the right y-axis).[7] Each of the three group members is represented by a different symbol – a circle, a triangle and a cross.

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 a "perfect" collusive outcome in the Chat 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 treatment on how subjects slowly manage to coordinate on lower efforts.

Panel 3 shows a group from the Chat treatment. In this case, behavior in the second stage is surprising: 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

their relative performance.

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

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

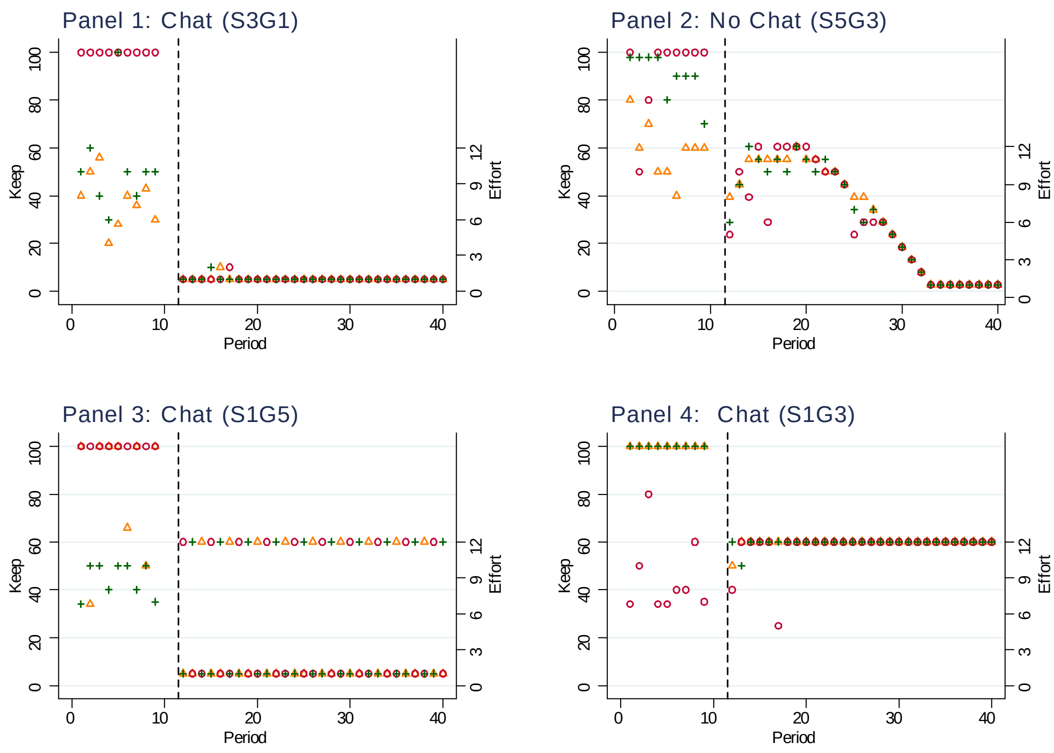

Figure 1: *Examples of group giving and investment decisions (S denotes session number and G group number).*

Period	Price vector	Keep (min, max)	Give to 1	Give to 2
1.	$(1,1,1)$	69.64 (33,100)	15.61	14.75
2.	$(1,\frac{1}{2},\frac{1}{2})$	73.93 (20,100)	13.14	12.93
3.	$(1,\frac{3}{4},\frac{3}{4})$	72.27 (0,100)	13.71	14.02
4.	$(1,\frac{5}{4},\frac{5}{4})$	71.88 (20,100)	14.24	13.88
5.	$(1,\frac{3}{2},\frac{3}{2})$	70.28 (20,100)	14.98	14.75
6.	$(1,1,\frac{2}{3})$	72.31 (30,100)	16.44	11.25
7.	$(1,1,\frac{3}{4})$	73.51 (25,100)	15.35	11.14
8.	$(1,\frac{3}{4},\frac{1}{2})$	77.48 (25,100)	12.56	9.95
9.	$(1,\frac{5}{4},\frac{3}{4})$	72.32 (25,100)	16.65	11.03

Table 2: *Giving rates.*

play. Although this does not allow the subjects to reach the maximal group payoff, this form of coordinating still leads to high payoffs relative to the one-shot Nash outcome. About 20% of groups in the Chat treatment exhibit a pattern like this at least part of the time.

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

4.2 Categorizing Social Preference Types from Giving Menus

Table 2 summarizes the mean choices of our subjects under all 9 price vectors in treatments: 1) Chat and 2) No Chat.[8] We will analyze the Robot treatment in section 4.5.

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 two treatments (Chat and No Chat) most of the participants (80.95%) are categorized as Other-Regarding. The balance of 19.05% of subjects are categorized as Selfish.[9]

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

[8] 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.

[9] Andreoni and Miller (2002), found 23% of their subjects can be classifed as perfectly selfish and Fisman et al. (2007) found that was the case for 26% of their sample.

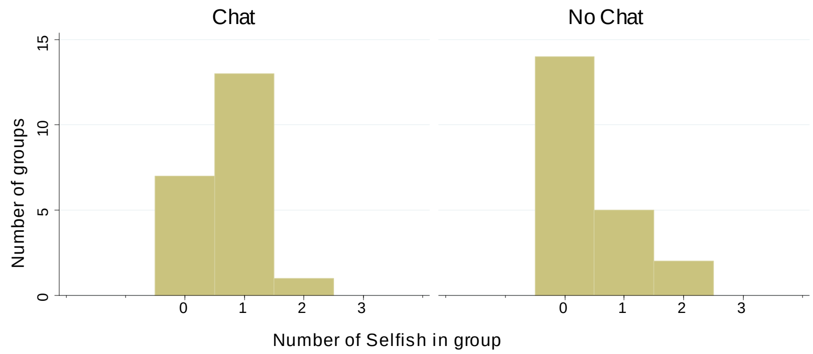

Figure 2: *Allocation of Selfish across groups.*

Chat and No Chat 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.

To test the robustness of our results, we use an alternative classification where a subject is classified as Selfish when he or she kept more than 90% of the endowment on average in dictator menus 1-9. Using this classification we observe groups with 0, 1, 2 and 3 Selfish group members under all three treatments. Our results from tables 4 and 5 are qualitatively robust to this alternative classification.

4.3 Social Preferences and Effort

Figure 3 provides a summary of effort choices over time by treatment. In both 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 in the Chat treatment (dashed line). When communication was absent (No Chat treatment), average effort stays close to the one-shot Nash equilibrium prediction for the Selfish type (dotted line).

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 two treatments.

Figure 4 gives a first overview of our findings. Consider first panel a). We compare the average effort of subjects categorized as Selfish with the average effort of subjects categorized as Other-Regarding. We see that for both treatments, average effort is higher for subjects categorized as Selfish, although a t-test rejects equality only for the No Chat treatment (p-values: $p<0.60$ in Chat and $p<0.01$ in No Chat).[10] In the No Chat treatment, average efforts are similar to the one-shot Nash equilibrium efforts

[10] This is consistent with Erkal, Gangadharan and Nikiforakis (2011) in that selfish individuals tend to exert higher levels of effort in tournaments.

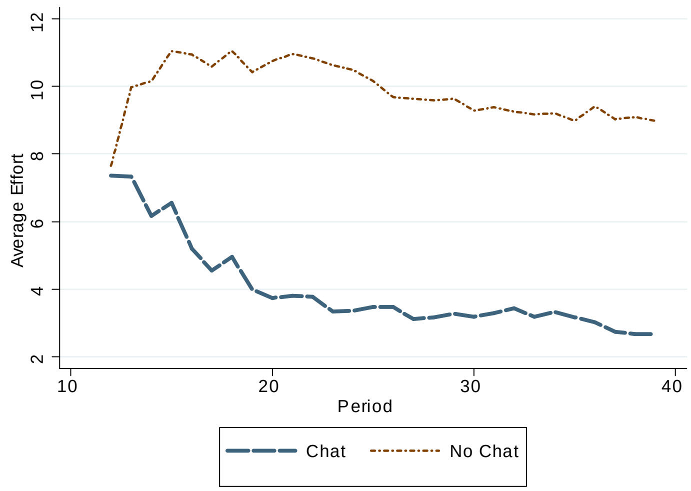

Figure 3: *Average effort by treatment over time.*

(i.e., efforts of 10 with $\rho = 0$) rather than to a collusive outcome and Other-Regarding subjects provide lower efforts on average.

In panel b) we consider average group effort as a function of the number of Selfish players within a group. When communication was not possible, we observe that each additional Selfish group member modestly increases average group effort though none of these increases reach statistical significance. When communication is possible, there is a pronounced increase in average effort when comparing a group with two versus fewer selfish group members, while groups with no Selfish seem to produce slightly higher efforts than groups with one Selfish. Also here differences do not reach statistical significance.

We further explore differences in group effort choices as a function of the number of Selfish subjects controlling for a number of group characteristics through regression analysis in Table 3. We use as the dependent variable the 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 3 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 treatment and column 2 reports the results for the No Chat treatment. In the Chat treatment, we do not find a significant effect of Selfish group members. In contrast, when communication is not possible (No Chat treatment), each Selfish group member increases average group effort by approximately .9 units on average, which equals a 9% increase over our baseline mean effort of roughly 9.7 per period.

Overall, these results suggest that, absent communication, average efforts are consistent with one-shot Nash equilibrium strategies. When communication is introduced, however, efforts seem to follow the collusive outcome and results are somewhat

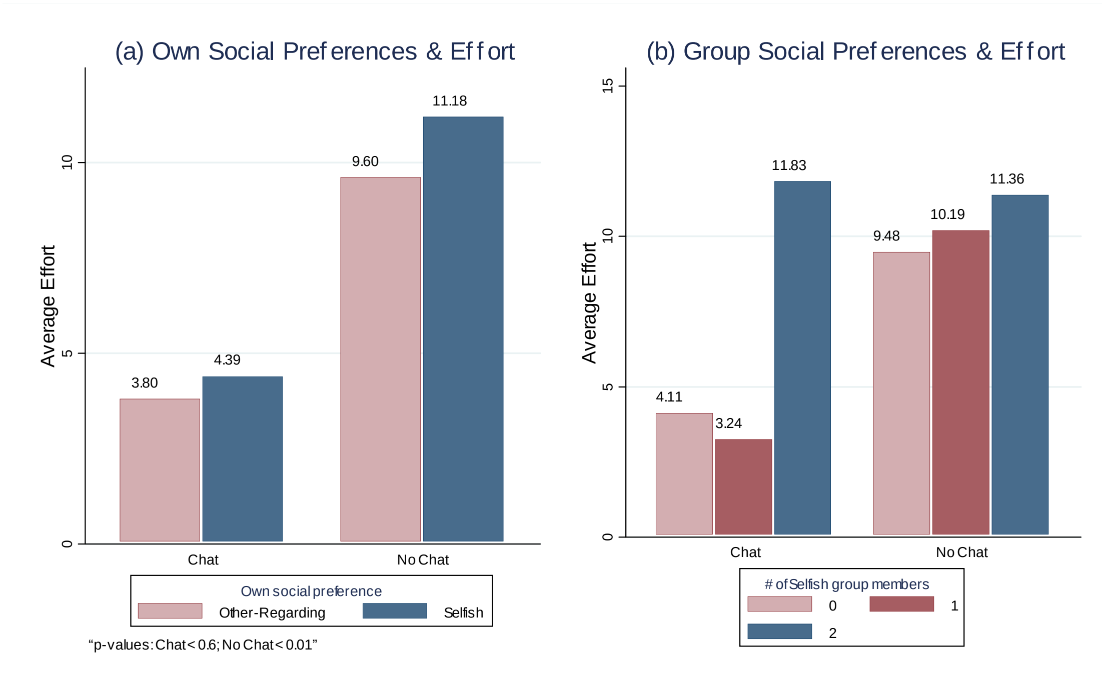

Figure 4: *Overview of effects of social preferences on effort.*

surprising: The presence of one Selfish individual leads to lowest aggregate efforts. This is due to Selfish individuals being more likely to lead suggesting coordination. We devote Section 4.4 to the detailed analysis of the Chat treatment. In the next subsection, we explore further the effect of group composition on efforts in the No Chat treatment.

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

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 3: *Effect of 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 4: *Effect of own and others social preferences on own effort (No Chat).*

4.3.1 No-Chat 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 treatment, clustering standard errors at the group level.[11] The dependent variable is individual effort and the explanatory variables are: Selfish and the number of other Selfish individuals in each group (# Other Selfish). Table 4 reports our results. We find further evidence that Other-Regarding subjects choose significantly less effort. Controlling for group composition, these subjects choose 1.5 fewer units of effort. The group composition effect on the other hand, is positive but insignificant. Thus, absent communication, Other-Regarding subjects depress efforts relative to Selfish subjects.

We performed a number of robustness checks. We do not include lagged effort choices in our main specification due to the issue of inconsistent estimates. 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 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 analogous results with this measure of selfishness. We also consider an alternative classification of social types: we classify Selfish subjects as those who 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 it is still significant. However, for the group level regressions although the sign is still correct, the coefficient estimates of these small samples are no longer significant.

We now summarize our primary result thus far.

[11] 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.

Result 1: *Absent communication, Other-Regarding subjects depress efforts relative to Selfish subjects.*

This result is consistent with the idea that groups comprising Other-regarding individuals attain lower efforts than groups with at least one Selfish individual. Overall, the difference between Selfish and Other-Regarding individuals' average efforts within each treatment, suggest that depressed effort is not the result of coordination. These results seem instead to be driven by one-shot considerations. They are consistent with Other-Regarding individuals internalizing the negative externality of their effort choice and exerting one-shot competitive efforts below those of Selfish individuals.

When communication is possible, there are now two channels through which individuals can coordinate on low efforts: 1) encouraging low efforts through chat and 2) effectively choosing low efforts and punishing otherwise. We now explore the possible effects of each channel.

4.4 Chat and Leadership

In the Chat 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 the No Chat treatment. We elicit this measure of "leadership" from the chat messages. We label a "Min-Effort Leader" as a subject that is the first to propose coordinating on the minimum effort case (i.e., for all group members to provide effort of 1).[12] We identify 13 Min-Effort Leaders (21%) among the 63 subjects (21 groups) in the Chat treatment.[13]

We start by providing a breakdown of the social preferences of the subjects we identified as leaders. Figure 5 shows the distribution of social preference types in the sample of Min-Effort Leaders and Non-Min-Effort Leaders. We observe that Selfish individuals are more likely to be leaders. A Pearson chi–squared test shows this difference is significant at the 5% level (p=0.03). Thus we find that 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 treatment, is it still true that low efforts are related to group members' social preferences similarly as in the

[12] We initially collected two other categories of leadership. A "Failed Leader" to denote a subject that 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. We also considered a "First Leader," which was the first subject to propose coordination of efforts. However, this latter category has little explanatory power and so we omit it from our analysis.

[13] We also had both a research assistant from Erasmus University Rotterdam and from Northwestern University independently code the leadership variables. The instructions given to the RAs are provided in the appendix. The correlations between the alternative leadership dummies and the ones we use in the paper are for Northwestern: 0.88 for whether a Min-Effort Leader exists (on a period/group level) and 0.82 for the subject being a Min-Effort Leader (subject level); and for Rotterdam 0.52 for whether a Min-Effort Leader exists and 0.56 for the subject being a Min-Effort Leader. For both of these classifications, we find similar results in our following analysis.

	(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)
Min-Effort Leader Exists		-5.709***	-3.661***
		(0.637)	(0.423)
Min-Effort Leader		0.0784	0.107
		(0.350)	(0.338)
MELeader*Selfish			-2.729***
			(0.678)
MELeader*#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	0.10/0.03	0.21/0.74	0.21/0.78

Standard errors in parentheses

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

Table 5: *Effect of social preferences on individual effort controling for leadership (Chat treatment).*

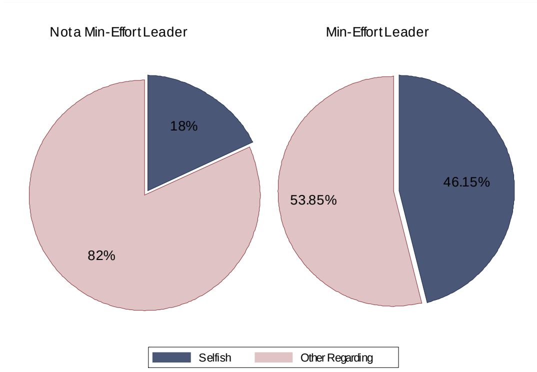

Figure 5: *Social preferences of leaders.*

No Chat treatment? Table 5 reports the results of a random effects model for the Chat treatment. Column 1 shows a regression without considering leader emergence, analogous to the one in Table 4 for the No Chat treatment. The dependent variable is individual effort and the explanatory variables are: Selfish and the number of other Selfish individuals in each group (# Other Selfish). In column 2 we add as a control whether a Min-Effort Leader has emerged (i.e., a dummy variable that takes on a value of one once a Min-Effort Leader emerged in the group) and whether the subject herself is a Min-Effort Leader (i.e., a time-invariant dummy variable that takes on value of one for all subjects who are classified as a Min-Effort Leader). 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 after controlling for the effect of social preferences running through leadership emergence, social preferences lead to significantly lower group efforts. The effect is slightly larger in magnitude than in the No Chat treatment. Precisely, a Selfish subject puts in 2 units effort more per period than an Other-Regarding subject after controlling for the emergence of a coordination leader. Furthermore, the presence of an additional Selfish group member increases a subject's own effort by 2 units per period also controlling for leader emergence.

Column 3 includes interactions of social preference measures and the emergence of a leader. We find that social preferences depress efforts when a Min-Effort Leader has not emerge in a group. Other-regarding subjects depress their effort relative to selfish ones by about 30%. 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 Min-Effort Leader is insignificant. Thus, Min-Effort Leaders do not lead also by good example: i.e., they only lead through suggesting low effort by chat message and not through actually initiating lower effort themselves. In other words, when a group has a Min-Effort Leader, all group members produce lower effort, but the Min-Effort Leader herself does not produce less effort than any other group member.

We conclude that social preferences are an important determinant of group effort also in the Chat 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 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 before a coordination leader emerges. 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. Thus, our analysis yields two more results:

Result 2a: *Selfish subjects are more likely to lead others to coordinate on low efforts.*

Result 2b: *Without the emergence of a coordination leader, Other-Regarding subjects depress efforts relative to Selfish subjects. When a leader emerges, there are no differences in effort choices between Other-Regarding and Selfish subjects.*

4.4.1 Propensity to "Collude"

Thus far we have been focussing on the relationship between social preferences and depressed efforts. Depressed efforts can of course also be a consequence of 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 consistent with coordination on minimum efforts: That is, 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 who gets the maximal payoff. This latter form of coordinating on low efforts is only witnessed in the Chat treatment where subjects were allowed to coordinate via chat.

# 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 6: *Propensity to "collude" by number of Selfish group members in the Chat treatment.*

Table 6 reports the proportion of groups achieving the "collusive" outcome in the Chat treatment. Here, we separate groups by the number of Selfish members (groups with $0, 1$, or 2 Selfish members). Similar to our results on efforts from 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.

Comparing the results in Table 6 to Figure 4 leads to an interesting observation. Even though groups with one Selfish are more likely to collude, average effort is quantitatively not very different to a group with no Selfish (3.2 vs. 4.1). As already explained in Section 4.4 the reason is that in the "pre-collusion phase" groups with no Selfish members put in lower efforts than groups with one Selfish member (average effort is 5.4 in a group of only Other-Regarding vs. 7.5 in a group with one Selfish prior to the emergence of a Min-Effort Leader). 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 while Other-Regarding have a tendency to put in lower efforts even absent collusive motives. Thus, we summarize our final primary result:

Result 3: *With communication, the propensity to "collude" is greater with one Selfish group member than with no Selfish group members*

For the No Chat treatment, coordinating on a "collusive outcome" was more difficult, since subjects were not able to chat. As shown in Table 7, 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 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 7: *Propensity to "collude" by # of Selfish in the No Chat treatment.*

One might object at this point that individuals we categorize as Selfish are the ones who 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. We do not find, however, that subjects with a background in Economics or Business, who we would expect to have a better sense of the optimal strategy, are more likely to be Min-Effort 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.

4.5 Robot Treatment

This treatment is similar to the No Chat 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."[14] In particular, we programmed 42 robot subjects who react to past effort decisions by approximating what real subjects did in the No Chat 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. Crucial to this treatment is that it is no longer the case that 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 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, subjects' decisions can reflect beliefs about the past subjects' social preferences. This is, in fact, 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, 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. To sum up, if our categorization is valid, we do not expect to see any difference between "Selfish" and "Other-Regarding" subjects in this treatment, while if our categorization

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

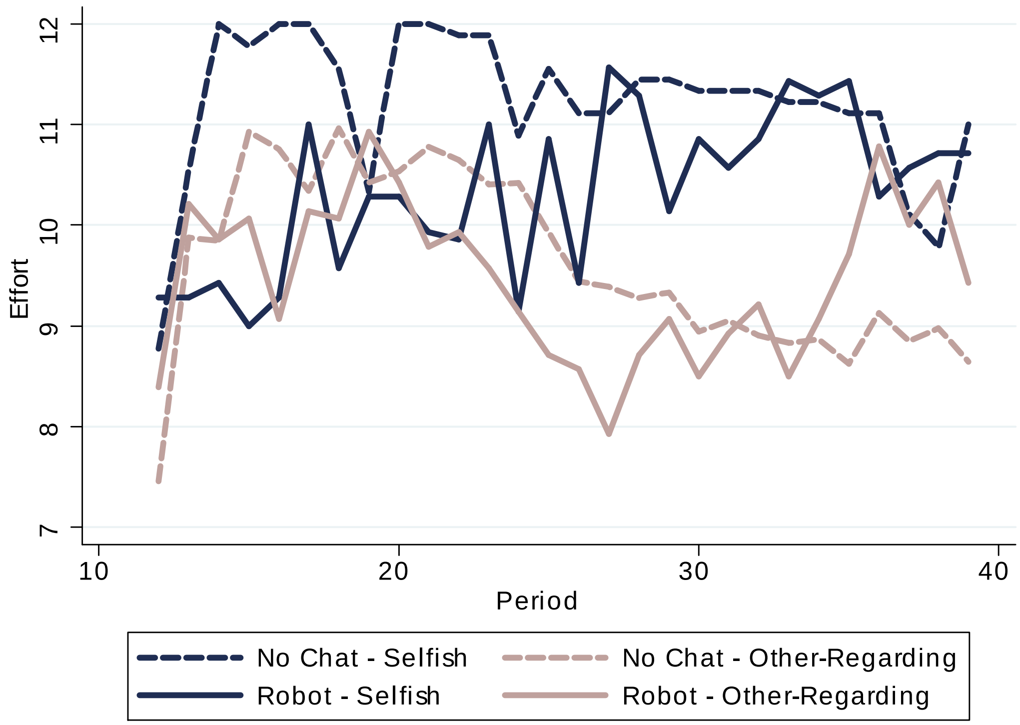

Figure 6: *Comparing efforts between Selfish and Other-Regarding types over time.*

only reflects a correlation with other decision-relevant factors such as beliefs or skill we would still observe a significant difference between the two subject groups.

We first compare subject behavior for the No Chat 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 0.21), 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 0.16). Thus, efforts of all social preference types in the Robot treatment converge towards the efforts of Selfish subjects in the No Chat 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 treatment (i.e., Selfish in the No Chat treatment vs. Selfish in the Robot treatment, p-value .7315; Selfish in the No Chat 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 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 human 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.

	No Chat 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 8: *Effect of social preferences on individual effort No Chat vs. Robot treatment.*

If we instead analyze individual rather than average aggregate effort choices, which may mask individual behavior, we find the same pattern of similar effort choices across social preference types. Table 8 reports the results of regressing individual effort on own and group members' social preference types for the No Chat and the Robot treatment. The coefficient estimate for Selfish is half the value as in the No Chat treatment and is no longer significant, though we do note the sample size is smaller in the Robot treatment.

Result 4: *When a subject's action affects a machine's success rather than a human's success, Selfish and Other-Regarding subjects behave similarly*

Overall, the results from the Robot treatment, although suggestive, provide further evidence that social preferences matter in creating and sustaining non-competitive efforts.

5 Conclusion

We studied how a relatively newly explored 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, subjects categorized as Selfish are more likely to coordinate their group members to minimal efforts, when communication is available. Second, before the emergence of such leaders, subjects categorized as other-regarding exert lower levels of effort—an average of over 30% 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. Finally, when communication is not available, groups of Other-Regarding subjects produce the lowest levels of effort. Since we find little evidence of collusive outcomes, this is again consistent with the idea that Other-Regarding individuals internalize their efforts' negative externality imposed on other subjects' 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 provides further evidence that other-regarding individuals are indeed depressing their efforts as they internalize the negative externality of higher effort.

Our findings suggest that for organizations attracting more other-regarding workers (e.g., firms engaged in corporate social responsibility or non-profit firms), relative performance incentives are likely to not be as effective as for other organizations. For firms using relative incentive pay, screening workers for particular positions 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.

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] Backes-Gellner, U. and K. Pull (2013) “Tournament Compensation Systems, Employee Heterogeneity, and Firm Performance,” *Human Resource Management*, 52 (3), 375-398.

[3] 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.

[4] Barany, I. (1992). “Fair Distribution Protocols or How the Players Replace Fortune,” *Mathematics of Operations Research*, 17, 327–340.

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

[6] Casas-Arce, P. and F.A. Martínez-Jerez (2009) “Relative Performance Compensation, Contests, and Dynamic Incentives,” *Management Science*, 55(8), 1306-1320.

[7] Charness, G. (2000). “Self-serving cheap talk: A test of Aumann’s conjecture.” *Games and Economic Behavior*, 33(2), 177-194.

[8] 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.

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

[10] 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.

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

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

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

[14] 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.

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

[16] Fischbacher, U. and S. Gächter (2010) "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments," *The American Economic Review*, 541-556.

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

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

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

[20] 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.

[21] Gächter, 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.

[22] Gächter, S., and Thöni, C. (2005) "Social Learning and Voluntary Cooperation among Like-Minded People," *Journal of the European Economic Association*, 3 (2-3), 303-314.

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

[24] Kidd, M., A. Nicholas and B. Rai. (2013) "Tournament outcomes and prosocial behaviour," J*ournal of Economic Psychology* 39, 387–401.

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

[26] Koszegi, B. (2013) "Behavioral Contract Theory," *Journal of Economic Literature*. Forthcoming.

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

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

[29] Loch, C. H., and Wu, Y. (2008) "Social Preferences and Supply Chain Performance: An Experimental Study," *Management Science*, 54 (11), 1835-1849.

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

[31] Riyanto, Y. and Zhang, J. (2013) "The impact of social comparison of ability on pro-social behaviour," *The Journal of Socio-Economics*, Volume 47, December, Pages 37–46.

[32] Rey-Biel, P., with R. Sheremeta and N. Uler. (2012)."(Bad) Luck or (Lack of) Effort? Sharing Rules in the US and Europe" *Working Paper*

[33] 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 price 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 2 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 9. 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 9: *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.[15] For treatments 1 and 2 we find that, for our subject pop-

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

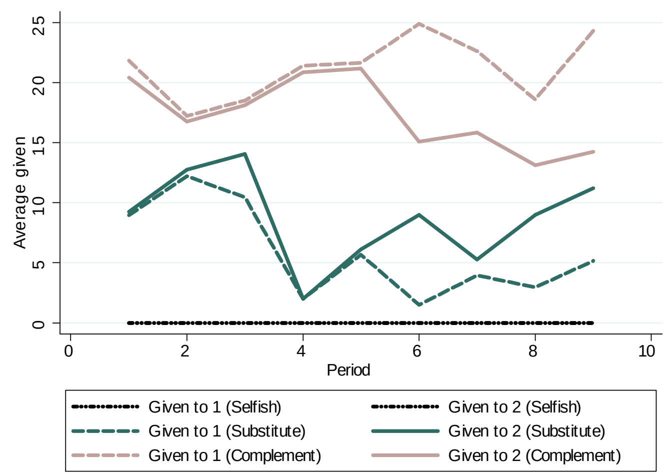

Figure 7: *Giving rates by social preference types.*

ulation, 19% are Selfish, 65% are Complements and 16% are Substitutes. Consistent with Andreoni and Miller (2002), hereafter AM, we find that 19% of subjects are (perfectly) Selfish, whereas AM find that 23% of subjects are perfectly Selfish. 7.1% 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 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 2 . Thus, as archetypal types would do, Complements react by allocating more to individual 1 while Substitutes react by allocating more to individual 2.

Table 10 is analogous to Table 3 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 treatment by approximately .8 units. In the Chat treatment, a linear regression again does not yield significant results; this is to be expected given the discussion in the main text of the confound of leadership. We will again consider the effect of social preferences on leadership and explore whether it

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

differs by Complements and Substitutes.

Table 11 is analogous to Table 4. Here, we present the results of a random effect panel regression model for the No Chat 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 treatment. Figure 8 reports the distribution of social preferences among Non-Min-Effort Leaders and Min-Effort Leaders as defined in Section 4.4. As before, Selfish are significantly more likely to become Min-Effort 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).

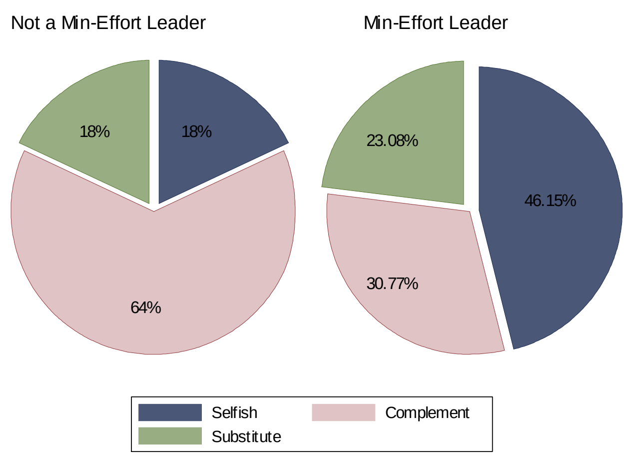

Figure 8: *The distribution of social preferences among Min-Effort Leaders and non-Min-Effort Leaders.*

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 5 for the Chat treatment.

We report these results in Table 12. The first column does not control for the emergence of a Min-Effort Leader and whether or not an individual turns out to be a

	Chat	No Chat
	Avg Effort (Grp/Sess)	Avg Effort (Grp/Sess)
# Compl.	-0.593	-0.873**
	(1.582)	(0.389)
# Subst.	-1.742	-0.856
	(2.009)	(0.685)
Constant	5.952	12.06***
	(4.017)	(0.942)
Observations	21	21
Adjusted R^2	-0.036	0.030

Standard errors in parentheses

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

Table 10: *Group composition and average group effort.*

	(1)		(2)	
	Effort		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 11: *Effect of own and others social preferences on own effort (No Chat).*

	(1)		(2)	
	Effort		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)
Min-Effort Leader Exists			-5.690***	(0.636)
Min-Effort 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 12: *Effect of social preferences (extended categorization 1) on individual effort controling for leadership (Chat treatment).*

Min-Effort 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 Min-Effort Leader and controlling for being a Min-Effort 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 difference 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 13 summarizes the two menus and the decisions of subjects in Treatments 1 and 2. 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 and 2, 67% of subjects are classified as Jealous.

Menu (Allocation)	Mean	Percent where rate=2
10 (20,40,40)	1.794	76%
11 (2,49,49)	1.259	54%

Table 13: *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:[16]

- Disinterested: not Jealous and Selfish (8%)
- Benevolent: not Jealous and Other-Regarding (25%)
- Spiteful: Jealous and Selfish (11%)
- Inequity Averse: Jealous and Other-Regarding (56%)

Table 14 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 3. In the Chat treatment, we do not find any significant effect of these social preferences types. In the No Chat 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.

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-Min-Effort Leaders (left panel) and Min-Effort Leaders (right panel) for the Chat treatment. As can be seen, Spiteful individuals have the highest propensity of becoming a Min-Effort 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 Min-Effort 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 Min-Effort Leader.

Finally, controlling for the emergence of a leader, as in Table 5, we can separate the relation of social preferences and leadership emergence from general effort choices. Table 15 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

[16]Population proportions are for Treatments 1 and 2.

	Chat Avg Effort (Grp/Sess)	No Chat Avg Effort (Grp/Sess)
# Spiteful	-4.822	1.488***
	(3.274)	(0.421)
# Inequ. Av.	-4.766	-0.807
	(2.945)	(0.489)
# Benev.	-4.614	-0.761
	(3.101)	(0.512)
Constant	17.73*	11.79***
	(8.834)	(0.934)
Observations	21	21
Adjusted R^2	0.087	0.017

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Table 14: *Group effort and Inequality Aversion (omitted category: Desinterested).*

analysis. 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.
- 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.

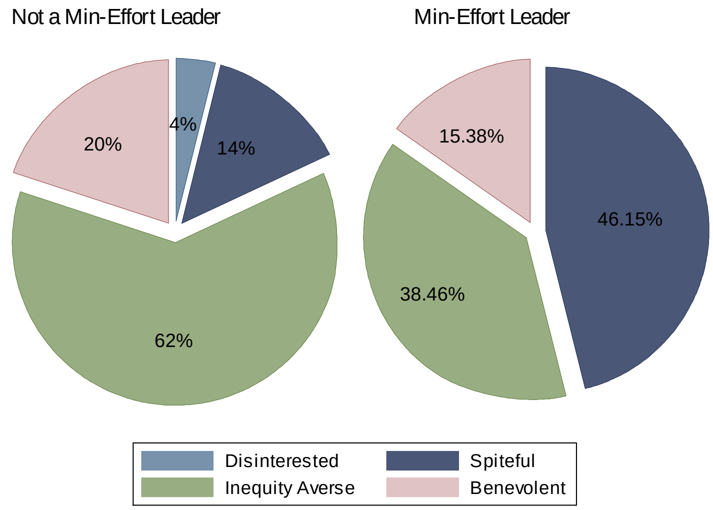

Figure 9: *Distribution of social preferences among non-Min-Effort Leaders and Min-Effort Leaders under extended categorization two.*

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 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 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

	(1)		(2)	
	Effort		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)
Min-Effort Leader Exists			-5.316***	(0.633)
Min-Effort 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 15: *Effect of social preferences (extended categorization 2) on individual effort controling for leadership (Chat treatment).*

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 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 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 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 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

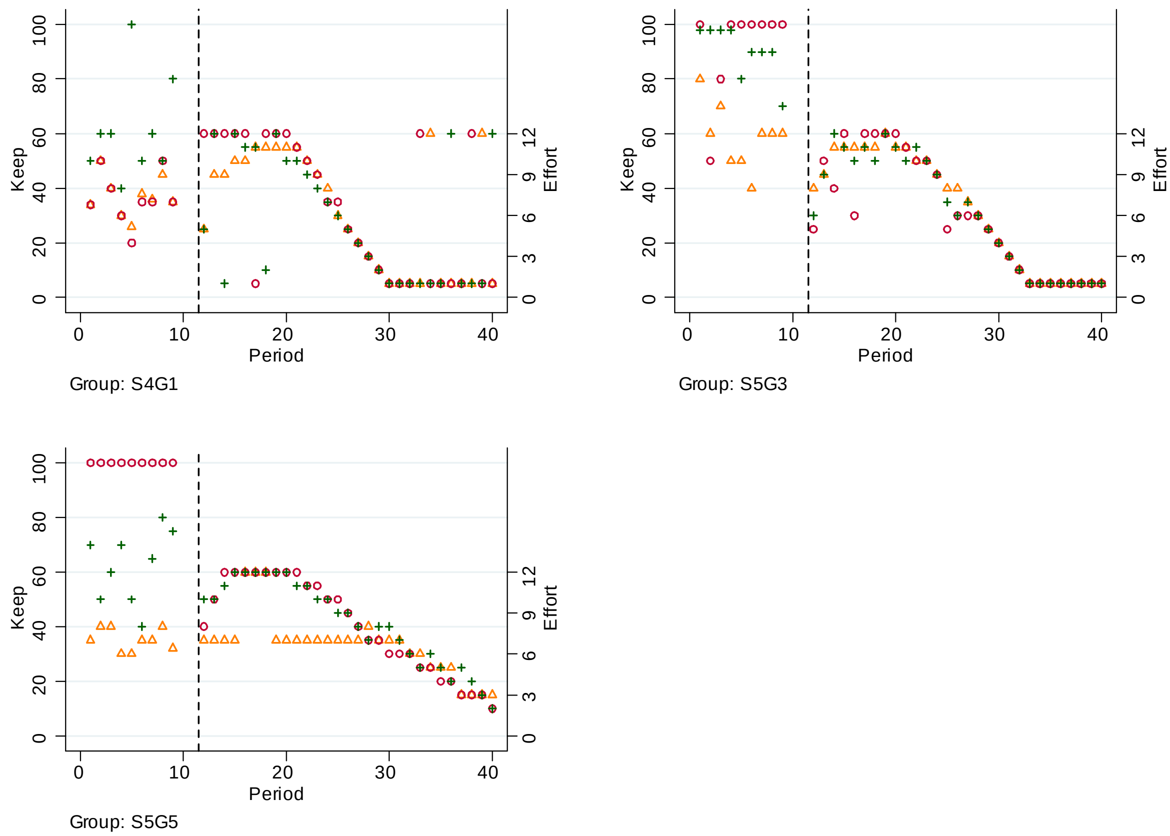

Figure 10: *Choices of groups classified as "colluding."*

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 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 16 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 16: *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 outcomes to the distribution of actual group outcomes of human subjects in the No Chat treatment. Table 17 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 17: *Randomly matched groups (simulations).*

6.4 Leader Classification Details

Attached file

6.5 Instructions for Subjects

Attached file