# Online Appendix: Do People Who Care About Others Cooperate More? Experimental Evidence from Relative Incentive Pay

Pablo Hernandez-Lagos
New York University Abu Dhabi

Dylan Minor
Northwestern University

Dana Sisak
Erasmus University Rotterdam
& Tinbergen Institute

This version: January 2017

**Abstract**

In this online appendix we explore three alternative social preference categorizations. We also provide details on the Robot treatment as well as provide results on the effect of social preferences on individual effort including lagged individual effort choices. Furthermore, we provide the details of the model used to derive the hypotheses of the paper. In the last sections, we present examples of group decisions and coordination ("collusive") outcomes. A copy of the experimental instructions and the instructions for Research Assistants to categorize leaders are attached as separate files.

| | No Chat | Chat | Robot | Total |
|---|---|---|---|---|
| Other-Regarding | 43 (68%) | 38 (60%) | 14 (67%) | 95 (65%) |
| Selfish | 20 (32%) | 25 (40%) | 7 (33%) | 52 (35%) |
| Total | 63 | 63 | 21 | 147 |

Table 1: *Distribution of Selfish and Other-Regarding subjects by treatment (90% cutoff classification).*

# 1 Online Appendix

## 1.1 Broader Social Preference Classifications

In this supplementary appendix we explore three alternative social preference categorizations. First we relax the stringent categorization of Selfish in the main manuscript and consider an individual as Selfish when he or she kept 90% or more of the endowment. For the other two alternative classifications we extend the number of possible social preference categories. 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 these categorization procedures, as well as some additional analysis using these expanded categories.

**Less Stringent Classification of Selfish**

For the main analysis we use a relatively conservative definition of a Selfish individual. Only an individual that keeps 100% of his or her endowment is classified as Selfish. In this section we relax this definition in order to explore the robustness of our results. In particular, we define a Selfish individual now as an individual that, on average, keeps 90% or more of his or her endowment in the dictator menus of period 1 to 9.

Table 1 presents the distribution of Selfish and Other-Regarding according to this new definition across treatments. Note that now 35% of all subjects are categorized as Selfish, an increase of 14 percentage points. This is reflected in all three treatments – subjects are more evenly divided between Selfish and Other-Regarding.

| # Selfish | No Chat | Chat [with Leader, without Leader] | Robot | Total |
|---|---|---|---|---|
| 0 | 9 (43%) | 3 (14%) [2,1] | 7 (33%) | 19 (30%) |
| 1 | 6 (28%) | 13 (62%) [9,4] | 9 (43%) | 28 (44%) |
| 2 | 4 (19%) | 3 (14%) [1,2] | 5 (24%) | 12 (19%) |
| 3 | 2 (10%) | 2 (10%) [1,1] | 0 | 4 (6%) |
| Total | 21 | 21 | 21 | 63 |

Table 2: *Distribution of groups by number of Selfish for each treatment (90% cutoff classification).*

Table 2 presents the distribution of groups with differing numbers of Selfish individuals in the three treatments. Note that under this alternative classification, now also groups with three Selfish group members exist in both the Chat and the No-Chat treatment. In general, more groups exist now with a majority of Selfish individuals than before.

In the rest of this section, we replicate the main regression analysis of Section 5 of the main manuscript for the alternative classification using a 90% cutoff for Selfish. Table 3 summarizes the results from the group-level analysis. The dependent variable is average group effort, averaged over all periods of play. Column 1 to 3 show both treatments pooled together, the Chat treatment and the No Chat treatment respectively, controlling only for the number of Selfish individuals in a group. While we do not find a significant effect of the number of Selfish on group effort for either treatment separately, the coefficient on number of Selfish is marginally significant in the pooled analysis in column 1. This is in contrast to our stricter classification, where we did find a significant effect for the No Chat treatment, but not the pooled analysis.

Columns 4 and 5 control for the emergence of a Min-Effort Leader in a group analogous to Table 7. The results are qualitatively similar. On average, controlling for the existence of a Min-Effort Leader in a group, more Selfish group members lead to higher group effort. Each Selfish group member increases group effort by about 1 unit as can be seen in column 4. Enabling the effect of social preferences of group members to depend on whether a Min-Effort Leader exists, as shown in column 5, we find that only in groups where no Min-Effort Leader emerged, do social preferences affect average group effort (joint test of significance of #Selfish and MELeaderGr*#Selfish results in a p-value of $p = 0.968$ and a coefficient of negligible magnitude). This is consistent with our earlier results.

Table 4 replicates the individual-level analysis of Section 5 of the main manuscript for the alternative classification. Columns 1 to 3 present the effect of own and group members' social preferences on individual effort. As with the more stringent definition of Selfish, a subject's social preferences do not matter for individual effort in the Chat treatment (column 1). Surprisingly, the group members' social preferences become

| | (1) Pooled | (2) Chat | (3) No Chat | (4) Chat | (5) Chat |
|---|---|---|---|---|---|
| # Selfish (alt) | 0.876* | 1.566 | 0.441 | 1.094* | 2.414*** |
| | (0.447) | (0.995) | (0.325) | (0.594) | (0.711) |
| Chat | -6.096*** | | | | |
| | (0.685) | | | | |
| Min-Effort Leader in Group | | | | -4.228*** | -1.206 |
| | | | | (0.885) | (1.272) |
| MELeaderGr∗#Selfish (alt) | | | | | -2.441** |
| | | | | | (0.954) |
| Constant | 8.992*** | 2.076* | 9.407*** | 5.254*** | 3.439*** |
| | (0.544) | (1.033) | (0.450) | (0.866) | (0.993) |
| Observations | 42 | 21 | 21 | 21 | 21 |
| $R^2$ | 0.646 | 0.175 | 0.074 | 0.637 | 0.739 |

Standard errors in parentheses

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

Table 3: *Effect of social preferences on group effort (90% cutoff classification, all periods).*

|  | (1) Chat | (2) No Chat | (3) Pooled | (4) Chat | (5) Chat |
|---|---|---|---|---|---|
| Period | -0.133*** | -0.0538* | -0.0932*** | -0.0752*** | -0.0762*** |
|  | (0.0276) | (0.0294) | (0.0209) | (0.0254) | (0.0252) |
| Selfish (alt) | 1.473 | 0.850** | 0.498 | 1.183** | 1.641*** |
|  | (0.987) | (0.397) | (0.531) | (0.503) | (0.510) |
| # Other Selfish (alt) | 1.612* | 0.237 | 0.894* | 1.363*** | 2.054*** |
|  | (0.966) | (0.319) | (0.505) | (0.463) | (0.405) |
| Chat |  |  | -6.421*** |  |  |
|  |  |  | (0.645) |  |  |
| Chat*Selfish (alt) |  |  | 0.887 |  |  |
|  |  |  | (1.263) |  |  |
| Min-Effort Leader Exists |  |  |  | -5.452*** | -3.633*** |
|  |  |  |  | (0.630) | (0.548) |
| Min-Effort Leader |  |  |  | 0.322 | 0.374 |
|  |  |  |  | (0.277) | (0.239) |
| MELeaderE*Selfish (alt) |  |  |  |  | -1.312*** |
|  |  |  |  |  | (0.464) |
| MELeaderE*# OthSelf (alt) |  |  |  |  | -1.708*** |
|  |  |  |  |  | (0.361) |
| Constant | 5.524*** | 10.81*** | 11.52*** | 7.159*** | 6.425*** |
|  | (1.489) | (0.632) | (0.588) | (0.702) | (0.655) |
| Observations | 1827 | 1827 | 3654 | 1827 | 1827 |
| $R^2$ within/between | 0.1/0.17 | 0.03/0.06 | 0.07/0.62 | 0.21/0.73 | 0.22/0.78 |

Standard errors in parentheses

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

Table 4: *Effect of social preferences on individual effort (90% cutoff classification, all periods).*

marginally significant now. In the No Chat treatment, the own social preference type matters for effort. Selfish individuals on average expend about .8 more units of effort which is slightly less than with the more stringent categorization. Their group members' social preferences on the other hand do not affect effort choices (column 2). Pooling both treatments together in column 3 we find as before that subjects in the Chat treatment, on average, expend lower effort: Other-Regarding effort is about 6.4 units lower and Selfish effort about 5.5 units lower than in the No Chat treatment. Furthermore, the number of Selfish group members is marginally significant. The own social preference type, on average, does not affect effort neither in the Chat ($p = 0.138$) nor in the No-Chat treatment.

Columns 4 and 5 present the results of the Chat treatment when additionally controlling for the emergence of a Min-Effort Leader. Controlling for the emergence of a Min-Effort Leader in a period, Selfish individuals choose higher effort on average than do Other-Regarding ones. Also having additional Selfish group members increases individual effort significantly. In column 5 we see that this is only the case in the absence of a Min-Effort Leader or before one emerges. Consistent with the results in Section 5.3 using our more conservative classification, social preferences matter before individuals manage to coordinate. Once the collusive outcome is reached, everyone cooperates and social preferences cease to affect effort.

Thus, overall we find that our results are relatively robust to this less stringent classification of Selfish. In particular, we find support for Result 1 and 4, while we do not find contradictory evidence to Result 2. Regarding Result 3 we still find that Selfish are more likely to be Min-Effort leaders (18% of Other-Regarding vs. 24% of Selfish are categorized as Min-Effort Leaders), though this difference becomes smaller and ceases to be significant. Finally, regarding Result 5, we again observe that groups with one Selfish group member are most likely to coordinate on minimum effort though also this difference is attenuated (67% of groups with no Selfish, 69% of groups with one Selfish, 33% of groups with 2 Selfish and 50% of groups with three Selfish successfully coordinate on minimum effort).

**Complements vs. Substitutes**

We use decision menus 1 to 9 (see Section 3 of the main paper 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 1. We denote player $i$'s monetary payoff as $\pi_i$ 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.

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

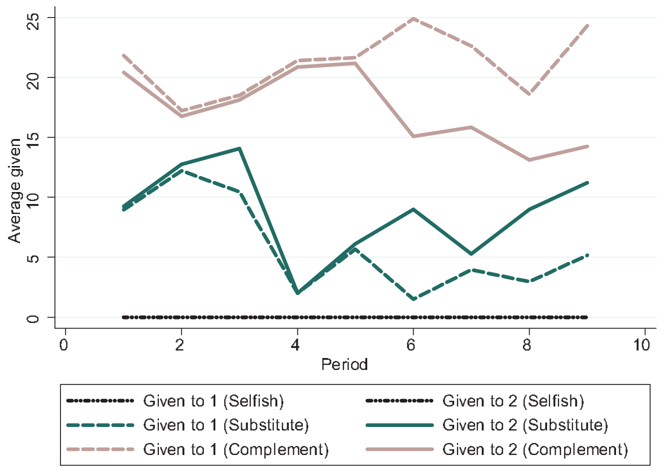



Figure 1: *Giving rates by social preference types.*

archetype's decision. We classify subjects as the archetype whose decision is closest to the subject's decision.[1] For treatments 1 and 2 we find that, for our subject population, 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 1 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

[1]Since we only use relative giving rates between the other two group members, our classification does not account for the intensity of social preferences. We can control for intensity separately by including the overall giving rate of a subject.

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

Table 5: *Overview of social preference types.*

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 of the main paper. Thus, as archetypal types would do, Complements react by allocating more to individual 1 while Substitutes react by allocating more to individual 2.

Table 6 provides details on the group level analysis 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 paper of the confound of leadership. We will again consider the effect of social preferences on leadership and explore whether it differs by Complements and Substitutes.

Table 7 presents 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 2 reports the distribution of social preferences among Non-Min-Effort Leaders and Min-Effort Leaders as defined in Section 5.3 of the main manuscript. 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).

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 for the Chat treatment.

We report these results in Table 8. 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 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

| | Chat<br>Avg Effort (Grp/Sess) | No Chat<br>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 6: *Group composition and average group effort.*

| | (1)<br>Effort | | (2)<br>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 7: *Effect of own and others social preferences on own effort (No Chat).*

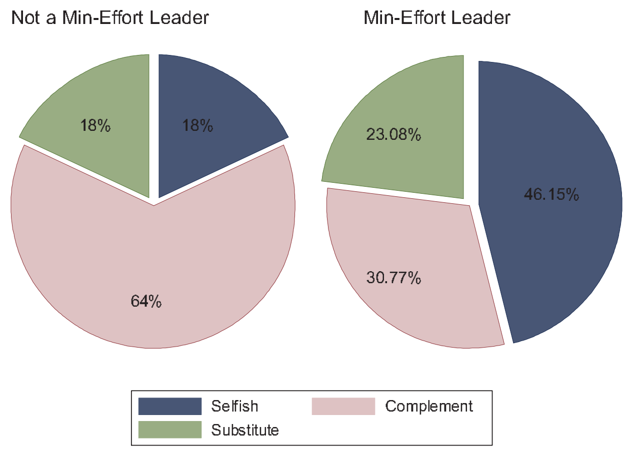



Figure 2: *The distribution of social preferences among Min-Effort Leaders and non-Min-Effort Leaders.*

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 9 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.

Using the category of Selfish/Other-Regarding as well as Jealous/Non-Jealous we construct 4 new social preference categories:[2]

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

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

| | (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 8: *Effect of social preferences (extended categorization 1) on individual effort controlling for leadership (Chat treatment).*

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

Table 9: *Average exchange rate chosen in menu 10 and 11.*

- Inequity Averse: Jealous and Other-Regarding (56%)

Table 10 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. 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 3 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

| | Chat<br>Avg Effort (Grp/Sess) | No Chat<br>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 10: *Group effort and Inequality Aversion (omitted category: Disinterested).*

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, we can separate the relation of social preferences and leadership emergence from general effort choices. Table 11 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 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.

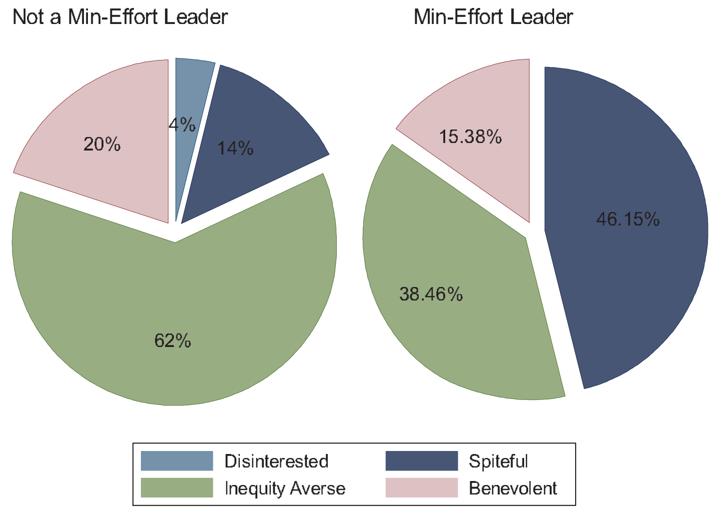



Figure 3: *Distribution of social preferences among non-Min-Effort Leaders and Min-Effort Leaders under extended categorization two.*

- 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.

## 1.2 Robot Treatment

This treatment is similar to the No Chat treatment in the sense that subjects cannot communicate but are permitted 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 each other, we paired them to two simulated subjects we call "robots." In particular, we programmed 42 robot subjects who react to past effort decisions by approximating what human 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 human subject did in previous No Chat treatments. Further below we present more technical details on the workings of the robots as well as how we tested them for the interested reader. Crucial to this treatment is that subject's effort choices no longer impose a negative externality on other players, since the robots receive no payoffs. Thus, the fundamental difference between the No Chat and the Robot treatment is that the latter

| | (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 11: *Effect of social preferences (extended categorization 2) on individual effort controling for leadership (Chat treatment).*

attempts to "turn off" subjects' social preferences since their actions no longer affect any other human. Note, however, that social preferences are not completely absent: 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.

We first compare subject behavior for the No Chat treatment and the Robot treatment graphically. Figure 4 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, and then 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). Interestingly, efforts of all social preference types in the Robot treatment converge towards the efforts of Selfish subjects in the No Chat treatment.

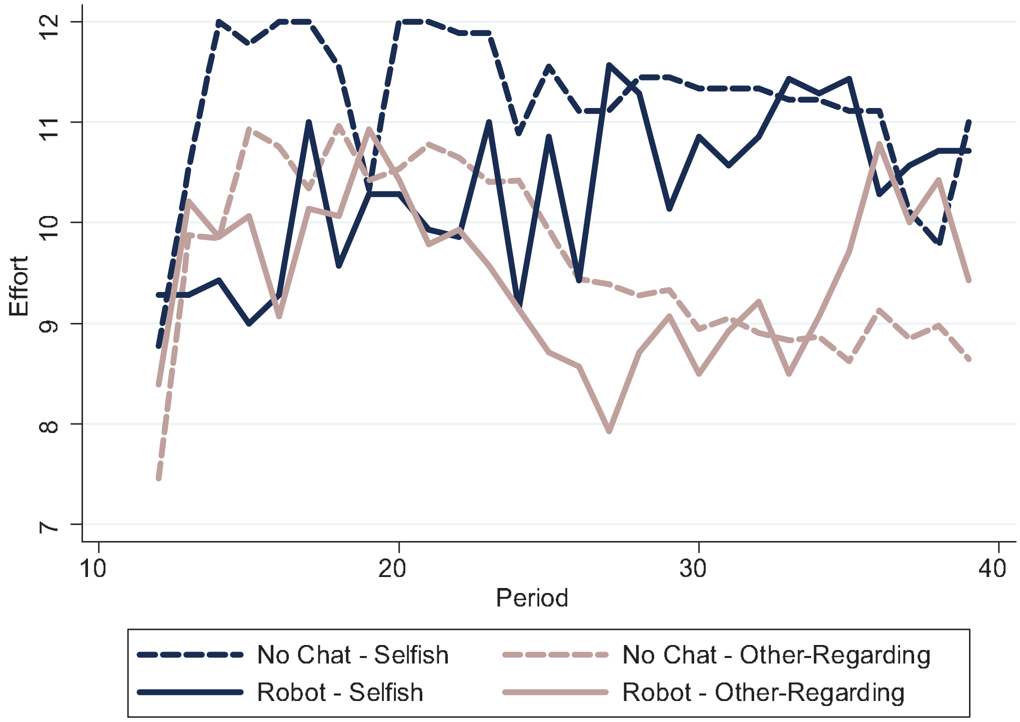



Figure 4: *Comparing efforts between Selfish and Other-Regarding types over time.*

Thus, while our 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 forgot that they were playing robot subjects and began behaving as if they were 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 that 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. Another possibility would be that Selfish and Other-Regarding differ in dimensions other than their social preferences, for example their skill of playing the game. This would only be consistent with our data in case that these differences become important over time, i.e. skill differences matter only with experience. In this case, we possibly over-estimate the effect of social preferences in later periods. In short, we find evidence for social preferences being the reason for differences in

| | All Periods | | Periods 30-40 | |
|---|---|---|---|---|
| | No Chat | Robot | No Chat | Robot |
| Period | -0.0538* | 0.0168 | -0.0303 | 0.109* |
| | (0.0294) | (0.0285) | (0.0448) | (0.0578) |
| Selfish | 1.478*** | 0.824 | 1.871** | 1.260 |
| | (0.401) | (0.813) | (0.765) | (0.981) |
| # Other Selfish | 0.569 | -0.280 | 0.858 | 0.0303 |
| | (0.412) | (0.996) | (0.678) | (1.081) |
| Constant | 10.85*** | 9.152*** | 9.717*** | 5.732** |
| | (0.502) | (0.685) | (1.805) | (2.411) |
| Observations | 1827 | 609 | 693 | 231 |
| $R^2$ within/between | 0.032/0.095 | 0.003/0.049 | 0.0025/0.0628 | 0.0353 /0.0616 |

Standard errors in parentheses

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

Table 12: *Effect of social preferences on individual effort No Chat vs. Robot treatment.*

behavior, though our results suggest that other factors possibly may play a role in later periods as well.

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 12 reports the results of regressing individual effort on individual's and group members' social preference types for the No Chat and the Robot treatment for all periods and periods 30-40. The coefficient estimate for Selfish is smaller in magnitude than in the No Chat treatment and is no longer significant, though we do note that the sample size is smaller for the Robot treatment.

These suggestive results from the Robot treatment provide evidence at least consistent with the idea that social preferences matter in creating and sustaining non-competitive efforts.

**Technical Details Robot Treatment**

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 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 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 that the human subjects experienced. For each of these 21 groups, we then ran 1000 repetitions of each group interacting over 29 periods. Table 13 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:$\leq 4$ |
| 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 13: *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 14 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 $\leq 4$ | 0.10 | 0.02 | 0.02 | 0.02 |
| 2: $\leq 4$, 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: $\leq 4$ | 0.13 | 0.10 | 0.07 | 0.06 |

Table 14: *Randomly matched groups (simulations).*

## 1.3 Lagged Effort

Here we present the individual level analysis of the main paper including lagged effort choices. Clearly, an individual´s effort choice will not only be determined by his or her social preference type, but also the history of effort choices of all group members. Thus we explore how taking into account past behavior affects the results of our individual level regression analysis and in particular our estimates of the effect of social preferences.

Table 15/Table 16 presents the results analogous to Table 6/Table 8 in the main manuscript, now including lagged own and other effort.

Other's lagged effort choices are significant and important predictors of individual effort choices in most specifications. Own last period effort is a significant predictor in the No-Chat treatment only. Our previous social preference parameters are still significant, although attenuated since we are now controlling for past choices.

## 1.4 Extremes Only Analysis

Table 17 presents the results of a random effects panel regression where we only compare subjects that on average kept less than 51 (Other-Regarding) with individuals that kept everything (Selfish). Thus we compare individual effort of these Selfish and Other-Regarding subjects controling for Period.

| | All Periods | | | Periods 30-40 | | |
|---|---|---|---|---|---|---|
| | Chat | No Chat | Pooled | Chat | No Chat | Pooled |
| Period | 0.00354 | -0.0350*** | -0.0164*** | -0.00855 | 0.0109 | 0.00202 |
| | (0.00675) | (0.00499) | (0.00550) | (0.0129) | (0.0173) | (0.0107) |
| L.Effort | 0.0580 | 0.509*** | 0.233*** | 0.0542 | 0.605*** | 0.293*** |
| | (0.0416) | (0.0812) | (0.0751) | (0.0358) | (0.131) | (0.100) |
| L.OAEffort | 0.436*** | 0.143*** | 0.329*** | 0.467*** | 0.156*** | 0.347*** |
| | (0.0606) | (0.0472) | (0.0531) | (0.0622) | (0.0603) | (0.0605) |
| L.OBEffort | 0.404*** | 0.136* | 0.300*** | 0.431*** | 0.160 | 0.300*** |
| | (0.0586) | (0.0765) | (0.0624) | (0.0751) | (0.103) | (0.0873) |
| Selfish | 0.226 | 0.581*** | 0.755*** | 0.120 | 0.425 | 0.713* |
| | (0.188) | (0.189) | (0.266) | (0.125) | (0.279) | (0.367) |
| # Other Selfish | 0.244 | -0.00943 | 0.0268 | 0.171 | -0.148 | -0.0799 |
| | (0.170) | (0.134) | (0.143) | (0.112) | (0.126) | (0.119) |
| Chat | | | -0.952*** | | | -0.246 |
| | | | (0.287) | | | (0.233) |
| Chat*Selfish | | | -0.606 | | | -0.651 |
| | | | (0.374) | | | (0.402) |
| Constant | -0.0249 | 2.992*** | 1.721*** | 0.290 | 0.273 | 0.336 |
| | (0.299) | (0.600) | (0.453) | (0.513) | (0.793) | (0.500) |
| Observations | 1764 | 1764 | 3528 | 693 | 693 | 1386 |
| $R^2$ within/between | 0.32/0.98 | 0.23/0.9 | 0.25/0.94 | 0.2/0.94 | 0.00/0.93 | 0.03/0.91 |

Standard errors in parentheses

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

Table 15: Effect of own and other's social preferences on own effort including lagged effort.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | | All Periods | | Per. 30-40 |
| Period | 0.00354 | -0.00452 | -0.00724 | -0.0156 |
| | (0.00675) | (0.00825) | (0.00788) | (0.0126) |
| L.Effort | 0.0580 | 0.00338 | -0.00638 | 0.00359 |
| | (0.0416) | (0.0423) | (0.0426) | (0.0315) |
| L.OAEffort | 0.436*** | 0.381*** | 0.371*** | 0.416*** |
| | (0.0606) | (0.0670) | (0.0713) | (0.0743) |
| L.OBEffort | 0.404*** | 0.351*** | 0.342*** | 0.380*** |
| | (0.0586) | (0.0553) | (0.0534) | (0.0781) |
| Selfish | 0.226 | 0.650*** | 0.897** | 0.741** |
| | (0.188) | (0.240) | (0.363) | (0.353) |
| # Other Selfish | 0.244 | 0.698*** | 1.087*** | 0.871** |
| | (0.170) | (0.216) | (0.318) | (0.346) |
| Min Effort Leader Exists | | -1.605*** | -1.021*** | -0.517** |
| | | (0.374) | (0.267) | (0.231) |
| Min Effort Leader | | 0.107 | 0.0947 | -0.0799 |
| | | (0.0974) | (0.104) | (0.0742) |
| MELeaderE*Selfish | | | -0.843** | -0.617* |
| | | | (0.368) | (0.356) |
| MELeaderE*#OthSelf | | | -1.063*** | -0.759** |
| | | | (0.333) | (0.345) |
| Constant | -0.0249 | 1.368*** | 1.432*** | 1.188* |
| | (0.299) | (0.420) | (0.442) | (0.662) |
| Observations | 1764 | 1764 | 1764 | 693 |
| $R^2$ within/between | 0.32/0.98 | 0.33/0.96 | 0.34/0.96 | 0.21/0.94 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 16: Effect of social preferences on individual effort controlling for leadership and including lagged effort (Chat treatment).

| | Chat | | No Chat | |
|---|---|---|---|---|
| Period | -0.0916*** | (0.0311) | -0.0543 | (0.0337) |
| Selfish | 0.572 | (0.782) | 1.698*** | (0.534) |
| Constant | 6.201*** | (1.245) | 10.90*** | (0.576) |
| Observations | 783 | | 841 | |
| $R^2$ within/between | 0.06/0.01 | | 0.04/0.17 | |

Standard errors in parentheses
$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 17: Effect of social preferences on individual effort (extremes classification).

# 2 Calculations: Hypotheses Section

Recall payoffs for individual $i$ are $\pi_i = \frac{x_i}{\bar{x}} \times 15 - x_i + 12$ in each round. The utility of subject $i$ is a combination of her payoffs and the payoffs of the other subjects in her group:

$$u_i = \rho_s \pi_i + \rho_o \sum_{k \neq i} \pi_k,$$

where $\rho_s$ is the weight placed on her own payoff and $\rho_o$ is the weight placed on the payoffs of others. We assume the following:

1. $\rho_s, \rho_o \in [0, 1]$, and $\rho_s + 2\rho_o = 1$.[3] Note $\rho_s = 1$ means subject $i$ is Selfish and $u_i = \pi_i$; $\rho_o > 0$ means subject $i$ is Other-Regarding.

2. $\rho_s > 2\rho_o$. (To focus on unique interior solutions in the analysis of the stage game.) Note that assumptions 1 and 2 imply $1/2 < (\rho_s - \rho_o)/\rho_s \leq 1$. This means that Other-Regarding subjects care more about their own payoff than the payoffs of the two other subjects combined. To ease notation, let us define $\Delta \equiv (\rho_s - \rho_o)/\rho_s$.[4]

3. Other-Regarding subjects have identical preferences. That is, $\rho_o$ is the same across Other-Regarding subjects.

4. $\rho_s$ and $\rho_o$ are common knowledge. We assume complete information and perfect monitoring. This simplification allows us to build our hypotheses drawing upon the theory of infinitely repeated games of complete information (e.g., Abreu,

[3] This assumption only serves to normalize the utility of an Other-Regarding subject to be comparable to a Selfish subject. Assuming weights adding up to an arbitrary number does not entail a qualitative change in the results of this section as long as the other assumptions hold.

[4] This also implies that $\rho_s > 1/2$, which is consistent with the results in Fisman et al. (2007, Figure 6) where the average "giving" parameter is above 1/2 in three person matchings.

1988).[5] In particular, we center our analysis on the fact that coordination can only be supported if players know deviations will be punished by switching from a high-payoff to a low-payoff continuation equilibrium.

Let us compare the effort choices by Selfish and Other-Regarding subjects when subjects' equilibrium profile is to play always stage-game Nash efforts. Before we proceed, note that an Other-Regarding subject 1 chooses $x_1$ to maximize

$$\rho_s\left(\frac{x_1}{x_1+x_2+x_3}W-x_1\right)+\rho_o\left(\frac{x_2}{x_1+x_2+x_3}W-x_2\right)+\rho_o\left(\frac{x_3}{x_1+x_2+x_3}W-x_3\right)$$

$$\iff (\rho_s x_1+\rho_o x_2+\rho_o x_3)\left(\frac{1}{x_1+x_2+x_3}W-1\right)$$

and a Selfish subject 1 chooses $x_1$ to maximize

$$x_1\left(\frac{1}{x_1+x_2+x_3}W-1\right).$$

Let us now express analytically the stage-game Nash efforts for each group configuration separately.

**1 Other-Regarding subject and 2 Selfish subjects**

As it is standard in these games, the best response functions in the stage game come from the First Order Conditions. Under Assumptions 1-4 (Second Order Conditions hold directly) stage-game Nash efforts are given by (without loss of generality, subject 1 is Other-Regarding)

$$\begin{aligned} x_1 &= \max\{\sqrt{\Delta W(x_2+x_3)}-(x_2+x_3)\,,0\} \\ x_2 &= \max\{\sqrt{W(x_1+x_3)}-(x_1+x_3)\,,0\} \\ x_3 &= \max\{\sqrt{W(x_1+x_2)}-(x_1+x_2)\,,0\} \end{aligned}$$

[5] A theoretical model that more closely relates to our experimental design is an indefinitely repeated game with incomplete information—because social preferences are private information. Such models, however, have received little attention arguably because of the technical challenge of tracking the evolution of beliefs over time (Bonatti, Cisternas and Toikka, forthcoming). Players may have incentives to manipulate others' beliefs (e.g., build a reputation) in addition to the incentives to sustain mutually beneficial outcomes through the threat of punishment (Forges 1992, Aumann et al 1968). Although results for the particular type of competition in the paper do not exist, the extant literature on oligopoly competition with privately known costs shows that first-best collusion can be exactly achieved given sufficiently little discounting (see, e.g., Athey and Bagwell 2008). With cheap talk communication, any payoff profile lying in the Pareto frontier that dominates an appropriately defined minmax value can be approximately attained in a perfect Bayesian equilibrium provided players are sufficiently patient (Escobar and Toikka, 2013). In other words, with communication it is possible to have coordination on the Pareto-optimal outcome even with incomplete information.

where $W = 3 \times 15 = 45$. (In what follows we suppress the $\max\{., 0\}$ because Assumption 2 allows us to consider interior solutions.) Stage-game Nash efforts solve the system of equations above and their analytical (interior) solutions are:

$$\begin{aligned} x_1 &= (2\Delta - 1)\frac{2\Delta}{(1+2\Delta)^2}W \\ x_2 = x_3 &= \frac{2\Delta}{(1+2\Delta)^2}W. \end{aligned}$$

**2 Other-Regarding subjects and 1 Selfish subject**

Stage-game Nash efforts are given by the solution to the following system of equations (Other-Regarding subjects are labeled 1 and 2):

$$\begin{aligned} x_1 &= \sqrt{\Delta W(x_2 + x_3)} - (x_2 + x_3) \\ x_2 &= \sqrt{\Delta W(x_1 + x_3)} - (x_1 + x_3) \\ x_3 &= \sqrt{W(x_1 + x_2)} - (x_1 + x_2). \end{aligned}$$

Stage-game Nash equilibrium efforts are given by

$$\begin{aligned} x_1 = x_2 &= \frac{2W}{\left(\frac{2-\Delta}{\Delta} + 2\right)^2} \\ x_3 &= \frac{2W}{\left(\frac{2-\Delta}{\Delta} + 2\right)^2}\left(\frac{2-\Delta}{\Delta}\right). \end{aligned}$$

**3 Other-Regarding subjects**

The game is symmetric, so stage game Nash equilibrium efforts are $2\Delta W/9$.

**3 Selfish subjects**

The game is also symmetric; stage-game Nash efforts are $2W/9$.

## Coordination and minimum continuation probability

In order to create Figure 1 in the main manuscript, we need to calculate the minimum continuation probability to sustain the grim-trigger equilibrium. The minimum continuation probability, $\delta^{min}$, is given by

$$\max_k \left\{ \frac{D^k - C}{D^k - P^k_{Ns}} \right\}.$$

Using the equilibrium expressions above, we compute the values of $D^k$, $C$ and $P^k_{Ns}$ for each type of player $k$ within each group configuration $Ns$. $C$ does not depend on the social preference type of the group composition. All subjects expend one and thus

$$C = \frac{1}{3}W - 1$$

$D^k$ only depends on whether the individual is Selfish or Other-Regarding. Assuming all other group members expend effort of one, while the deviator best-responds to this, we get

$$\begin{aligned}
D^o &= \left(\rho_s\left(\sqrt{2\Delta W} - 2\right) + 2\rho_o\right)\left(\frac{W}{\sqrt{2\Delta W}} - 1\right) \\
&= \rho_s\left(\sqrt{W} - \sqrt{2\Delta}\right)^2 \\
D^s &= (\sqrt{2W} - 2)\left(\frac{W}{\sqrt{2W}} - 1\right) \\
&= (\sqrt{W} - \sqrt{2})^2
\end{aligned}$$

Finally, $P^k_{Ns}$ depends both on the type of individual as well as the group members' types. Thus we go case by case.

**1 Other-Regarding subject and 2 Selfish subjects**

In this case $Ns = 2$. The subjects' stage-game Nash efforts are

$$\begin{aligned}
x_1 &= (2\Delta - 1)\frac{2\Delta}{(1+2\Delta)^2}W \\
x_2 = x_3 &= \frac{2\Delta}{(1+2\Delta)^2}W,
\end{aligned}$$

Stage-game utilities are given by

$$\begin{aligned}
P^o_{2s} &= \left(\rho_s \frac{2\Delta(2\Delta - 1)}{(1+2\Delta)^2}W + 2\rho_o \frac{2\Delta}{(1+2\Delta)^2}W\right)\left(\frac{W}{\sqrt{\Delta W(\frac{4\Delta}{(1+2\Delta)^2}W)}} - 1\right) \\
&= \frac{\rho_s}{(1+2\Delta)^2}W \\
P^s_{2s} &= \frac{2\Delta}{(1+2\Delta)^2}W\left(\frac{W}{\sqrt{\Delta W(\frac{4\Delta}{(1+2\Delta)^2}W)}} - 1\right) \\
&= \frac{1}{(1+2\Delta)^2}W
\end{aligned}$$

**2 Other-Regarding subjects and 1 Selfish subject**

Stage-game Nash equilibrium efforts are given by (subjects 1 and 2 are Other-Regarding)

$$\begin{aligned} x_1 = x_2 &= \frac{2W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2} \\ x_3 &= \frac{2W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\left(\frac{2-\Delta}{\Delta}\right), \end{aligned}$$

Stage-game utilities are given by

$$\begin{aligned} P^o_{1s} &= \left((\rho_s+\rho_o)\frac{2W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}+\rho_o\frac{2W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\left(\frac{2-\Delta}{\Delta}\right)\right)\left(\frac{W}{\sqrt{W\left(\frac{4W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\right)}}-1\right) \\ &= \frac{W\left(\frac{2-\Delta}{\Delta}\right)}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\left(\rho_s+\rho_o\left(\frac{2}{\Delta}\right)\right) \\ P^s_{1s} &= \frac{2W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\left(\frac{2-\Delta}{\Delta}\right)\left(\frac{W}{\sqrt{W\left(\frac{4W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\right)}}-1\right) \\ &= \frac{2W}{\left(\frac{2-\Delta}{\Delta}+2\right)^2}\left(\frac{2-\Delta}{\Delta}\right)^2 \end{aligned}$$

**3 Other-Regarding subjects**

Stage-game Nash equilibrium efforts are $2\Delta W/9$ and stage-game utilities are given by

$$P^o_{0s} = \frac{W}{9}(3-2\Delta)$$

**3 Selfish subjects**

Stage-game Nash equilibrium efforts are $2W/9$ and stage-game utilities are given by

$$P^s_{3s} = \frac{W}{9}$$

The values of $\delta^{min}$ for each group configuration and for different $\rho_s$ are plotted in Figure 1 in the paper.

**Credible Punishment**

To justify Hypothesis 4b in the paper, we argue that Selfish subjects have the least to lose (compared to Other-Regarding subjects) if the outcome is the stage-game Nash forever. Under our assumptions, coordinating on minimal efforts yields $C = 1$ for any type of subject. Utilities from stage-game Nash, $P^s_{Ns}$ and $P^o_{Ns}$, however, differ. In this appendix we show that $P^s_{Ns} > P^o_{Ns}$, that is, the utility in the punishment phase of a Selfish subject is higher than the utility of an Other-Regarding subject.

Consider heterogeneous groups, i.e., $Ns = 1$ and $Ns = 2$. Note first that the utility of an Other-Regarding subject (without loss of generality, subject 1) is $P^o_{Ns} = (\rho_s x_1 + \rho_o x_2 + \rho_o x_3)\left(\frac{1}{x_1+x_2+x_3}W - 1\right)$ and the utility of a Selfish subject in the same group (without loss of generality, subject 2) is $P^s_{Ns} = x_2\left(\frac{1}{x_1+x_2+x_3}W - 1\right)$. From these expressions, comparing $P^o_{Ns}$ and $P^s_{Ns}$ is the equivalent to comparing $(\rho_s x_1 + \rho_o x_2 + \rho_o x_3)$ and $x_2$ in a group with $Ns$ Selfish members:

$$\begin{aligned} P^o_{Ns} &< P^s_{Ns} \\ \iff \rho_s x_1 + \rho_o x_2 + \rho_o x_3 &< x_2 \\ \rho_s x_1 + \rho_o x_3 &< (1-\rho_o)x_2 \end{aligned}$$

Note that $x_1 < x_2$ (because $x_1$ is the effort of an Other-Regarding subject and $x_2$ is the effort of a Selfish subject in a group in which every subject plays stage-game Nash efforts) and, either $x_3 < x_2$ if $Ns = 1$ (subject 3 is Other-Regarding) or $x_3 = x_2$ if $Ns = 2$ (subject 3 is Selfish). Under Assumption 1, the result follows.

# 3 Examples of Decisions

We begin with some examples to illustrate subjects' behavior. Figure 5 illustrates the patterns of decisions across time. In the first stage (periods 1 to 9), we 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).[6] 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 (i.e., the effort choice stage).

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

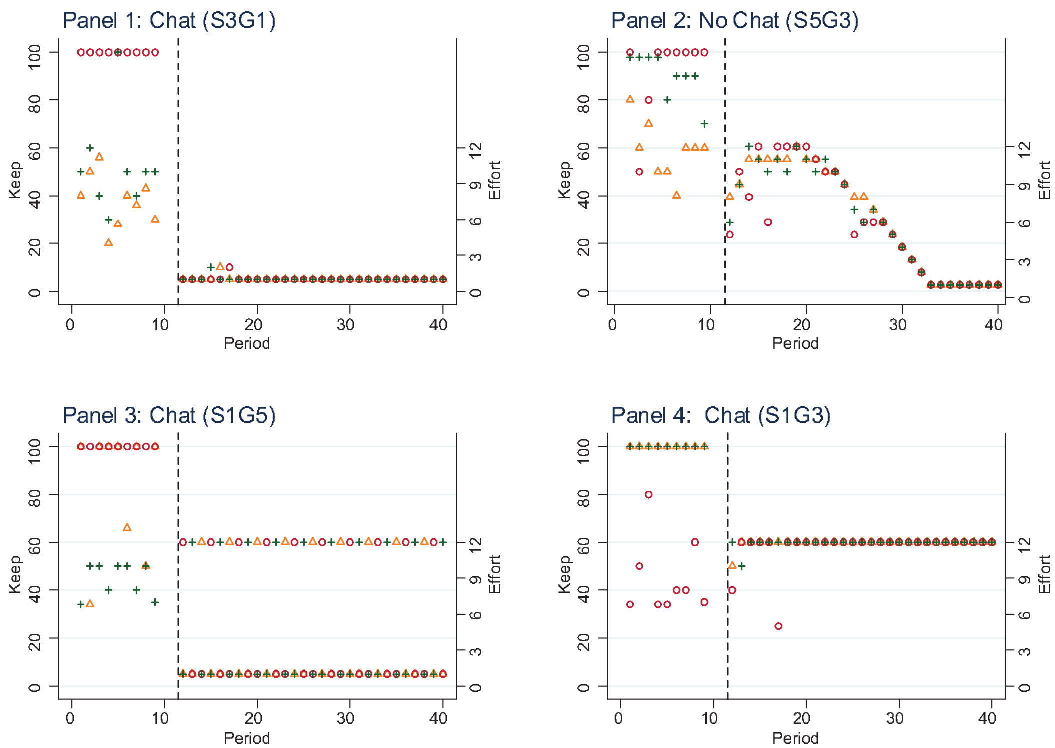



Figure 5: *Examples of group giving and investment decisions (S denotes session number and G group number).*

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 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 Subjectively Categorized Collusion

Figure 6 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.

# 5 Leader Classification Details

Attached file.

# 6 Instructions for Subjects

Attached file.

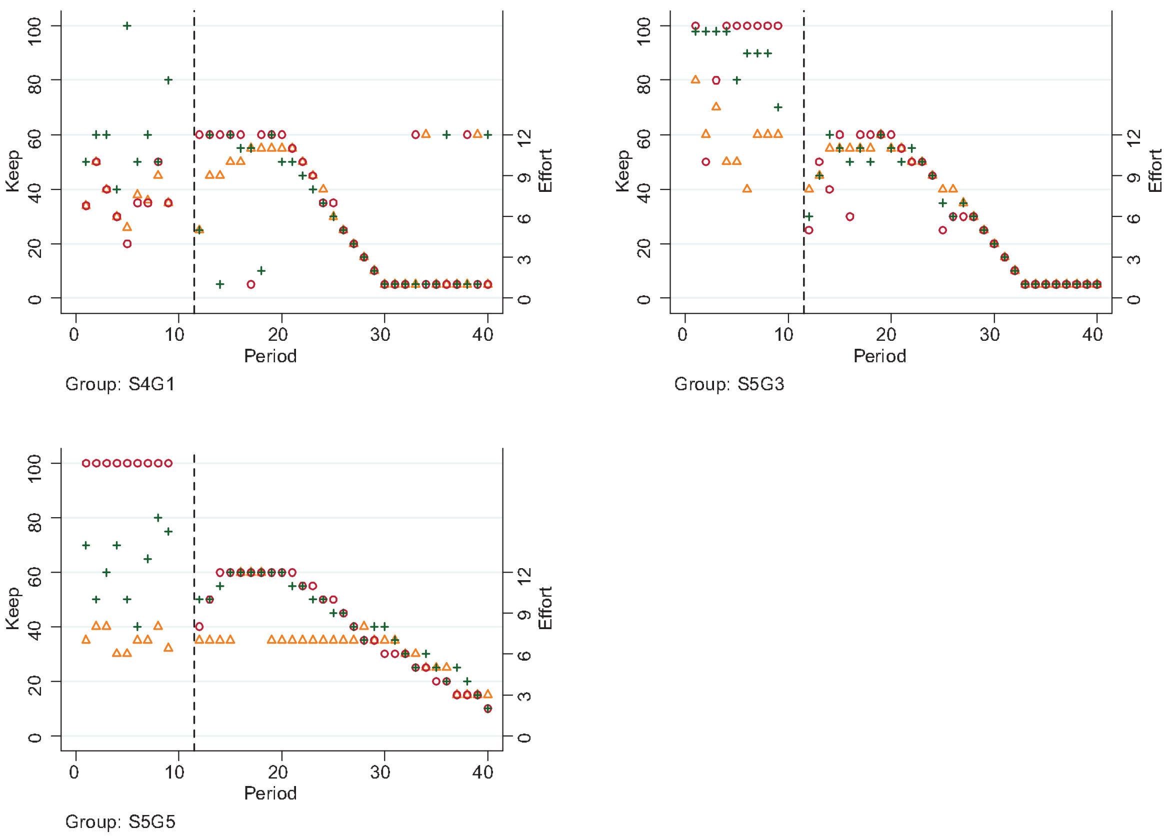


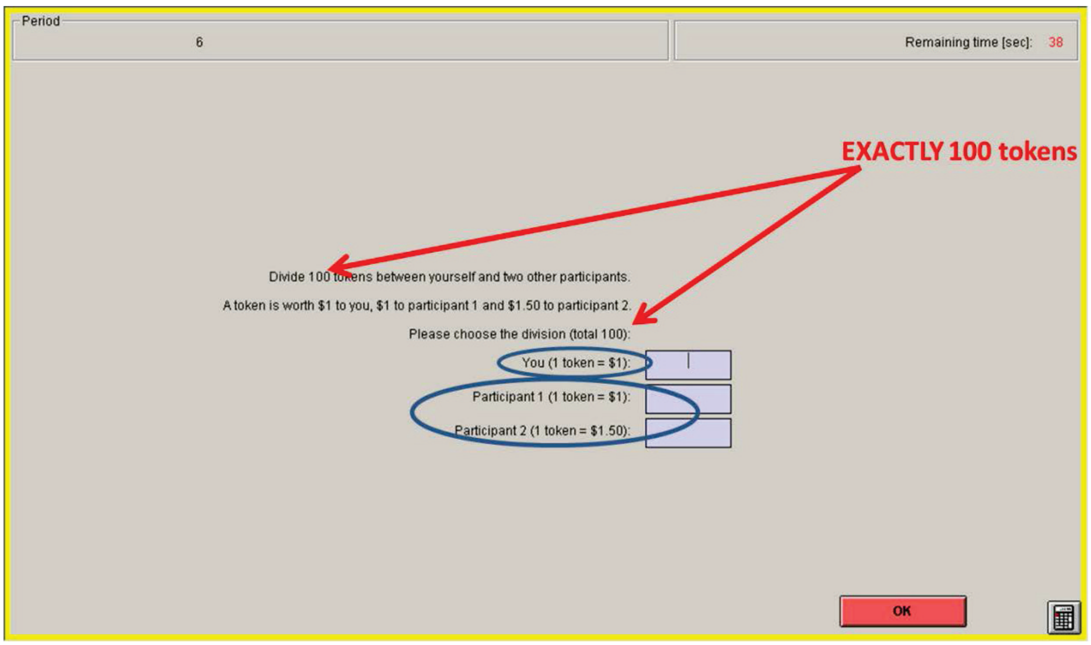

Figure 6: *Choices of groups classified as "colluding."*

*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>**

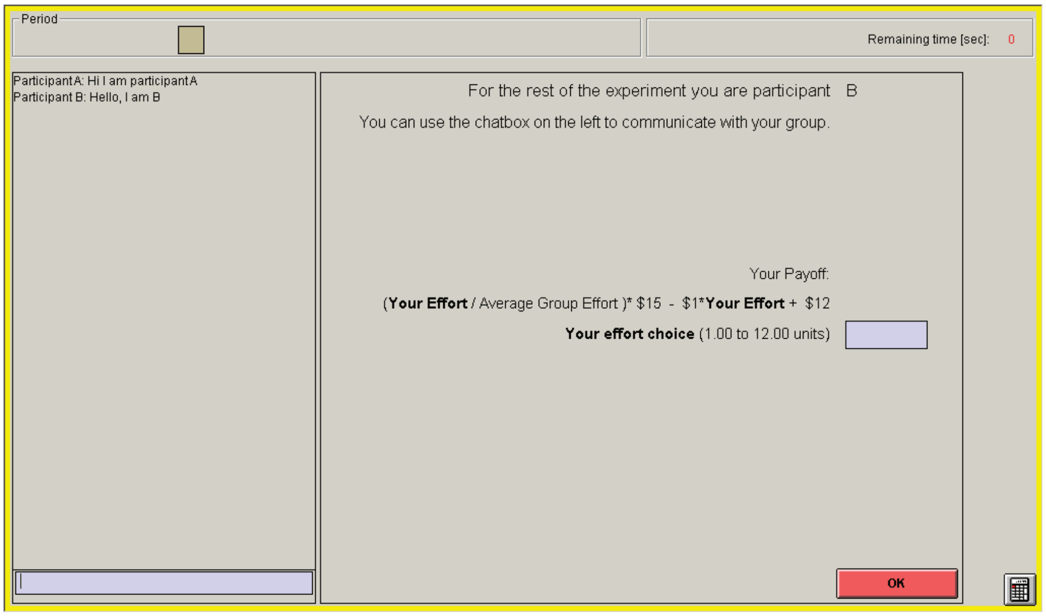
Screens 10-11 will be similar to the screen shown on the following Figure:


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:

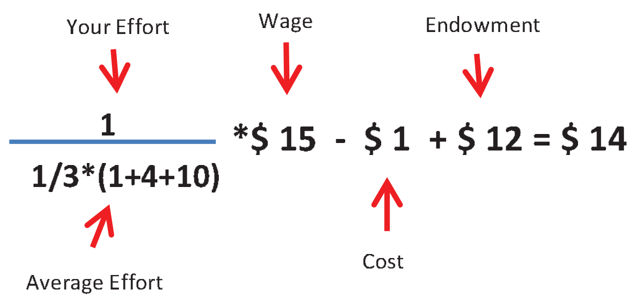



The total number of periods for this stage is unknown to all group members. Instead, there will be a 95% chance you will continue for another period.

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:



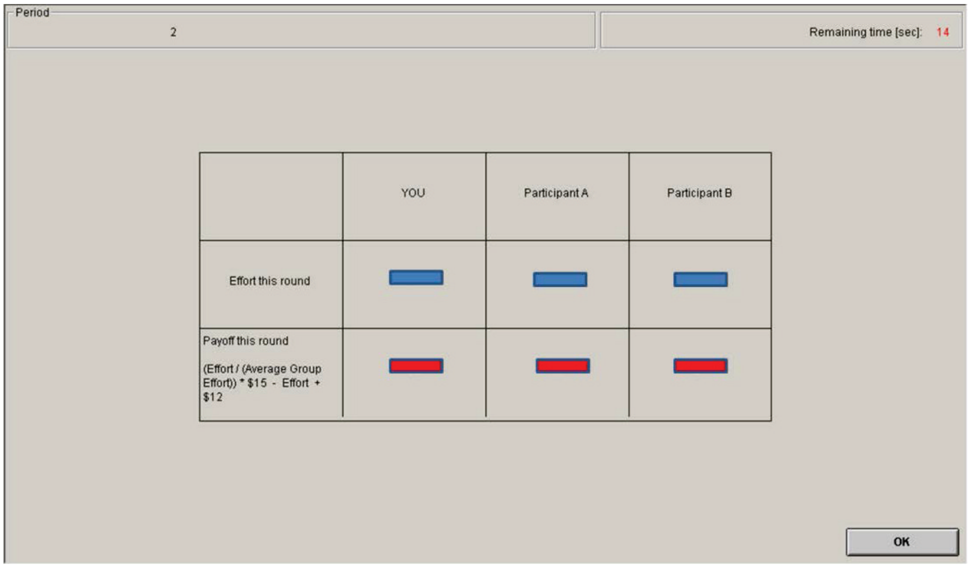

Notice that your effort DIVIDED by the average effort is equal to 1/5, so your relative compensation in this example would be 1/5 of $15 = $ 3. Hence, your total payoff would be $ 3 - $ 1 + $ 12 = $ 14.

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!

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 |