Exp Econ (2017) 20:809–835
DOI 10.1007/s10683-017-9512-9

ORIGINAL PAPER

Do people who care about others cooperate more? Experimental evidence from relative incentive pay

Pablo Hernandez-Lagos[1] · **Dylan Minor**[2] · **Dana Sisak**[3,4]

Received: 19 February 2015 / Revised: 17 January 2017 / Accepted: 19 January 2017 /
Published online: 28 February 2017

Abstract We experimentally study ways in which social preferences affect individual and group performance under indefinitely repeated relative incentives. We also identify the mediating role that communication and leadership play in generating these effects. We find other-regarding individuals tend to depress efforts by 15% on average. However, selfish individuals are nearly three times more likely to lead players to coordinate on minimal efforts when communication is possible. Hence, the other-regarding composition of a group has complex consequences for organizational performance.

Keywords Social preferences · Relative performance · Cooperation · Leadership

JEL Classification M52 · D03 · C7 · C9

1 Introduction

Relative performance incentives are a common feature of the workplace. An interesting property of relative pay is that a worker's performance also affects his or her co-workers' compensation; in particular, it imposes a negative externality. An

Electronic supplementary material The online version of this article (doi:10.1007/s10683-017-9512-9) contains supplementary material, which is available to authorized users.

✉ Pablo Hernandez-Lagos
pablo.hernandez@nyu.edu

[1] New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

[2] Northwestern University, Evanston, IL, USA

[3] Erasmus University Rotterdam, Rotterdam, The Netherlands

[4] Tinbergen Institute, Rotterdam, The Netherlands

increase in one's own performance will not only increase one's own compensation, but inevitably also decrease a co-worker's expected pay. How this externality affects the incentives of a worker will crucially depend on whether a worker incorporates this reduction in her own effort decision. However, it will also depend on other features of the workplace environment, including how easily communication can foster coordination, the likelihood of someone leading such coordination, and the expectation of future interactions.

In this paper, we consider how these forces might blunt the effectiveness of relative performance incentives. In particular, we use a controlled laboratory environment to examine two primary channels through which agents may reduce effort under relative incentives. The first one is "other-regarding" concerns: some agents may incorporate other agents' payoffs into their own effort choice. Thus, other-regarding agents should respond differently to relative incentives compared to "selfish" agents. Even though the fact that individuals have heterogeneous degrees of other-regardingness (e.g., see Andreoni and Miller 2002; Fisman et al. 2007) is well-documented in the literature, we know little about the effect of other-regarding concerns on the effectiveness of relative performance pay.

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 agent's behavior (e.g., see Dal Bó 2005). We consider this channel because the social preference composition of the group affects the severity of future punishment, hence the extent to which coordination can be sustained.

Important to exploring the channel of indefinitely repeated incentives is accounting for the role of communication and leadership. Coordinating on low efforts seems likely to be driven by the ease with which communication can happen (e.g., see Cooper et al. 1992). In addition, the potential for coordination may stimulate leadership emergence (e.g., see, Hermalin 2012; Kreps 1986). Although leaders make mutually beneficial outcomes focal in simple coordination games, we know little about their effect on agents' behavior in indefinitely repeated interactions within a relative performance incentive structure. Leaders in our setting are important since they can direct individuals towards low effort outcomes and their emergence may well be linked to social preferences.

Our experimental results indicate that overall groups with more other-regarding players tend to depress total efforts. When communication is not possible, outcomes in which all group members choose minimal efforts rarely occur and the average effort level is close to the stage game equilibrium prediction. Nonetheless, other-regarding group members depress their effort by around 15% relative to selfish group members. Thus, our results are consistent with other-regarding individuals internalizing the externality they impose without engaging in long-term strategic behavior.

With communication, a coordinating leader may emerge. In our particular setting, we label "leader" as any individual who suggests that the group should 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 other-regarding subjects depress their effort relative to selfish ones by about 50% before a leader emerges. We also find that selfish individuals are 2.7 times

more likely than other-regarding individuals to successfully lead their groups to the minimal effort outcome.

This implies that the effect of social preferences on work performance under relative incentives is complex. On the one hand, other-regarding workers have a tendency to depress effort, apparently through the 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.

We see the contributions of this paper as threefold. First, we document for the first time how individual social preferences affect behavior when facing relative performance incentives in indefinitely repeated settings. Second, we explore how the composition of a group in terms of individual social preferences affects outcomes. Third, we identify how communication and endogenous leadership mediate these effects as well as how social preferences relate to the emergence of coordinating leaders.

2 Literature

The significant body of literature that documents different degrees of social preferences (for example Andreoni and Miller 2002; Fisman et al. 2007; DellaVigna 2009) has led researchers to investigate their effects on public good contributions and other pro-social behaviors (e.g. Loch and Wu 2008; Dreber et al. 2014; Bowles and Polania-Reyes 2012; Kőszegi 2014). 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 et al. 2013; Erkal et al. 2011; Rey-Biel et al. 2012; 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 (i.e., 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 many important settings, such as the workplace. For indefinitely repeated settings it is not clear a priori whether other-regarding concerns will depress efforts due to internalizing the negative externality imposed on others or will instead increase efforts due to more lenient punishment in the case of a deviation, which makes sustaining a collusive outcome harder. Consequently, the effects of social preferences seen in indefinitely repeated games could be quite different from those captured through the other types of settings commonly found in the extant literature.

The importance of group composition in a dimension other than the degree of other-regardingness has been previously explored. Casas-Arce and Martínez-Jerez (2009) for example, find that relative performance incentives (tournaments in their setting) are less effective than piece rates when participants have heterogeneous abilities. 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 et al. (2005) allude to the role of social preferences in indefinitely repeated (or at least long-term) interactions. Although the core of Bandiera et al. (2005) is to compare workers' productivity under piece rate and relative incentives, they also document two results that are related to this paper. First, Bandiera et al. (2005) compare fruit pickers with the aforementioned incentive schemes in two different settings: one that allows peer monitoring and another one that does not. They find that relative compensation leads to lower productivity only when monitoring is allowed. They conclude that monitoring, not social preferences, drives down effort in their setting. The authors keep their monitoring technology and relative incentives constant throughout their study; they also do not exogenously vary their subjects' exposure to altruism. Second, Bandiera et al. (2005) find that workers with *social ties* depress effort. Social ties could capture social preferences; but they could also capture the salience of punishment should one "defect" from low efforts. As a result, although this study clearly showed that social ties can reduce efforts, it is unclear whether social preferences can do the same. 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 to identify the link between social preferences and behavior, both as a function of individual preferences and group composition.

Indefinitely repeated settings have been another important area of research: Pareto improvements over the one-shot Nash equilibrium can be obtained as equilibrium outcomes if the value of the future is high enough.[1] However, the fact that cooperation (or "collusion" in the context of competition) can be an equilibrium outcome does not guarantee that subjects will cooperate (Dal Bó and Fréchette 2011, 2014).[2] In fact, it has been documented that the majority of the time individuals do not achieve the Pareto-optimal outcomes (e.g., Palfrey and Rosenthal 1994 find cooperation rates from 29 to 40% in public goods games, and Dal Bó (2005) found cooperation rates of 38% in indefinitely repeated prisoner's dilemmas). Further, there has been a great variety of outcomes in this literature, some of which deviate from standard economic models (see Fudenberg et al. 2012). Our paper complements this work by documenting the role of individual and group social preferences on outcomes in indefinitely repeated games.

Although theoretically cheap talk communication does not rule out equilibria, empirically it has been found to facilitate coordination in indefinitely repeated games (Fonseca and Normann 2012; Embrey et al. 2013). One channel through which communication helps equilibrium selection in games of coordination is

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

[2] There is a fairly large experimental literature on collusion, mostly focused on exploring the effect of monitoring (see e.g. Aoyagi and Fréchette 2009; Duffy and Ochs 2009) and strategic uncertainty (see e.g. Blonski (2004)). Our focus is on the role of group composition in terms of social preferences on cooperation. For an updated survey on cooperation in infinitely repeated games see Dal Bó and Fréchette (2014).

through a leader, as argued by Kreps (1986) and Hermalin (2012). The theoretical economics literature on leadership has focused on how pre-imposed self-regarding leaders coordinate (e.g. Bolton et al. 2013), motivate (e.g. Rotemberg and Saloner 1993, 2000), and signal information through their actions (e.g. Hermalin 1998). The role of leaders in these studies is to overcome individuals' incentives to act against the interest of the group. Meanwhile, the experimental literature has focused on whether leaders foster cooperation in social dilemmas, mostly from Hermalin (1998) leading-by-example perspective. These studies have found that leaders indeed spur cooperation, often through reciprocity from followers.[3] To our knowledge one study, Koukoumelis et al. (2012), explores leadership through communication in a social dilemma. In their study, the authors exogenously assign the role of "communicator" to one group member in a finitely repeated voluntary contribution game. They find that this one-way "free-form" communication has a large positive effect on contributions. A growing experimental literature studies leaders without pre-imposed salience or authority in finitely repeated interactions (see e.g. Bruttel and Fischbacher 2010; Gächter et al. 2012; Kocher et al. 2013; Arbak and Villeval 2013). Also focusing on social dilemmas, this literature has documented that emergent leaders are motivated by efficiency concerns, social image or generosity, and generally contribute more than non-leaders. Our work complements this literature in that we explore the endogenous emergence of leaders in indefinitely repeated settings, and how this phenomenon relates to social preferences. In addition, whereas we primarily study leadership through communication, most of the other papers study leadership influence through actions and authority.

Finally, our work also contributes to the literature on communication in games with multiple equilibria (e.g. Cooper et al. 1992; Ledyard 1994; Seelya et al. 2007; Cooper and Kühn 2014); 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 min from reading instructions to subject payment, which averaged approximately $16 per subject. Participants were not allowed to take part in more than one session. The treatments were programmed and conducted using *z-Tree* developed by Fischbacher (2007).

We had the dual purpose of identifying subjects' social preferences and measuring their choices when facing a relative performance incentive scheme. In order to achieve

[3] See, for example, Meidinger and Villeval (2002), Gächter and Renner (2005), Güth et al. (2007), and Moxnes and Heijden (2003).

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.[4] 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 online appendix for this paper. Finally, since we follow the categorization of Andreoni and Miller (2002), we are thus exploring 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 1–9, to classify participants as "Selfish" or "Other-Regarding", consistent with our intended meaning used in Sect. 4. An archetypal Selfish type is only interested in his own monetary payoff and thus should never allocate any tokens to his or her group members. Thus we classify as Selfish all subjects that throughout rounds 1–9 do not allocate any tokens to another group member. The remainder of subjects are classified as Other-Regarding. We consider various other possible classifications in the analysis found in our online appendix; however, they provide little additional insight to this simple classification.

For the second stage, participants were again randomly matched with two other players for the remainder of the experiment.[5] The purpose of this stage was to give players the possibility to collude by jointly providing low levels of effort.[6] Thus, we

[4] 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 individual 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" interchangeably to represent non-selfish behavior.

[5] Since we randomly allocate subjects in the second stage, we do not guarantee an equal distribution of possible treatments. In particular, ex-post we find fewer subjects are Selfish, which results in fewer groups dominated by Selfish subjects. Thus, in Section 1.1 of the online appendix we loosen our definition of Selfish to obtain more balance between treatments, and find results consistent with our original results.

[6] We note also that since the relative performance stage game is played after the dictator game, it is possible that the dictator game could influence choices in the relative performance stage. However, since we are measuring the *relative* effort choices of Selfish and Other-Regarding players, for this potentially to be a problem it is this relative difference that must be influenced.

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 et al. (2012). In particular, our random draw resulted in 29 periods of relative-performance-pay play, which was then fixed for all subjects, in all treatments.

A subject′s per period payoff during this stage 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 $x_i \in [1, 12]$.[7] 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 et al. (2005).[8] Note these figures are in Berkeley Bucks $, converted at $66.6 Berkeley Bucks to 1 US$, which is how it was presented to subjects.[9] 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. Subjects were paid the sum of their earnings over all periods for this stage.[10]

The 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 the 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.

For the final stage, subjects completed a risk aversion test as in Holt and Laury (2002), and a basic demographic questionnaire.

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.

[7] 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.

[8] Note that this is mathematically the same as a Tullock contest played by risk-neutral individuals. That is, the principal has a total pool of 45 Berkeley Bucks to distribute across workers based on their relative performance.

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

[10] While paying for all rounds is consistent with our intended framework of an indefinitely repeated workplace interaction, it is also in line with standard practice in the experimental literature on infinitely repeated games, see e.g. Dal Bó and Fréchette (2014) or Sherstyuk et al. (2013). Furthermore, Sherstyuk et al. (2013) find that cooperation rates in a standard prisoner's dilemma are not significantly different when paying subjects cumulatively for all periods or paying for the last period only. Paying a random period, on the other hand, leads to lower cooperation rates and thus is consistent with random payment inducing a present bias.

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. We report the results of this treatment in the online appendix.

4 Hypotheses

Before turning to our results, we develop several hypotheses to guide our ensuing analysis. To ease exposition, we use the label Selfish to mean those individuals that only value their own payoff. In addition, we use the label Other-Regarding to denote those individuals that value both their own payoff and some fraction of their partners' payoffs.[11]

Let us start by studying the stage game. 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 ρ_s is the weight placed on her own payoff and ρ_o is the weight placed on the payoffs of others. We assume the following:

1. $\rho_s, \rho_o \in [0, 1]$, and $\rho_s + 2\rho_o = 1$.[12] 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 $\Delta \equiv (\rho_s - \rho_o)/\rho_s$.[13]
3. Other-Regarding subjects have identical preferences. That is, ρ_o is the same across Other-Regarding subjects.
4. ρ_s and ρ_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

[11] From now on we use the capitalized form of selfish and other-regarding to refer to our categorization.

[12] 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.

[13] 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.

Table 1 Nash equilibrium effort choices in the stage game by Selfish and Other-Regarding subjects across group's social preference composition

	3 Selfish	2 Selfish	1 Selfish	No Selfish
Other-Regarding	–	$(2\Delta - 1)\frac{2\Delta W}{(1+2\Delta)^2}$	$\frac{2W}{(\frac{2-\Delta}{\Delta}+2)^2}$	$\frac{2\Delta W}{9}$
Selfish	$\frac{2W}{9}$	$\frac{2\Delta W}{(1+2\Delta)^2}$	$\left(\frac{2-\Delta}{\Delta}\right)\frac{2W}{(\frac{2-\Delta}{\Delta}+2)^2}$	–
Average effort	$\frac{2W}{9}$	$\frac{2\Delta W}{3(1+2\Delta)}$	$\frac{2W}{3(\frac{2-\Delta}{\Delta}+2)}$	$\frac{2\Delta W}{9}$

$W = 15 \times 3 = 45$

(e.g., Abreu 1988).[14] 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.

Table 1 shows the stage-game Nash effort choices (the details of the calculations are in the online appendix). The first row shows the effort choices of Other-Regarding subjects for each group composition, the second row shows the effort choices of Selfish subjects for each group composition, and the third row shows the average effort across subjects for each group composition. Simple algebra confirms that Other-Regarding subjects' effort choices are decreasing in the number of Selfish individuals in the group. Let us call x^o_{0s} the highest of these efforts—the effort choice of an Other-Regarding subject in a group with no Selfish individuals. It is also direct to check that Selfish subjects choose lowest effort when all the group members are Selfish. Let us call x^s_{3s} this effort level. Note that $x^s_{3s} > x^o_{0s}$. As a result, Other-Regarding subjects choose lower efforts than Selfish subjects across groups when in each group the stage-game Nash efforts are played. This logic leads to our first hypothesis which compares individual behavior across social preferences.

Hypothesis 1 Given that the stage game is played, Selfish subjects exert more effort than Other-Regarding subjects.

Group composition also affects average effort choices. From the third row of Table 1 it follows that average efforts in a group are increasing in the number of Selfish subjects in the group. As a result, we hypothesize to observe the following:

[14] 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 et al. 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.

Hypothesis 2 Given that the stage game is played, average group effort choices are increasing in the number of Selfish group members.

In indefinitely repeated games, achieving Pareto-dominant outcomes is a well-known theoretical possibility—provided fixed partners and not too much discounting. It proves difficult, however, to sustain the Pareto-dominant outcome experimentally absent communication (see, e.g., Fonseca and Normann 2012). Evidence suggests that subjects usually either play always the stage-game Nash efforts or revert to stage-game Nash efforts after a few attempts to coordinate (Dal Bó and Fréchette 2014). Thus we expect Hypotheses 1 and 2 to apply when subject cannot communicate, i.e. in the No Chat treatment. Communication, however, is an important aspect in the real world and may serve as a mechanism to facilitate coordination on minimal (cooperative) efforts. In fact, Cooper and Kühn (2014) find that communication increases coordination in a related experiment. Therefore, we should expect that, with communication, chosen efforts are on average lower than chosen efforts when communication is not possible.

Hypothesis 3 Absent communication, average group effort choices are consistent with the stage game equilibrium. When communication is possible, group effort choices are reduced as subjects attempt to cooperate.

The effect of social preferences on the likelihood of cooperation can take different forms. First, we study how social preferences of group members affect the minimum continuation probability necessary to sustain the cooperative outcome. The minimum continuation probability can be used as a measure of the likelihood of cooperation as a lower number implies more "cases" where cooperation is sustainable in equilibrium. Second we discuss how social preferences may affect communication itself.

To study the possibility of cooperation, let us consider grim-trigger strategies in which each subject chooses the minimal effort in round 1 and continues cooperating until there is a defection, then each subject chooses the stage-game Nash effort forever.[15] We focus on this strategy profile not because we believe that subjects necessarily behave as this strategy prescribes, but rather to use it as a benchmark to state the next hypotheses of this paper.[16] A subject's decision to coordinate using a grim-trigger strategy depends on his own social preferences and the number of Selfish subjects in the group. Let us denote C^k_{Ns}, D^k_{Ns}, and P^k_{Ns} the per-period utilities from coordination on minimal efforts, best unilateral deviation and (stage-game Nash) punishment, respectively, for a subject $k = \{\text{Other-Regarding, Selfish}\}$ in a group with Ns Selfish subjects. Note that C^k_{Ns} does not depend on k or Ns, hence

[15] In principle, coordination can also occur on efforts different from (1, 1, 1). For example, for a group with one Selfish and two very Other-Regarding subjects (ρ_s close to 0.5) the joint utility maximizing outcome is for the Selfish individual to put in an effort slightly larger than 1. We focus on (1, 1, 1) as it is (a) joint profit-maximizing, (b) a Pareto-optimal outcome from the players' perspective but the worst outcome from the principal's point of view, (c) arguably very salient. Furthermore, this is also the most prevalent "collusive outcome" observed in our data.

[16] Cabral et al. (2014) develop their hypotheses using trigger strategies for the same reason. Dal Bó and Fréchette (2014) find that grim-trigger strategies (and also "always defect" and "tit-for-tat") are consistent with a great majority of observed behavior from previous studies.

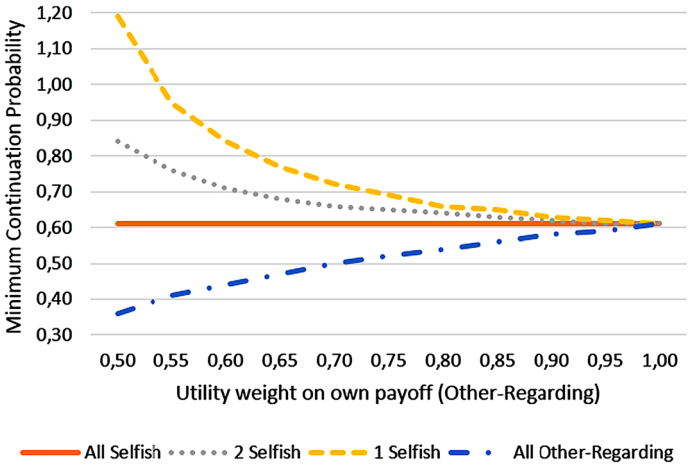

Fig. 1 Minimum continuation probability $\delta^{\min}$ needed to sustain cooperation

$C^k_{Ns} = C$. Similarly, D^k_{Ns} does not depend on the number of Selfish individuals, only on the social preferences of the subject, hence $D^k_{Ns} = D^k$ (the analytical expressions for D^k, C and P^k_{Ns} are given in the online appendix). Provided a continuation probability of δ this is an equilibrium in a group with Ns Selfish subjects if and only if

$$\delta \geq \delta^{\min} = \max_k \left\{ \frac{D^k - C}{D^k - P^k_{Ns}} \right\} \tag{1}$$

Inequality (1) makes explicit the core idea of this paper. For a fixed k, the harsher the punishment (i.e. the lower the per-period utility from the punishment phase, P^k_{Ns}) the smaller is the right-hand side of (1), which means that each k subject requires a lower continuation probability to cooperate. Given our previous analysis, one would be tempted to think that the right-hand side of (1) is decreasing in the number of Selfish subjects as they choose higher efforts in the punishment phase. As it turns out, this relationship is non-monotonic. Figure 1 plots the minimum continuation probability needed to sustain cooperation for a given value ρ_s, the Other-Regarding player's weight of another's payoff, for each of the four possible group configurations.

For example, fix the weight that an Other-Regarding subject puts on own payoffs at $\rho_s = 0.75$. The minimum required continuation probability is lowest (0.52) if there are no Selfish members in the group. This value, however, is highest (0.69) when there is only one Selfish member in the group. Note this is the case for any value of ρ_s.[17] Intuitively, the lone Selfish faces weak punishment in a group with two Other-Regarding and thus has a strong incentive to deviate. Therefore, we

[17] Note that in the group with 1 Selfish, if the two Other-Regarding subjects' weights on own payoff are close to 0.5, coordination is not an equilibrium even with a continuation probability of 95%.

should expect that the cooperative outcome is *harder* to sustain when there is only one Selfish subject in the group.

Hypothesis 4a Coordination on minimal effort is **least** likely for a group with only one Selfish player.

In stating Hypothesis 4a we are equating the success of communication with the size of the minimum continuation probability. Communication will be more likely to produce cooperation in groups with a lower $\delta^{\min}$. But social preferences may affect the success of communication and thus cooperation also through different channels more directly. Although theoretical details about the role of communication and social preferences on behavior remain elusive in settings like ours (see, e.g., Fonseca and Normann 2012), experimental evidence suggests that the effect of communication may depend on the group's social preference composition. In our setting, theory prescribes that both Other-Regarding and Selfish subjects could use the chat to communicate that they will be using grim-trigger strategies. Coordination will occur (provided it is an equilibrium) because such messages create incentives to the sender to fulfill them, if others consider the message credible (e.g., such messages are "self-committing", see Aumann 1990; Farrell and Rabin 1996, p. 114). However, although coordination yields utility C each period for everyone and forever, the utility each period from playing always stage-game Nash efforts, P^k_{Ns}, depends on the social preferences of the subject (k) for each group composition (Ns). Specifically, $P^s_{Ns} > P^o_{Ns}$ for $Ns = 1, 2$, i.e., Selfish subjects fare better than Other-Regarding subjects in the punishment phase.[18] In repeated games, successful communication should specify arguments in favor of coordination that require agreements not just on a single action but on entire contingent plans (see, e.g., Farrell and Rabin 1996; Cooper and Kühn 2014). Selfish subjects therefore may use the chat to convey they indeed have the least to lose if the outcome is stage-game Nash forever so their messages should be more credible. Credible threats of punishment have been found to be the most effective type of communication as they are associated with significantly more coordination (Cooper and Kühn 2014). In our setting, every player can use the chat to communicate, but promises of punishment made by Selfish subjects should be more credible than the promises made by Other-Regarding subjects, everything else equal. This logic, however, leads to a different prediction: Cooperation should be *easier* to sustain in groups with at least one Selfish subject. Thus, by means of initiating messages advocating for cooperation, Selfish subjects may lead others towards coordination more easily. This leads to our alternative hypothesis regarding the relationship of social preferences and likelihood of cooperation.

[18] To see this, note that u_i of an Other-Regarding subject can be written as $u_i = \left(\rho_s x_i + \rho_o \sum_{j\neq i} x_j\right)\left(\frac{W}{\sum_j x_j} - 1\right)$ and the utility of a Selfish subject k as $u_k = x_k\left(\frac{W}{\sum_j x_j} - 1\right)$. Writing the utility this way is useful because the term on the right is the same across subject types. Thus, comparing P^o_{Ns} and P^s_{Ns} is equivalent to comparing $(\rho_s x_i + \rho_o \sum_{j\neq i} x_j)$ and x_k in the same group. The result follows from the fact that the effort of an Other-Regarding subject x_i is less than x_k, one of the x_j is equal to x_k, and Assumption 1. The analytical expressions are in the online appendix.

Hypothesis 4b Coordination on minimal effort is **more** likely for a group with at least one Selfish player than a group with no Selfish players.

The broad results on communication in infinitely repeated settings preclude us from providing a clear prediction as to which of the two forces dominates. Thus, ultimately, we treat the question of which group composition is most frequently associated with coordination as an empirical one when communication is allowed.

5 Experimental results

We begin by classifying subjects by their social preferences derived from their giving behavior. We then use this classification to study the relationship of individual social preferences as well as group composition of social preferences and effort choices. We study the likelihood of observing a "collusive" outcome and how it relates to social preferences in Sect. 5.4. In Sect. 5.5 we present a short summary of the robustness checks we conducted.

5.1 Categorizing social preference types from giving menus

Table 2 summarizes the mean choices of our subjects under all 9 price vectors in the Chat and No Chat treatments. Vector (a, b, c) represents 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.

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. In Sect. 5.5 we report the results of an alternative classification, where a subject is classified as Selfish if he or she kept at least 90% of the endowment on average over the 9 rounds. Furthermore, we explore other, more detailed, categorizations of social preferences. Details for all of these can be found in the

Table 2 Giving rates

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	$\left(1,\frac{1}{2},\frac{1}{2}\right)$	73.93 (20, 100)	13.14	12.93
3	$\left(1,\frac{3}{4},\frac{3}{4}\right)$	72.27 (0, 100)	13.71	14.02
4	$\left(1,\frac{5}{4},\frac{5}{4}\right)$	71.88 (20, 100)	14.24	13.88
5	$\left(1,\frac{3}{2},\frac{3}{2}\right)$	70.28 (20, 100)	14.98	14.75
6	$\left(1,1,\frac{2}{3}\right)$	72.31 (30, 100)	16.44	11.25
7	$\left(1,1,\frac{3}{4}\right)$	73.51 (25, 100)	15.35	11.14
8	$\left(1,\frac{3}{4},\frac{1}{2}\right)$	77.48 (25, 100)	12.56	9.95
9	$\left(1,\frac{5}{4},\frac{3}{4}\right)$	72.32 (25, 100)	16.65	11.03

Table 3 Distribution of Selfish and Other-Regarding individuals by treatment

	No Chat	Chat	Total
Other-Regarding	54 (86%)	48 (76%)	102 (81%)
Selfish	9 (14%)	15 (24%)	24 (19%)
Total	63	63	126

Table 4 Distribution of groups by number of Selfish for each treatment

# Selfish	No Chat	Chat [with leader, without leader]	Total
0	14 (67%)	7 (33%) [3, 4]	21 (50%)
1	5 (24%)	13 (62%) [10, 3]	18 (43%)
2	2 (9%)	1 (5%) [0, 1]	3 (7%)
3	0	0	0
Total	21	21	42

online appendix. Using our Baseline categorization, and taking together the two treatments (Chat and No Chat) most of the participants (81%) are categorized as Other-Regarding. The remaining subjects (19%) are categorized as Selfish.[19]

As described in Sect. 3 subjects were randomly allocated into groups without regard to their social preference type. Table 3 shows the distribution of Selfish subjects across treatments. We observe both Selfish and Other-Regarding in each treatment. Table 4 shows the distribution of groups with different numbers of Selfish group members in the second stage. Since subjects were allocated randomly and Selfish subjects are relatively rare we do not observe groups with only Selfish group members in the 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.

5.2 Social preferences and effort

Figure 2 provides a summary of effort choices over time by treatment. In both the Chat and the No Chat treatment we observe average effort of around 8 units at the beginning of the relative incentives stage. As hypothesized in Sect. 4, 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 (i.e., 10) for the Selfish type (dotted line).

How do individual social preferences and group composition relate to efforts? To find an answer to this question we exploit the random allocation of subjects into

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

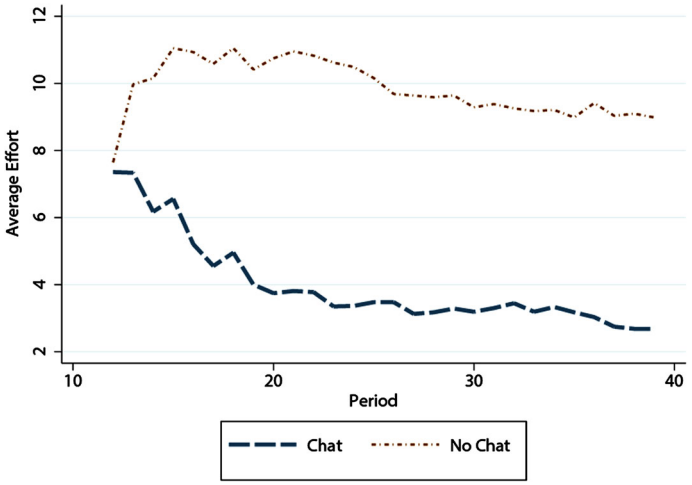

Fig. 2 Average effort by treatment over time

groups. We compare behavior of groups with different numbers of Selfish and Other-Regarding individuals in each of the two treatments.[20]

Figure 3 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).[21] Thus, we find that in the No Chat treatment, average efforts are similar to the stage game Nash equilibrium efforts (i.e., efforts of 10 for Selfish individuals) rather than to a cooperative outcome of (1, 1, 1) 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 Selfish group members versus those with fewer Selfish members; however, likely due to only one group with two Selfish members in the data, the difference does not reach statistical significance. Meanwhile, groups with only one Selfish member generate the lowest average effort.

We further explore differences in group average effort choices as a function of the number of Selfish subjects through regression analysis in Table 5. We use as the dependent variable the group effort averaged over all rounds of play (at stage 2, our relative performance stage) in columns 1, 2 and 3, and averaged over the final

[20] See the online appendix for examples of group giving and effort choices.

[21] This is consistent with Erkal et al. (2011) in that selfish individuals tend to exert higher levels of effort in tournaments.

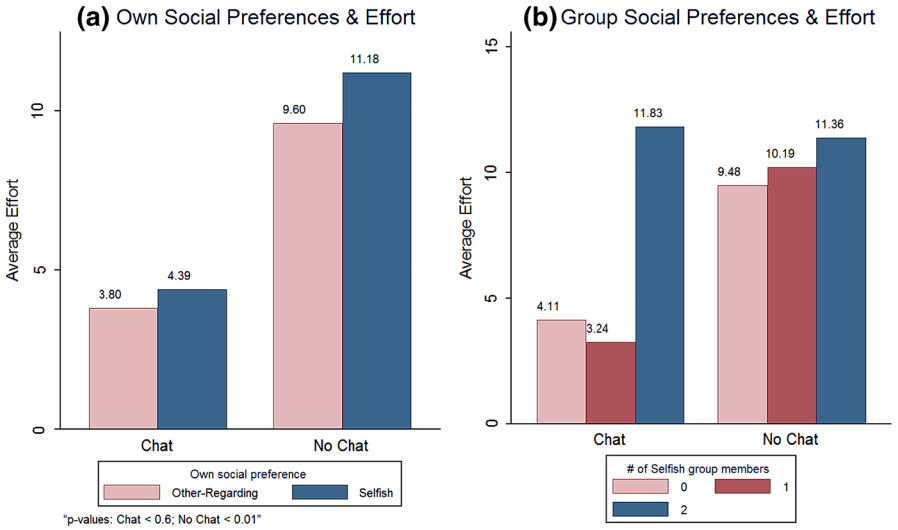

Fig. 3 Overview of effects of social preferences on effort

Table 5 Effect of group's social preference composition on group effort

	All periods			Periods 30–40		
	Pooled	Chat	No Chat	Pooled	Chat	No Chat
# Selfish	0.950	1.063	0.872**	1.438*	1.791	1.196*
	(0.699)	(1.626)	(0.379)	(0.786)	(1.610)	(0.687)
Chat	−6.159***			−6.532***		
	(0.670)			(0.812)		
Constant	9.420***	3.180***	9.453***	8.552***	1.769*	8.656***
	(0.484)	(1.022)	(0.440)	(0.764)	(0.925)	(0.793)
Observations	42	21	21	42	21	21
Adjusted R^2	0.606	−0.012	0.081	0.532	0.051	0.029

Standard errors in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

periods, periods 30–40 in columns 4, 5 and 6. Pooling both treatments together (columns 1 and 4) we find no effect of social preferences on average group effort over all periods, while there is a positive, marginally significant effect of the number of Selfish group members on group effort in the last periods. Each Selfish group member increases average group effort by about 1.4 units. Consistent with the literature on non-binding communication as a coordination device (see e.g. Cooper and Kühn 2014), the possibility to communicate significantly reduces average group effort by about 6.2 units relative to the No Chat treatment over all periods and 6.5 units in late periods only. This result supports Hypothesis 3. Splitting the data by

treatment, we do not find a significant effect of Selfish group members in the Chat treatment (columns 2 and 5). This is likely the result of greater effort from a group with two Selfish members cancelling out the reduced effort of the groups with only one Selfish member. In contrast, when communication is not possible (No Chat treatment, column 3), each Selfish group member increases average group effort by approximately 0.9 units, which equals a 9% increase over our baseline mean effort of roughly 9.7 per period. This effect is slightly larger in the last periods (column 6).

To disentangle the effect of an individual's social preferences from group composition effects, we estimate a random effects model, clustering standard errors at the group level.[22] The dependent variable is individual effort in a given period and the explanatory variables are: Selfish and the number of other Selfish individuals in each group (# Other Selfish). We control for # Other Selfish since, as given in Sect. 4, we expect Selfish and Other-Regarding group members to influence efforts differently.[23] This means that we are exploring the effect of individual social preferences conditional on how many other Selfish players are in one's group.

Table 6 reports our results. For the Chat treatment, all social preference variables remain insignificant (columns 1 and 4). On the other hand, we find further evidence that Other-Regarding subjects choose significantly less effort in the No Chat treatment both overall and in the last periods only (columns 2 and 5). Controlling for group composition, these subjects choose 1.5 fewer units of effort on average and 1.9 fewer units in the last periods. The group composition effect on the other hand, is positive but insignificant. Thus, absent communication, Other-Regarding subjects depress efforts relative to Selfish subjects, but only through the channel of individual social preferences. Finally, when pooling both treatments (columns 3 and 6), we again observe that the possibility of communication significantly reduces individual effort by about 6–6.5 units for both Selfish and Other-Regarding. Also, consistent with the analysis by treatment, Selfish individuals provide significantly more effort in the No Chat treatment than Other-Regarding ones, while this difference does not reach statistical significance in the Chat treatment. Testing for joint significance of the coefficients of Selfish plus Chat * Selfish the p values equal $p = 0.443$ for all periods and $p = 0.208$ for periods 30–40.

Overall, these results suggest that, absent communication, average efforts are consistent with stage game Nash equilibrium efforts, as hypothesized in Section 4, Hypothesis 3. This provides our first two primary results, which are consistent with our first two hypotheses:

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

[22] 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.

[23] We also conduct regressions where the effect of the number of other Selfish may depend on one's own social preference type as the theoretical model predicts. We did not find that this is the case and thus present the simpler specification here.

Table 6 Effect of own and others social preferences on own effort

	All Periods			Periods 30-40		
	Chat	No Chat	Pooled	Chat	No Chat	Pooled
Period	−0.133***	−0.0538*	−0.0932***	−0.0743**	−0.0303	−0.0523*
	(0.0276)	(0.0294)	(0.0209)	(0.0318)	(0.0448)	(0.0273)
Selfish	1.069	1.478***	1.434***	1.793	1.871**	1.788**
	(1.596)	(0.401)	(0.450)	(1.569)	(0.765)	(0.812)
# Other Selfish	1.060	0.569	0.805	1.790	0.858	1.307
	(1.581)	(0.412)	(0.800)	(1.575)	(0.678)	(0.863)
Chat			−6.063***			−6.490***
			(0.665)			(0.787)
Chat * Selfish			−0.479			−0.213
			(1.432)			(1.545)
Constant	6.628***	10.85***	11.82***	4.370***	9.717***	10.37***
	(1.471)	(0.502)	(0.554)	(1.630)	(1.805)	(1.233)
Observations	1827	1827	3654	693	693	1386
R^2 within/betw.	0.10/0.03	0.03/0.1	0.07/0.6	0.03/0.81	0.00/0.06	0.01/0.51

Standard errors in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Result 2 Absent communication, average group effort increases in the number of Selfish group members

When communication is introduced, however, efforts seem to more closely follow the cooperative outcome and results are somewhat surprising: The presence of one Selfish individual leads to lowest aggregate efforts. This is due to Selfish individuals being more likely to lead by suggesting coordination on low efforts, as we describe it in the next section.

5.3 Effort choices, 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 coordinating leader can then overcome the equilibrium selection problem. From the content of 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).[24] Thus a group can at most possess one Min-Effort Leader. We identify 13 Min-Effort Leaders (21%)

[24] 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.

among the 63 subjects (21 groups) in the Chat treatment and thus 13 out of 21 groups exhibit such a leader.[25]

How is leadership related to social preferences? We find that 7 of the 13 Min-Effort Leaders are Other-Regarding while 6 are Selfish. That means that 15% of all Other-Regarding are Min-Effort Leaders while 40% of all Selfish are Min-Effort Leaders. Thus, considering the likelihood of a given individual becoming a leader, Selfish individuals are more likely to be Min-Effort Leaders than Other-Regarding ones. A Pearson chi–squared test confirms that this difference is significant at the 5% level ($p = 0.03$).

Do social preferences affect outcomes in the Chat treatment beyond the likelihood of a Selfish subject emerging as a coordinating leader? To answer this question, we again perform group and individual level regressions, now controlling for the emergence of a Min-Effort Leader. Table 7 presents the results of the group level analysis. Group average effort, averaged over all periods (columns 1–2) and periods 30–40 (columns 3–4) is regressed on the number of Selfish in a group, whether the group exhibited a Min-Effort Leader and their interaction. Groups in which a Min-Effort Leader emerged have significantly lower effort levels. Furthermore, overall, an additional Selfish group member increases efforts by about 2 units, significantly so in the last 11 periods. Controlling for whether a Min-Effort Leader emerged in a group (columns 2 and 4) we find that an additional Selfish group member increases average effort only when no Min-Effort leader emerged in a group. In groups with a Min-Effort Leader the number of Selfish group members does not affect average effort. In these groups, the sum of coefficients on number of Selfish and Min-Effort Leader in Group equals -0.441 for all periods and -0.356 for the later periods and p values are 0.657 for all periods and 0.506 for late periods, respectively.

To complement the group-level analysis, Table 8 reports the results of a random effects model exploring individual effort choices. Column 1 shows a regression without considering leader emergence, analogous to the results reported in Table 6. In column 2 we add a control for whether a Min-Effort Leader has emerged and whether the subject herself is 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 influencing 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,

[25] 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 online 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.

Table 7 Effect of social preferences on group effort controlling for emergence of a Min-Effort Leader (Chat treatment)

	All periods		Periods 30–40	
# Selfish	1.607	2.827***	2.356**	3.971***
	(0.961)	(0.851)	(0.973)	(0.719)
Min-Effort Leader in group	−4.786***	−2.448*	−4.974***	−1.879**
	(1.040)	(1.166)	(1.051)	(0.809)
MELeaderGr * # Selfish		−3.268**		−4.326***
		(1.297)		(0.889)
Constant	5.754***	4.992***	4.444****	3.435***
	(0.775)	(0.730)	(0.799)	(0.632)
Observations	21	21	21	21
R^2	0.640	0.724	0.690	0.822

Standard errors in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 8 Effect of social preferences on individual effort controlling for leadership (Chat treatment)

	(1) All periods	(2)	(3)	(4) Per. 30–40
Period	−0.133***	−0.0725***	−0.0728***	−0.0567**
	(0.0276)	(0.0250)	(0.0245)	(0.0234)
Selfish	1.069	2.054***	2.797***	3.857***
	(1.596)	(0.737)	(0.687)	(0.612)
# Other Selfish	1.060	2.067***	2.864***	3.846***
	(1.581)	(0.694)	(0.600)	(0.734)
Min-Effort Leader Exists		−5.709***	−3.661***	−2.713***
		(0.637)	(0.423)	(0.422)
Min-Effort Leader		0.0784	0.107	0.0109
		(0.350)	(0.338)	(0.0589)
MELeaderE * Selfish			−2.729***	−3.555***
			(0.678)	(0.645)
MELeaderE * #OthSelf			−2.800***	−3.539***
			(0.562)	(0.779)
Constant	6.628***	7.353***	6.911***	5.589***
	(1.471)	(0.741)	(0.789)	(1.032)
Observations	1827	1827	1827	693
R^2 within/betw.	0.10/0.03	0.21/0.74	0.21/0.78	0.03/0.81

Standard errors in parentheses

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

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 emerged in a group. At the beginning of the game Selfish subjects expend an additional 2.8 units of effort relative to Other-Regarding subjects. Also, having an additional Selfish group member increases individual effort by about the same amount. Once a leader emerges there is no difference between Selfish and Other-Regarding choices. Formally, testing the joint significance of the coefficient on Selfish and the coefficient on the interaction MELeaderE * Selfish gives a p value of 0.87 and a joint coefficient of negligible magnitude of 0.068. Testing the joint significance of the coefficient on # Other Selfish and MELeaderE * # OthSelf gives a p value of 0.87 as well. 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. Column 4 reports estimates from only the last 11 periods of play and finds results similar to those reported in column 3.

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 use communication to coordinate the group on a cooperative outcome, consistent with Hypothesis 3. Such a "leader" tends to be a Selfish individual. This explains why the presence of only one Selfish individual is associated with lower efforts in the Chat treatment (see Fig. 3b). On the other hand, controlling for the existence of a leader in the group, Other-Regarding subjects have a tendency to put in lower effort than their Selfish counterparts before a coordination leader emerges, exactly as in the No Chat treatment, suggesting these individuals internalize the externality their effort inflicts on their group members. From a principal's perspective these results suggest that in a work environment where communication is possible a heterogeneous social-preference group leads to the lowest work effort: adding a Selfish subject to an otherwise Other-Regarding group of workers could promote the emergence of a leader to coordinate on low efforts.[26] Finally, once a coordination leader emerges and is successful, both Selfish and Other-Regarding workers are providing the same minimal effort, which means that there is no longer a difference between their efforts. Thus, our analysis yields two additional results:

Result 3 Consistent with Hypothesis 4b, Selfish subjects are more likely to lead others to coordinate on low efforts.

Result 4 Without/before the emergence of a coordination leader, Other-Regarding subjects depress efforts relative to Selfish subjects, consistent with Hypothesis 1. When a leader emerges, there are no differences in effort choices between Other-Regarding and Selfish subjects

[26] We note that we do not observe the other possible homogenous group of only Selfish members. Thus our comparison for homogeneous is for those groups only containing Other-Regarding members. We suspect that in practice this unobserved group in our experiment is a rarely occurring group.

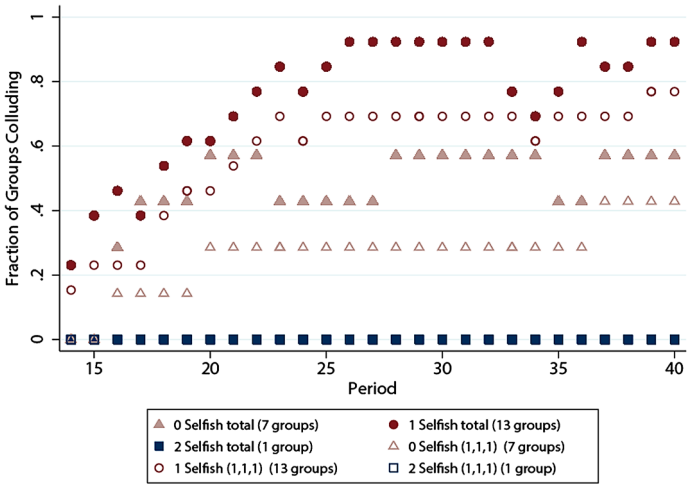

Fig. 4 Fraction of groups achieving (1, 1, 1) for 3 successive rounds of play (*hollow symbols*) and (1, 1, 1) or alternating (1, 1, 12) for 3 successive rounds (*solid symbols*) by number of Selfish group members for the Chat treatment

5.4 Propensity to cooperate

Thus far we have been focusing on the relationship between social preferences and depressed efforts. Depressed efforts can of course also be a consequence of coordination (or collusion if we interpret our design as a competitive setting). While we are naturally unable 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)]. 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.[27]

Figure 4 depicts the dynamics 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 Sect. 5.3, when chat is available, groups with 1 Selfish member are more likely to exhibit collusive outcomes 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 cooperative outcome than groups with no Selfish members. Note though that the fraction of groups choosing the turn-taking strategy (1, 1, 12) is

[27] Analyzing the chat messages reveals two reasons for the occurrence of this coordinated strategy. Some groups were of the opinion that this was in fact the profit maximizing strategy to take. For other groups taking turns on choosing maximal effort was used to "make things even" after one subject deviated from the collusive outcome of (1, 1, 1).

similar for groups with one or no Selfish group member, which means that this outcome does not seem to be related to social preferences.

Comparing the results in Fig. 4 with the results in Fig. 3 leads to an interesting observation. Even though groups with one Selfish member are more likely to cooperate, average effort is quantitatively not very different from a group with no Selfish (3.2 vs. 4.1). As already explained in Sect. 5.3 the reason for this 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 complex ways when communication is possible: Selfish individuals play an important role in facilitating coordination while Other-Regarding have a tendency to put in lower efforts even without coordination (Hypothesis 1). Thus, we summarize our final primary result, which is consistent with Hypothesis 4b, but contradicts Hypothesis 4a:

Result 5 With communication, coordination on minimal effort is more likely for a group with only one Selfish player.

For the No Chat treatment, coordinating on a "collusive outcome" was more difficult, since subjects were not able to chat. As shown in Table 9, we find for this setting that only 1 out of 21 groups end up with minimum efforts in the last 3 periods and only if the group has no Selfish members. One other group with no Selfish group members managed to sustain (1, 1, 1) for 3 periods during the course of the game, but then reverted back to higher effort. If we expand the definition of "collusive" outcome to include two subjective cases of "collusion" (we report their behavior in the online appendix), then we find one additional group with no Selfish members and one additional group with 1 Selfish member successfully "collude" by the end of the game. Again, it seems that collusion is not a main driver of behavior in this treatment and results seem more consistent with the predictions of the stage game.

5.5 Robustness checks

We performed a number of robustness checks on which we report in more detail in the online appendix. First, our regression results using our baseline classification are robust to clustering standard errors at the individual level instead of the group level in our individual-level analysis. Second, we relaxed the definition of Selfish to

Table 9 Propensity to "collude" by # of Selfish in the No Chat treatment

# 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

include all individuals who kept 90% or more on average. Using this classification, 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 is now 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.

We also explored an additional continuous social preference measure. In particular, we conducted individual-level regressions using the average endowment keep in rounds 1–9 directly in our regressions. Our results are qualitatively unchanged. 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. Furthermore, none of these additional controls show consistent patterns throughout the analysis. Finally, in order to create a starker contrast between groups, we compare average efforts of Selfish to Other-Regarding individuals who on average *kept less than 50% of their endowment*. To do this we ran a random effects regression of individual effort on Period and an indicator whether that individual is Selfish for the Chat and No Chat treatment, dropping all subjects that kept between 50 and 100% of their endowment. Consistent with our baseline results in the Chat treatment there is no significant effect (not controlling for leadership) while in the No Chat treatment Selfish individuals expend around 1.7 units more effort than Other-Regarding individuals.

Since the environment we study is dynamic with fixed matching, subjects can respond to past effort choices of their group members. Controlling for the social preferences of the group members can account for some of this path dependence in our analysis, though it is clearly imperfect. Thus, we finally conduct our analysis including lagged effort choices of all group members. Both own and others' lagged effort are significant and important predictors of individual effort choices. Nonetheless, our previous social preference parameters are still significant, although attenuated since we are now controlling for past choices.

6 Conclusion

We studied how an important dimension of worker heterogeneity affects the performance of those subject to relative performance incentives. 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 complex way. First, subjects categorized as Selfish are more likely to coordinate their group members to exercise 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 heterogeneous 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 people's payoffs.

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 unlikely to be as effective as they are 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 that 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.

Acknowledgements 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 Bó, Josse Delfgaauw, Robert Dur, Dirk Engelmann, Sacha Kapoor, Martin Kolmar, John Morgan, Felix Vardy, Bauke Visser and two anonymous referees. The authors thank the UC Berkeley Xlab (protocol 2011-04-3183) for financial support. Dana Sisak gratefully acknowledges the financial support of the Swiss National Science Foundation through Grant PBSGP1-130765.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica*, *56*(2), 383–396.

Athey, S., & Bagwell, K. (2008). Collusion with persistent cost shocks. *Econometrica, 76*(3), 493–540.

Aumann, R. (1990). Nash equilibria are not self-enforcing. In J. J. Gabszewicz, J. F. Richard & L. A. Wolsey (Eds.), *Economic decision making: Games, econometrics and optimisation*. Amsterdam: Elsevier.

Aumann, R., Maschler, M., & Stearns, R. E. (1968). *Repeated games of incomplete information: An approach to the non-zero-sum case, Ch IV*. Princeton: Mathematica ST-143.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.
Aoyagi, M., & Fréchette, G. (2009). Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic Theory*, *144*(3), 1135–1165.
Arbak, E., & Villeval, M. C. (2013). Voluntary leadership: Motivation and influence. *Social Choice and Welfare*, *40*(3), 635–662.
Backes-Gellner, U., & Pull, K. (2013). Tournament compensation systems, employee heterogeneity, and firm performance. *Human Resource Management*, *52*(3), 375–398.
Bandiera, O., Barankay, I., & Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, *120*(3), 917–962.
Blonski, M., & Spagnolo, G. (2004). *Prisoners' other dilemma*. SSE/EFI working paper 437.
Bolton, P., Brunnermeier, M. K., & Veldkamp, L. (2013). Leadership, coordination, and corporate culture. *The Review of Economic Studies*, *80*(2), 512–537.
Bonatti A., Cisternas, G., & Toikka, J. (2016). Dynamic oligopoly with incomplete information. *The Review of Economic Studies*. doi:10.1093/restud/rdw049.
Bowles, S., & Polania-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*, *50*(2), 368–425.
Bruttel, L., & Fischbacher, U. (2010). *Taking the initiative. What motivates leaders?* TWI Research Paper Series 61, Thurgauer Wirtschaftsinstitut, Universität Konstanz.
Cabral, L., Ozbay, E. Y., & Schotter, A. (2014). Intrinsic and instrumental reciprocity: An experimental study. *Games and Economic Behavior*, *87*, 100–121.
Casas-Arce, P., & Martínez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science*, *55*(8), 1306–1320.
Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1992). Communication in coordination games. *The Quarterly Journal of Economics*, *107*(2), 739–771.
Cooper, J. D., & Kühn, K.-U. (2014). Communication, renegotiation, and the scope for collusion. *American Economic Journal: Microeconomics*, *6*(2), 24778.
Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *The American Economic Review*, *5*, 1591–1604.
Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *The American Economic Review*, *101*(1), 411–429.
Dal Bó, P., & Fréchette, G. R. (2014). *On the determinants of cooperation in infinitely repeated games: A survey*. Available at SSRN 2535963.
DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–372.
Dreber, A., Fudenberg, D., & Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior and Organization*, *98*, 41–55.
Duffy, J., & Ochs, J. (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior*, *66*(2), 785–812.
Embrey, M. S., Fréchette, G. R., & Stacchetti, E. (2013). *An experimental study of imperfect public monitoring: Renegotiation proofness vs efficiency*. Working paper.
Escobar, J. F., & Toikka, J. (2013). Efficiency in games with Markovian private information. *Econometrica*, *81*(5), 1887–1934.
Erkal, N., Gangadharan, L., & Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, *101*(3), 3330–3348.
Farrell, J., & Rabin, M. (1996). Cheap talk. *The Journal of Economic Perspectives*, *10*(3), 103–118.
Fehr, E., & Fischbacher, U. (2002). Why social preferences matter—The impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal*, *112*, C1–C33.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.
Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review, 100*(1), 541–556.
Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*(5), 1858–1876.
Fonseca, M. A., & Normann, H. T. (2012). Explicit vs. tacit collusion—The impact of communication in oligopoly experiments. *European Economic Review*, *56*(8), 1759–1772.

Forges, F. (1992). Repeated games of incomplete information: Non-zero-sum. *Handbook of game theory with economic applications, 1*, 155–177.
Friedman, J. (1971). A noncooperative equilibrium for supergames. *The Review of Economic Studies*, *38*, 1–12.
Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or incomplete information. *Econometrica*, *54*, 533–554.
Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, *102*(2), 720–749.
Gächter, S., Nosenzo, D., Renner, E., & Sefton, M. (2012). Who makes a good leader? Cooperativeness, optimism and leading-by-example. *Economic Inquiry*, *50*(4), 867–879.
Gächter, S., & Thöni, C. (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association*, *3*(2–3), 303–314.
Gächter, S., & Renner, E. (2005). *Leading by example in the presence of free rider incentives*. CeDEx discussion paper, University of Nottingham.
Güth, W., Levati, M. V., Sutter, M., & Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics*, *91*(5), 1023–1042.
Hermalin, B. (1998). Towards an economic theory of leadership: Leading by example. *American Economic Review*, *88*(5), 1188–1206.
Hermalin, B. (2012). Leadership and corporate culture. In R. Gibbons & J. Roberts (Eds.), *Handbook of organizational economics*. Princeton: Princeton University Press.
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.
Kidd, M., Nicholas, A., & Rai, B. (2013). Tournament outcomes and prosocial behaviour. *Journal of Economic Psychology*, *39*, 387–401.
Kocher, M. G., Pogrebna, G., & Sutter, M. (2013). Other-regarding preferences and management styles. *Journal of Economic Behavior and Organization*, *88*, 109–132.
Kőszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature*, *52*(4), 1075–1118.
Koukoumelis, A., Levati, M. V., & Weisser, J. (2012). Leading by words: A voluntary contribution experiment with one-way communication. *Journal of Economic Behavior and Organization*, *81*(2), 379–390.
Kreps, D. M. (1986). Corporate culture and economic theory. In M. Tsuchiya (Ed.), *Technology, innovation, and business strategy*. Tokyo: Nippon Keizai Shumbunsha Press.
Ledyard, J. (1994). Public goods: A survey of experimental research. In J. Kagel & A. Roth (Eds.), *Handbook of experimental economics*. Princeton: Princeton University Press.
Loch, C. H., & Wu, Y. (2008). Social preferences and supply chain performance: An experimental study. *Management Science*, *54*(11), 1835–1849.
Meidinger, C., & Villeval, M. C. (2002). *Leadership in teams: Signaling or reciprocating?*. GATE working paper, 10-03, Lyon.
Moxnes, E., & van der Heijden, E. (2003). The effect of leadership in a public bad experiment. *Journal of Conflict Resolution*, *47*(6), 773–795.
Palfrey, T., & Rosenthal, H. (1994). Repeated play, cooperation and coordination: An experimental study. *Review of Economic Studies*, *61*(3), 545–565.
Riyanto, Y., & Zhang, J. (2013). The impact of social comparison of ability on pro-social behaviour. *The Journal of Socio-Economics*, *47*, 37–46.
Rey-Biel, P., Sheremeta, R., & Uler, N. (2012). *(Bad) luck or (lack of) effort? Sharing rules in the US and Europe*. Working paper.
Rotemberg, J., & Saloner, G. (1993). Leadership styles and incentives. *Management Science*, *39*, 1299–1318.
Rotemberg, J., & Saloner, G. (2000). Visionaries, managers and strategic direction. *Rand Journal of Economics*, *31*, 693–716.
Seelya, B., Van Huyck, J., & Battalio, R. (2007). Credible assignments can improve efficiency in laboratory public goods games. *Journal of Public Economics*, *89*(8), 1437–1455.
Sherstyuk, K., Tarui, N., & Saijo, T. (2013). Payment schemes in infinite-horizon experimental games. *Experimental Economics*, *16*(1), 125–153.