Employee types and endogenous organizational design: An experiment*

Antoni Cunyat†
University of Valencia and ERI-CIS

Randolph Sloof‡
University of Amsterdam and Tinbergen Institute

May 16, 2011

Abstract

When managers are sufficiently guided by social preferences, incentive provision through an organizational mode based on informal implicit contracts may provide a cost-effective alternative to a more formal mode based on explicit contracts and active monitoring. This paper reports the results from a stylized laboratory experiment designed to test whether subjects in the role of firm owner rely on the social preferences of other ('employee') subjects with whom they are matched when choosing which payoff version of a simple trust game these employee subjects should play ('the organizational mode'). Our main finding is that they do so, albeit in a different way than theory predicts. The importance of the first mover's social preferences for trusting behavior is recognized by the owner subjects, but the significant (first order) impact second movers' social preferences have on trusting behavior of first movers seems to be overlooked. JEL-codes: C91, J40, M50.

1 Introduction

A major research theme within organizational economics is how to motivate employees to exert well-directed effort. This issue is typically addressed using the principal-agent model as point of departure. In the standard version of this model the agent is assumed to solely care about his own monetary compensation and to dislike effort. Similarly so, the principal just wants to maximize her own net profit

*We would like to thank Enrique Fatás and LINEEX for their help and hospitality while we conducted the experiment. Programming support of Héctor Solaz is gratefully acknowledged. An earlier version of the paper has been presented at conferences in Amsterdam (EALE 2008) and Rotterdam; we benefitted from the participants' comments, with a special thanks to Michael Vlassopoulos. An associate editor and two anonymous referees provided very thoughtful and constructive remarks that improved the paper considerably. Antoni Cunyat acknowledges financial support from the Spanish Ministry of Education and Science, Project SEJ 2005-08054/ECON and from Conselleria d'Empresa, Universitat i Ciència of Valencian Government through project GV2006-201.

†Departament d'Anàlisi Econòmica. Av/ dels Tarongers s/n. Campus dels Tarongers. Edifici Departamental Oriental. 46022 Valencia (Spain). Antonio.Cunat@uv.es. http://www.uv.es/acunat.

‡University of Amsterdam, School of Economics, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands. r.sloof@uva.nl. http://www1.fee.uva.nl/pp/rsloof/.

and does not care about the agent's well-being. Given these assumptions it is derived how explicit incentive contracts should be optimally designed as to motivate the agent to put in sufficient effort.

Many empirical studies have found, however, that people may have alternative motivations that go beyond material self-interest. Fairness, altruism, empathy and a preference to react in kind to friendly or hostile actions of others (reciprocity) are among the various alternative motivations identified. The presence of such 'social preferences' may have profound implications for the provision of effort incentives. One of these is that they may make cost-effective alternatives based on non-enforceable, 'implicit' contracts feasible. Workers are more easily motivated to exert effort when they know that their manager cares about their personal well-being and thus will reward higher effort with a larger (non-contractible) bonus. In that case higher effort levels can be induced without having to bear the costs of a formal performance measurement and evaluation system. For this reason it may actually pay for firms to hire 'empathic' managers who do not solely care about profit maximization. The manager's personality type may help in overcoming a difficult incentive problem with the workers; see Rotemberg and Saloner (1993), Rotemberg (1994), and Hermalin (2001, Section 4.2) for formalizations of this intuitive idea.

Relying on social preferences as implicit-contract enforcement device requires that those who hire managers ('owners') recognize two important behavioral forces. First, the *direct* impact of the manager's social preferences on his *own* behavior should be well understood. Second, the *indirect* impact of the manager's social preferences on the behavior of those he manages should be appropriately recognized. Especially the latter behavioral force is crucial. It requires that owners understand that the manager's preferences are potentially key for worker behavior and that workers' own social preferences may be of second order. Intuitively, both selfish and more socially oriented workers alike are (much) stronger motivated by a manager committed to reward high effort than by a manager who always behaves opportunistically. Finally, owners should also recognize that an organizational mode based on informal implicit agreements is viable only if managers are sufficiently empathic. If not, a more formal mode characterized by explicit contracts and active monitoring by managers is likely to perform better.

In this paper we intend to test the (recognition of the) above behavorial forces. We do so in a stylized laboratory experiment, making use of simple trust games to capture in highly reduced form the main characteristics of the firm owner's decision problem: (i) to choose the optimal incentive scheme ('organizational mode') and (ii) to appoint the right manager capable of obtaining the best outcome for the owner, given the organizational mode chosen. The main focus is on how owner subjects take account of the social preferences of those with whom they are matched and to what extent they are able to optimize accordingly. In particular, we investigate to what degree subjects are able to backward induct in view of the observed social preference characteristics of others in their group.

[Insert Figure 1 about here]

The two trust games on which our experiment is based are depicted in Figure 1. Although subjects were confronted with an entirely neutral and abstract framing, for ease of presentation here we discuss these games in terms of the particular application that motivated their choice. In motivation game M on the l.h.s. a worker first decides whether to shirk or to work.[1] In case the worker shirks, he does not

[1] This game corresponds to the trust game used by Kreps (1990) to model corporate culture and also to a simplified

get a reward (on top of his wage). If, however, the worker exerts effort, the manager decides whether to reward him with a bonus or not. Because the bonus is non-contractible, a selfish manager will not pay it and, anticipating this, the worker will not work. But if the manager could credibly commit to pay the bonus (only) when the worker exerts effort, the worker would be motivated to do so. Therefore, in this organizational mode the main role for managers is to inspire and motivate the work force, by cultivating a culture that hard work will be rewarded by the organization.

The inspection game I on the r.h.s. in Figure 1 intends to capture in highly reduced form an organizational mode based on explicit incentive contracts and active monitoring. Here the role of managers is to supervise and to monitor that the workers do not deliver substandard work (cf. Calvo and Wellisz (1978)). Workers receive a given wage for putting in effort, but are fined whenever they are caught shirking. In game I the manager first decides whether to monitor or not. Two different versions of the game are considered, one for high (I_H) and one for low (I_L) inspection costs. If the manager monitors, the worker works for sure. If not, the worker chooses between shirking and working. Payoffs are such that selfish workers will shirk if not monitored. Realizing this, the manager commits to monitor in the first stage.[2]

Under selfish preferences the worker is predicted to work in mode I but not in mode M. In that case the inspection mode is more efficient and will be preferred by the owner of the firm. Things change when employees have social preferences. Putting a very empathic employee in the manager's position under mode M (i.e. as second mover) will then yield first best. This follows because, anticipating that he will be rewarded by the empathic manager, the worker puts in effort and does not shirk. And compared to structure I, organizational mode M saves on the costs of the monitoring technology.[3] The M-mode thus becomes relatively more attractive the more empathic the employees are. Moreover, because the M-mode relies on empathic management, the owner will put the more empathic employees within the work force in the manager's role whereas in the I mode this is not the case. A final intuitive prediction is that the more cost-effective the inspection system under mode I is, the less likely it is that the owner prefers mode M. Lower inspection costs are represented in Figure 1b by having payoffs of 440 (instead of 360) after the manager's decision to monitor.

The experiment that we use to test the behavioral forces behind these predictions has two parts. In part one subjects make 18 decisions in 9 different games that all have the same entry-reward structure as in Figure 1. These decisions are used to generate individual 'track records' (e_i,r_i), with e_i (r_i) the number of entry (reward) choices made. The track record is taken as a (imprecise) measure of a subject's

---

version of the game used by Rotemberg and Saloner (1993) to study the impact of leadership style on workers' incentives to innovate. In spirit the M-game also corresponds to the 'loose supervision' regime in the model of supervision and workgroup identity studied by Akerlof and Kranton (2005). The inspection game I (to be discussed shortly) then corresponds to their 'strict supervision' regime.

[2] Game I reflects a simplified version of an inspection game where the manager can commit to a particular inspection strategy. In a more general setup, the manager commits to a particular inspection probability, such that the worker is just induced to exert effort with probability one; see Section 5 in Avenhaus et al. (2002) for a full discussion and justification of this game. In their real effort experiment Dickinson and Villeval (2008) also use an inspection game in which the principal/manager can commit ex ante to a given monitoring technology. For simplicity, here we restrict the manager to inspection probabilities of either zero (no monitoring at all) or one (always monitor).

[3] This explains why the maximum joint payoffs for the worker and the manager under M (1100) are higher than the maximum joint payoffs under I (980). Apart from installation and operational monitoring costs, another type of monitoring costs is that the worker dislikes being monitored, because it gives him the negative feeling of being controlled (cf. Frey (1993) and Falk and Kosfeld (2006) who find that being monitored may lower work morale). Even if he works, his payoffs in the inspection mode therefore depend on being monitored (either 360 or 440) or not (490). In the Appendix we provide an elaborate justification of the particular parameterization depicted in Figure 1.

preference type. In part two subjects either take the role of owner or of employee and in each period firms consisting of one owner and two employees are formed based on a strangers design. Given the observed track records of her two employees, the owner decides which employee becomes manager and who gets the role of worker. In the neutral frame that we consider this reduces to deciding in which position – first or second mover – the two other group members will play the trust game at hand. After that the two employees take decisions in the trust game (organizational mode) that applies. In the first ten periods the trust game is exogenously given (either game M or game I from Figure 1), with first 5 times one game and then 5 times the other. In the final 5 periods the owner first chooses the trust game before she assigns roles. Employees' payoffs are as in Figure 1 and the owner's payoffs equal those of the manager (but are private information to her).

Our main findings are as follows. First, in both trust games of Figure 1 subjects with the first mover role are more likely to 'enter' the higher the r_i value is of the second mover with whom they are matched. This confirms that the more 'empathic' second movers are, the higher the willingness of first movers to enter the reciprocal relationship. Second, subjects with the role of firm owner seem to overlook this mechanism when assigning their employees to different roles. They mainly look at the employee's own track record, in particular e_i, to form expectations about how he would behave in role A. As a result, in the M-game they typically assign the manager role (second mover) to the employee with the *lower* r_i-value, i.e. to the *less* empathic employee. Third, given reverse role allocation under the M mode, also the choice between trust games deviates from equilibrium predictions. Nevertheless, subjects in the role of owner in general do correctly realize that the social preferences of their employees make the M-mode relatively more attractive. Overall we conclude that subjects do take account of the social preference characteristics of others, but mainly to the extent to which these are expected to affect the others' *own* behavior. The *indirect* impact of one subject's social preferences on the behavior of *other* subjects appears not appropriately recognized; subjects in the role of owner do not seem to backward induct sufficiently in view of others' social preferences.

Many laboratory experiments have already been conducted that relate to the above discussed issues of endogenous organizational design. In a series of influential papers, for instance, Ernst Fehr and various coauthors have studied the firm's choice of optimal incentive contracts. A main common finding is that social preferences can serve as a cost-effective contract enforcement device and contracts may therefore deliberately be left incomplete.[4] This suggests that in practice firms may prefer implicit contracts over explicit incentive contracts, although under selfish preferences the latter would be optimal.[5] Other experiments focus on the relation between workers' social preferences and their self-sorting into different pay for performance schemes (see e.g. Cabrales et al. (2010), Dohmen and Falk (2011), Eriksson and Villeval (2008) and Teyssier (2008)). Also the effect of monitoring on worker behavior has already been studied in the lab; see e.g. Dickinson and Villeval (2008) and Schweitzer and Ho (2005). Compared to these experiments, we study (implicit and explicit) contracts and monitoring in highly reduced form. The main contribution of our study is that we explicitly relate these reduced form organizational choices

[4] See e.g. Fehr et al. (1997, 2007), Fehr and List (2004) and Fehr and Schmidt (2000, 2004, 2007).

[5] Here it must be acknowledged though that recent field experiments cast doubts on the extent to which laboratory gift exchange findings can be generalized to actual practice. List (2006) and Gneezy and List (2006), for example, find field evidence that the importance of social preferences in one-shot labor relationships is limited. See DellaVigna (2009, Section 2.3.5) for a concise overview of recent field experiments on gift exchange and Fehr, Goette, and Zehnder (2009) and List (2009) for more elaborate discussions.

to the observed characteristics ('track records') of the employees that are to be affected by these instruments. Another novel and important feature is that we explore the endogenous allocation of roles within organizations (games).

This paper proceeds as follows. Assuming that subjects may care about the well-being of others, we first derive the formal hypotheses that are put to the test. Section 3 presents the details of our experimental design and Section 4 reports the results. The concluding discussion puts our main findings into perspective and briefly discusses what can be learnt from our stylized laboratory experiment about actual firm behavior.

2 Theoretical predictions and hypotheses

Our experiment is based on the two games depicted in Figure 1.[6] These games have the same general decision structure, which is reflected in Figure 2 below. Player A first chooses between Out and Enter. If player A enters, player B subsequently chooses between Reward and No reward. Payoffs are such that choosing No reward yields B the most in monetary payoffs, whereas Reward corresponds to sacrificing to reward A for the 'kind' choice to enter. From $d > c > b > a$ it immediately follows that, if both players are selfish, (Out, No reward) is the unique subgame perfect equilibrium.

[Insert Figure 2 about here]

Social preferences may lead players away from the inefficient (Out, No reward) outcome. Various alternative (to purely self-centered money maximization) motivations have been identified in the literature – like fairness, altruism, empathy and reciprocity – and a number of theoretical models have been developed to capture these in a formal way. Prominent examples include Fehr and Schmidt (1999)´s model of inequality-aversion, Charness and Rabin (2002)´s model of quasi-maximin preferences, and Rabin (1993)´s model of intention-based reciprocity (see also Dufwenberg and Kirchsteiger (2004)). Although these models can lead to quite different predictions in particular situations, a common theme they share is that social preferences may be efficiency enhancing. It is this common aspect that we want to emphasize here.[7]

To illustrate the impact alternative (non-selfish) motivations may have, we capture them in a very simple and stylized way. Let π_i and π_j denote player i's and j's monetary payoffs. Following Charness and

[6] In the Appendix we discuss a basic reduced form model of endogenous organizational design that underlies these two specific games. This Appendix also discusses our choice of parameters and thereby motivates the monetary payoffs the subjects earn in different roles.

[7] Our experiment thus neither should be taken as an attempt to discriminate between various types of social preferences, nor as providing a test of a particular version of social preferences per se. Although in this section we use quasi-maximin preferences to derive and illustrate the main implications in a parsimonious way, similar predictions would have been obtained under relevant alternative specifications. For instance, incorporating Dufwenberg and Kirchsteiger (2004)'s type of intention-based reciprocity motivations leads to similar predictions as in hypotheses H1 through H3 below. In particular, Sloof and Sonnemans (ming, Appendix B) show that the probability with which player B chooses to reward weakly increases with his reciprocal attitude Y_B – see their Lemma 1 which shows that this probability equals $p = \max\left\{0, \min\left\{\left[\frac{d-b}{d-c} - \frac{2}{Y_B\cdot(c-a)}\right], 1\right\}\right\}$ – and that the probability with which player A enters increases with p (Lemma 2). Combining these two results yields H1. Hypotheses H2 and H3 are almost a direct consequence of H1.

Rabin (2002), we assume that player i's preferences take the following form (with $i \neq j$ and $i, j \in \{A, B\}$):

$$\begin{aligned} U_i(\pi_i, \pi_j) &= \rho_i \cdot \pi_j + (1 - \rho_i) \cdot \pi_i \quad \text{if} \quad \pi_i > \pi_j \\ &= \sigma_i \cdot \pi_j + (1 - \sigma_i) \cdot \pi_i \quad \text{if} \quad \pi_i \leq \pi_j \end{aligned} \tag{1}$$

In this specification, parameter ρ_i gives the weight player i attaches to the other player's payoffs when she herself is ahead in the payoff distribution. Parameter σ_i reflects the corresponding weight when she is behind in this regard. Without any restrictions on ρ_i and σ_i utility function (1) can capture a range of different motivations. Charness and Rabin (2002) use the results of a variety of simple games with a similar decision structure as in Figure 2, to estimate the values of ρ_i and σ_i. They find that on average players do not care about other players' payoffs when they are behind, but put a positive weight on the well-being of others when they are ahead. In line with their estimates we therefore assume that $0 < \rho_i < 1$ and that $\sigma_i \leq 0$. Note that the inequality-aversion model of Fehr and Schmidt (1999) is a special case of this (additionally requiring $|\sigma_i| \geq \rho_i$). Our assumptions are also in line with Hermalin (2001, Section 4.2), who assumes that a player suffers from 'remorse' only if he is ahead (in terms of payoffs).

Assuming preferences as in (1), the game is again easily solved by backward induction. Player B will choose to reward whenever $\rho_B \geq \rho^* \equiv \frac{d-c}{d-a}$.[8] Anticipating this, player A enters only when ρ_B exceeds this threshold, independent of the level of σ_A and ρ_A. Hence the predicted outcome is Out when $\rho_B < \rho^*$ and (Enter, Reward) in case $\rho_B \geq \rho^*$. (Outcome (Enter, No reward) is thus never observed on the equilibrium path.) This establishes that when player B cares sufficiently about A's well-being, the inefficient outcome Out is avoided. Note that ρ_i is the key parameter here. Following Rotemberg and Saloner (1993), we say that a player is more *empathic* the higher his ρ_i is. This yields our first main hypothesis.

H1 The more empathic player B is (i.e. the higher ρ_B is), the more likely it is that player A enters.

Players A and B represent the work force of a given firm. The role of owner of this firm is captured by player C, who decides on role assignment. In particular, given the empathy characteristics of her two employees as reflected by the ρ_i-values, she decide which employee gets role A and who gets role B. For ease of exposition we assume that, unlike her two employees, player C is selfish.[9] Her monetary payoffs are equal to those of the manager and therefore depend on the version of the general game in Figure 2 that is played. In the I-games the payoffs of player C equal those of player A, in the M-game they correspond to those of player B.[10] Because in equilibrium the outcome either equals Out or (Enter, Reward) and A and B thus always obtain the same (either b or c), the predictions regarding role assignment are actually the same for these two games. Player C prefers to assign role B to the more empathic employee, because this maximizes the probability that the more efficient outcome (Enter, Reward) is obtained.[11]

[8] Here we assume that when B is indifferent, he chooses to reward. This assumption is inessential for the comparative statics hypotheses (H1 through H3 below) we are interested in.

[9] Footnote 11 below explains that this assumption is immaterial for our hypotheses.

[10] In the experiment player C's payoffs are private information to her, so the observed behavior of A and B cannot be driven by the actual payoffs player C obtains. This justifies that (1) does not include π_C.

[11] Note that role assignment effectively makes a difference only when $\rho_i < \rho^* < \rho_j$ for employees i and j. Then the outcome is (Enter, Reward) when employee j is put in the B-role and Out when employee i is put in the B-role. Because $c > b$ a selfish player C prefers the former assignment. Also note that on the equilibrium path all players earn exactly

H2 Players C will assign the role of player B to the more empathic employee within her work force (i.e. to the one with the higher ρ_i-value).

A direct corollary of the above prediction is that managers (players A) in the inspection mode will on average be more selfish than managers (players B) in the motivation mode.

Apart from role assignment, player C may possibly also choose between game M and game I. These two games differ in the values of the payoff parameters a through d. Here we simply work with the numbers in Figure 1, but the reasoning applies more generally. In both games, the only two possible equilibrium outcomes are Out and (Enter, Reward). Player C therefore prefers game M only if the latter outcome is expected in that game. If not, player C prefers game I. The latter follows because the payoff of 280 from outcome Out in game M falls short of both 360[440] and 490 in game $I_H[I_L]$. From the above it follows that only when the empathy level of player B exceeds the critical threshold of $\rho^* = \frac{1}{3}$ $\left(= \frac{730-550}{730-190}\right)$, outcome (Enter, Reward) is expected in the M-game. Therefore, only if player C has an employee within her work force with an empathy level that exceeds this threshold, she prefers game M.[12] Hence we obtain that C is more likely to choose game M the more empathic her employees are.

H3 The more empathic the work force is (i.e. the higher ρ_i of the more empathic employee), the more likely it is that player C prefers the M-game over the I-game.

Hypotheses H1 through H3 are the main predictions we want to test. Yet another interesting aspect to consider is the role of inspection costs in the I-game. Higher inspection costs correspond to lower payoffs attached to the Out outcome in this game (i.e. $b = 440$ in I_L versus $b = 360$ in I_H). If player C would be perfectly informed about her employees' empathy parameters ρ_i, variations in these payoffs would not affect the predicted outcomes (as long as these exceed the payoffs of the corresponding Out outcome in the M-game). However, it seems reasonable to assume that parameter ρ_i is employee i's private information. The owner may have a good idea about what the value of ρ_i is, but she may not be completely sure about its exact value. She is therefore unsure whether outcome Out or (Enter, Reward) will result in each of the two games. Other things equal, her expected payoffs of choosing the I-game are then lower the lower the value of parameter b in that game is. Player C is thus more likely to choose game M over I_H than over I_L.

Clearly, also in the experiment employees do not observe the level of empathy of their colleagues precisely, and neither do so owners C. Based on an observable track record of past choices, however, an estimate r_i of a player's empathy level ρ_i can be obtained. Exactly how individual track records are generated in the experiment is explained in the next section.

---

the same (either b or c). In the absence of payoff differences social preferences like in (1) effectively correspond to selfish preferences. Hence exactly the same predictions follow when player C is also guided by social preferences similar to those in (1).

[12] More generally this reasoning applies whenever $b^I > b^M$ and $c^I < c^M$ (with the superscripts of these payoff parameters referring to the type of game). As explained in the Appendix, the first inequality reflects the idea that the value of the worker's effort exceeds the overall costs of a formal monitoring system. Restriction $c^I < c^M$ derives from the natural assumption that installing a monitoring technology brings about (fixed) investment costs, even when it is not actively used in the end.

Table 1: Overview of sessions and treatments (in part 2)

session	rounds $1-5$	rounds $6-10$	rounds $11-15$
1	I_L	M	I_L versus M
2	M	I_L	M versus I_L
3	I_H	M	I_H versus M
4	M	I_H	M versus I_H

3 Experimental design

We ran four sessions in total, which differed according to (the order of) the treatments considered. In each session we kept the two types of games (game M and game I) fixed. Between sessions we varied the particular version of the inspection game, having either the one representing low inspection costs (I_L) or the other one with high inspection costs (I_H). We also varied the order in which the motivation game and inspection game were considered. Table 1 provides an overview. All sessions were run in May 2007 at the LINEEX laboratory of the University of Valencia. Overall 180 subjects participated, with 45 subjects per session. The subject pool consisted of undergraduate students at the University of Valencia. The vast majority of them (88%) were students in Economics or Business, 57% were male. They earned on average 24.5 euros in somewhat less than 2 hours, including a show up fee of 7 euros.

At the beginning of each session subjects were informed that the experiment consisted of two parts. They were also informed that possibly some of the choices they made in part 1 would become observable to some other participants in part 2. In particular, the instructions for part one explained that:[13]

> "...It may happen that in part two some other participants get some information about the decisions you made in part one. It may also happen though that none of your part one decisions will ever become known to any other participant..."

Apart from this information, subjects were kept ignorant about the actual content of part 2 until that part actually started. The above announcement has the clear disadvantage that it may influence subjects' decisions in part 1. We considered it necessary though, in order to avoid any potential impression of deception. Moreover, if we would not make the announcement, subjects would be surprised at the start of part 2 when their choices of part 1 became known, and might think that it is quite likely that another "surprise" would follow. This might then affect their behavior in part 2.

In part 1 subjects made decisions for a series of nine extensive form games that all have the same decision structure as in Figure 2. We used a neutral frame for the entire experiment, with A's choosing between A1 (Out) and A2 (Enter) and B's choosing between B1 (No reward) and B2 (Reward). Table 2 provides an overview of the games used. Subjects first made 9 decisions as player A, after that they made 9 *conditional* (on entry) decisions as player B. They did not get any feedback about the actual outcomes of these games. They were just informed that at the end of the experiment one of the 18 choices made in part 1 would be randomly selected and paid (see below for a more detailed explanation).

The nine different games of part 1 have been chosen as follows. The first three games are just (five times) upscaled versions of the M-game, the I_H-game and the I_L-game, respectively. These games differ

[13] The experiment was conducted in Spanish. An English translation of the complete instructions can be found at the journal's website as online supplementary material.

Table 2: Overview of the nine games in part one

game	A stays out (b, b)	If A enters, B chooses (a, d) vs. (c, c)	sacrifice $d - c$	reward $c - a$	ρ^*
I	(1400,1400)	(950,3650) vs. (2750,2750)	900	1800	0.33
II	(1800,1800)	(200,2900) vs. (2450,2450)	450	2250	0.17
III	(2200,2200)	(200,2900) vs. (2450,2450)	450	2250	0.17
IV	(1400,1400)	(200,2900) vs. (2450,2450)	450	2250	0.17
V	(1800,1800)	(950,3650) vs. (2750,2750)	900	1800	0.33
VI	(2200,2200)	(950,3650) vs. (2750,2750)	900	1800	0.33
VII	(1400,1400)	(1250,3950) vs. (2600,2600)	1350	1350	0.5
VIII	(1800,1800)	(1250,3950) vs. (2600,2600)	1350	1350	0.5
IX	(2200,2200)	(1250,3950) vs. (2600,2600)	1350	1350	0.5

Remark: It holds that $\rho^* = \frac{sacrifice}{sacrifice+reward} = \frac{(d-c)}{(d-c)+(c-a)}$.

in two important ways. First, they correspond to different ratios of the amount player B has to sacrifice in order to give player A a particular reward. According to the theory discussed in Section 2, player B is only willing to give this reward if his empathy parameter ρ_i exceeds threshold ρ^*. This threshold differs between the M-game and the two I-games, see the final column in Table 2. Second, the M and I games also differ in the amount player A forgoes by choosing to enter.

The remaining six games have been chosen using games I through III as starting point. Game IV combines the payoffs of the 'stay-out' option of the M-game with the sacrifice-reward values of the two I-games. Games V and VI do so vice versa. The final three games combine the stay-out payoffs of games I through III with equal sacrifice-reward values such that an ρ^* of one half results.

We used part 1 to generate a 'track record' for each individual. Such a track record consisted of a pair (e_i, r_i), with $e_i, r_i \in \{0, 1, ..., 9\}$ the number of enter choices subject i made as player A and the number of 'reward' choices s/he made as player B respectively. Under the assumptions of the theory spelled out in Section 2, only decisions made as a second mover can serve as a proxy for social preferences. We therefore (only) use r_i as an estimate of a subject's empathy level ρ_i.

Part two of the experiment consisted of 15 periods. Subjects first learned their roles, being either an owner ("person C") or an employee ("group member"). Subjects kept the same role throughout all 15 periods. Roles were assigned as follows. In each session we ranked the 45 subjects on the basis of the number of 'reward' choices r_i in their track record. Subjects with rank 16 to 30 were assigned the role of owner and the remainder the role of employee. This procedure – unknown to the subjects – secured that we had enough variation in empathy types among employees.

At the beginning of a period, firms (called "groups" in the experiment) consisting of one owner and two employees were exogenously formed. Matching was based on a stranger design. In each period each owner was matched to two different subjects from the lowest and highest tercile in the ranking of r-choices, respectively. We made sure that each subject met each other subject only once. Subjects were thus never confronted with the same firm member again and were explicitly informed about that.

The 15 periods were divided into three blocks of five. In the first two blocks the game to be played was exogenously given (see Table 1 above). Here the owner first decided, on the basis of the observed track records (e_1, r_1) and (e_2, r_2) of her two assigned employees, which role to assign to each of them (either A or B). Also the two employees observed the track record of each other. After employees were assigned their roles, they played the game that applied, making decisions for the role assigned. This determined their period payoffs, as given in the respective extensive form games. The owner received a period payoff equal to those of player A in the I-game and equal to those of player B in the M-game. To easily remind owners about this fact, we labelled the I-game as game "Azul" and the M-game as game "Blanco" (and we printed these games against the corresponding background color). The complete setup of the game was common knowledge, except for the payoffs of the owner, which were known to players C only. We did so to prevent that employees' decisions were guided by empathic feelings towards player C. In the final five periods 11 to 15, the owner first chose which game to play (after being informed about her/his employees' track records). Once a game had been chosen, the order of decisions was as before.

Except for the decisions made within their own firm in a given period, subjects did not get any information on how the other subjects behaved in part 2. Although they may have recorded the decisions made by previous firm members in earlier periods, this information is of limited value because they would never meet with the same other subject again. The observable track records they obtained from part 1 were thus the main clue they could use to predict the behavior of other subjects within their firm. The track record of owners was never made public. Therefore, for one third of the subjects the decisions made in part one remained private information throughout the experiment.

Payoffs were determined in the following way. From the first part one game was selected at random. For this particular game subjects were then randomly coupled in pairs and were randomly assigned roles.[14] The individual payoffs that resulted from these pairings gave the earnings for part one. To this amount we added the overall payoffs from part two. The conversion rate was such that 500 points in the experiment corresponded with 1 euro in money. Apart from that, subjects received a show up fee of 7 euros.

The experiment was computerized using the z-tree programming package (cf. Fischbacher (2007)). Subjects started with written instructions for part one, which were also read aloud by the experimenter. At the end of part one subjects received new instructions for part two. Before part two started, subjects played one practice period.[15] After finishing part two subjects filled in a short questionnaire. Having completed this, the experimental points earned were exchanged for money and subjects were paid individually and discreetly.

4 Results

In this section we first describe the distribution of individual 'track records' generated by the choices subjects made in part one. Next we look at how players A and B behave in the three games M, I_H and I_L at hand. We are particularly interested in testing our first hypothesis that A's are more likely to enter the higher the r-value of player B they are coupled with. The final subsection looks at the 'organizational design' choices made by players C. It is tested whether they assign the B-role to the employee with the higher r-value and whether they are more likely to choose the M-game when this r-value is higher.

[14] Because we had an odd number of subjects within each session (45), we actually assigned 22 subjects the A-role and 23 subjects the B-role. The decision of one randomly selected A-subject was then used twice to determine the payoffs of two different B-roles.

[15] The practice period resembled an actual period in the sense that subjects were given feedback on the decisions of others (in the same way as in an actual period). Yet it was explicitly emphasized at the start of the practice period that this period was unpaid and only meant to get acquainted with the screens and the strategic situation. The matching corresponded to the one of a randomly selected actual period of part two.

Table 3: Overall distribution of individual track records and consistency of r-choices

	number of e choices											Consistency
# **of r's**	0	1	2	3	4	5	6	7	8	9	Total	of r choices
0	**10**	11	9	4	4	3				1	42	0
1	1	**2**	4	5	5	4		2			23	1
2			**2**	5	1	7	1		2		18	1.78
3		2	5	**3**	6	4	3		3	1	27	1.63
4			3	6	**13**	5	4			2	33	2.30
5		1	2	3	4	**6**	1				17	2.47
6			1		1	2		1	1		6	1.33
7					1	2	1				4	2
8			1			1					2	1
9					2	1	1		1	**3**	8	0
Total	11	16	27	26	37	35	11	3	7	7	180	All 1.31

Remark: In the left hand panel, numbers on the diagonal where $e_i = r_i$ appear in bold. In the right hand panel the average level of consistency is based on the individual levels of consistency that range from zero (fully consistent) to four (fully inconsistent).

4.1 Individual track records

In part one each subject makes nine entry decisions as player A and nine (conditional) reward choices as player B. From these 18 choices, an individual track record (e_i, r_i) results. The left hand panel in Table 3 gives the overall distribution observed for the 180 subjects in our experiment.

On average subjects choose to enter 3.76 times as player A and to reward on average 2.83 times as player B. As the frequency distribution makes clear though, there is quite some heterogeneity. Most observations are scattered around the diagonal, suggesting that the number of entry and reward choices are correlated. Indeed, for our full sample of 180 subjects the Spearman rank correlation between e and r choices equals 0.48 and is highly significant ($p < 0.001$). This also holds when we compute correlations for each of the four sessions in isolation. Moreover, many entries in Table 3 are above the diagonal. This indicates that subjects typically choose to enter somewhat more often as player A than they choose to reward as player B, which is corroborated by formal signrank tests.[16]

There are some minor differences in the observed track records across sessions. Comparing the number of r-choices by means of a Kruskal-Wallis test, we do not find a significant difference (at the 5% level) between the four sessions ($p = 0.0723$). For the number of enter choices there are some differences though ($p = 0.0131$). Both ranksum and Kolmogorov-Smirnov tests reveal that subjects in session 3 have a lower e-value than those in sessions 1 and 4. In the former the average equals 2.87, in the latter two 4.13 and 3.96, respectively. (In session 2 the average e-value is comparable to sessions 1 and 4 and equals 4.09. Yet because in this session the variation is higher than in the other ones, no significant difference with session 3 is found.) But even in session 3 the average e-value (2.87) exceeds the average r-value (2.44). The main observation of a substantial correlation between e and r-choices with (slightly) higher e-choices thus applies to all sessions.

Table 2 in the previous section reports the threshold value ρ^* for each game implemented in part one. A subject that makes fully consistent choices should (as player B) reward either 0, 3, 6 or 9 times.

[16] Only in session 3 we do not find a significant difference (at the 5%-level) in indivivual e and r scores. But for the three other sessions we do, as well as overall.

The column labelled 'Total' in Table 3 reveals that we do indeed observe spikes at 0 and (to a smaller extent) 3 reward choices in the frequency distribution, but also that deviations from either 0, 3, 6 or 9 times occur relatively often. Moreover, even if a subject chooses to reward in (say) 3 out of the 9 choices made as player B, these choices may not be consistent across games (e.g. Table 2 predicts that the three choices to reward should be made in games II, III and IV then). To explore the consistency of choices, we construct the following measure. For each individual subject we compute the minimum number of changes in the sequence of reward choices made that would be needed to make the sequence consistent. For instance, a subject who chooses to reward in games II and III only obtains a score of 1, because the sequence of choices is made consistent by (hypothetically) changing the no reward choice in game IV into a choice to reward. Our consistency measure thus ranges from zero (fully consistent) to four (entirely inconsistent).[17] The right hand panel of Table 3 reports the average level of consistency for groups of subjects with the same r_i value. Overall, the average level of consistency equals 1.31. Of the 180 subjects 35% are fully consistent while 26% make only one inconsistent choice. The majority of the subjects thus behaved close to being consistent. The frequencies for consistency values of 2, 3 and 4 equal 18%, 16% and 6% respectively.[18]

As explained in the previous section, after part one we sorted the subjects on the basis of their r-scores and assigned those in the second tercile the role of player C. In sessions 1 and 4 these were subjects with $r_i = 2$ to $r_i = 4$, whereas in sessions 2 and 3 these where subjects with $r_i = 1$ to $r_i = 3$ and with $r_i = 1$ to $r_i = 4$, respectively. Recall that in every period players C were assigned one employee from the low-r group and another one from the high-r group.

4.2 Employees' choices

To get a first impression of how players A and B behaved in part two, Figure 3 displays the relative frequencies of the observed outcomes in the three different games. In periods 1 to 10 the game was exogenously given whereas in the last five periods it was endogenously chosen by player C. The predominant outcome in game I_L is that player A chooses Out. In the very few instances that A chooses to enter, player B chooses to reward (R) about equally often as not to reward (N). The latter also applies for game I_H, but there player A enters somewhat more frequently. Finally, in the motivation game player A enters (much) more often than in the two inspection games. But there player B also appears less willing to reward, especially when the motivation game is endogenously chosen.

[Insert Figure 3 about here]

The first hypothesis we want to test is whether player A's are more likely to enter the higher the r-value of player B they are coupled with. We do so by estimating a random effects probit model of the probability that A chooses to enter, where the data are clustered at the session level. The six specifications

[17] The maximum of four immediately follows from observing that never rewarding and always rewarding are consistent sequences, so our consistency measure is always weakly below $\min\{r_i, 9 - r_i\}$.

[18] Table 3 reveals that subjects who make 4 or 5 reward choices are the ones with the highest average of inconsistent choices. Indeed, all fully inconsistent (i.e. consistency score of 4) subjects are from this group as well as the majority of those with a consistency score of 3. A plausible explanation here is that these subjects are confused and therefore follow a $50 - 50$ rule.

reported in Table 4 include various combinations of players B's and A's track record characteristics. Apart from that, two dummies for respectively the I_H-game and the M-game are incorporated (so I_L serves as baseline), together with a time trend 'period'. Additional dummies are included to take account of the order in which the exogenous games are considered within a session and whether the game played is endogenously chosen by player C or not. Our main interest lies in the track record characteristics. Because subjects' e_i and r_i values are highly correlated, these are studied both in isolation and jointly. In specifications (1) through (3) only the track record characteristics of player B are incorporated, in specification (4) through (6) those of player A are included as well.

Before turning to the results obtained, it is important to point out that the coefficient estimates in Table 4 do not suffer from a sample selection bias. For sure, employees' roles are not assigned randomly but are endogenously chosen by players C. When Cs take the track records of their two employees into account in making this assignment choice (as the theory predicts), the track records of the two employees A and B are no longer exogenous. The crucial thing to note here, however, is that selection has been based on *observables*. As has been observed in the literature, selection on observables per se does not lead to a sample selection bias in the estimates.[19] For such a bias to occur the (for the experimenter) unobserved characteristics of C that drive her assignment decision should be correlated with the (for the experimenter) unobserved characteristics of A that drive his entry decision. Given that all the information subjects within a matching group have about each other comes from the experimenter (viz. the track records) and the experiment is based on a stranger design, so reputational concerns are absent by construction, there is no compelling reason at all for such a correlation to exist. Player A's entry decision can thus be studied independently from C's assignment choice. For the very same reason it can also be analyzed separately from C's game choice in periods 11 to 15.

In all specifications of Table 4 that include r_B (first row), the r-value in B's track record is a highly significant determinant of A's decision to enter. The higher r_B is, the larger A's entry propensity. This finding is in line with our first hypothesis H1. Another consistent pattern is that entry is also significantly more likely in the I_H and the M-game as compared to the I_L-game.[20] The time trend 'period' is always significantly negative, indicating that the propensity to enter decreases over time. In all specifications the 'endo' dummy and its interaction with the game dummies are insignificant. Only for the I_H-game the sum of 'endo' and 'I_H·endo' appears significant (in all specifications according to a Wald $\chi^2(1)$ test), indicating that players A are more likely to enter when the I_H-game is endogenously chosen by player C than when this game is exogenously given.[21]

With regard to the other track record characteristics, the second specification includes e_B rather than r_B. Given the high correlation between e_i and r_i, it is not surprising to find that e_B significantly increases A's inclination to enter. Once r_B and e_B are both included, however, the coefficient belonging

[19] See e.g. Vella (1998, Section II) and Cameron and Trivedi (2005, Section 16.5.7). Paraphrasing the words of Vella (1998, p. 129): "When the relationship between the work [assignment] decision and the wage [entry decision] is purely through the observables, however, one can control for this by including the appropriate conditioning variables in the wage [entry decision] equation. Thus, sample selection bias will not arise purely because of differences in observable characteristics." Vella continues to explain that differences in observables across two samples "...is by no means necessary, or even indicative, of selection bias."

[20] This not only follows from the significance of the I_H-game and M-game dummies in isolation, but also from testing sums of these coefficients and the relevant coefficients belonging to the interaction terms as well.

[21] Because player A does not know the payoffs of player C (cf. Section 3), this difference cannot be explained by A having empathic feelings towards player C. Player B's reaction does not provide a convincing explanation either, because B's rewarding behavior appears not to vary with the game being exogenously given or endogenously chosen (cf. Table 5).

Table 4: Random effects probit estimations of A choosing Enter

	(1)	(2)	(3)	(4)	(5)	(6)
r_B	0.088***		0.092***	0.172***	0.107***	0.172***
	(0.017)		(0.021)	(0.030)	(0.023)	(0.030)
e_B		0.062***	-0.011	-0.016	-0.011	-0.016
		(0.023)	(0.029)	(0.029)	(0.029)	(0.029)
r_A				0.106***		0.103***
				(0.027)		(0.030)
e_A					0.044*	0.008
					(0.025)	(0.027)
I_H-game	1.098***	1.233***	1.093***	0.992***	1.099***	0.996***
	(0.339)	(0.333)	(0.339)	(0.345)	(0.341)	(0.345)
M-game	1.672***	1.671***	1.664***	1.573***	1.629***	1.569***
	(0.325)	(0.322)	(0.326)	(0.331)	(0.327)	(0.331)
period	-0.081***	-0.080***	-0.081***	-0.086***	-0.084***	-0.087***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
First-I	0.584*	0.745**	0.578*	0.384	0.523	0.379
	(0.341)	(0.336)	(0.342)	(0.350)	(0.345)	(0.350)
endo	0.283	0.308	0.283	0.387	0.336	0.394
	(0.386)	(0.382)	(0.386)	(0.389)	(0.387)	(0.389)
I_H·First-I	-0.763**	-0.905**	-0.769**	-0.455	-0.662*	-0.444
	(0.387)	(0.381)	(0.388)	(0.401)	(0.394)	(0.403)
M·First-I	-0.651*	-0.753*	-0.648	-0.455	-0.549	-0.442
	(0.394)	(0.389)	(0.394)	(0.402)	(0.399)	(0.404)
I_H·endo	0.415	0.372	0.417	0.377	0.404	0.376
	(0.396)	(0.389)	(0.396)	(0.398)	(0.396)	(0.398)
M·endo	0.083	0.065	0.081	-0.013	0.035	-0.019
	(0.329)	(0.324)	(0.329)	(0.331)	(0.330)	(0.332)
constant	-1.669***	-1.692***	-1.636***	-2.064***	-1.820***	-2.084***
	(0.358)	(0.367)	(0.369)	(0.389)	(0.386)	(0.395)
Log L	-454.015	-463.572	-453.947	-446.418	-452.413	-446.373
N (clusters)	900 (4)	900 (4)	900 (4)	900 (4)	900 (4)	900 (4)
LR-chi2	125.366***	106.251***	125.501***	140.560***	128.569***	140.648***

Remark: Dummy First-I equals one for sessions 1 and 3 where the I-game is played first (and zero otherwise) while dummy endo equals one for periods 11 through 15 in which the game is endogenously chosen. Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Estimates based on clustering on sessions. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

to e_B becomes insignificant while the one belonging to r_B remains highly significant. This shows that player B's r_B value affects the entry decision of player A while, ceteris paribus, his e_B value does not.

Specification (4) through (6) reveal that a similar conclusion applies regarding A's characteristics; only r_A matters and (ceteris paribus) e_A does not. The significance of r_A suggests that player A's expectation about B's reaction to A's entry choice partially depends on how A would have behaved were he in B's position. Interestingly, in columns (4) and (6) the coefficient belonging to r_B is significantly larger than the one belonging to r_A ($p < 0.001$ and $p = 0.007$, respectively). Together with the insignificance of both e_A and e_B, this implies that player A's entry decision is more strongly affected by B's track record characteristics than by A's own track record.[22] Player A's own social preferences are thus of second order. We summarize our main finding from Table 4 in Result 1.

Result 1. The higher r_B is, the more likely it is that player A enters. Player B's main track record characteristic r_B is a stronger determinant of A's entry choice than player A's own track record characteristic r_A (or e_A) is.

We also calculated random effects probit estimates of the probability that B chooses to reward entry, see Table 5. We did so mainly as a consistency check of whether B's with a higher r_B-value as measured in part one are (ceteris paribus) indeed more likely to reward entry in part two. This is important to verify, because the r_B-values are obtained from decisions made under the strategy method whereas in part two of the experiment subjects make real (or 'hot') decisions instead of contingent ones. Moreover, also the way in which subjects were rewarded for their decisions differed between parts one and two (cf. Section 3). It could potentially be the case that the payment mechanism has important consequences for behavior. Finally, subjects were informed that their part one choices would potentially become public in part two. This may have made them more willing to reward in part one than in part two. In that case r_B may be interpreted as an upper bound of B's actual empathy level (ρ_B).

The estimates obtained reveal that r_B has explanatory power regarding the probability that B rewards entry; in the first column of Table 5 (only) the coefficient belonging to r_B is significant.[23] As one would expect, however, also player B's e_B value has such explanatory power, see column (2). When both r_B and e_B are included in the regressions (see columns (3) through (6)), their independent effects cannot be isolated and both coefficients become insignificant. This suggests that e_B and r_B are equally informative about player B's likely behavior. It is also found that the track record characteristics of A have no impact on B's reward decision. The behavior of player B is independent of whether the game is exogenously given or endogenously chosen by player C as well.[24]

The findings regarding the impact of r_B are consistent. The higher r_B in the track record of player B is, the more likely it is that he will reward entry. This in turn makes it more attractive for A to enter, in

[22] To properly take account of differences in average size and variation in r_A and r_B, a cleaner comparison is to first standardize the track record variables (with mean 0 and standard deviation 1) before comparing their coefficients. If we do so, exactly the same conclusion is obtained; the coefficient belonging to the standardized r_B variable is significantly larger than the one belonging to the standardized r_A variable.

[23] To verify whether inconsistency makes r_B a less good predictor of reward behavior in part 2, we also considered an alternative specification that added the interaction term $I_{Incon} \cdot r_B$ to specification (1). Here dummy I_{Incon} equals 1 only if the subject has an inconsistency level that exceeds one (cf. Table 3). The coefficient of the interaction term is small and insignificant (0.013 with standard error 0.041) while the coefficient belonging to r_B remains largely unaffected (0.068 with s.e. 0.032). Hence a higher level of inconsistency does not make r_B a poor predictor of future behavior.

[24] This follows from the insignificance of the coefficients belonging to 'endo', 'I_H·endo' and 'M·endo', and of the appropriate sums of these coefficients (Wald $\chi^2(1)$ tests).

Table 5: Random effects probit estimations of B choosing Reward

	(1)	(2)	(3)	(4)	(5)	(6)
r_B	0.071**		0.061	0.084	0.048	0.087
	(0.030)		(0.040)	(0.056)	(0.042)	(0.056)
e_B		0.076*	0.021	0.017	0.024	0.019
		(0.042)	(0.055)	(0.056)	(0.056)	(0.056)
r_A				0.032		0.064
				(0.053)		(0.060)
e_A					-0.039	-0.063
					(0.046)	(0.052)
I_H-game	0.101	0.193	0.100	0.094	0.045	-0.000
	(0.960)	(0.953)	(0.961)	(0.954)	(0.978)	(0.972)
M-game	-0.990	-1.014	-0.993	-1.002	-1.059	-1.119
	(0.966)	(0.962)	(0.968)	(0.961)	(0.986)	(0.980)
period	-0.083	-0.083	-0.083	-0.082	-0.090	-0.093
	(0.056)	(0.056)	(0.056)	(0.056)	(0.056)	(0.057)
First-I	-0.755	-0.718	-0.768	-0.829	-0.808	-0.955
	(1.020)	(1.014)	(1.022)	(1.020)	(1.037)	(1.039)
endo	0.779	0.874	0.794	0.852	0.792	0.907
	(0.883)	(0.890)	(0.884)	(0.889)	(0.888)	(0.895)
I_H·First-I	-0.076	-0.085	-0.042	0.060	-0.115	0.049
	(1.051)	(1.048)	(1.057)	(1.064)	(1.073)	(1.077)
M·First-I	0.783	0.788	0.804	0.866	0.853	1.008
	(1.123)	(1.120)	(1.126)	(1.125)	(1.140)	(1.144)
I_H·endo	-0.709	-0.784	-0.718	-0.782	-0.682	-0.789
	(0.878)	(0.883)	(0.879)	(0.884)	(0.884)	(0.888)
M·endo	-0.481	-0.546	-0.487	-0.565	-0.430	-0.553
	(0.758)	(0.763)	(0.759)	(0.769)	(0.764)	(0.772)
constant	0.535	0.479	0.490	0.334	0.776	0.634
	(1.029)	(1.032)	(1.038)	(1.064)	(1.104)	(1.105)
Log L	-130.521	-131.623	-130.451	-130.275	-130.090	-129.526
N (clusters)	237 (4)	237 (4)	237 (4)	237 (4)	237 (4)	237 (4)
LR-chi2	28.337***	26.135***	28.479***	28.831***	29.199***	30.328***

Remark: Dummy First-I equals one for sessions 1 and 3 where the *I*-game is played first (and zero otherwise) while dummy endo equals one for periods 11 through 15 in which the game is endogenously chosen. Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Estimates based on clustering on sessions. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

line with what we observe in Result 1. Player B's track record thus contains valuable information about how he is going to behave, which is actually used by player A to guide her entry decision. In the next subsection we investigate whether this mechanism is recognized as such by player C when deciding on role allocation and game choice.

Another observation that follows from our findings is that, *given* the employees' r-values, their e-values do not provide useful additional information about their likely future behavior. More specifically, regarding player B's likely behavior e_B and r_B are equally informative whereas player A's entry choice is mainly guided by r_B and (to a lesser extent) r_A, and not by e_A or e_B. The informational value of the individual track records thus lies in the r-values. The e-values are informative only to the extent that they are correlated with the r-values.

4.3 Allocation of roles and game choice

In part 2 player C makes two types of choices. First, in every period she decides on role assignment. Observing the track records of her two employees, player C decides who gets the role of player A and who becomes player B. Moreover, in periods 11 to 15 player C also chooses, before role assignment, the game that is going to be played. These two different choices are discussed in turn.

4.3.1 Role assignment

In studying the assignment choices we focus on the data of the first ten periods only where the game is exogenously given. We do so because in periods 11 to 15 C's assignment decision cannot be treated independently from C's choice of game. These interdependent choices are studied in the next subsection.

Figure 4 provide an overview of the assignment decisions in each of the three different games. This figure reveals whether the employee who obtained role B is the one with the higher r-value in his track record ($r_A < r_B$; see the black bars) or whether this is the other way around ($r_A > r_B$; light dotted bars), and similarly so for the e-values.[25] For the observations on the diagonal from lower-left to upper-right it holds that the ranking of r-values and e-values in the two roles coincide. For the off diagonal observations these rankings are opposite of each other.

[Insert Figure 4 about here]

From the observed assignment patterns it immediately follows that our second hypothesis is rejected; players C do not predominantly assign the role of player B to the employee with the higher r-value in his track record. In the two inspection games it does hold that C is (weakly) more likely to assign role B to the employee with the higher r-value. In the motivation game this is actually the other way around. There player C is more likely to assign role A to the higher r-value within her work force.[26] Given the

[25] Owing to our role assignment procedure based on the ranking of the r_i-values, the two employees within a group never had the same r-value. This does not hold for the e_i values though, explaining why we have $e_A \geq e_B$ (versus $e_A < e_B$) instead of $e_A > e_B$ in Figure 4.

[26] To account for the multiple assignment decisions per player C, we formally test this as follows. For each individual player C we compute the relative frequency of $r_B > r_A$ for each (exogenous) game separately. We then test whether these frequencies are significantly different from 50% by means of signrank tests. In the I_L-game there is no significant difference ($p = 0.1124$) while for the I_H-game this frequency is significantly above 50% ($p = 0.0034$) and in the M-game significantly below ($p = 0.0010$).

fairly high correlation between subjects' r and e-values, the order of e-values among the two employees typically corresponds with the order of r-values. Assignment on the basis of relative e-values thus often (but not always) coincides with allocation on the basis of relative r-values.

Comparing assignment patterns of r-values across games, we find that these do not differ between the two inspection games. But they are significantly different for the motivation game; the relative frequency of $r_B > r_A$ (per individual player C) is significantly lower in the motivation game as compared to the frequencies in the inspection games.[27] As a result of this, managers in the M-game (players B) have on average an r-score that does not differ significantly from managers in the I-game (players A).[28] This finding contrasts with the theoretical prediction that managers in the M-game will be more empathic than managers in the I-game.

Result 2. In the inspection games Cs (weakly) more often assign role B to the high-r employee; from 57% in game I_L to 63% in game I_H. In contrast, in only 35% of the cases where the motivation game is played, players C assign the role of player B to the employee with the higher r-value.

Result 2 is opposite to what we expected. Especially in the M-game it is important for player C to stimulate entry, because the payoff after Out is low, and assigning the high-r employee to role B appears an effective instrument to do so (cf. Result 1). Yet the majority of player Cs does not do this.

A potential explanation why Cs in game M tend to assign role B to the low-r employee is that they naively assume that A's and B's decisions are (only) driven by their own e_A and r_B values, respectively. In particular, Cs may overlook that A's actual entry decision is mainly guided by the value of r_B.[29] If Cs indeed have such "naive" expectations, they would prefer to assign the high-e employee to role A and the low-r employee to role B. This follows because C gets the same as player B in game M and thus is necessarily better off when A enters instead of staying out (and she is best off when B chooses no reward in reaction). This could explain the direction in the assignment patterns we observe; high (e_i, r_i) types typically get role A whereas low (e_i, r_i) types usually get role B. Moreover, one would expect naive Cs to focus predominantly on relative e-values in order to stimulate entry, because stimulating no reward in reaction to entry is useful only when entry can be induced.

For the two inspection games matters are less clear under "naive" expectations. Surely, C then prefers to give role B to the high-r employee. This maximizes the probability that B chooses to reward after A enters. But given that C gets the same as A in these games, entry is now risky for player C. She may end up with the lowest payoff when B decides not to reward A's entry choice. It is therefore a priori unclear whether a naive C wants to increase the probability of entry by assigning the high-e employee to role A, or whether she prefers to avoid the worst outcome (Enter, No reward) in this game by giving

[27] These conclusions are based on comparing the relative frequencies of $r_B > r_A$ of individual player C's across games. For I_L versus M and I_H versus M signrank tests (for matched pairs) yield p-values of 0.0133 and 0.0009, respectively. For I_L versus I_H a ranksum test (unmatched data) gives $p = 0.4740$.

[28] This follows from comparing (per individual player C) the average value of r_B under game M with the average value of r_A under games I_L and I_H, respectively. Using signrank tests we obtain a p-value of 0.4715 for M versus I_L and of 1.000 for M versus I_H.

[29] One potential reason for this might be that players C fail to focus on the fact that not only she observes the track records of her employees', but also the employees themselves do so of each other. Note, however, that subjects were explicitly informed about this aspect in the instructions (which were both supplied on paper and read aloud). Even so, it is conceivable that a design in which a strong emphasis is put on the fact that all group members observe the track records of the two employees (by e.g. explicitly reminding the subjects of this at the beginning of every new period) would lead to different results.

Table 6: RE probit estimations of (lower id) employee getting role B

		I-game			M-game	
	(1)	(2)	(3)	(4)	(5)	(6)
$r_1 - r_2$	0.045**		0.057**	-0.068***		-0.032
	(0.020)		(0.027)	(0.014)		(0.020)
$e_1 - e_2$		0.026	-0.023		-0.115***	-0.081***
		(0.026)	(0.035)		(0.022)	(0.030)
$(r_1 - r_2) \cdot I_H$	0.017		-0.041			
	(0.028)		(0.040)			
$(e_1 - e_2) \cdot I_H$		0.115**	0.143**			
		(0.047)	(0.067)			
I_H-game	0.003	0.019	-0.028			
	(0.148)	(0.152)	(0.154)			
First-I	0.073	0.050	0.038	0.051	-0.131	-0.076
	(0.301)	(0.302)	(0.304)	(0.302)	(0.307)	(0.309)
period	-0.026	-0.022	-0.024	-0.011	0.012	0.004
	(0.052)	(0.052)	(0.053)	(0.052)	(0.053)	(0.053)
constant	0.112	0.081	0.143	-0.028	0.018	0.017
	(0.437)	(0.439)	(0.442)	(0.188)	(0.190)	(0.190)
Log L	-199.040	-198.880	-196.518	-195.049	-192.641	-191.339
N (cluster)	300 (4)	300 (4)	300 (4)	300 (4)	300 (4)	300 (4)
LR-chi2	17.595***	17.915***	22.639***	24.175***	28.991***	31.595***

Remark: Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Estimates based on clustering on sessions. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

the low-e employee role A. Under naive expectations one thus expects that allocation decisions may be driven by both the relative r-values and the relative e-values of the two employees.

Given the substantial correlation between e_i and r_i, it is difficult to identify precisely their separate effects on the probability of obtaining role B. But the random effects probit estimates in Table 6 provide suggestive evidence. These estimates are calculated as follows. For each allocation decision of player C we focus on the employee with the lower subject id in player C's current group of employees. Because subject id's are allocated at random, this corresponds to a random selection of one of the two employees player C has in a given period. For these employees we estimate the probability that they are assigned role B (with the data clustered at the session level). We do so for the inspection games and the motivation game separately. Two main explanatory variables are considered: (i) the difference in r-values between the lower id employee and the other employee $(r_1 - r_2)$ and (ii) the difference in e-values between them $(e_1 - e_2)$. Because these two differences are highly correlated (just as the levels are),[30] we consider specifications in which only one of the two is included and a specification where they are both included. In the specifications for the inspection game the track record differences are interacted with a dummy indicating game I_H, to test whether the level of inspection costs matter. As before we also include a time trend 'period' and a dummy for potential sequence effects (i.e. 'First-I') as controls.

For the two inspection games opposite results are found. In the I_L-game players C appears mostly

[30] The Spearman rank correlation equals 0.5304 and is highly significant ($p < 0.0001$).

concerned with getting the high r-value in role B; in both specifications (2) and (3) $e_1 - e_2$ is insignificant while $r_1 - r_2$ is significant in both (1) and (3). In game I_H player C seems to focus mainly on getting the low e-value in role A. This follows because in (3) the sum of coefficients belonging to $e_1 - e_2$ and $(e_1 - e_2) \cdot I_H$ is significant (Wald test: $\chi^2(1) = 4.52$ and $p = 0.033$) whereas the sum of coefficients belonging to $r_1 - r_2$ and $(r_1 - r_2) \cdot I_H$ is not ($\chi^2(1) = 0.27$ and $p = 0.6067$).[31] Note, however, that given the high correlation between e_i and r_i, assigning the higher r-value to role B (as in I_L) typically coincides with assigning the lower e-value to role A (as in I_H). Therefore, for many allocation decisions this shift in focus between I_L and I_H does not matter.[32] For the motivation game we find that both $r_1 - r_2$ and $e_1 - e_2$ are highly significant when considered in isolation, but only the coefficient of the latter remains significantly negative when both are included. This suggests that in this game C's are mostly concerned with getting the high-e employee in role A.

These patterns make sense when players C have naive expectations as described earlier. In the two inspection games players C will be mostly concerned with avoiding the very unattractive outcome (Enter, No reward), and they think they can do so by having the higher r-value in role B or the lower e-value in role A. In the motivation game players C particularly would like to stimulate entry. Naive Cs think that this is best accomplished by assigning the higher e-value to role A.[33]

The above discussion suggests that (naive) players C make suboptimal allocation decisions, especially in the motivation game. Result 1 indicates that entry is best stimulated by assigning the higher r-value to role B. On the other hand, such an allocation also stimulates B to reward entry and – conditional on entry – C would be better off if B chooses no reward instead. Moreover, A's entry decision also appears to be significantly positively affected by A's own r_A value, suggesting that an empathic employee may already be inclined to enter even when his manager is not particularly empathic. To assess the overall effect we therefore investigate how player C's profits vary with her allocation decision. Table 7 presents random effects regression estimates of player C's profit for each of the three games separately (with the data clustered at the session level). Here the data is again restricted to the first ten periods in which the game is exogenously given. As explanatory variables we include the track record characteristics of C's two employees, with subscript 1 (2) again referring to the employee with the lower (higher) subject id.[34] We also include a sequence dummy and a time trend as controls. The impact of C's allocation decision is measured by a dummy $I_{\{r_A<r_B\}}$, equal to one only if the high-r employee is assigned role B.

In regard to the two inspection games the allocation dummy $I_{\{r_A<r_B\}}$ appears insignificant. We find some evidence that a more empathic work force lowers profits under the I_L game, but employees' track record characteristics do not affect profits in the I_H game. An explanation for the former finding is that a more empathic work force not only makes outcome (Enter, Reward) more likely, but also the worst possible outcome (Enter, No reward).[35] For the M-game we observe that the e-values of the two

[31] In specification (1) $(r_1 - r_2) + (r_1 - r_2) \cdot I_H$ is significant ($\chi^2(1) = 9.69$, $p = 0.0019$), just as $(e_1 - e_2) + (e_1 - e_2) \cdot I_H$ in (2) is ($\chi^2(1) = 13.33$, $p = 0.0003$).

[32] Nevertheless, it remains rather puzzling that in game I_H the owner focuses more than in game I_L on assigning the low-e employee to role A. Because owners should rather prefer entry in I_H than in I_L, one intuitively would expect this to be the other way around.

[33] An alternative possibility is that Cs do realize that in theory As should take Bs preferences into account, but do not trust As to be rational. Moreover, for the reasons discussed immediately below Result 1, Cs may expect hypothesis H1 to fail in practice. For instance, Cs may believe that As simply take r_B as an upper bound on empathy level and fear defection in part 2 where reward choices are not publicly recorded.

[34] Like in Tables (4) through (6) we also considered specifications that only included r_i values and specifications that only included e_i values. In all these specifications the relevant dummy $I_{\{r_A<r_B\}}$ was insignificant.

[35] This follows from running separate RE probit estimates of the probability of outcome (Enter, Reward) and (Enter, No

Table 7: Random effects regressions of player C's profit

	I_L-game	I_H-game	M-game
r_1	-0.646	2.147	18.950***
	(4.480)	(5.964)	(7.162)
e_1	-3.006	0.007	8.861
	(3.641)	(6.814)	(6.392)
r_2	-8.107*	4.430	12.898*
	(4.735)	(6.528)	(7.303)
e_2	-7.161*	3.587	6.272
	(3.718)	(6.733)	(6.911)
I-first	-39.003	-50.561	116.431
	(49.754)	(64.329)	(86.737)
period	-6.481	-8.380	-29.249***
	(5.046)	(6.543)	(8.196)
$I_{\{r_A<r_B\}}$	14.629	-28.578	-2.904
	(15.264)	(19.649)	(24.750)
constant	531.724***	402.408***	391.704***
	(56.179)	(76.210)	(74.588)
Overall R^2	0.113	0.044	0.086
N (clusters)	150 (2)	150 (2)	300 (4)
Wald-chi2	18.02**	6.47	27.62***

Remark: Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. In all three specifications a Lagrange multiplier test for random effects is insignificant. Wald-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

Table 8: Game choices and assignment by session

Session	Game I_L		Game I_H		Game M		Total
	$r_A > r_B$	$r_A < r_B$	$r_A > r_B$	$r_A < r_B$	$r_A > r_B$	$r_A < r_B$	
1	13 (17%)	26 (35%)			24 (32%)	12 (16%)	75 (100%)
2	12 (16%)	23 (31%)			36 (48%)	4 (5%)	75 (100%)
1 & 2	25 (17%)	49 (33%)			60 (40%)	16 (11%)	150 (100%)
3			11 (15%)	13 (17%)	35 (47%)	16 (21%)	75 (100%)
4			5 (7%)	11 (15%)	38 (51%)	21 (28%)	75 (100%)
3 & 4			16 (11%)	24 (16%)	73 (49%)	37 (25%)	150 (100%)

employees do not significantly affect player C's profits. The r_i-values, on the other hand, significantly increase profits. Also here the allocation dummy is insignificant. Apparently the two opposing forces as described in the previous paragraph cancel out and we do not obtain evidence that players C make suboptimal allocation decisions in the motivation game.

Result 3. In the inspection game with low inspection costs (I_L) players C focus on getting the high-r value in role B whereas for high inspection costs (game I_H) they focus on assigning the low-e employee to role A. In the motivation game Players C are mostly concerned with getting the high-e employee in role A. In all types of games profit levels are largely insensitive to the assignment of the employees.

4.3.2 Game choice

We finally look at the choices players C make in the final five periods of part 2. Player C then decides on both the organizational mode and on role assignment and we therefore first consider these decisions jointly. Table 8 provides an overview of the actual choices made. With respect to assignment this table indicates whether the employee with the higher r-value obtains role A ($r_A > r_B$) or whether s/he obtains role B ($r_A < r_B$).

From Table 8 it can be observed that when player C chooses between I_L and M, she is about equally likely to choose either game. But when the choice is between I_H and M, she chooses the motivation game in around 73% of the cases. In line with theoretical predictions, therefore, C is more likely to choose M over I_H than M over I_L.

With regard to assignment patterns the findings are the same as before. Taking the two inspection games together players C allocate role B to the high r-value employee in overall 64% of the cases (73 out of 114) when this game is endogenously chosen. This compares well with the 60% observed for the exogenous I-games (cf. Result 2). For the motivation game the corresponding percentages are 28% for the endogenous treatments and 35% in the exogenous games. Hence, also when the motivation game is endogenously chosen, player C is more likely to assign role B to the low-r employee. These findings corroborate Result 2.

In order to explore which employees' characteristics drive game choice Table 9 reports random effects probit estimates of the probability that game M is chosen. In these estimates r_{High} (r_{Low}) refers to the r-value of the employee with the higher (lower) r in his track record. Variables e_{High} and e_{Low}

reward), respectively.

Table 9: Random effects probit estimations of C choosing game M

	(1)	(2)	(3)	(4)	(5)	(6)
r_{High}	0.090*		0.057	0.063	0.057	0.062
	(0.047)		(0.048)	(0.049)	(0.049)	(0.049)
e_{High}		0.149***	0.135***	0.138***	0.135***	0.144***
		(0.047)	(0.049)	(0.049)	(0.050)	(0.050)
r_{Low}				0.186		0.210
				(0.129)		(0.137)
e_{Low}					-0.002	-0.030
					(0.052)	(0.055)
r_C	-0.032	0.010	-0.012	-0.069	-0.011	-0.069
	(0.083)	(0.081)	(0.084)	(0.093)	(0.085)	(0.094)
e_C	0.004	0.027	0.019	0.019	0.019	0.017
	(0.045)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
M versus I_H	0.386**	0.598***	0.586***	0.640***	0.585***	0.633***
	(0.168)	(0.184)	(0.185)	(0.189)	(0.186)	(0.189)
$\pi_M - \pi_I$	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
period	-0.032	-0.035	-0.035	-0.035	-0.035	-0.035
	(0.054)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
First-I	-0.215	-0.031	-0.058	-0.056	-0.057	-0.046
	(0.164)	(0.174)	(0.175)	(0.176)	(0.176)	(0.177)
constant	0.084	-0.547	-0.679	-0.674	-0.677	-0.636
	(0.792)	(0.833)	(0.844)	(0.848)	(0.846)	(0.851)
Log L	-182.069	-178.852	-178.151	-177.104	-178.150	-176.957
N (clusters)	300 (4)	300 (4)	300 (4)	300 (4)	300 (4)	300 (4)
LR-chi2	25.457***	31.889***	33.293***	35.386***	33.294***	35.680***

Remark: Dummy First-I equals one for sessions 1 and 3 where the I-game is played first (and zero otherwise). Standard errors in parentheses. ***/**/* indicates significance at the 1/5/10% level. Estimates based on clustering on sessions. LR-chi2 reports the test statistic from testing that all coefficients (except the constant) are zero.

are defined similarly. Apart from the track record characteristics of player C's two employees, we also include as explanatory variables a dummy equal to one if and only if I_H is the alternative to game M and a variable $\pi_M - \pi_I$ measuring the difference in average realized profits player C obtained from the two games in the first ten periods (where these games were exogenously given). Intuitively, one would expect that C is more likely to choose game M when I_H is the alternative and when game M yielded her higher profits than game I did in the past. We also include a time trend 'period', a sequence dummy $First - I$ and player C's own track record characteristics as controls.

The estimates in Table 9 paint a consistent pattern regarding game choice. From the employees' characteristics, only e_{High} matters. The higher e_{High} is, the more likely it becomes that player C chooses the M-game. The other variables in the employees' track records are typically insignificant. The estimates thus do not lend support to the hypothesis that the higher the value of r_{High} within player C's work force is, the more likely it is that the M-game is chosen (cf. H3). The single two other important determinants of game choice are the profit difference $\pi_M - \pi_I$ and the M versus I_H dummy. Not surprisingly, the better (relative) experience player C has had with game M in the past, the more likely she is to choose this game over the inspection game. The same applies when the alternative to the M-mode is worse, i.e. I_H instead of I_L.

Result 4. (i) Players C's choice between the motivation game and the inspection game is not guided by the r_i-values in her employees' track records. From these records only the highest e-value e_{High} matters; the higher e_{High}, the more likely it is that C chooses game M. (ii) player C is more likely to choose game M over game I_H than over game I_L.

Overall the following general picture emerges from our findings. For players C the motivation game is attractive only if entry can be induced in this game. They naively think that this can be best accomplished by allocating role A to the employee with the highest e_i-value. Player Cs are therefore likely to choose game M only when this e_{High}-value is relatively high (Result 4). In case e_{High} is low Cs are more likely to choose the inspection game and will assign this player the role of B (Result 2). A rationale for this assignment is that player C hopes to avoid the subgame perfect Nash equilibrium outcome (Out, No reward) in this way. Allocation decisions are naive in the sense that Cs seem to overlook the significant impact of B's track record (in particular r_B) on A's entry choices, especially in game M where this relationship is of vital importance (Result 1). They therefore mainly look at the employee's own track record to form expectations about how he would behave in role A. This leads to allocation choices that differ from equilibrium predictions (Result 2). Yet, because player A's behavior appears to be affected to some extent by his own characteristics as well (in line with Cs "naive" expectations!), profit levels appear largely insensitive to assignment (Result 3). Although game choice is mainly guided by the employees' e_i-values, players C in general therefore do correctly realize that social preferences make organizational mode M more attractive than mode I_H.

5 Concluding discussion

A number of theoretical papers have argued that a manager's social preferences may act as commitment device that makes non-enforceable, 'implicit' incentive contracts with workers feasible; see e.g. Rotemberg and Saloner (1993), Rotemberg (1994), and Hermalin (2001). The underlying intuitive idea is that when managers (sufficiently) empathize with their employees, they are (much) less inclined to renege on implicit informal agreements that hard work will be rewarded. This in turn motivates selfish and more prosocial workers alike to put in high effort. Given its reliance on social preferences only, this type of incentive provision may potentially serve as a cost-effective alternative to a more formal incentive scheme based on explicit contracts and active monitoring. Crucial for the informal mechanism to work though is that those who hire managers ('owners') understand that these managers' preferences are (potentially) key for worker behavior. Moreover, owners should also realize that an organizational mode based on informal implicit agreements is viable only if managers are sufficiently empathic. Otherwise a more formal mode with explicit contracts and active monitoring is predicted to perform better.

This paper reports the results from a stylized laboratory experiment designed to test whether subjects in the role of firm owner recognize the above behavioral forces. We make use of simple trust games that capture in highly reduced form the main characteristics of the owner's decision problem. The main questions of interest are (i) whether (and how) owner subjects take the observed social preferences of other ('employee') subjects with whom they are matched into account when choosing which payoff version of the trust game these employee subjects should play ('the organizational mode') and (ii) whether owner subjects allocate the more 'empathic' employee to the second mover role in the trust game at hand.

Our main findings are that owner subjects do take the social preferences of their employees into account when choosing between games and when making allocation decisions, but in a different way than theory predicts. In the trust game representing the informal ('motivational') mode they often assign the second mover role to the *less* empathic employee. The choice between games appears to be mainly guided by the social preferences of the intended candidate for the first mover role. Overall our results indicate that the importance of the first mover's social preferences for trusting behavior is recognized by the owner subjects, but also that the significant (first order) impact second movers' social preferences have on trusting behavior of first movers seems to be overlooked. Owner subjects are thus unable to fully backward induct in view of the observed social preference characteristics of their employee subjects. Nevertheless, they do correctly realize that the (potential) existence of 'strong' social preferences within their work force makes the trust game reflecting the informal mode relatively more attractive.

The observation that owner subjects do not backward induct sufficiently is in line with earlier experimental findings that people tend to analyze extensive form games in a forward rather than in a backward manner. Johnson et al. (2002) consider a three-round alternating offer bargaining game in which the pie up for division shrinks over time. The actual pie sizes in each of the three rounds were hidden in boxes on the computer screen, which could be opened by moving the cursor into the box. The computer software recorded which box was opened, for how long, the order in which the boxes were opened etc.. Strikingly, most subjects focused on the first round box and did not sufficiently look ahead. In a non-negligible fraction of observations, subjects did not even open the round two and round three boxes. In a classroom experiment Rubinstein (1999) also finds that people have a natural tendency to analyze extensive form games forward rather than backwards. Our results suggest that these findings carry over to situations in which people have to form expectations about how others will behave in a sequential game. In our experiment owner subjects tend to look forward at the first (trusting) decision taken by player A, overlooking the fact that this decision is affected by player A's expectation about B's likely response and for which the revealed social preferences of B provide relevant information.

Clearly, given the highly stylized nature of our lab experiment, this paper only provides one very specific piece of evidence that may potentially prove helpful in understanding firm structure and behavior in practice. We test the recognition of the behavioral forces that underly the game-theoretical analyses of using a manager's social preferences as implicit contract enforcement device in (arguably) the simplest setting possible. We find that owner subjects do take account of others' social preferences, but only to the extent to which these are expected to affect a person's own behavior. This indicates that recognition of the indirect impact of one person's social preferences on another person's behavior with whom he interacts is not innate and points at a potential cognitive bias that may also be relevant in organizational practice (cf. Camerer and Malmendier (2007)). In particular, our results suggest that – regarding incentive provision for workers – firm owners may underestimate the importance of managers' social preferences and overestimate the importance of workers' own social preferences in providing implicit incentives.

As pointed out by Camerer and Weber (2007), the mismatch in scope between simple lab experiments that use randomly matched subjects that are unfamiliar to each other and much more complex real life organizations in which (self-) selected and highly experienced agents interact personally, leads to heightened concerns about generalizability of lab results. Indeed our stylized experiment ignores a variety of aspects that are likely to be important drivers of behavior in the field. For instance, owners will typically have much more experience, know their employees personally, engage in repeated interaction with them and usually have much more flexibility in the kind of (explicit or implicit) contracts they can offer.[36] These concerns limit the inferences that can be made from our experiment for actual firm behavior. At the same time, however, our findings do suggest that also in practice social preferences may be exploited differently than the established theoretical models predict.[37]

Appendix: basic model of endogenous organizational design

In the experiment subjects are confronted with the simple games depicted in Figure 1. To motivate our choice of monetary payoffs in these games, we consider in this Appendix a bare bone reduced form model of endogenous organizational design. Because the main purpose here is to justify our parameter choices, in this model we abstract away from the owner's assignment decision.

A firm consists of three agents: the owner who owns the firm, a manager hired to run the firm on her behalf and a worker doing the productive work. The worker can either put in low effort ('shirk') or high effort ('work'). In the former case the value of his productivity equals v_0 whereas in the latter case it is v_1 (here all parameters are positive). The worker's disutility of putting in high effort equals g. Therefore, a selfish worker will shirk if no additional measures are taken.

One way to motivate the worker to put in high effort is to set up a performance monitoring system. We assume that such a system, when fully implemented, always induces the worker to work. It brings about three types of costs though. First, there are the costs k of setting up and installing the monitoring technology. Investments in technological equipment and organizational procedures are needed to allow accurate measurement of the worker's productivity.[38] Second, h denotes the firm's inspection costs. Even with the monitoring technology in place, scarce resources like the manager's time need to be devoted to monitor the worker. Third, the worker dislikes being monitored because it gives him the feeling of being controlled (cf. Frey (1993), Falk and Kosfeld (2006)), leading to a disutility of d. We assume that the overall costs of the formal monitoring system fall short of the net benefits of getting the worker to work:

$$k + h + d < v_1 - v_0 - g. \tag{A1}$$

Therefore, in the absence of alternative incentive instruments, the firm would benefit from using a formal monitoring system. As in the main text we will refer to this as the 'inspection mode', or I-mode in short. In regard to compensation we assume that the worker receives a fixed wage w_I. The manager is paid on

[36] The latter need not always be the case though. In actual firms bonus contracts may not be feasible because of minimum wage requirements that effectively restrict the range of bonuses the firm is willing to pay. Likewise, in reality informal bonus contracts may effectively be legally limited by anti-discrimination laws, because verifiable performance evidence is needed to justify the bonus (and the cost advantage relative to a formal contract is lost).

[37] Camerer and Weber (2007, p. 3) observe that "If obvious distinctions between large firms and small experiments, like incentives or experience and selection of agents, are expected to make a difference in behavior, then the way in which those distinctions matter should be part of the theory." In Section II.C of their paper they discuss a number of arguments in favor of doing simple laboratory experiments on organizations.

[38] These costs are equivalent to the investments in verification technology required under explicit contracts in the experiments of Fehr and Schmidt (2000) and Fehr et al. (2007).

the basis of performance pay, getting a share $f_I \in (0,1)$ of the firm's net profits (while the owner gets the remainder).

An alternative way to motivate the worker is to promise him a bonus whenever he puts in high effort. Because effort itself is non-contractable, this bonus payment cannot be made part of a formal contract though. The incentive system thus relies on an *implicit* contract that the promise will be kept. This type of organizational design is labelled as the motivation mode, or M-mode in short. Here the worker receives a wage w_M and is promised a bonus b_M on top of that if he exerts high effort. The manager gets a fraction $f_M \in (0,1)$ of firm profits. This performance pay scheme gives a selfish manager an incentive to renege on the promised bonus payment.

Overall the game model of Figure A1 results. First the owner chooses the organizational mode. If the M-mode is chosen, the worker moves next by deciding whether to shirk or to work. Only if the worker works, the manager decides whether to pay the promised bonus or not. (Here the implicit assumption is that the manager never wants to reward shirking with a bonus.) In the I-mode the manager moves before the worker does. The manager either commits to monitor or not to do so. In the former case the worker is assumed to work, because the disutility of working falls short of the costs of getting caught shirking. If the manager does not monitor, the worker chooses between shirking or working. The players' payoffs then follow from the assumptions made above.

[Insert Figure A1 about here]

If players are selfish, the predicted outcome is easily determined by backwards induction. A selfish manager will not pay the bonus in the M-mode (given $f_M \cdot b_M > 0$). Anticipating this, a selfish worker will shirk under this organizational design. In the inspection mode a selfish worker will shirk if not monitored by the manager, therefore the manager will monitor him.[39] The outcome is that the worker does work under this mode, at the expense of the overall costs of the monitoring technology $(k+h+d)$. Given assumption $(A1)$, under selfish preferences the I-mode is more efficient than the M-mode is. Hence assuming the owner obtains (part of) these efficiency gains, she chooses the I-mode over the M-mode.[40] It would be more efficient, however, if the worker could be motivated to work in the M-mode, as this would save the overall costs of the monitoring technology.

Of course, in a fully fledged model the compensation parameters w_M, w_I, b_M, f_M and f_I would be endogenous. Here we just make the following simplifying assumptions. First, mainly for practical reasons we focus on the case in which $f_M = f_I = \frac{1}{2}$.[41] The share fraction is thus the same for the two organizational modes, such that the owner does not simply prefer one mode over the other because she can pay the manager less. Second, with respect to the wage and bonus payments w_M, w_I and b_M we assume that:

$$w_M = \frac{f_M \cdot v_0}{1+f_M};\ w_I = \frac{f_I \cdot (v_1 - k - h) + g + d}{1+f_I};\ \text{and}\ b_M = \frac{f_M \cdot (v_1 - v_0) + g}{1+f_M} \tag{A2}$$

[39] Note that $f_I \cdot (v_1 - w_I - k - h) > f_I \cdot (v_0 - w_I - k)$ follows from inequality $(A1)$.

[40] In particular this requires $(1-f_M) \cdot (v_0 - w_M) < (1-f_I) \cdot (v_1 - w_I - k - h)$.

[41] A share fraction of one half is convenient in the experiment, because the owner's (i.e. player C's) earnings then simply correspond to those of the second mover (player B) in the M-mode and the first mover (player A) in the I-mode.

The wage level w_M ($= v_0/3$) ensures that all firm members earn the same when the worker chooses to shirk in the M-mode. Similarly so, w_I is set such that all firm members get the same when the manager monitors in the I-mode. Finally, b_M makes that all members earn the same when the manager pays the bonus in the M-mode after the worker decided to work. Effectively, payoff differences are minimized in the three most relevant outcomes and potential efficiency gains are shared equally.

Under assumption ($A2$) and $f_M = f_I = \frac{1}{2}$, the resulting payoffs in the two modes follow directly from the exogenous production technology parameters appearing in inequality ($A1$). The payoffs appearing in Figure 1 then result from making the following choices:

$$v_0 = 840;\ v_1 = 1740;\ g = 90;\ k = 180;\ h = 260\,[100] \text{ and } d = 130\,[50]$$

where for h and d the first value refers to I_H and the second to I_L.

References

Akerlof, G. and R. Kranton (2005). Identity and the economics of organization. *Journal of Economic Perspectives 19*, 9–32.

Avenhaus, R., B. von Stengel, and S. Zamir (2002). Inspection games. In R. Aumann and S. Hart (Eds.), *Hanbook of Game Theory*, Volume 3, Chapter 51, pp. 1947–1987. North-Holland.

Cabrales, A., R. Miniaci, M. Piovesan, and G. Ponti (2010). Social preferences and strategic uncertainty: An experiment on markets and contracts. *American Economic Review 100*, 2261–2278.

Calvo, G. and S. Wellisz (1978). Supervision, loss of control, and the optimum size of the firm. *Journal of Political Economy 86*, 943–952.

Camerer, C. and U. Malmendier (2007). Behavorial economics of organizations. In P. Daimond and H. Vartiainen (Eds.), *Behavioral Economics and its Applications*, Chapter 7, pp. 235–281. Princeton University Press.

Camerer, C. and R. Weber (2007). Experimental organizational economics. forthcoming in Handbook of Organizational Economics.

Cameron, A. and P. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics 117*, 817–869.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature 47*, 315–372.

Dickinson, D. and M.-C. Villeval (2008). Does monitoring decrease work effort? the complementarity between agency and crowding-out theories. *Games and Economic Behavior 63*, 56–76.

Dohmen, T. and A. Falk (2011). Performance pay and multi-dimensional sorting: Productivity, preferences and gender. *American Economic Review 101*, 556–590.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*, 268–298.

Eriksson, T. and M.-C. Villeval (2008). Performance-pay, sorting and social motivation. *Journal of Economic Behavior and Organization 68*, 412–421.

Falk, A. and M. Kosfeld (2006). The hidden costs of control. *American Economic Review 96*, 1611–1630.

Fehr, E., S. Gächter, and G. Kirchsteiger (1997). Reciprocity as a contract enforcement device: experimental evidence. *Econometrica 65*, 833–860.

Fehr, E., L. Goette, and C. Zehnder (2009). A behavioral account of the labor market: The role of fairness concerns. *Annual Review of Economics 1*, 355–384.

Fehr, E., A. Klein, and K. Schmidt (2007). Fairness and contract design. *Econometrica 75*, 121–154.

Fehr, E. and J. List (2004). The hidden costs and returns of incentives - trust and trustworthiness among ceos. *Journal of the European Economic Association 2*, 743–771.

Fehr, E. and K. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*, 817–868.

Fehr, E. and K. Schmidt (2000). Fairness, incentives and contractual choices. *European Economic Review 44*, 1057–1068.

Fehr, E. and K. Schmidt (2004). Fairness and incentives in a multi-task principal agent model. *Scandinavian Journal of Economics 106*, 453–474.

Fehr, E. and K. Schmidt (2007). Adding a stick to the carrot? The interaction of bonuses and fines. *American Economic Review 97*, 177–181.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*, 171–178.

Frey, B. (1993). Does monitoring increase work effort? The rivalry with trust and loyalty. *Economic Inquiry 31*, 663–670.

Gneezy, U. and J. List (2006). Putting behavioral economics to work: Field evidence on gift exchange. *Econometrica 74*, 1365–1384.

Hermalin, B. (2001). Economics and corporate culture. In C. Cooper, S. Cartwright, and P. Earley (Eds.), *The International Handbook of Organizational Culture and Climate*. John Wiley and Sons Ltd.

Johnson, E., C. Camerer, S. Sen, and T. Rymon (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory 104*, 16–47.

Kreps, D. (1990). Corporate culture and economic theory. In J. Alt and K. Shepsle (Eds.), *Perspectives on Positive Political Economy*. Cambridge University Press.

List, J. (2006). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. *Journal of Political Economy 114*, 1–37.

List, J. (2009). Social preferences: Some thoughts from the field. *Annual Review of Economics 1*, 563–579.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*, 1280–1302.

Rotemberg, J. (1994). Human relations in the workplace. *Journal of Political Economy 102*, 684–717.

Rotemberg, J. and G. Saloner (1993). Leadership style and incentives. *Management Science 39*, 1299–1318.

Rubinstein, A. (1999). Experience from a course in game theory: pre- and postclass problem sets as a didactic device. *Games and Economic Behavior 28*, 155–170.

Schweitzer, M. and T. Ho (2005). Trust and verify: Monitoring in interdependent relationships. In J. Morgan (Ed.), *Experimental and Behavioral Economics*, Volume 13. JAI Press.

Sloof, R. and J. Sonnemans (forthcoming). The interaction between explicit and relational incentives: An experiment. *Games and Economic Behavior*. DOI: 10.1016/j.geb.2011.03.006.

Teyssier, S. (2008). Experimental evidence on inequity aversion and self-selection between incentive contracts. GATE working paper W.P. 08-21.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources 33*, 127–172.