Computable and Experimental Economics Laboratory

Matteo Ploner

Personal Autonomy in Trust-Based Interactions. An Experimental Analysis

CEEL Working Paper 1-07

Computable and Experimental Economics Laboratory

Via Inama, 5 38100 Trento, Italy

http://www-ceel.economia.unitn.it
tel. +39.461.882246

Personal Autonomy in Trust-Based Interactions

An Experimental Analysis

Matteo Ploner*

Abstract

The paper experimentally investigates the interactions between restrictions to personal autonomy and reciprocity in a Principal-Agent relationship. Previous experimental contributions have shown that actions aimed at restricting decisional autonomy are likely to reduce reciprocity in trust-based relationships. Results in our experiment, which is a modified version of the Investment Game, differ from previous findings and conform more to standard economic predictions. Principals in our interaction do not support the self-determination of agents. On the other side, agents do not show any positive reciprocity when allowed to freely determine their behavior in the game.

JEL classification: C72, C91, D23, M50
Keywords: Principal-Agent relationship, Trust, Reciprocity, Self-Determination, Incentives

*Computable and Experimental Economics Laboratory (CEEL)
University of Trento, Italy
e-mail address: mploner@economia.unitn.it

Financial support was provided by CEEL which is acknowledged for this but also for the technical support provided before, during and after the experiment. Special thanks are due to Luigi Mittone, Ivan Soraperra and Marco Tecilla but also to the staff of the laboratory which provided a valuable contribution to this paper. The value of the support of Dominique Cappelletti is incommensurable and not limited to the present paper. Comments from participants in seminars at the Max Planck Institute for Economics, Jena (Germany) and at the University of Trento (Italy) are greatly appreciated. An earlier version of the paper was presented at the "Trust, Reciprocity and Social Capital: The 2006 Ratio Colloquium for Young Social Scientists". The usual caveats apply, however.

1 Introduction

A large set of economic interactions is characterized by conflicting interests of the different parties involved in the exchange. The standard economic approach conceives extrinsic rewards as the only mechanism to align divergent interests. In particular, the Principal-Agent paradigm provides a manageable framework to deal with situations of conflict involving risk sharing and incentives (among earlier contributions see, Shavell, 1979; Holmstrom, 1979). In a principal-agent relationship one party, the principal, derives utility from the outcomes following actions undertaken by another party, the agent. Furthermore, as the action to be undertaken by the agent is costly to the agent, the principal will endow the agent with an extrinsic, usually monetary, reward in order to "convince" the agent to deliver the expected effort. Among the measures providing incentives in situations of this kind, the most relevant seem to be *monitoring, evaluation and contracting* (Prendergast, 1999). The standard economic approach to agency issues has traditionally been characterized by the assumption of strict self-interest and full rationality (Grossman and Hart, 1983). However, recent contributions in the domain of Behavioral Economics have enlarged the scope of the utility function to encompass also "psychological" sources of motivation springing from social interactions. Among the most successful investigations into such social preferences it is possible to identify models accounting for *equity*(among others, Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and *reciprocity*(among others, Rabin, 1993). For the purposes of the present contribution it is of particular interest to consider the issue of reciprocity in some detail. Preferences for reciprocity are mainly affected by the process leading to an outcome and not by the outcome itself. In particular, when intentions backing an action are perceived as benevolent, attitudes towards the individual undertaking that action will be benevolent too. With explicit reference to the domain of the present analysis, an action is evaluated as benevolent if it is based on genuine trust. As summarized by (James, 2002, p. 291), "A trusts B means that A expects that B will not exploit a vulnerability that A has created for himself by taking the action".

Trust is an important element for economic relationships as it favors efficiency (on this aspect see, among others, Arrow, 1974) and economic growth (La Porta *et al.*, 1997; Putnam, 1993). However, as argued by Slovic (1999), there are psychological aspects which render trust a fragile social artifact. Moreover, once distrust comes into play it is very difficult to reestablish trust as subjects who proved to be unreliable in the past are not given the opportunity to show their loyalty in further interactions. A major source of fragility of trust is the fact that its replication relies on reciprocal attitudes of the counterparts. Indeed, it has been shown that trust can persist only in environments where subjects systematically refrain from exploiting trustful behavior (Güth and Kliemt, 1994). The work of Fehr and Gächter (1998) provides a survey of various empirical and theoretical works supporting the idea of reciprocity as a common trait of human beings. The authors make an important distinction between *positive* and *negative* reciprocity. The former refers to the desire to reward positive actions, while the latter refers to retaliation against unfaithful individuals. From their review it emerges that the *Investment Game* (Berg *et al.*, 1995) is an effective design to experimentally investigate trust-based interactions. This experimental design also provides a fundamental reference for the experiment

presented here.

The concept of reciprocity has been widely investigated with reference to labor markets within the theory of efficiency wages. Since the seminal work of Akerlof (1982), the theory has been object of intensive study. At an empirical level, the theory has been investigated using the so-called Gift Exchange Experiment Fehr *et al.* (1993). In a standard gift-exchange experiment, firms (principals) make offers about wages and workers (agents) can decide whether to accept an offer or not. Once the offer is accepted a firm and a worker are matched and the worker can freely set the level of effort to be provided. Wages are a cost for the firm and a source of value for the worker and the opposite holds for effort levels. The standard economic prediction is that workers set their effort at the minimum level independently of the wage offer. Firms anticipate this and accordingly decide to set wages at the minimum possible level. Contrary to this prediction, experimental work provides evidence of a positive correlation between offers made by the firm and performance of the agent. Firms offer wages higher than the market clearing level to obtain higher levels of effort and agents appear to repay the firm's "gift".

The focus of the present study is on the interaction between extrinsic motivational sources related with practices restricting personal autonomy and other determinants of behavior like trust and reciprocity*. Aim of control practices in a Principal-Agent relationship is to reduce opportunistic behavior and thus reduce risk borne by the principal. Different monitoring practices can serve this purpose and obtain, under standard economic assumptions, an alignment of the conflicting interests of the parties involved in an agency relationship. However, works in the psychological literature have shown that when extending sources of motivation to encompass *intrinsic motivations* anomalous behavior may emerge. Intrinsic motivations refer to the desire to take an action for its own sake and not for the outcomes associated with the action (Ryan and Deci, 2000). Recent experimental works highlighted the importance of intrinsic motivations to behavior having a relevant economic content. Particular attention has been paid to the potential crowding-out of intrinsic motivations following the provision of extrinsic (monetary) rewards (Frey and Oberholzer-Gee, 1997; Gneezy and Rustichini, 2000). A direct consequence of substitution between different sources of motivations is the violation of the monotonic relation between incentives and effort. Concerning the motivational crowding-out issue, works in the economic literature have mainly focused on extrinsic incentives in the form of monetary rewards but psychological literature has enlarged the inquiry also to other extrinsic motivational means. Of particular interest here are studies dealing with control practices and their impact on decisional autonomy or self-determination. Autonomy implies free choice of the outcome and of the means to achieve it. Under control, intentions remain free while actions are bounded Deci and Ryan (1987).

About the impact of self-determination in a work organization, Deci *et al.* (1989) notice that support to autonomy by the management positively impacts on perception of the organization by the workforce. More focused on the interaction between monitoring and effort is the contribution of Frey (1993). From a cross-disciplines literature review and from empirical evidence the author concludes that the presence of a personalized relationship between the principal and the agent is likely to foster the crowding out of intrinsic motivations in favor of extrinsic motivations. The work of Brandts and Charness (2004) investigates

about the impact of contextual elements on gift-exchange experiments. In one of the experimental variations presented, a minimum wage is exogenously set. This restricts not only the opportunity set of the firm but also of the worker who, under the minimum wage condition, is "forced to be generous". When the minimum wage is exogenously set, effort provided by the workers is generally lower than when no restriction is set and, furthermore, also the likelihood of observing high wages sharply decreases. The work of Falk and Kosfeld (200x) also deals with incomplete contracts but introduces a different kind of restriction on agents' decisional autonomy. In their experiment, the principal has the option of bounding actions of the agent through the provision of a mandatory minimum level of performance. Under standard economic assumptions the restriction is always preferred by the principal who ensures herself against opportunistic behavior. However, the authors find that due to "control-averse" behavior the agents tend to offer a lower performance when their decisional autonomy is bounded than when they are let free to choose positive effort without any lower bound restriction. From this it follows that *ex post* principals preserving agent's decisional autonomy, the large majority in the experiment, are made better off than those relying on the control device. The authors extend results from the experiment to control practices in general and observe that signals of distrust associated with control may be counterproductive. The authors implicitly assume that the minimum wage restriction conveys only a signal of distrust and no other relevant information. However, one can ask whether the voluntary provision of a minimum wage is subject to alternative interpretations. In particular, it seems plausible to argue that agents may interpret the minimum wage as an acceptable level of effort from the principal point of view and thus reduce their feelings of guilty.

The present paper addresses the same research question of Falk and Kosfeld (200x) but employs a radically different experimental design*. The basic interaction structure employed is provided by the Investment Game as presented by Berg *et al.* (1995). In order to address the causal relationship between control practices introduced by the principal and reciprocity attitudes shown by the agent the following modification is introduced: before any transfer between the to parties takes place, the principal can decide whether to adopt a strategy restricting the decisional autonomy of the agent at a given cost. When control is not implemented, the two parties move to the following stage of the game (i.e., standard Investment Game) with the principal and the agent playing the role of the trustor and the trustee, respectively. When control is enforced the agent is asked to state the conditional repayment for each amount potentially sent by the principal in the Investment Game. Then, the principal chooses the amount to send to the agent who is given no further chance to renegotiate the repayment commitment. The amount sent by the principal is multiplied by a positive factor and payoffs are calculated on the basis of the stated intentions of the agent.

The results of the experiment show that very few principals decide to warrant autonomy to the matched agent. In general, the agents do not reward the autonomy supportive option undertaken by the principal. In fact, they behave in a very opportunistic way when playing the investment game by sending back very low amounts. Given the lack of reciprocity of the agents, decisions of the principals turn out to be correct. In fact, they are relatively better off when implementing the control strategy than when warranting decisional autonomy

to the agents.

The remaining of the paper is organized as follows. Section 2 illustrates the experimental design; section 3 discusses different preference structures of the agents and outlines the associated behavioral predictions in the game; section 4 analyzes the data collected in the experiment; section 5 includes the discussion of the results and directions for future research.

2 Experimental Design

2.1 The Experiment

The basic strategic structure of the game employed in the experiment resembles real-life situations in which complete contracts are not at disposal and trust is required to reach the socially desirable outcome. As an example consider a situation in which an employer (i.e., the principal) has to decide how much to invest in the training of an employee (i.e., the agent). The decision is risky for the employer as long as there are no opportunities to restrict the employee's action after the completion of the training. Thus, the employee has the incentive to exploit the opportunity to be trained by the current employer and then be hired by another firm at a salary which accounts for the increase in productivity following the training but not for its cost. The current employer opens a vulnerability for herself by "sponsoring" the training of the employee. An interaction of this kind is well captured by the Investment Game (Berg *et al.*, 1995). In the Investment Game two parties, the trustor and the trustee, are endowed with an initial amount of money. The trustee is given the opportunity to invest the initial endowment in a revenue-generating asset. Profits are not appropriated by the trustor but are transferred to the trustee. At this stage the trustee can freely decide whether to return some of the revenues to the trustor. The standard economic prediction (i.e., the sub-game perfect equilibrium) for the game is that the trustor anticipates that the trustee will return nothing. This gives no incentive to the Trustor to invest any portion of the initial endowment. The current design builds on the experimental framework of Berg *et al.* by providing the principal with a device to circumvent vulnerability. Before entering this investment the principal is offered the opportunity of knowing in advance, at a cost c, the future actions of the agent. The availability of the detection option and the actual choice of the principal are known to the agent[1].

[Figure 1 to be inserted here]

Figure 1 provides a graphical description of the game which will be henceforth identified as *The Detection Game*. The experimental game is a sequential game with complete and perfect information. The first decisional node belongs to the principal (P) who has to decide whether to detect ($D = 1$) or not ($D = 0$), at a cost c, the intentions of the agent in the subsequent interaction. When a principal decides to warrant the decisional autonomy to the matched agent, and thus not to know in advance her actions in the subsequent interaction, the two

[1]The control mechanism employed in the experiment captures informational features of monitoring mechanisms which allow to infer future performances of a subject from current characteristics of the same. An example of control practices of this kind is provided by medical tests on the habits and health status of the employers

players move towards a standard Investment Game. After having chosen to set $D = 0$, the principal selects the amount x of her endowment E she intends to invest. The amount sent is multiplied by an exogenously given factor m and added to initial endowment E of the agent. Then, the agent decides how much she wants to return to the principal from the amount received mx. After having collected the decisions of the agent the payoffs of the players are computed.

When a principal decides to restrict the autonomy of the agent, and thus to know in advance her future actions, an amount c is withdrawn from the initial endowment E of the principal. Under the bounding strategy the agent is asked to fulfill a vector, henceforth called repayment vector (see Table 1), where binding associations between each possible amount sent by the principal and the amount returned by the agent $(y|x)$ are made. Amounts reported in the vector cannot be renegotiated at any further stage of the experiment. When filling the repayment vector each agent commits to actually undertake the action reported in the vector. The fulfilled vector is sent to the principal who then chooses the amount she is going to send to the agent. Given this interaction structure the monetary payoffs of the trustor and of the trustee are equal to $\Pi_1 = E - x + y$ and $\Pi_2 = E - y + mx$, respectively. The game is played only once in order to rule out possible confounding factors associated with repetition of the game.

2.2 Experimental Procedures

The computer-based experiment was run at the Computable and Experimental Economics Laboratory (CEEL) of the University of Trento. A client-server infrastructure purposely built using the programming language Borland© Delphi© was employed. Participants were undergraduate students of the University of Trento and most of them had a training in Economics. Three identical sessions were run with 16 participants per session (N = 3x16 = 48 Subjects).

Before the interaction, instructions[2] were read loudly by the experimenter and participants were free to ask for clarifications. Each participant was endowed, independently from the role in the experiment, with 10 units of Experimental Currency Unit (ECU). At the end of the experiment, each ECU was exchanged with €0.33 and participants were paid accordingly. A show-up fee of €2 was assigned to each participant and this made the maximum gain from the participation to the experiment equal to €12. On average the experiment lasted 40 minutes. Anonymity among the participants was warranted during and after the game; matching between the subjects was randomly determined and no communication among the participants was allowed. A control questionnaire was submitted to the participants before the beginning of the interaction so as to prevent mistakes due to a lack of understanding of the experimental instructions. The experiment started only after all participants correctly answered all the question.

[Table 1 to be inserted here]

A strategy method was employed to collect responses of the agents. In more details, agents had to fill in two distinct forms in two sequential screen-shots before knowing the actual choice of the principal in correspondence to the first

[2]see Table 4 for an English version instructions. Instructions for the experiment were written in Italian

decisional node. A representation of the repayment vectors is reported in Table 1. While agents were asked to fill in both vectors in Table 1, the actual vector presented to the principal was made dependent on her choice in correspondence to the first decisional node in the Detection Game. When opting for detection, a vector similar to the one reported in the upper panel of Table 1 and containing "promised" repayments from the matched agent was sent to the principal. When autonomy of the agent was preserved nothing was submitted to the principal before collecting her choices in the sub-game represented by the Investment Game.

3 Behavioral Predictions

As presented in section 3, under standard economic assumptions of rationality and self-seekingness two distinct sub-game perfect equilibrium outcomes can be identified in the Detection Game. One of them is that resulting from the composed decision of the principal to avoid the implementation of detection (i.e., $D = 0$) and submit nothing in the following node corresponding to the choice in the Investment Game ($x = 0$). Given these principal's action profile, the opportunity set of the agent results in an empty set, which necessarily implies a null amount sent back by the agent ($y = 0$).

In correspondence to the alternative equilibrium outcome, the principal decides to pay the amount c and to undertake the detection option ($D = 1$). Under this bounding condition, the best action for the agent is to fully exploit the gains associated with the multiplier factor m by inducing the principal to send all the available endowment ($E - c$). At this end, the agent will commit herself to a repayment equal to $E - c + \epsilon$ when facing the maximum amount sent by the principal. Given that $y \in \{0, \ldots, mx\}$, ϵ will be equal to 1. The additional condition $(y|x \neq 9) - (E - c) < 1$ allows the agent to obtain the best outcome. When the two conditions above are respected, the principal will have no incentives to choose an amount $x \neq 9$. In sum, the equilibrium outcome will be reached through the following actions: the principal decides to implement detection ($D = 1$) and to submit the whole available endowment ($x = E - c$), while the agent provides the minimal reward ($y = E - c + 1$) to the amount sent by the principal. Given the parameters employed in the experiment (i.e., $m = 2$ and $E = 10$) the couple of payoffs π of the principal and the agent obtained under the former equilibrium are $\pi_P = 10$ and $\pi_A = 10$, respectively. Those obtained in the latter are equal to $\pi_A = 10$ and $\pi_B = 18$. What can be observed with reference to the two equilibrium's outcomes is that the latter delivers a situation which is Pareto superior and, thus, socially preferable. However, higher risk is associated to the strategy leading to this equilibrium when accounting for potential "anomalous" behavior of the agents.

Previous contributions have already shown that individuals are generally characterized by reciprocity concerns. What emerges from the relevant literature is that reciprocity is a function of both the actions undertaken by the counterpart and the contextual factors affecting intrinsic motivations. These behavioral aspects can be investigated within the simple interaction structure presented in section 2.1 *via* the introduction of a modified value function of the agent. The main innovation in the behavioral value function is the fact that the agents obtain some value from returning to the principal a proportion of

the amount sent by the same. The source of value is the reciprocation of a choice taken by another individual in a condition of trust. In addition, it is assumed that reciprocity concerns are present only when decisional autonomy is warranted to the agent. Aim of this assumption is to capture the positive interaction between self-determination and reciprocity.

Under the simple behavioral characterization examined here the value function of the agent can be represented as $V_A = E - y + mx - (1 - D)(y - rx)^2$. The value function differs from the payoff function π_A because of the component $-(1 - D)(y - rx)^2$. This source of value captures the psychological cost associated with repayments which are higher/lower than the perceived kindness of the action undertaken by the principal[3]. Perceived kindness is measured by the factor rx where x is the monetary amount sent by the principal and r is a multiplier factor defining the reciprocity value attached by the agent to the amount received. When $r > 1$ the agent is a positive reciprocator, while an $r < 1$ registers the absence of positive reciprocity concerns. In the limit case of $r = 0$, the equilibrium behavior of an agent characterized by a value function V_A is the same as the behavior of an agent concerned with the mere monetary payoffs.

To obtain a testable behavioral prediction the assumption that agents are characterized by positive reciprocity (i.e., $r > 1$) is employed. Given a continuously differentiable agent's value function V_A and $x_{ij} \in [0, E]$ and $y_j \in [0, mx_{ij}]$, where x_{ij} is the amount sent by principal i to agent j and y_j is the amount returned by agent j, it can be easily shown that the best reply function for any agent j is equal to $y_j^* = rx_{ij} - 1/2$. From this it follows that the best action for the principal is to send the whole available endowment when $r > 1$ and to send nothing when $r < 1$. Thus, given the agent's value function V_A and the assumption of positive reciprocity characterizing the agents, the unique equilibrium outcome in the behavioral characterization of the game is defined by the following actions: the principal does not implement the detection device and sends the whole available amount to the agent (i.e., $x^* = E$). Then, the agent returns to the principal an amount $y^* = rx^*$ which is bigger than E.

While the previous equilibrium behavior is derived with explicit reference to a game played in the continuum of choices available to the agent and the principal, the experiment is played with integer numbers. However, the equilibrium outcomes identified provide a useful guidance also for the interaction in the game. Indeed, in the experimental game a principal will choose to preserve agent's autonomy and submit all her endowment whenever she believes that a positive correlation between self-determination and reciprocity exists and the agent has a predisposition to positive reciprocity.

The behavioral predictions defined above will be checked against data collected in the experiment. Given that the alternative assumptions of self-regarding rationality and reciprocity predict distinct equilibrium outcomes, it will be possible to determine which of the two assumptions is closer to the actual nature of subjects in the experiment.

[3]An implicit assumption of the modeling strategy undertaken here is that neither the principal nor the agent is concerned of the relative final allocation of payoffs. This assumption is supported by the work of McCabe *et al.* (2003) which shows that behavior in trust games is mainly driven by reciprocity concerns and not by considerations about payoffs allocation

4 Data Analysis

4.1 Descriptive Statistics

Some descriptive statistics about choices of the principals and payoffs of the two parties in the game are reported in Table 2[4]. The choices of the agent are better detailed in Figure 2.

[Table 2 to be inserted here]

4.1.1 Principals

As reported in Table 2, only 6 (i.e., 25% of the total sample) principals choose to preserve self-determination of the matched agent. Thus, the vast majority of principals prefer to gain control over agent's actions. The amount sent by the principals under the autonomy supportive condition is on average quite low (1.50 ECU) and smaller than the average amount sent under the detection condition (4.31 ECU). Principals are, on average, relatively better off when choosing detection than when moving directly to the sub-game represented by the Investment Game (profits are 10.46 vs 8.83 ECU, respectively).
Table 2 highlights also that, independently of the presence of autonomy in decision making, a negative median return on investment is registered. Concerning responses of the principals to the repayment commitment of the agent, it can be observed that 12 principals out of 13 (92.30%) choose to submit the amount x maximizing their payoff given the repayment vector submitted by the agent. This signals a strong commitment to profit maximization and a good understanding of the mechanisms of the game.

4.1.2 Agents

Figure 2 provides a representation of the choices of the agents under the two detection conditions. The X-axis reports the hypothetical amount x submitted by the principal, while the Y-axis reports the amount y to which the agent commits herself. Individual choices appear on the graph as jittered empty circles[5] and boxplots provide the usual synthetic description of the distribution of choices. In correspondence to the average amount y conditional to x a filled circle is depicted. The 45° dashed line provides a graphical reference to identify the reciprocity content of each observation. In particular, values located above the line are characterized by positive reciprocity (i.e., $y > x$).

[Figure 2 to be inserted here]

From the upper panel in Figure 2, it can be observed that when no detection is present the values of y are not very sensitive to the different values of x.

[4] A malfunctioning of the computer hardware of one of the terminals caused a loss in the data. In particular, given the nature of the software employed for the experiment, 5 observations referring to the amounts sent by the principal and to profits of both players are missing. Concerning choices of the agent, only the observation corresponding to the computer causing the malfunctioning is missing from the dataset, that 23 observations for each x under the two detection condition are collected. It is important to remark here that the data analysis is not heavily affected by the loss in the data as it focuses mainly on choices of the agents.

[5] A small white noise has been added to each observation to allow a 4better appraisal of individual choices

Median repayment switches from 0, when the hypothetical amount sent is lower than 3, to 1 in correspondence to values of x higher than 3. This signals a very low level of reciprocity concerns among the agents when playing the Investment Game. In correspondence to the total endowment submitted by the agent (i.e., $x = 10$), only 1 observation out of 23 (4.34%) is located in the positive reciprocity area. A similar pattern is registered also in correspondence to the other values of x. At a more detailed level it can be observed that, independently of the value of x, most of the observations are clustered in correspondence to the rational selfish equilibrium value $y = 0$. Taken together these observations lead to the conclusion that positive reciprocity is nearly absent under the autonomy supportive condition and, moreover, agents show a behavior which is very close to the rational self-seeking prediction as outlined in section 3.

In the lower panel of Figure 2 the choices of the agent under the detection condition are reported. From the figure it emerges that the choices of the agents are quite sensitive to offers of the principals. Median and average return in correspondence to each potential offer of the principal are quite close to the positive reciprocity threshold represented by the dashed 45° line. Pairwise non-parametric tests (Wilcoxon signed rank test) reveal that differences between choices under the two alternative detection conditions are statistically significant, at least, at the conventional 5% level for $x > 1$.
Recalling behavioral predictions illustrated in section 3, it is interesting to focus on agent's behavior in correspondence to the maximum amount sent by the principal (i.e., $x = 9$). The average amount sent back is very close to the median value of the distribution (8.00) and equal to 7.82 ECU. On average, values registered are lower than expected under rationality assumptions (i.e., 10 ECU). Parametric and non-parametric tests strongly reject the null hypothesis that values registered are on average equal to 10 (t-test, $p.value = 0.024$; Wilcoxon signed rank test, $p.value = 0.013$).

4.2 Regression Analysis

The regression analysis reported in Table 3 provides a description of agent's behavior in the experiment. The dependent variable in the model is represented by the decisions of the agent (y_i). The explanatory variables employed in the regression are the amount sent by the principal for each amount y under the two detection condition (x_{ij}^D) and a dummy variable capturing the presence of detection ($D = 1$ if detection is present and $D = 0$ otherwise). Given that several data points are collected in correspondence to the same agent (as choices are collected through repayment vectors), a control on clustering of errors at the individual level is introduced. Given that the support of agent's decisions is defined over integer numbers, a generalized linear model (Poisson family) is estimated.

[Table 3 to be inserted here]

From the estimation reported in Table 3 it emerges that the amount sent by the principal has a positive and significant impact on the amount returned by the agent. This testifies the presence of some form of reciprocity in the population of agents but, as evidenced above, the amount returned to the principal is on average lower than the amount sent by the principal. The impact of detection

on the amount sent by the agent is positive and significant. The same holds also for the interaction term between the two explanatory variables. Agents under detection systematically return more than agents provided with full autonomy at the decisional stage. The coefficient of the interaction term evidences how the reciprocity attached to the amount sent by the principal is stronger under detection than in the complementary condition. Overall, very low levels of repayment are registered but, contrary to what expected under the behavioral prediction, higher levels emerged in the autonomy restrictive condition.

5 Discussion and Conclusion

The kind of interaction experimentally investigated here provides a manageable framework to study the impact of self-determination on choices involving trust and reciprocity. A strong pattern emerging from the data collected in the laboratory is that a small minority of principals decides to warrant decisional autonomy to the matched agent. Furthermore, undertaking the autonomy supportive opportunity is not associated with full trust in the counterpart as very low amounts of endowment are submitted to the agent in the standard investment game following the detection decisional node. However, this cautious behavior turns out to be the appropriate one given the overall negative returns registered in the investment game as a consequence of the opportunistic behavior of the agents. In the case of restriction of agent's autonomy, despite a positive correlation between amount sent by the principal and the response of the matched agent, the overall amount sent back by the agents is not enough to deliver positive returns to the investment undertaken by the principals. With reference to the behavioral predictions outlined in section 3, it can be noticed that the amount to which the agent commits herself when the principal sends the whole amount available is lower than what expected under standard economic assumptions. The gap between observed behavior and the rationality benchmark may be due either to a lack of understanding of the strategic structure of the game or to negative reciprocity. An agent endowed with negative reciprocity may be willing to incur a cost in order to punish behavior which is perceived as unfair (Fehr and Fischbacher, 2004). With explicit reference to the experimental context, while it is not possible to dismiss *a priori* the negative reciprocity hypothesis, it can be argued that the ratio between the cost borne by the agent and the cost faced by the principal is very high. This evidence casts some doubts on the negative reciprocity interpretation. Behavior of the agents forces a consideration also about opportunity costs associated with choices of the principals. When accounting for anomalous behavior of the agents the strategy undertaken by the vast majority of the principals is riskier than the alternative rational behavior identified in section 3 (i.e., not to implement detection and invest nothing). The distribution of payoffs in the experiment reveals that the risk borne by the majority of principals is not rewarded. Indeed, the median outcome is lower than what they would have obtained choosing the alternative rational strategy.

The outcomes of this experiment radically differ from previous contributions finding positive reciprocity in correspondence to the choice of preserving decisional autonomy of the counterpart. Further research is needed to address the origins of these differences but some directions for future inquiry can be

identified here. Restrictions to decisional autonomy are not introduced by setting a lower bound to wages or effort levels like in previous study (respectively, Brandts and Charness, 2004; Falk and Kosfeld, 200x) but by a manipulation of the sequence of moves in the game. Compared to the mere introduction of a minimum threshold, the treatment implemented in the experiment induces a more articulated strategic reasoning. It may be that higher levels of cognitive load crowd-out other-regarding concerns (Güth *et al.*, 2005) and this leads to a substitution of reciprocity with self-oriented strategic concerns. This could provide a partial explanation to the observed differences between our and previous findings. However, it must also be observed that setting a lower bound is likely to induce anchoring to that value. At this aim, the control treatment reported by Falk and Kosfeld does not dismiss the case that the lower bound set by the principal is interpreted by the agent as an acceptale level of effort. If this was the case the result of lower reciprocity under conditions of control would not be caused by limitations of self-determination but by a reduction in the feelings of guilty of the agents.

To summarize, behavior observed in the laboratory is very far from the autonomy-triggered reciprocity prediction and much closer to behavior observed under standard economic assumptions. The departure from the behavioral benchmark accounting for social preferences is mainly due to the fact that the principals correctly forecast the opportunistic behavior of the agents in the trust-based interaction. Nevertheless, it is important to notice that agents do not fully exploit the efficiency gains in the game. This sub-optimal behavior may be due either to cognitive limits in the representation of the strategic structure of the interaction or, less likely, to a strong preference for negative reciprocity following the introduction of detection. Further research is needed to disentangle these two sources of behavior.

6 Appendix

Figure 1: The Detection Game

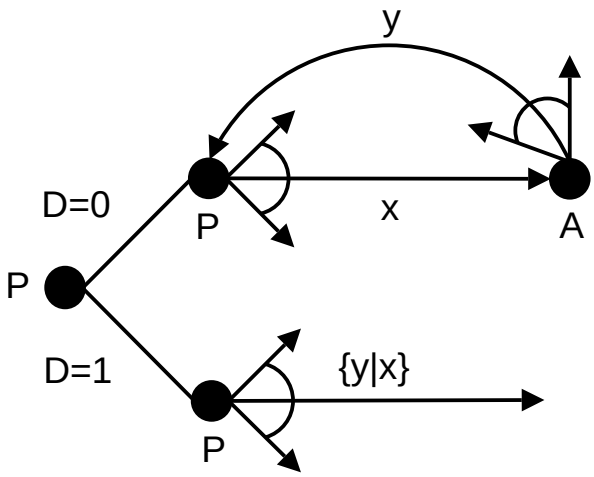

Figure 2: Distribution of choices under alternative detection condition

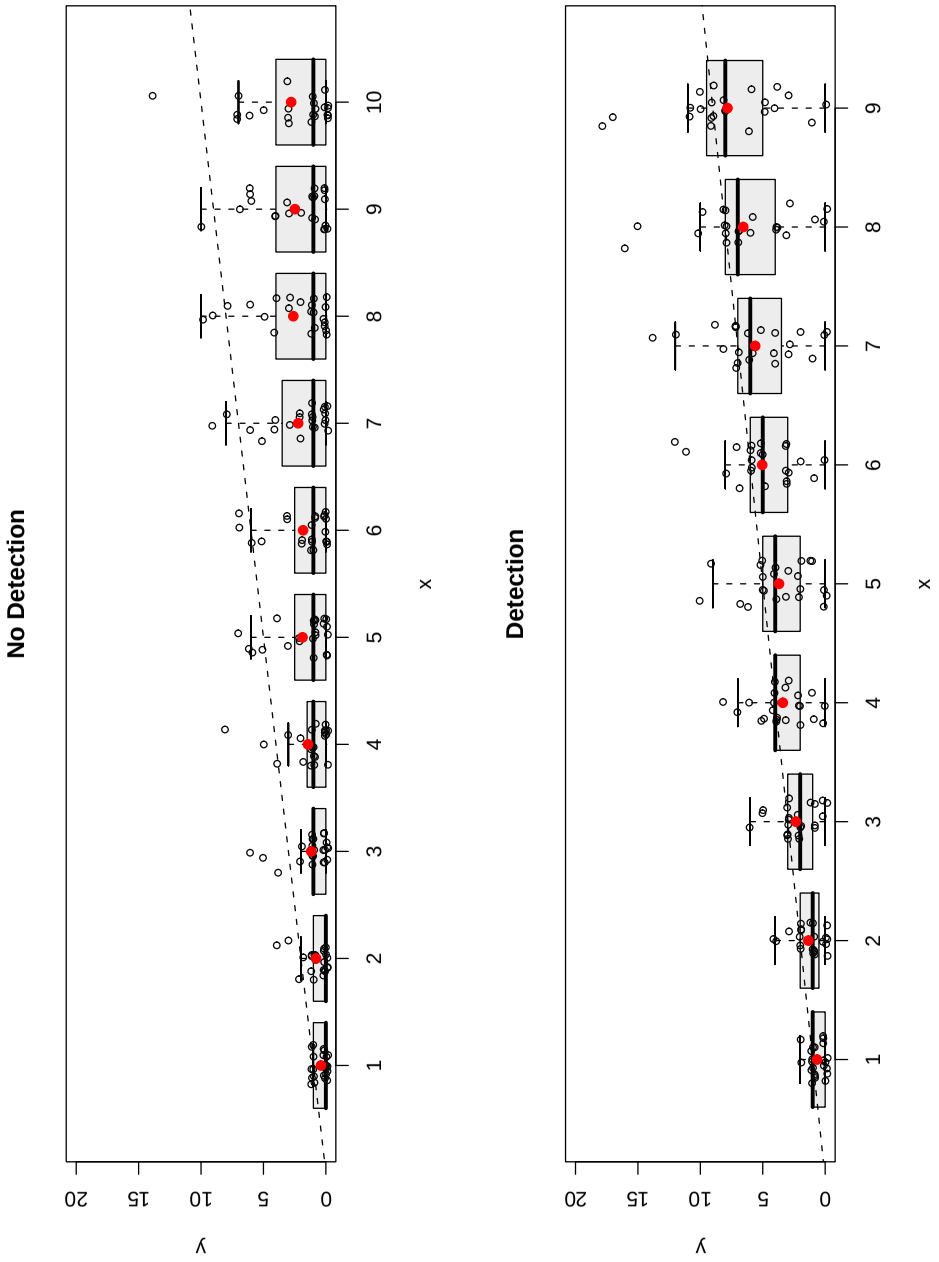

Table 1: Repayment Vectors

	x	1	2	3	4	5	6	7	8	9	10
No Detection ($D = 0$)	y										

	x	1	2	3	4	5	6	7	8	9
Detection ($D = 1$)	y									

Table 2: Descriptive Statistics

Subject	Variable	Detection	N^*	Mean	Median	Std. dev.
Principal	*Choice*	Yes	13.00	4.31	3.00	3.84
		No	6.00	1.50	1.00	1.87
	Profits	Yes	=	10.46	9.00	3.18
		No	=	8.83	9.50	1.60
Agent	*Profits*	Yes	=	12.85	11.00	3.46
		No	=	12.67	11.50	3.44

* Total observations expected are N=24. Missing observations are due to an hardware failure in session 2 of the experiment. See footnote in section 4 for more details)

Table 3: Generalized Linear Mixed-Effects Regression on Agent's Choices (Poisson Family)

$$y \sim offer + detection + offer * detection$$

	Coefficient	Std. Error	z	$Pr(\,\vert z\vert)$
x	0.181	0.023	7.896	2.89e-15
detection	0.583	0.184	3.176	1.49e-03
$x * detection$	0.055	0.027	2.013	4.41e-02

Table 4: Intructions (Translation from Italian)

Dear Participant,
this is neither an IQ test nor a test aimed at measuring your skills but an interaction in which you will be asked to choose according to your preferences. You will receive a reward at the end of the experiment which is proportional to the outcome of the interaction below described. Moreover, independently from the interaction, you will receive a payment of €2. Before starting the experiment you will be asked to answer some questions aimed at verifying your understanding of the experiment. Your answer to the questionnaire will not affect the payoff you are going to earn but the experiment will not start until all the participants have answered all the questions correctly. Amounts presented in the experiment are expressed in ECU (Experimental Currency Units). At the end of the experiment amounts expressed in ECU are converted in € at an exchange rate of €1 each 3 ECU. Two different roles are present in the experiment and they are referred to as Subject A and Subject B. You will be randomly assigned one of the two roles at the beginning of the experiment and you will maintain that role throughout all the experiment. Each Subject A is randomly and anonymously matched with a Subject B. Only one interaction will take place during the experiment. The experiment is based on the following interaction scheme:

1. Subject A and Subject B in a couple both receive 10 ECU
2. Subject A can decide how many of her 10 ECU send to Subject B
3. The ECU sent by A are multiplied by two and assigned to B
4. B decides how many of the ECU previously sent by A and multiplied by the factor of 2 send to A (please notice that the amount sent to A can be at most equal to the amount received from A and multiplied by a factor of 2)
5. The interaction ends with computation of the ECU belonging to A and B, the conversion in € and actual payment of the amount due to Subject A and Subject B. The number of ECU of A at the end of the interaction are equal to the 10 ECU assigned as initial endowment decreased by the ECU sent to B and increased by the ECU received from B. At the end of the experiment B will detain a number of ECU equal to the 10 ECU assigned as initial endowment increased by the ECU received from A multiplied by two and decreased by the ECU sent to A.

With respect to the interaction structure previously described a variation is introduced in the experiment: At the beginning of the experiment, before phase 2) starts, Subject A can decide whether to pay 1 ECU and monitor the repayment intentions of B in what above is defined phase 4). Subject B has to fill a vector in which is specified for each possible amount of ECU sent by A the repayment intentions of B. In more details, in two distinct screenshots Subject B must state how much she plans to return to Subject B if A decides of NOT buying the right to monitor intention and if A decides to buy the right. When A decides to buy the information the correspondent vector filled by B will be shown to A who will thus decide how much send to B knowing in advance the amount returned by B. When A decides of NOT buying the information no vector is shown to A who has to choose how much send to B without having any information on intentions of B. It is important to remark that under this condition the information provided by B remain hidden and the decision of A is not affected by the information provided by B in any way.
The experiment ends with the computation of the ECU belonging to Subject A and to Subject B after the interaction previously described and their conversion in €. Are there any questions?

References

Akerlof, G. A. (1982). Labor Contracts as Partial Gift Exchange. *Quarterly Journal of Economics*, pages 543–69.

Arrow, K. J. (1974). *The Limits of Organizations*. Norton Company.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity and social history. *Games and Economic Behavior*, **10**, 122–142.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity and competition. *The American Economic Review*, **90**(1).

Brandts, J. and Charness, G. (2004). Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence. *The Economic Journal*, (114), 684–708.

Deci, E. L. and Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, **53**(6), 1024–1037.

Deci, E. L., Connell, J. P., and Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, **74**(4), 580–590.

Falk, A. and Kosfeld, M. (200x). Distrust - The Hidden Cost of Control. *American Economic Review*.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, (25), 63–87.

Fehr, E. and Gächter, S. (1998). Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review*, **15**, 845–859.

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

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics*, **108**(2), 437–459.

Frey, B. S. (1993). Does monitoring increase work effort? the rivalry with trust and loyalty. *Economic Inquiry*, pages 663–.

Frey, B. S. and Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *he American Economic Review*, **87**(4), 746–755.

Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, **115**, 791–810.

Grossman, S. J. and Hart, O. D. (1983). An Analysis of the Principal-Agent Problem. *Econometrica*, **51**(1), 7–46.

Güth, W. and Kliemt, H. (1994). Competition or co-operation: On the evolutionary economics of trust, exploitation and moral attitudes. *Metroeconomica*, **45**(2), 155–187.

Güth, W., Levati, M. V., and Ploner, M. (2005). On the social dimension of time and risk preferences: An experimental study. Discussion Paper 26-2005, Max Planck Institute of Economics, Jena, available at `http://econpapers.repec.org/paper/esidiscus/2005-26.htm`.

Holmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics*, **10**(1), 74–91.

James, H. S. (2002). The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior and Organizations*, **47**, 291–307.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. W. (1997). Trust in large organizations. *The American Economic Review*, **87**(2), 333–338.

McCabe, K. A., Rigdon, M. L., and Smith, V. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior and Organization*, **52**, 267–275.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, pages 7–63.

Putnam, R. (1993). *Making Democracy Work: Civic Tradition in Modern Italy*. Princeton University Press, Princeton, NJ.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, **83**(5), 1281–1302.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, **25**, 54–67.

Shavell, S. (1979). Risk Sharing and Incentives in the principal and Agent Relationship. *The Bell Journal of Economics*, **10**(1), 55–73.

Slovic, P. (1999). Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield. *Risk Analysis*, **19**(4).