An Experimental Investigation of Fairness and Reciprocal Behavior in a Simple Principal-Multiagent Relationship*

Alessandro Rossi†, Massimo Warglien‡

First version: September 10, 1999
This version: March 24, 2000

Abstract

Issues of fairness within the agency framework have traditionally been investigated – both theoretically and experimentally – within two alternative approaches: a "vertical", hierarchical framework (studies of fairness in the agency relationship between one principal and one single agent) and a "horizontal", agent–to–agent framework (studies of reciprocity in peers' interactions under alternative incentive schemes). We explore in the laboratory a game which integrates vertical and horizontal relationships and allows to investigate how principal's fairness

*Paper presented at the CEEL 2000 Workshop at the university of Trento, March 2000. Authors would like to thank all the participants at the CEEL 2000 Workshop and at the 1999 ESA Regional Meeting, Grenoble 1999, and Rachel Croson, Manfred Königstein, Teck Ho, Alessandro Narduzzo, Tibor Neugebauer, Paolo Patelli, Karim Sadrieh, Rupert Sausgruber and Enrico Zaninotto for their helpful comments and suggestions on earlier drafts of this paper. Michele Lorenzini and Marco Tecilla provided unvaluable help during the development of the experimental software. Support from Enel is gratefully acknowledged.

†Department of Management and Computer Science and CEEL (Computable and Experimental Economics Laboratory), University of Trento, Via Inama 5, I–38100 Trento; ph: +39–0461–882372, fax: +39–0461–882124, email: arossi@cs.unitn.it, http://www.cs.unitn.it/rock

‡Department of Business Administration, University of Venice, Ca'Bembo, S. Trovaso 1075, I–30123 Venezia; ph: +39–041–2578745, email: warglien@unive.it

affects cooperation between two interdependent agents performing a simple production game. We set a 2–stage game where, in the second stage, agents play a prisoner's dilemma game and, in the first stage, the principal can withdraw any share of the output generated by the agents in the second stage. Despite theory predicting that no fairness should be observed by the principal and no cooperation should be observed by agents, our experimental findings show that agents are to some extent sensitive to principal's fairness. When the principal plays unfair (fair) agents are frequently observed to jointly defect (cooperate). Thus, fairness considerations on principal's actions may act as a coordination device for the agents and reciprocal behavior may, as a result, affect their propensity towards cooperation or defection in the game.

JEL Classification: C72, C92.

Keywords: Principal–agent theory, Prisoner's Dilemma, Team production, Reciprocity, Fairness, Experimental economics.

1 Introduction

Since our early school days (not to speak about family life), we have to learn the subtleties and complexities of reciprocity, fairness and envy in hierarchical relationships. Is it fair to share precious knowledge with your lazy school mate during a written test? Is the teacher fair in giving the same bad marks to the copying and the copied student? How to behave during next test? Despite the obvious (and painful) behavioral relevance of such questions, issues of fairness in hierarchical contexts have been empirically addressed only in recent years, mostly in a framework of agency relationships and contract design. A growing body of research in experimental economics has investigated the behavioral consequences of alternative types of incentive schemes on workers productivity (e.g.: group versus individual schemes, absolute versus relative evaluation methods),

the effectiveness of economic and non economic contract enforcement devices on work effort levels, or the impact of incentives on the intertemporal behavior and the risk attitude of decision makers.[1]

Much of the evidence gathered in the laboratory on these and related issues has shown that the behavior of subjects in agency relationships is significantly affected by relative and distributive concerns (how to share a rent, how to divide a common outcome). Subjects seem to take into account the way other players behave and perform systematic comparisons of the payoffs earned by others. Their concern for fairness results in "fair play" and reciprocal behavior (costly punishment of others' unfair behavior and costly reward of others' fair behavior).

Virtually all of these studies have focused on the emergence of reciprocating behavior in two basic dyadic relational schemes:

- *reciprocity within the vertical agency relationship between a principal and a single agent.* These studies investigate, under various conditions, the existence of reciprocal norms influencing the agency relationship between a principal and one single agent;
- *reciprocity within the horizontal agent–agent relationship under alternative compensation schemes.* These studies highlight that in team compensation and peer-to-peer working relationships relative payoff considerations may be of crucial im portance in affecting job performances (consider, for instance, the impact of relative evaluation or group incentive schemes, or the effect of information about peers compensation on job performance).

These two streams of research have been, so far, investigated independently. However, in many contexts of empirical relevance – especially in organizations – "vertical" relationships between a principal and an agent are intertwined with "horizontal" relations within the same hierarchical level. Nevertheless, very little research has jointly addressed

[1] For a comprehensive survey of experimental research on these topics we point the reader to existing surveys on these topics [Gächter and Fehr, 1999, Rossi, 1999].

these two dimensions of interaction. Not only is empirical research on "triangular" principal-agent relationships substantially absent. Theory is missing as well. To our knowledge, only a few theoretical studies [Mookherjee, 1984, Itoh, 1994] have developed the principal–agent framework in a multi–agent setting. Triangular features are similarly overlooked by economic theories of reciprocity.[2]

This paper may be regarded as an exploratory attempt to blend the vertical and the horizontal agency relations in a laboratory setting in order to analyze the emergence of "triangular" reciprocity. In particular, we explore whether and to what extent a principal's fairness affects cooperation between interdependent agents performing a simple production task.

In order to make it easier to interpret the experimental results, we have kept the experimental scheme as simple and familiar as possible. Basically, the experiment consists in a two-stage game in which in the first stage a principal decides which share of the pie generated by his/her agents he/she will keep for him/herself and which share will correspondingly be distributed to the agents; in the second stage, agents generate the pie by playing a production game in which the relative payoffs of the agents have a Prisoner's Dilemma structure, but their absolute value is determined by the unilateral choice of the principal in stage 1. Actually, one may synthetically think of it as a prisoner's dilemma embedded in a dictator game. Not surprisingly, we find that the principal-dictator's fairness strongly affects agents' behavior. Generous principals foster mutual cooperation between agents, while greedy ones induce more joint defection.

The paper is organized as follows: next section summarizes the main experimental literature on fairness in agency relationships and contract design. Section 3 introduces our experiment. Section 4 presents the main experimental results. Further developments of our research are shortly discussed in Section 5.

[2] By the way, see a short discussion in the concluding section of Rabin [1993].

2 Previous studies

Recalling the classification introduced in the previous section, in the following we will briefly review the evidence gathered on the "vertical" dimension of fairness in the experimental literature on agency, and then we will turn to the "horizontal" dimension reviewing some experiments in team and contract theory.

Reciprocity within the vertical agency relationship between a principal and a single agent. Fehr, jointly with other scholars [Fehr et al., 1993, 1998b, Fehr and Falk, forthcoming, Fehr et al., 1998a, Fehr and Tougareva, 1995] has conducted experiments based on the so–called "Gift Exchange Game", that is, a two–stage game similar to a sequential social dilemma, which can be summarized as follows: the first–stage is a wage determination game in which workers (agents) and firms (principals) trade for stipulating job contracts with each other (according to a particular labor market structure); in the second–stage, workers who have successfully concluded a contract with a firm must choose an effort level. Theoretical predictions suggest that workers should exhibit minimal effort levels (because efforts above that level are increasingly costly) no matter the wage they receive. Since firms are aware of this, they should respond by paying the competitive (zero rent) wage corresponding to the minimum effort level. Experimental findings, however, show that average wages are substantially above the competitive wage corresponding to the minimum effort level, and effort levels are higher than the minimum level. Moreover, workers's wages contain substantial amounts of rent (wages are much higher than the competitive wage corresponding to the workers' observed effort levels). These results seem to suggest that principals actually do take into account fairness motives when offering a contract to agents, and that agents react to fair wages showing working efforts higher than the minimum level. Further studies have then investigated to what extent the reciprocal attitude of principals and agents may be able to mitigate the *contract enforcement problem* [Fehr and Gächter, 1998, Fehr et al., 1997]. This problem is typical of agency relationships because many employment contracts are incomplete and workers have some effort discretion; thus,

whenever firms have limited enforcement technology and deal with rational and purely selfish workers, they are unable to enforce the efficient effort level, and can only achieve a minimal effort level (below the efficient one). On the other hand, if one assumes that principals and agents may be influenced by reciprocal concerns, their behavior may result in an outcome of the contract enforcement game different from Nash equilibrium. As a matter of fact, behavioral evidence confirms that reciprocal motives within the agency relationship may be regarded as a successful device in raising effort levels above the minimum [Fehr and Gächter, 1998], and that reciprocity alone may be more effective than many traditional contract enforcement devices such as incentive contracting, fines and monitoring [Fehr et al., 1997].

Two additional experimental studies are worth to mention. Keser and Willinger [forthcoming] implemented in the laboratory a standard textbook principal–agent game with hidden action. In this setting, the principal states a contract that specifies the agent's wage contingent to the observed ex–post profits and the agent has to decide whether to accept or decline the contract. In case of acceptance, then, he/she has to decide whether to perform a high level of effort or a low one (less expensive). The agent's chosen effort level affects the magnitude of the expected profits of the principal. The simplicity of this setting allows the authors to contrast clearly theoretical predictions to experimental findings. On the one hand, theoretical analysis have traditionally almost neglected the strategic nature of the game and, as a result, the problem has commonly been treated as an individual maximization problem for the principal, where the agent's behavior is taken into account introducing two constraints: the (*i*) participation constraint and the (*ii*) incentive compatibility constraint. These constraints assure, respectively, that (*i*) the wage schedule offered to the agent is chosen so that the expected wage is at least not lower than the side option in the labor market and that (*ii*) the expected wage, given that the agent performs the action that maximizes the expected wealth of the principal, is at least not lower than the expected wage in the case when he/she performs any other available action. As a result, at the equilibrium

the principal offers to the agent a contract that makes him/her indifferent in participating to the firm or not, and, in case of participation, to act in the interest of the principal or not. In other words, theory assumes that the agent joins the firm and performs in the best interest of the principal in a setting in which the former has no sensible gain at all in choosing so, while the latter is the only one to benefit of all the additional profits. Experimental evidence tells a different story: principal offers are far from the theoretical levels and are much more fair. In this regard, principals seem to clearly understand the ultimatum–nature of the game and avoid offers too unfair, because they may be rejected by agents.

Finally, Anderhub et al. [1999] investigate a similarly simple principal–agent game with no hidden action and deterministic profits function where the agent's contract consists in a fixed component (base pay) and a return share on firm's profits. They show that agents tend to reject unfair contracts and that fair contracts are reciprocated (efforts level are higher than the optimal ones conditional to the accepted contract).

Reciprocity within the horizontal agent–agent relationship under alternative compensation schemes. While no experimental study on team compensation, to our knowledge, have explicitly focused on the issue of fairness, it is still possible to interpret some results of this body of research as an evidence that subjects' behavior in this setting is affected by distributive concerns. Team literature has both theoretically and experimentally shown that free–riding may be the outcome of many team work interactions. This is clearly the case when the so–called egalitarian revenue sharing rule is introduced. That is, a group incentive mechanism, parallel to the voluntary mechanism in the public good literature, that assigns to each participant of the group the same share of the produced outcome, regardless of the individual contributions. Experimental evidence shows that overtime people tend to lower their contribution to the group outcome and to shirk, behaving according to the dominant strategy of the game.

Some contributions have suggested modifications in the incentive mechanism so minimize the de–motivation induced by the egalitarian

sharing. For instance, Hölmstrom [1982] has devised a forcing contract mechanism, a simple modification of the sharing rule that makes the distribution of the produced outcome among the team workers contingent upon the fulfillment of a production target and that prevents agents from free–ride. Nalbantian and Schotter [1997] proved evidence that theoretically equivalent incentive compatible devices are not equivalent in the laboratory and did clearly show that, while competitive incentives largely enhances productivity, other incentive compatible mechanism are ineffective in avoiding free–riding.

But free–riding may be lessened in other ways: some studies have explicitly investigated the role of peer pressure and direct control from coworkers as a determinant of motivation in team workgroup (see, for instance, Kandel and Lazear [1992] and Plott and Casari [1999] for an experiment in a common pool setting).

Moreover, the fact that subjects in a team work setting do care about the relative distribution of earnings within the participants pool is reflected by some studies. For instance, Croson [1999] shows that introducing feedback on effort levels of other participants in a team may result in imitation dynamics where participants converge towards the same contribution level (and where everyone earns the same payoff). In comparison with a control treatment that had no such feedback, groups in the information treatment had similar effort levels on average but much more higher variance. This higher variance was due to high between–group variance and low within–group variance. That is, some teams coordinated into high levels of effort (cooperating) while other teams locked into low levels (shirking). Hence, the simple introduction of feedback on effort chosen by other participants in the group change dramatically the outcome of the game. A possible interpretation of this difference may be that the introduction of feedback makes other subject's behavior and earnings more transparent than in the control treatment, where one can just make inferences on others' behavior based on the total group contribution. In this treatment, then, convergence towards a common contribution level (cooperation or free-riding) is fostered by a dynamical process of adjustment of individual

contributions based on easy comparisons of how other individuals are behaving and how much they are earning.

Other studies put in evidence that reciprocal behavior may be elicited by the degree of saliency of effort levels. Real task experiments within the standard egalitarian revenue sharing rule, have proved to elicit behavioral responses somehow different from standard experiments with simulated task. When the task environment is modeled so that people perform a real production task, subjects seem to feel obliged to reciprocate to high levels of contribution of other teammates and free riding is not the typical outcome of the experiment [London and Oldham, 1977, van Dijk et al., 1998].

Finally, in some asymmetric settings, where some subjects benefit of some advantages relative to others, disadvantaged subjects may be more focused in acting so that advantaged subjects do not earn considerably more than themselves, regardless of maximizing their individual expected payoff. This is the case of many bargaining situations (such as the ultimatum game, where responders decline unfair offers even if this result in losses for them) and in contract theory this is the case of some experiments on tournaments. Bull et al. [1987], for instance, examined the behavior of experimental subjects of 2–person rank–order asymmetric tournaments.[3] Theoretical analysis predicts that, in equilibrium, the effort level of one contestant should be inversely proportional to his/her cost effort and, as a consequence, the advantaged contestant should have a larger probability to win the prize because of its favorable cost schedule. The authors, on the other hand, proved evidence that disadvantaged subjects may show effort levels systematically higher than equilibrium predictions, the reason being in perceiving the setting as unfair and exploited by the other participant. As a result they may act as if they were concerned to "steal" the prize of the tournament to the advantaged contestant, although the expected payoff for the disadvantaged subject, in doing so, is much lower than at the

[3] These are tournaments in which participants show different skills or attitudes towards effort. The player with a higher (lower) effort cost than another is called the "disadvantaged" (advantaged) player.

equilibrium, due to his/her high costs of effort.

3 A Prisoner Dilemma with a dictator

3.1 The model

Consider this simple production setting with one principal and two agents: the two agents are involved in a simple production task where each of them has to decide on the allocation of his/her working effort. More precisely, each agent has to decide whether he/she is going to help (or collaborate with) the other agent or not. The decision of one agent affects both his/her production level and the one corresponding to the other agent: while helping efforts of one agent increase the other's production level, on the other side they decrease the agent's own production amount. Moreover, if both the agents decide to provide help, they are both better off (with respect to their production levels) but, regardless of what the other agent is going to do (provide help or not), the production level for one agent is always higher when he/she is not providing help because he/she can concentrate more effort on his/her own production task.

Produced units are placed in a market with excess demand by the firm owner, the principal. Without loss of generality we can assume that each produced unit is worth 1 experimental currency unit for the principal. He/she is the residual claimant of the value of units produced by the two agents. Agents' remuneration for the production is governed by a simple piece rate rule, whose rate per unit is identical for the two agents and is decided by the principal.

Agents cannot decide to terminate the contract with the firm (this means that no side market option in introduced in the model and the participation of workers in the firm is not investigated here).

The described production task is modeled as a two–stage game, that runs as follows: in the first stage the principal publicly announces which share $1 - (W/100)$ of the output value (that has to be produced in the second stage by the two agents) is to be attributed to him/herself as his/her own

payoff in the round. Alternatively, one can interpret $W/100$ as the piece rate, the per unit remuneration assigned to agents by the principal.[4] The domain for W is any integer number between 1 and 100. The decision of the principal is binding to the agents. In the second stage, then, each agent has to decide between two alternative strategies (A and B) that result in different individual output values, as shown in Figure 1.

(q_1, q_2)	A	B
A	$60, 60$	$10, 70$
B	$70, 10$	$20, 20$

Figure 1: The firm production function: relationship between agents' decisions and agents' individual output levels.

This structure of output may be thought as the simplest way to model task interdependency of two agents in a production setting: if they both choose to cooperate or help each other (strategy A) they are both better off, while defection or restraining from helping efforts towards the other agent (strategy B) is the dominant strategy (for an agent concerned in maximizing his/her own production level).

Hence, the agents' relative payoff structure in the game clearly recall a prisoner's dilemma game where absolute payoffs depend on W as depicted in Figure 2.

$\text{Payoff}(A_1)$, $\text{Payoff}(A_2)$	A	B
A	$(W/100)60, (W/100)60$	$(W/100)10, (W/100)70$
B	$(W/100)70, (W/100)10$	$(W/100)20, (W/100)20$

Figure 2: Agents' payoff conditional to piece rate W and agents' behavior.

Finally, the principal's payoff, depending on the agents' strategies, is determined as in Figure 3.

Using a standard backward induction argument it is clear that, whatever the principal decides in the first stage of the game, in the second stage the

[4] Language in subjects' instructions where kept as neutral as possible and we explicitly avoided terms as "piece rate" or "remuneration".

	A	B
A	$(1 - W/100) \times 120$	$(1 - W/100) \times 80$
B	$(1 - W/100) \times 80$	$(1 - W/100) \times 40$

Figure 3: Principal's Payoff conditional to piece rate W and agents' behavior.

agents should defect (avoid helping efforts), since they face a standard prisoner's dilemma game (whose payoffs are a linear transformation of the production levels of Figure 1). As a result, the first stage of the game somehow recalls the structure of a dictator game, where the principal may decide to retain the largest possible share of the pie. Hence, the unique Nash equilibrium for the one shot game is for the principal to choose $W = 1$ and for the two agents to play strategy B.

3.2 The experimental design

The experimental design is very simple, and consists in a two experiments that were played sequentially by a population of 54 college undergraduates recruited at the University of Trento (Italy) during July 1999 (30 of them were undergraduates in Economics). Subjects were recruited through announcements on bulletin boards in the Faculty of Economics and were asked to show up at the Computable and Experimental Economic Laboratory. The announcements claimed that subjects would have been engaged in an experiment lasting about 1 hour and would have been able to gain up to a maximum of 50000 Italian liras (approximately equal to 25 US dollars). During the experiment subjects earned experimental points that were at the end converted in Italian liras at the rate of 15 Italian liras per experimental point and were paid to the subjects in addition to a show up fee of 10000 Italian liras (approximately equal to 5 US dollars). The exchange rate was known in advance by all subjects. Their average final payoff was of about 34000 Italian liras (approximately equal to 17 US dollars) for subjects in the role of principals and of about 16000 Italian liras (approximately equal to 8 US dollars) for subjects in the role of agents, amounts which seemed more than sufficient to motivate them

during the experiment.

The subjects were randomly divided in groups of 3 subjects who remained anonymously grouped during the entire experiment; the role of principal or agent was also randomly assigned. Subsequently, subjects were seated in front of computer terminals. After that an experimental administrator had read the experiment instructions[5] and answered aloud to any question,[6] the experiment begun. Interaction between subjects were reduced to the minimum during the experiment: each subject could see some two other participants in the room but not their terminal monitors and verbal communication was not allowed at all. Since one group could finish the experiment earlier than the others, participants were asked to remain quietly seated at their desk and to fill a payment form needed for the payment of the experiment.

The experiment consisted in the repetition of 15 identical rounds of the game presented in Section 3.1. The number of repetitions were considered a reasonable length of time to allow learning to take place (if any was to occur). Each round was thus organized:

- *First stage.* The subject in the role of the principal is asked to type a number between 1 and 100, corresponding to the value to assign to variable W;
- *Second stage.* Each of the two subjects having the role of agent are communicated the value of variable W and are asked to choose between strategy A or B;
- *End of round.* each of the three subjects is given information on the decision taken and the payoff earned by all the participants of the group.

At the end of the 15 rounds, subjects where told that they had to participate to another experiment (experienced treatment), where groups

[5] A translation from Italian of instructions is in Appendix A.

[6] Each subject was revealed his/her role in the experiment, the principal role or the agent role, only after all questions were answered, so that, in asking questions to the administrator, subjects could not signal to other participants their role.

were randomly reshuffled while everyone kept the role held in the previous experiment (novice treatment).

The total payoff of each subject at the end of the experiment was then equal to the sum of the payoffs earned by the subject during the 30 rounds, plus the show up fee.

4 Experimental Results

Figure 4 shows the plot overtime of the average piece rate W and the average joint production levels $(q_1 + q_2)$ in each of the two treatments; Table 1 presents some summary statistics on piece rates and on the percentage of agents cooperating in each experimental session.[7]

Table 1: Descriptive Statistics per session

			novice	treatment	experienced	treatment
session	groups	subjects	av. W	coop. rate	av. W	coop. rate
1	4	12	23.5	0.492	24.5	0.458
2	4	12	20.5	0.375	17.6	0.225
3	4	12	36.7	0.533	35.8	0.308
4	3	9	23.6	0.555	23.4	0.322
5	3	9	-[7]	-[7]	30.8	0.411
ALL	18	56	26.3	0.484	26.3	0.343

The equilibrium prediction is fulfilled in a relative low number of observations (around 12% in the novice treatment and 20% in the experienced treatment). The time series of piece rate do not show any significant trend towards the equilibrium in both treatments. Both treatments presents also the same average piece rate value ($\overline{W} = 26.3$).

The observed average of agents cooperating (the plot is not presented here, but can be easily inferred from Figure 4, since the average rate of

[7] It was not possible to report the results for session 5, novice treatment, because of a bug in the experimental software that overwrote results of the experienced treatment in the novice treatment logfile.

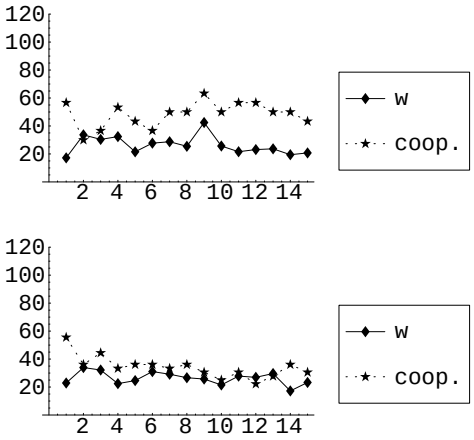

Figure 4: Average piece rate W and average production levels for the novice and the experienced treatments.

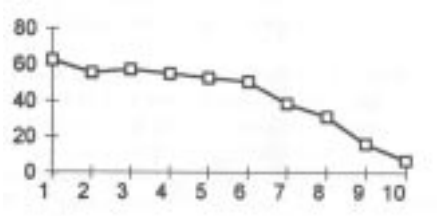

Figure 5: Average cooperation rates in a 10-round prisoner' dilemma (partners condition).

cooperation is a linear transformation of the average production levels) appears to be constant overtime in the novice treatment and to show a slightly decreasing trend in the experienced treatment. These graphical inferences are confirmed by statistical tests: the decline in the cooperation rate between the first (56.6%) and the last (43.3%) period in the novice treatment is not significant, while the same difference (from 55.5% to 30.5%) is significant in the experienced treatment (randomization test, $\alpha = 0.05$). The average cooperation rates are, respectively, about 48% in the novice treatment and about 34% in the experienced treatment.

Although it was not possible to test whether the difference between the two treatments was significant, due to the lack of sufficient independent

observations, it is possible to compare this observed cooperation levels with the experimental evidence on standard iterated prisoner's dilemma. Figure 5 reports the average cooperation rate in a 10–round prisoner's dilemma in the experiment run by Andreoni and Miller [1993]. The difference in cooperation rates with respect to our experiment is evident: subjects in the novice treatment show no convergence towards the equilibrium over the 15 rounds, while in Andreoni and Miller [1993] the cooperation rate converges close to zero in the 10 rounds. In the experienced treatment subjects show a slight downward trend but still cooperation levels are sensibly higher that in the standard prisoner's dilemma game.

As it is reasonable to expect, a close investigation of data reveals that the experimental behavior of agents is heavily affected by the behavior of the principal. In accordance with common sense, but contradicting equilibrium predictions, the level of cooperation between agents responds to the generosity of the principal. When the principal increases the piece rate, agents do not decrease the joint production level in 84% (81%) of the observations in the novice (experienced) treatment, and when the principal decreases the piece rate, agents do not increase the joint production level in the 82% (91%) of the observations in the novice (experienced) treatment.

This pattern of behavior may be viewed, in both novice and experienced triples of players, in Figure 6. When the principal tends to choose values of W near the equilibrium, agents coordinate on (B,B) in the majority of observations. As the principal selects higher values of W, more pairs of agents tend to coordinate on (A,A) (in this case, all three subjects are better off than in equilibrium). Thus, principal's fairness matters and affects the mutual relationships of agents. For values of W lesser than 65-75, there is a neat monotonic mapping from the piece rate W to the output achieved. Notice, however, that for high levels of W the correlation between piece rate and output breaks down. This effect may be not significant since there are very few observations in the right tail of the histograms plotted in Figure 6 ($W > 70$ only in 8% (4%) of the observations in the novice (experienced) treatment). Moreover, as the history of individual sequences of runs reveals (see Figures 6-7), the

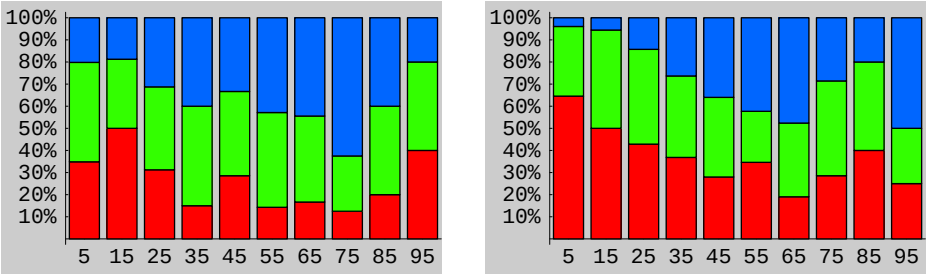

Figure 6: Histogram plot of the frequency of productivity of agents corresponding to different classes of piece rate W decided by the principal in the novice and in the experienced treatment. Top, middle and bottom bars are respectively related to the frequency of observations where both the agents cooperate, one agent only cooperates and the other one defects, both agents defect.

breakdown in some observation is maybe due to the fact that most "ultrafair" piece rates have been offered by principals after a sequence of highly unfair moves – which might impair their credibility.

Despite this "ultrafairness" bias, the overall link between agents' performance and principals' fairness is significant for both the novice and expert populations of players: the Spearman rank order correlation coefficient is equal to $r_s = 21\%$ for novices and equal to $r_s = 36\%$ for experts (both significant at the $\alpha = 0.001$ level). The increased rank order correlation between the treatments is due to a large extent to the increased frequency of (B,B) responses to low piece rates. Thus, it seems that players have learned a strategy of coordinated reciprocation to the principal unfair moves, and this may explain as well the descendent trend in average production levels of the agents in the experienced treatment. This suggests that the impact of fairness considerations is not a temporary phenomenon, to be dissolved by a better understanding of the game structure, but instead it is deeply connected to the way subjects understand and interpret the game itself.

As a matter of fact, one may argue that the link between the level of the piece rate W and cooperation rates could be explained in many other ways

alternative to this interpretation in terms of coordinated reciprocal behavior of the two agents against the principal. In particular, one may suggest that the attitude of the agents to cooperate in a prisoner's dilemma game may be affected by absolute size effects in the payoff structure, so that instances of the prisoner's dilemma game with the same relative structure of the payoffs (same ratios) but with different absolute values may be played differently by experimental subjects. Put it in other words, maybe the experimental outcomes of a prisoner's dilemma game are not invariant to linear transformation of the payoffs.

Thus, we implemented as a control treatment a 2-person prisoner's dilemma game with payoffs of random absolute size, in order to test the existence of differences on the agents' attitude to cooperate. Payoffs for the two players of the game were, once again, the ones depicted in Fig. 2, but this time a computerized device, rather than another experimental subject (the principal), chose the value of W sampling at random from the uniform distribution $U \sim [1, 100]$.[8]

The results of the control treatment are collected in Fig. 7. The comparison of Fig. 7 with Fig. 6 clearly shows that, while the experimental outcomes of a prisoner's dilemma game are not invariant to linear transformation of the payoffs, the effects on cooperation rates of size increases of agents' payoffs are opposite to the ones of fairness depicted in the baseline treatment. As a matter of fact, the control treatment shows that, as the magnitude of the payoffs increases, more and more agents defect, while cooperation is a more frequent outcome when payoffs are

[8] The experimental design closely followed the one described for the principal–two–agent treatment, with the following differences: 36 first year undergraduate students (with no previous knowledge of game theory) were recruited; subjects were divided in three cohorts of 12 participants and then randomly matched in couples; only one experiment rather than two were run, since we didn't want to test for the role of experience given the simpler structure of the game; experimental points were converted at the rate of 40 Italian liras per point, and subjects earned on average 22000 Italian liras (approximately equal to 11 US dollars) for an experiment lasting, on average, around 35 minutes. Each round of the experiment run as follows: in the first stage the computer program extracts the random value of W and sends it to the two agents, in the second stage each of the two agents play a prisoner's dilemma game with the payoff collected in Fig. 2.

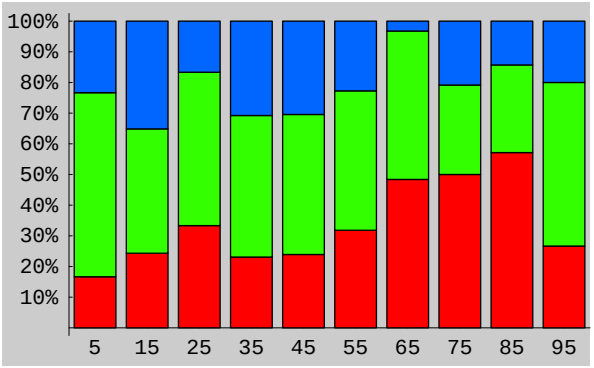

Figure 7: Histogram plot of the frequency of cooperation of agents corresponding to different classes of the random scale value W in the control treatment. Top, middle and bottom bars are respectively related to the frequency of observations where both the agents cooperate, one agent only cooperates and the other one defects, both agents defect.

smaller. In other words, the analysis of single play sequences (not showed here) suggests that many agents seems during the rounds of the experiment to signal cooperation when the payoffs of the game are smaller and to exploit the other agent through defection when the payoffs are higher.[9] Overall, the comparison of the results of the control treatment to the ones of the baseline treatment confirm us that in the baseline treatment the two agents reacts to the attitude of the principal to share the outcome of the game, performing coordinated action of reciprocal behavior to sanction (recompense) unfair (fair) offer to share the pie.

While the broad facts seem quite clear, understanding how principals' fairness affects relationships between agents deserves some caution. Simplifying a bit things, reciprocity has two faces, one positive and the other negative [Rabin, 1993]: I may be willing to sacrifice my own material well-being to help the kind other or I may be willing to sacrifice my own material well–being to punish the unkind other. In dyadic relationships, these two faces can be easily distinguished. This may not be the case with

[9]This result seems to be particular robust and interesting if one takes into account that the relative structure of the payoffs of Fig. 2 is such that the gain from defection is relatively small compared to the gain from mutual cooperation.

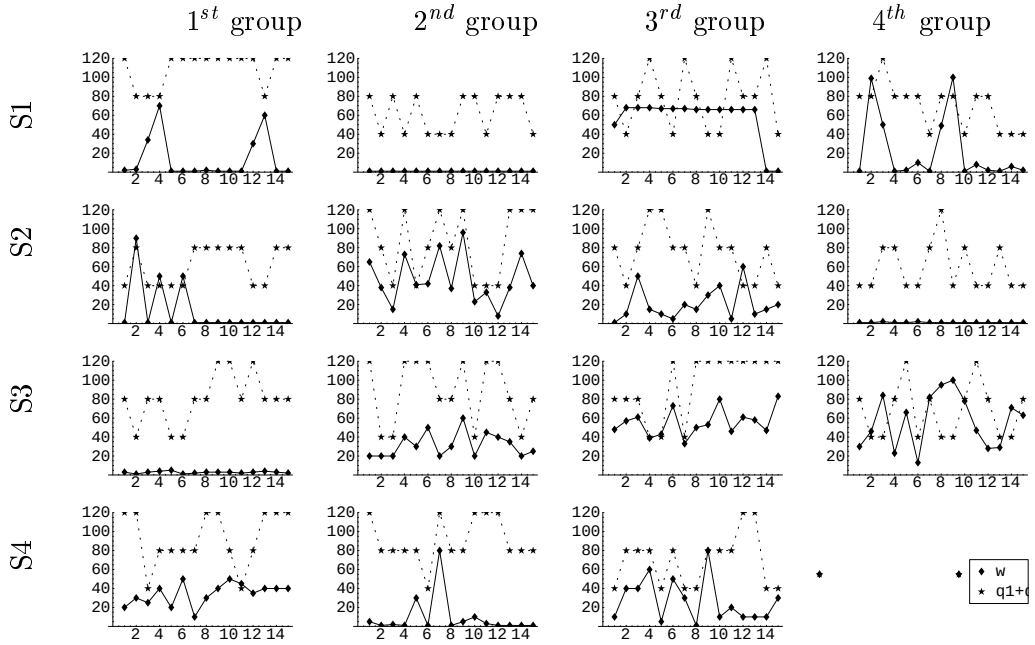

Figure 8: Piece rates and joint production levels series for each group, novice treatment.

triangular relationships. In particular, our experiment clearly shows that fair principals tend to generate positive reciprocity between pairs of agents – they act in each other's favor, as well as in favor of the principal, and show higher and more persistent cooperation rates than in a conventional prisoners' dilemma. On the other hand, interpreting how principals' unfairness affects relationships between agents is much harder. The experimental data suggest that unfair principals induce less cooperation between agents than the one usually observed in prisoners' dilemmas. But is this due to the fact that greedy principals generate greedy agents, or to the fact that agents unite their purposes in retaliating the principal? In other words, does hierarchical unfairness induce unkindness or mutualism between agents? Unfortunately, the structure of the game doesn't help much in directly discriminating between these two hypotheses. By deciding to produce B, an agent hurts the principal but at the same time does makes the other agent worse off. There is no way to infer an agent's intention from a single move.

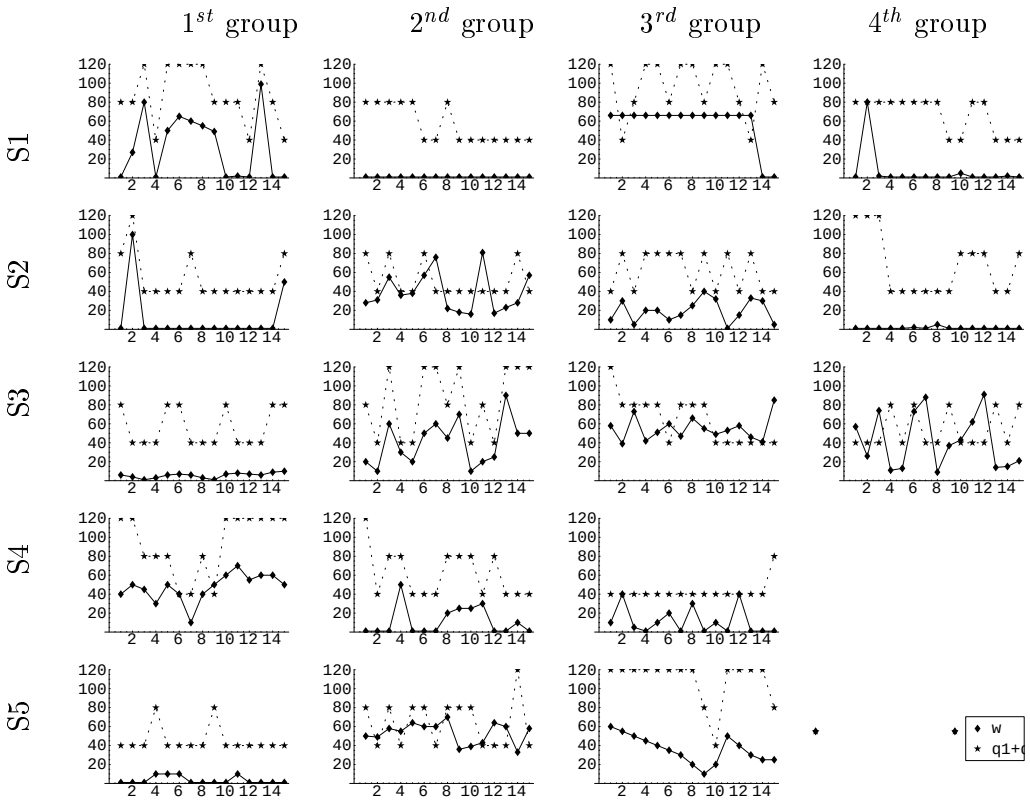

Figure 9: Piece rates and joint production levels series for each group, experienced treatment.

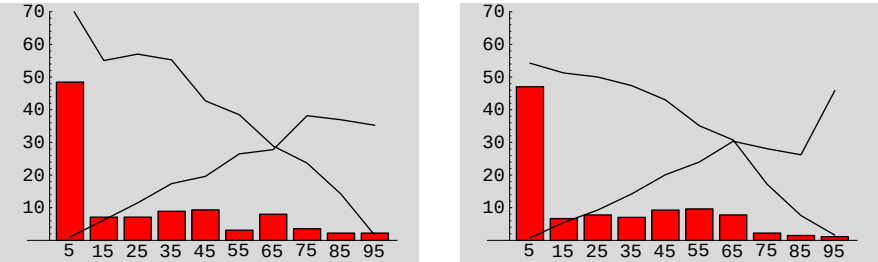

Figure 10: Relative frequencies of piece rate W and average payoff for a principal (downward data plot) and an agent (upward data plot) in the novice and in the experienced treatment.

Nevertheless, it is reasonable to argue that prompt shifts from (B, B) to (A, A) as principals increase piece rates may witness the existence of positive reciprocity (mutual help) between agents. A closer analysis of single play sequences (Figures 8–9) reveals that there are many such examples. In particular, in the experienced treatment, almost all instances of (A, A) emerging after mutual defection are responses to increases in piece rates, and some single play sequences show neat examples of well coordinated series of retaliation and reward by both agents (an almost perfect example is 2^{nd} group – S3 in the experienced treatment). However, the reverse doesn't always hold, and often increases in piece rates fail to elicit cooperation between agents. It isn't clear whether these counterexamples are due to mistrust against the principal or self–interest: thus the indirect evidence from single play sequences is only partially supportive of positive reciprocity between agents, and probably more accurate experimental design is needed to better discriminate between the two types of agents' reciprocity.

Another interesting problem is explaining the persisting high level of unfairness of principals (see Figure 10). If for example one takes as a benchmark usual laboratory behavior in bargaining games [Roth, 1995], the behavior of principals in our experiment seems unusually greedy. In about 50% of cases, principals take as much as possible. Actually, our principals' behavior bears more resemblance to that of players of a dictator game. In part, this may be explained by the persistence of some miscoordination in

agents' responses, that makes unfairness paying off on average. Given the observed distribution of agents' responses, taking as much as possible is still the best move for principals even in the experienced treatment (Figure 10). Although experienced agents succeed in reducing the steepness of the principals' average payoff curve, the level of coordinated retaliation by agents is not enough to transform such curve in a parable (even in the case of minimal piece rates, there is still a 30% of eveniences in which at least one agent plays A) . We also suspect that, as the number of other players increases, considerations of unilateral fairness (pure altruism) may dilute.

5 Discussion and further research

While we think that our experiment convincingly demonstrates that vertical and horizontal fairness interact in hierarchical triangles, much needs to be done to better understand the nature of such interaction. As we have already seen, in our experiment it is hard to discriminate between two different effects of principals' unfairness: negative versus positive reciprocity between agents. More accurate experimental design might be devised to separate those two effects. Furthermore, we think that less symmetric situations are useful to explore. For example, the principal might be able to differentiate agents' rewards, introducing asymmetries in incentives; asymmetries in agents' capabilities are case of interest as well. Also, the effects of information asymmetries deserve further investigation: fairness considerations may be significantly affected by different distributions of information among players.

Finally, we claim that our experiment suggests more prudence in the use of standard game–theoretic concepts in organization theory. While the use of non–cooperative games as a tool for modeling organizational phenomena has become widespread (and the prisoner's dilemma is especially abused!), little or no attention has been accorded to how behavior in such games may change when they are immersed in a hierarchical context. Our experiment shows that even when equilibria do not change, the hierarchical context may deeply affect actual agents' strategies. We think that much

useful understanding might be gained by systematically exploring how well–known games are played in hierarchical contexts.

A Experimental Instructions

A.1 Introduction

You are participating to an economic experiment. You are kindly asked to carefully read the instructions. Then you will be able to ask questions that will be openly answered. This experiment will last about one hour. If you follow the instructions closely and make decisions carefully, you can earn a considerable amount of money. During the experiment you can earn experimental points that at the end of the experiment will be converted into Italian liras (1 experimental point = 15 Italian liras) and will be added to the fixed amount of 10000 Italian lira. This will be your monetary payment for participating in the experiment.

A.2 Instructions

During the whole experiment you are anonymously matched with other two players in this room. One of the players is called *player* 1 (from now on, P_1) and the other two players are called *player* 2 and *player* 3 (P_2 and P_3). Matching will be performed at random by the computer program at the beginning of the experiment and will not revealed. Your identity during the experiment (P_1, P_2 or P_3) will be revealed after reading the instructions and after that all questions will have been answered.

The experiment involves the repetition for 15 times (rounds) of two stages, that will be described in a moment. At the end of each round, payoffs will be announced and then the next round will start. Your final payoff will be equal to the sum of payoffs earned in each of the rounds, plus the fixed payment of 10000 Italian lira.

Each round runs as follows:

First phase

Player P_1 decides and sends to players P_2 and P_3 the value to be assigned to W. W is a percentage number that can be chosen among all integer numbers between 1(%) and 100(%).

Second phase

Players P_2 and P_3 decide, independently and simultaneously, whether to undertake action A or action B.

End of round

The experimental software computes the quantities Q_2 and Q_3, produced by players P_2 e P_3, on the basis of their choices during the second phase according to the following table:

P_2's action \ P_3's action	A	B
A	$Q_2 = 60, Q_3 = 60$	$Q_2 = 10, Q_3 = 70$
B	$Q_2 = 70, Q_3 = 10$	$Q_2 = 20, Q_3 = 20$

Finally, the experimental software computes and sends to everyone the payoff earned by each player. Payoffs are computed according to the following formulas:

P_1**'s Payoff** $= (100 - W)(Q_2 + Q_3)$;

P_2**'s Payoff** $= WQ_2$;

P_3**'s Payoff** $= WQ_3$.

References

V. Anderhub, S. Gächter, and M. Königstein. Efficient contracting and fair play in a simple principal-agent experiment. Discussion Paper, 1999.

J. Andreoni and J. H. Miller. Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal*, 103:570–585, 1993.

C. Bull, A. Schotter, and K. Weigelt. Tournaments and piece rates: An experimental study. *Journal of Political Economy*, 95(1):1–33, 1987.

R. Croson. Feedback in voluntary contributions mechanisms: An experiment in team production. In R. M. Isaac, editor, *Research in experimental economics*, volume 7. JAI Press Inc., 1999.

E. Fehr and A. Falk. Wage rigidities in a competitive incomplete contracts market. *Journal of Political Economy*, forthcoming.

E. Fehr and S. Gächter. How effective are trust- and reciprocity-based incentives? In A. Ben-Ner and L. Putterman, editors, *Economics, Values ad Organizations*. Cambridge University Press, 1998.

E. Fehr, S. Gächter, and G. Kirchsteiger. Reciprocity as a contract enforcement device. *Econometrica*, 65(4):833–60, 1997.

E. Fehr, E. Kirchler, A. Weichbold, and S. Gächter. When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor Economics*, 16(2):324–51, 1998a.

E. Fehr, G. Kirchsteiger, and A. Riedl. Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, 108 (2):437–60, 1993.

E. Fehr, G. Kirchsteiger, and A. Riedl. Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42(1): 1–34, 1998b.

E. Fehr and E. Tougareva. Do high stakes remove reciprocal fairness? - evidence from russia. Discussion Paper, University of Zürich, 1995.

S. Gächter and E. Fehr. Experiments and labour economics. Discussion Paper, University of Zürich, 1999.

B. R. Hölmstrom. Moral hazard in teams. *Bell Journal of Economics*, 13 (2):324–40, 1982.

H. Itoh. Job design, delegation and cooperation: A principal-agent analysis. *European Economic Review*, 38(3-4):691–700, 1994.

E. Kandel and E. P. Lazear. Peer pressure and partnership. *Journal of Political Economy*, 100(4):801–18, 1992.

C. Keser and M. Willinger. Principals' principles when agent's actions are hidden. *International Journal of Industrial Organization*, forthcoming.

M. London and G. R. Oldham. A comparison of group and individual incentive plans. *Academy of Management Journal*, 20(1):34–41, 1977.

D. Mookherjee. Optimal incentive schemes with many agents. *Review of Economic Studies*, 51:433–46, 1984.

H. R. Nalbantian and A. Schotter. Productivity under group incentives: An experimental study. *American Economic Review*, 87(3):314–41, 1997.

C. Plott and M. Casari. Agents monitoring each others in a common-pool resource environment. Discussion Paper, 1999.

C. Prendergast. The provision of incentives in firms. *Journal of Economic Literature*, XXXVII:7–63, 1999.

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

A. Rossi. Incentives in managerial compensation: A survey of experimental research. Rock Group Working Paper, 1999.

A. E. Roth. Bargaining experiments. In J. Kagel and A. E. Roth, editors, *The Handbook of Experimental Economics*. Princeton University Press, 1995.

F. van Dijk, J. Sonnemans, and F. van Winden. Incentive systems in a real effort experiment. Discussion Paper, 1998.