QUADERNI

Università degli Studi di Siena
DIPARTIMENTO DI ECONOMIA POLITICA

GIORGIO CORICELLI

Sequence Matters: an Experimental Study of the Effects of Experiencing Positive and Negative Reciprocity

n. 369 – Novembre 2002

Abstract - This paper presents an experimental analysis of people's behavior in situations involving both positive and negative reciprocity. The experiment implements sequences of two types of extensive form games called Punishment games and Trust games. The contemporaneous use of these two types of games allows us to define an ideal framework for understanding the basic elements of reciprocal behavior. Results show that the level of trust and punishment are consistent with the view that emotions are involved.

PsycINFO Classification: 2360

JEL Classification: D63, C78, C91

Keywords: Reciprocity, trust, intentions, emotions, experiments.

I am grateful for financial support from the National Science Foundation (grant number SES9818561). I would like to thank James C. Cox, Amnon Rapoport, Luis González Morales, Astrid Hopfensitz and Rupert Sausgruber for their suggestions and comments. I am particularly grateful to Samuel Bowles for his valuable review of the paper.

Giorgio Coricelli, Department of Economics, University of Siena

1 Introduction

This paper studies human behavior in games with underlying reciprocity responses. We use the explication of reciprocity given by Fehr and Gächter: "Reciprocity means that in response to friendly actions, people are frequently much nicer and much more cooperative than predicted by the self-interest model; conversely, in response to hostile actions they are frequently much more nasty and even brutal" (cf. Fehr and Gächter, 2000 a, pp159.) In this definition there are two implicit concepts: positive reciprocity and negative reciprocity. People respond to friendly or hostile actions disregarding material incentives.

Experimental results from ultimatum bargaining games, trust games, gift exchange games, public goods experiments, and social dilemma games show evidence for reciprocal behaviors (negative and positive reciprocity). In the experimental literature we can find several explanations for the evidence of reciprocal behavior: (i) inequality aversion (Bolton and Ockenfels, 2000; Fehr and Schmidt, 1999); (ii) person-based response: people respond to the type of person they face (Levine, 1998); (iii) intention-based response: a desire to reward good intentions or punish bad intentions (Kagel et al., 1996; Brandts and Solà, 2001; Falk, Fehr, and Fischbacher, 2000); boundedly rational behavior (Roth and Erev, 1995; Gale, Binmore and Samuelson, 1995.)

Reciprocity has been modeled as evolutionarily stable behavior (Güth and Yaari, 1992; Bowles and Gintis, 1999.) In this sense, reciprocity represents the best strategy in terms of fitness, i.e., no other strategy can enter the population and perform better. Bolton (1997) shows how a 50-50 split in a bargaining problem is a rational strategy in an evolutionary framework, i.e., it is an evolutionarily stable strategy (ESS).

Young (1993 b) theoretically indicates 50-50 split as stochastically stable division in the bargaining problem characterized by two bargainers randomly matched from two classes of agents[1]. This unique stable equilibrium arises as a result of the evolutionary bargaining process when mobility between the two classes is allowed[2]. In this case 50-50 split is a "convention", meaning a self-enforcing and expected pattern of behavior (Lewis, 1969; Sugden, 1986).

The theory of Reciprocal Altruism (Trivers, 1971; Axelrod and Hamilton, 1981; Axelrod, 1984) is consistent with the evolutionary approach. This was noticed by Dawes and Thaler (1998): "An implication of reciprocal altruism is that individuals will be uncooperative in dilemma situations when there is no possibility of future reciprocity from others, as in situations of anonymity or interacting with people on a 'one shot' basis." But the experimental evidence is in favor of a more "radical" definition of reciprocity[3]: people engage in reciprocal behaviors even in single interactions (evidence of 40-60 percent of socially optimal contribution in "one shot" public good experiments; evidence of 50% trust and 75% reciprocal behaviors in the single play Trust Game by McCabe and Smith, 2000.) Therefore, reciprocal altruism gives a limited explanation of the observed cooperation (and reciprocal behaviors).

---

[1] See Young (1993 a) for a formal analysis of the stochastically stable equilibria. This analysis is based on graph theory. Stochastic stability is explained by the dynamics of the process of equilibria selection in a no-learning and no-reputation environment. The theory does not assume common knowledge over agents' utility functions.

[2] The stochastically stable division is the asymmetric Nash solution in the cases of homogeneous agents (in terms of utility function and information) in both classes. Uniqueness is achieved if precision of demand tend to 0. The asymmetry arises for the difference in the amount of information. The class with more agents has more information. When both classes are characterized by heterogeneous agents, the stochastically stable division is still a unique asymmetric Nash; but in this case the asymmetry arises for the difference in the distributions of types between the two classes, and not for their relative dimension.

[3] In analogy to the concept of *strong reciprocity*. See Gintis (2000) for an evolutionary model.

Our hypothesis is that human minds have a propensity to establish long-term reputations as cooperators and non-defectors, and consequently people are willing to incur the risk that their anonymous counterparts are like-minded persons (see Coricelli, McCabe, and Smith, 2000.) This hypothesis implies that some subjects in single-play extensive form games will opt, or respond to, moves whose intentions are to signal a desire for positive reciprocity, and the achievement of greater individual as well as social surplus than if each agent played opportunistically[4]. This perspective only partially explains the pattern of trust and reciprocal behavior in our experiment. In addition, we assume that experienced emotions matter.

The specific emotions that are related with negative and positive reciprocity are anger and elation. Anger (elation) is related with *counterfactual thinking* (see Lewis, 1973; Roese and Olson, 1995; Kahneman and Miller, 1986) that arises when individuals compare the obtained outcome with better (worse) outcomes that might have been realized (see Niedenthal et al., 1994; Zeelenberg et al., 1998). We use the terms upward or downward counterfactual when the obtained outcome is compared with better outcomes or worse outcomes, respectively. Anger and elation result from upward and downward counterfactual, respectively.

In the case of reciprocal interaction the obtained outcome depends on the decision of two or more agents. We refer to anger (or elation) if the individual will feel just partially responsible of her final outcome. Indeed, she will attribute the responsibility of her outcome to another individual or group of individuals[5]. The counterfactual reasoning allows the individual to focus on the alternative outcome that she would have obtained if

[4] Opportunism means that one is only interested in own monetary payoff. Standard game theory assume common knowledge of opportunism.

[5] For consideration about responsibility see Connolly and Ordoñez, 1997; and Zeelenberg et al., 2000.

the other individual (or individuals) had behaved differently. This process is based on the attribution of intentionality to the other individual's behavior (see Rabin, 1993; Brandts and Solà, 2001; and Falk et al., 1999.)

In the specific of our experimental framework distinct emotions are linked to the appraisals of responsibility for the determination of the final outcome, in accordance to the appraisal theory of emotion (see Frijda, 1986; Frijda et al., 1989). "When people perceive some other person to be the cause of their misfortune, they feel angry; when people perceive impersonal circumstances beyond human control to be the cause of their misfortune they fell sad; when they perceive themselves to be the cause of their misfortune they fell guilty" (cf. Keltner et al., 1993, pp 741.)

We assume that the experience of anger and elation affects individual choice. We name these emotions as "reciprocity-based" emotions. In particular, experienced reciprocity-based anger increases the willingness to punish (even if this is costly), and decreases trust. Even though we did not use a direct measure of emotional arousal, we strongly believe that our experimental results show the presence of reciprocity-based emotions and their effects to the levels of trust and punishment.

In the next section we illustrate the design and the procedure of our experiment; section 3 briefly reports the experimental details; section 4 reports our results; and section 5 concludes the paper with discussion.

2 Experimental design and hypotheses

Subjects participate in two-player extensive form games. The games are of two types: Punishment Games (P)[6] and Trust Games (T)[7] (see Fig.1). The subjects play two sequences of games. Each sequence is composed by three games of the same type (trust or punishment).

There are 6 subjects for each experimental session. The subjects are matched following the matching rules described in Fig. 2. These two types of games (P and T) are introduced in order to study negative and positive reciprocity. In the punishment game the first player can move left, ending the game and splitting an amount of money, or move right and possibly obtain a greater payoff for herself and a smaller payoff for the second player, if the second player does not punish her. Punishment occurs if the second player moves right, thus ending the game with zero payoff for both players.

The sequence of punishment games is characterized by an increase in the difference between the payoffs that the second player would have gotten if the first player had played non-opportunistically (moving left)[8], compared to the payoffs if the first player moves right and the second does not punish her (moving left). These differences are 26, 31, and 36 experimental dollars, for the first, second, and third punishment game.

[6] The punishment game is a truncation of an ultimatum game. In the two-person ultimatum game the first mover proposes how to divide an amount of money, a pie. The responder can either accepts or rejects the offer. If the second mover rejects the offer, both end up with nothing. Experimental results show a mean offer of 30-40% of the pie, and offers below 20% are almost always rejected.

[7] The trust game is a simplified version of an investment game (see Berg et al, 1995). The Investment Game is a two-person sequential game; the first mover (player 1) must choose how much of $10 to send to the second mover (player 2). Both players know that the amount sent will be tripled when it reaches player 2. Player 2 must then decide how much of the tripled money to send back to player 1 (and how much to keep). The money sent back to player 1 does not triple again. Game theory predicts that rational agents playing once will act as follows: player 2 has a dominant strategy to keep all of the tripled money sent by player 1; player 1 should infer this behavior by player 2, and therefore send nothing. In fact, this does not happen. Over 90% of the player 1s exhibit some degree of trust by sending some amount of money, and over a third of the player 2s show themselves to be trustworthy by sending back more than player 1 initially sent.

[8] For a definition of opportunistic behavior, see note n. 5 above.

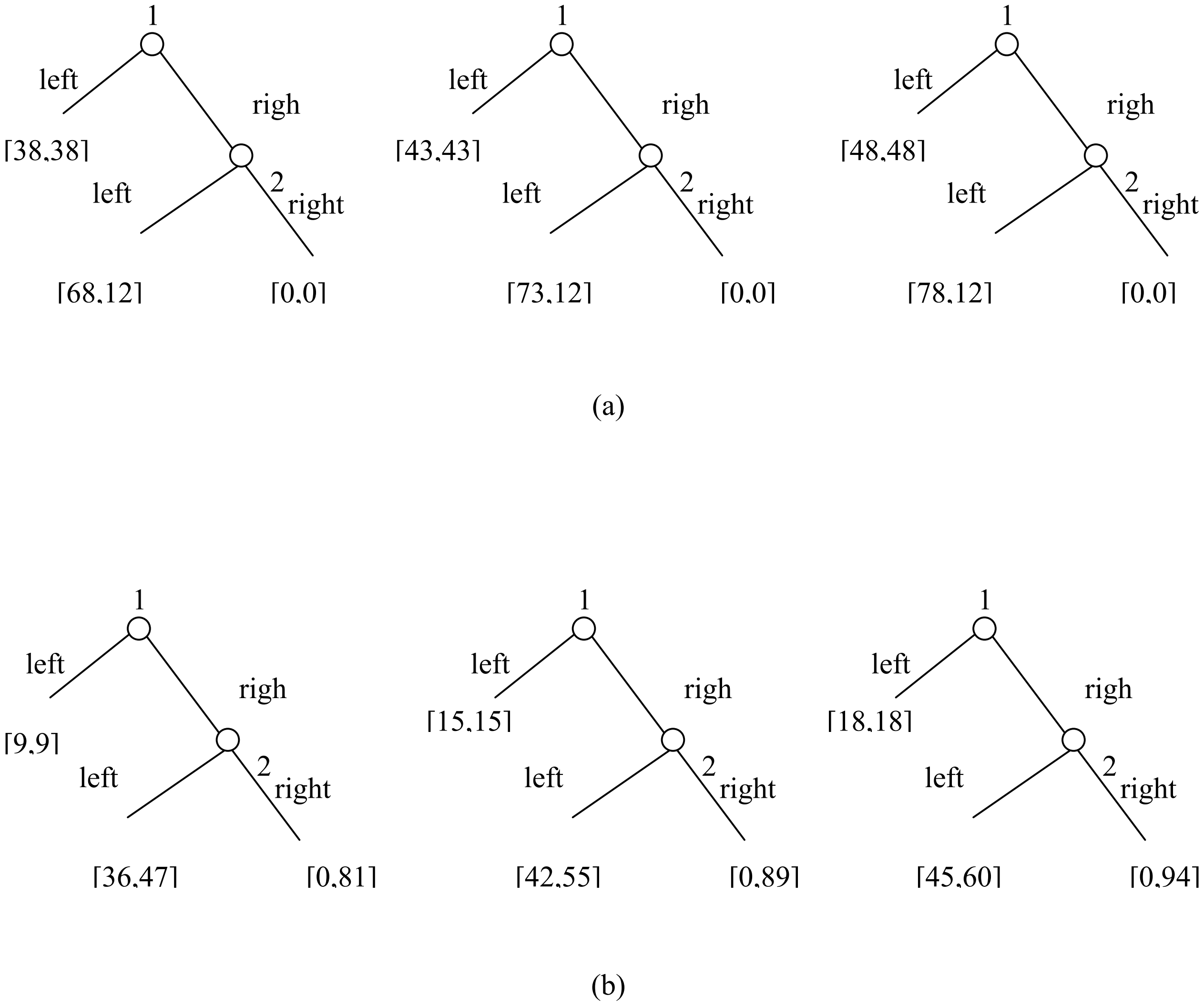


Fig. 1. (a) Sequence of extensive form punishment games: P1, P2, and P3, respectively.
(b) Sequence of extensive form trust games: T1, T2, and T3, respectively.

Note: In each game player 1 moves first by playing "left" or "right". If player 1 moves "left" the game is over. If player 1 moves "right" then player 2 can play "left" or "right", ending the game. Wherever the game ends, player 1 gets the first payoff and player 2 gets the second payoff.

In this sequence the cost of punishing (i.e., the difference between the second player's payoff of moving left and moving right) is the same (i.e., is equal to 12 experimental dollars in each game.) It is identical also the improvement in the first player's payoffs for moving right and the second player moving left (i.e., is equal to 30 experimental dollars for each one of the three games of this sequence.) The subgame perfect Nash equilibrium solution is always to move right for the first player and always to play left for the second player. Different behaviors indicate deviation from the standard game theoretical solution. If the first player moves left, ending the game, she behaves non-opportunistically. If this is the case, the second player can compare the obtained outcome with the worse outcomes she would have obtained if the first player had moved right (downward counterfactual.) This comparison would determine a positive emotion (elation) as a response to the first player's behavior. If the first player moves right, playing opportunistically, the second player would experience worse outcomes. In this case, the result of upward counterfactual induces anger.

The second sequence of games is composed by three trust games. In the trust game the first player can move left, ending the game with a small payoff for both players, or she can move right, giving to the second player the opportunity to choose between reciprocating and defecting. If the second player chooses left, she and the first player get more than the first possible outcome; whereas the game ends with her maximum payoff and with a zero payoff for the first player if she chooses right. In this game if the first player moves right, she trusts the second player to reciprocate (i.e., not to move right). The subgame perfect Nash equilibrium solution is always to move left for the first player and always to move right for the second player. In the sequence of trust games (see Fig.

1b) the potential benefit (in terms of increasing payoffs if the second player reciprocates) for the first player to move right (trust) is the same for the three games (i.e., is always equal to 27.) The difference in payoffs for the second player between moving right (defect) and moving left (reciprocate) is always equal to 34 experimental dollars. What varies through the sequence is the difference between the second player's payoff when she moves left (reciprocate) and her payoff when the first player has ended the game, moving left (these differences are: 38, 40, and 42 for the first, second and third trust game, respectively.) In other words, what varies is the benefit that the second player gets from the first player's trusting behavior. In this way first and second players' patterns of behavior through the three games are explained by their motivation in terms of reciprocity.

In the sequence of trust games the first player can experience anger from the second player's defecting behavior. If the second player goes right, the first player gets a payoff of zero experimental dollars. In this case, and also in the case in which she cannot trust the second player (she moves left), she might compare her payoff with a better non-obtained payoff (upward counterfactual.) When the first player moves right and second player moves left, the first player would compare her payoff with worse payoffs (downward counterfactual.) This would determine elation and appreciation of the reciprocal behavior of the second player.

The last part of the analysis focuses on the relationship between the two sequences. We assume that subjects are carrying over emotional experience from one sequence to another, and different emotional experiences determine different behaviors. For this reason we introduce two conditions. The subjects in condition 1 (punishment-trust) play first the three punishment games and then the three trust games; whereas in condition 2 (trust-punishment) subjects play first the three trust games and then the three punishment games. The matching protocol for condition 2 is described in Fig. 2b. This protocol is identical to the first one with the order of the two sequences inverted. Figs. 2a and 2b show that the position (first or second mover) of the players in the sequences is the same, and it is alternating, i.e., the subjects that were players one in the first sequence are second players in the second sequence, and the opposite for the others.

Position \ Type of Game	SEQUENCE 1			SEQUENCE 2		
	P1	P2	P3	T1	T2	T3
Player 1 Plaver 2	ID#1 ID#2	ID#1 ID#4	ID#1 ID#6	ID#2 ID#1	ID#4 ID#1	ID#6 ID#1
Player 1 Plaver 2	ID#3 ID#4	ID#3 ID#6	ID#3 ID#2	ID#4 ID#3	ID#6 ID#3	ID#2 ID#3
Player 1 Plaver 2	ID#5 ID#6	ID#5 ID#2	ID#5 ID#4	ID#6 ID#5	ID#2 ID#5	ID#4 ID#5

(a)

Position \ Type of Game	SEQUENCE 1			SEQUENCE 2		
	T1	T2	T3	P1	P2	P3
Player 1 Plaver 2	ID#1 ID#2	ID#1 ID#4	ID#1 ID#6	ID#2 ID#1	ID#4 ID#1	ID#6 ID#1
Player 1 Plaver 2	ID#3 ID#4	ID#3 ID#6	ID#3 ID#2	ID#4 ID#3	ID#6 ID#3	ID#2 ID#3
Player 1 Plaver 2	ID#5 ID#6	ID#5 ID#2	ID#5 ID#4	ID#6 ID#5	ID#2 ID#5	ID#4 ID#5

(b)

Fig. 2. (a) Matching Protocol. Condition 1. (b) Matching Protocol. Condition 2.

We formulate two behavioral hypotheses related with the difference between the two conditions.

Hypothesis 1 *The proportion of punish behavior of the second players in condition 2 is higher than in condition 1.*

Hypothesis 2 *The proportion of trust behavior of the first players in condition 1 is lower than in condition 2.*

These two hypotheses are based on considerations about subjects' emotional experience through the experiment. Having experienced opportunistic behavior from the others increases negative reciprocity, and decreases trust in positive reciprocity.

3 Experimental Details

Subjects do not know the matching protocol implemented. They know that they are always playing with a different opponent from game to game, but they do not know with whom they are playing. Anonymity is maintained; subjects are paid separately and with a double blind procedure. The experimenter knows just the ID number of the subject for data analysis purposes, but does not know the corresponding name. This procedure is made clear through the instruction. The games are not labeled. The subjects are informed on their position before playing a game (e.g. first player or second player). The subjects see all the payoff information, and players 2 see the move of players 1 before making a decision.

The subjects of our experiment were 78 undergraduate students from the University of Arizona. 48 subjects participated in condition 1, and 30 subjects participated in condition 2. They received 5 US dollars show up fees and two of the payoffs obtained during the experiment. The two payoffs, one from each sequence, were randomly chosen at the end of the experiment. This procedure has been introduced in order to avoid endowment effects. The payoffs were converted at a rate of 10 Experimental Dollars to 1 US dollar. The experiment was computerized.

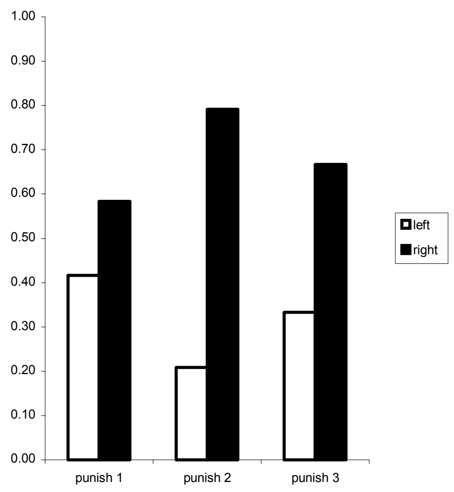
4 Results

4.1 Means choice proportion: Punishment games

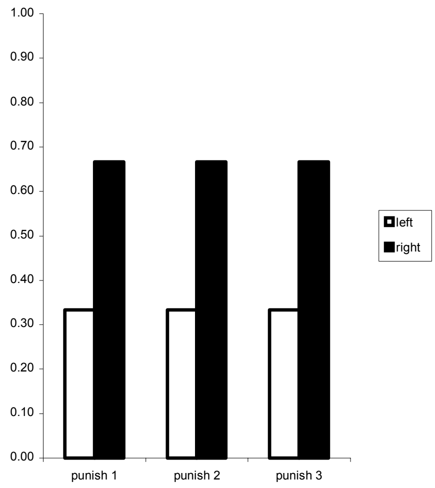
The result of the experiment shows that 68 percent of the first players in the punishment games had opportunistic behavior, i.e. they moved right; and 25 percent of second players responded punishing them (moving right). The mean proportion of opportunistic behavior of the first players in the punishment games does not change significantly along the sequence (see Figs. 3a and 3b.) This is true in both conditions (McNemar test, P1 vs. P2, p-value=0.23; P1 vs. P3, p-value=0.79; and P2 vs. P3, p-value=0.51). As shown in Figs. 3c and 3d, the proportion of punishment increases along the sequence (Kruskal-Wallis test=2.61, p-value=0.27). In particular, there is a significant increase from the first to the second punishment game, and then there is a drop of punishment in the last game, this proportion being anyway higher than the first game.

4.2 Means choice proportion: Trust games

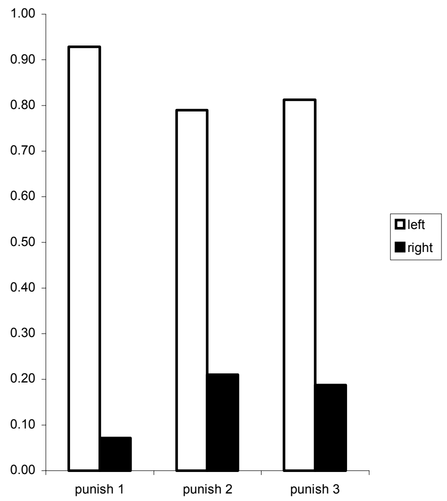
In the trust games 50 percent of the first players trusted the second players, and 45 percent of them reciprocated. Figs. 4a and 4b show mean choice proportions of first players for the sequence of trust games in condition 1 and condition 2, respectively. The



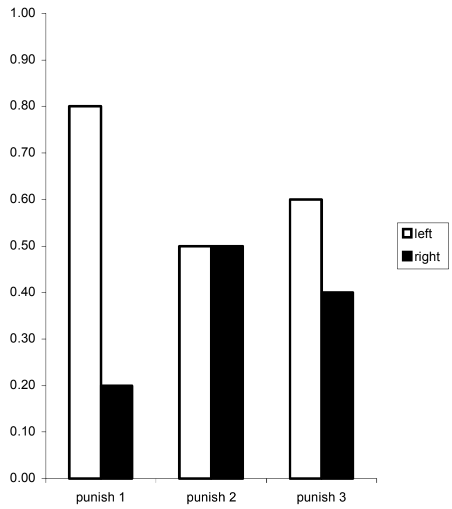
(a) First players in condition 1



(b) First players in condition 2

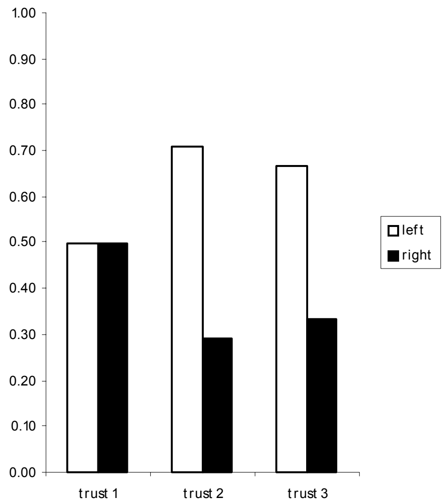


(c) Second players in condition 1

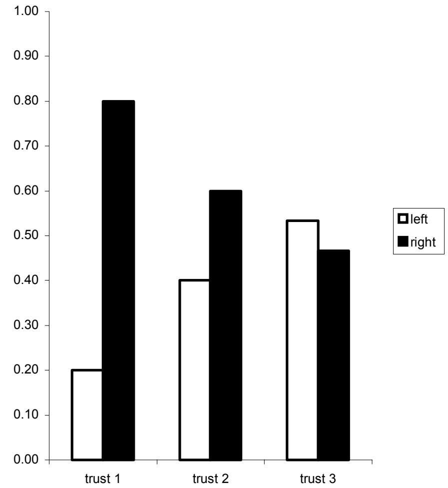


(d) Second players in condition 2

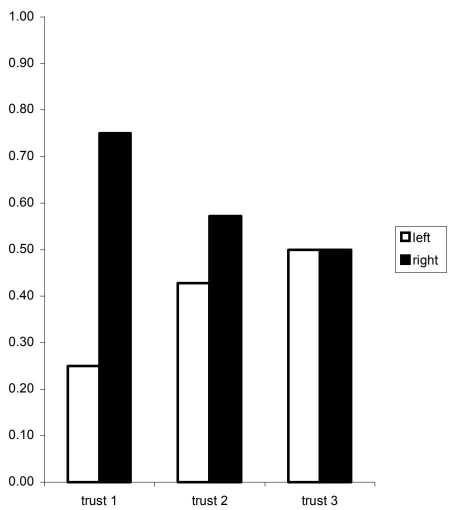
Fig. 3. Mean choice proportion of punishment games: (a) First players in condition 1; (b) First players in condition 2; (c) Second players in condition 1; (d) Second players in condition 2.



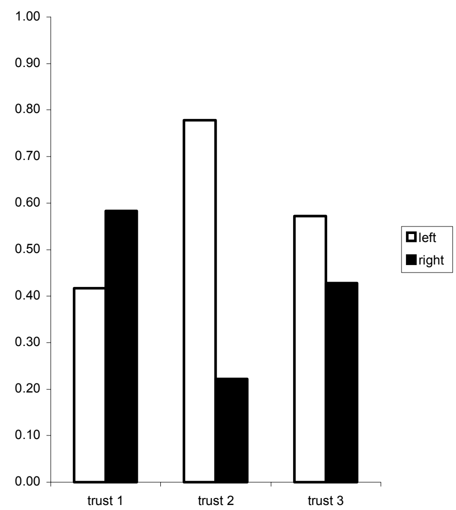
(a) First players in condition 1



(b) First players in condition 2



(c) Second players in condition 1



(d) Second players in condition 2

Fig. 4. Mean choice proportion of trust games: (a) First players in condition 1; (b) First players in condition 2; (c) Second players in condition 1; (d) Second players in condition 2.

level of trust significantly decreases from the first game of the sequence in both conditions (McNemar test, T1 vs. T2, p-value=0.02; T1 vs. T3, p-value=0.03), but does not varies from the second to the third game of the sequence (McNemar test, p-value=1). Figs. 4c and 4d show a significant increase in the proportion of second players' reciprocal behavior (moving left) along the sequence of trust games (Kruskal-Wallis test=3.51, p-value=0.17).

4.3 Interaction between first and second players' behavior

Table 1 presents a contingency table with the first players' behavior (rows) in the punishment game conditioned to the second players' earlier moves (columns). The chi-square test shows that we can reject (p-value=0.0025) the hypothesis of independence between rows and columns of this contingency table, i.e. first players' behavior depends from past (t-1) second players' behavior[9]. In particular the first players are responding to the second players' punishment behavior. Experiencing punishment significantly decreases opportunistic behavior.

Table 1

		Players 2	
		Not-punish (t-1)	**Punish (t-1)**
Players 1	**Not-opportunistic (t)**	3	6
	Opportunistic (t)	38	6

Chi-square=9.16
p-value=0.0025

Contingency table. The entries are numbers of subjects. Punishment games.
First players' behavior (rows) conditional to past (t-1) second players' behavior (columns). Note: Not-opportunistic=player 1 moves left at time t, opportunistic=player 1 moves right at t; not-punish=player 2 moved left at t-1, punish=player 2 moved right at t-1.

[9] The same test over the first players' behavior in the punishment game conditional to second players' past (t-2) behavior is not significant (Chi-square=1.28 , p-value=0.25).

Table 2 shows the result from the second players' behavior (rows) contingent to first players' behavior at time t-1 (columns). The second players responded to first players' opportunistic behavior (chi-square=9.16, and p-value=0.0025). In particular there is a significant increase in punishment as a response to past opportunistic behavior. It is interesting to notice that second movers never punished if they had experienced not-opportunistic behavior at time t-1.

Table 2

		Players 1	
		Not-opportunistic (t-1)	**Opportunistic (t-1)**
Players 2	**Not-punish (t)**	18	20
	Punish (t)	0	15

Chi-square=8.71
p-value=0.003

Contingency table. The entries are numbers of subjects. Punishment games. Second players' behavior (rows) conditional to past (t-1) first players' behavior (columns). Note: Not-punish=first player moves left at time t, punish=first player moves right at t; not-opportunistic=first player moved left at time t-1, opportunistic=first player moved right at t-1.

Table 3 shows the behavior of first players in the trust games. The contingency table represents first players' trusting behavior (rows) contingent to second players' reciprocal behavior at time t-1 (columns). We reject the hypothesis of independence between rows and columns (p-value=0.013), i.e. trust depends on earlier second movers' reciprocal behavior.[10]

[10] The same test over the first players' behavior in the trust game conditional to second players' past (t-2) behavior is not significant (Chi-square=1.42 , p-value=0.23).

Table 3

Players 1		Players 2	
		Reciprocate (t-1)	**Defect (t-1)**
	Not-trust (t)	2	11
	Trust (t)	17	10

Chi-square=6.17
p-value=0.013

Contingency table. The entries are numbers of subjects. Trust games.
First players' behavior (rows) conditional to past (t-1) second players' behavior (columns). Note: Not-trust=first player moves left at time t, trust=first player moves right at time t; reciprocate=second player moved left at time t-1, defect=second player moved right at time t-1.

Table 4 illustrates second players' behavior in the trust game. The chi-square test cannot reject the hypothesis of independence between the rows (second players' reciprocal behavior at time t) and the columns (first players' trusting behavior at time t-1). This indicates that second players in the trust games were no more likely to reciprocate if they have been trusted in earlier period.

Table 4

Players 2		Players 1	
		Not-trust (t-1)	**Trust (t-1)**
	Reciprocate (t)	4	13
	Defect (t)	7	6

Chi-square=1.75
p-value=0.1851

Contingency table. The entries are numbers of subjects. Trust games.
Second players' behavior (rows) conditional to past (t-1) first players' behavior (columns). Note: Reciprocate=second player moves left at time t, defect=second player moves left at time t; not-trusted=first player moved left at time t-1, trusted=first player moved right at time t-1.

4.4 Condition 1 vs. Condition 2. Support for hypotheses 1 and 2: Sequence matters

Table 5 reports Wilcoxon rank-sum test for differences of means of subjects' behavior in condition 1 and condition 2. The first line shows no significant difference between the mean proportions of first players moving left (not-opportunistic behavior) in the punishment games of condition 1 and condition 2 (p-value=0.79). The second line shows that punishment in condition 2 is higher than in condition 1 (p-value=0.09), this result is in support of hypothesis 1. The third line shows that trust is significantly higher in condition 2 than in condition 1 (p-value=0.049), in support of hypothesis 2. Finally, the fourth line shows no significant difference between reciprocity in condition 2 and condition 1 (p-value=0.218). These results show that having experienced a sequence of punishment games as second player or a sequence of trust games as first player affect behavior. In particular, it increases subject's sensitivity to other subjects' behavior. They tend to attribute to the others more responsibility and intentionality for their actions.

Table 5
Wilcoxon rank-sum test for differences between samples means proportion (μ) of behaviors in condition 1 (C1) and condition 2 (C2).

Samples	μ_{C1}	μ_{C2}	Normal statistic with correction, z	p-value
First players non-opportunistic behavior in punishment games	0.32	0.33	-0.26	0.79
Second players punishment behavior in punishment games	0.18	0.32	-1.68	0.093
First players trust behavior in trust games	0.38	0.62	-1.96	0.049
Second players reciprocal behavior in trust games	0.35	0.54	-1.23	0.218

4.5 Test for alternative explanations: Efficiency and inequality aversion

Our design allows us to distinguish consistent patterns of behavior in terms of efficiency (desire to maximize the total payoffs)[11] and inequality aversion[12]. As shown from the individual data (see Appendix A1, and A2), there is no significant evidence of consistent patterns of behavior motivated by efficiency and inequality aversion. Only five subjects followed the "efficiency pattern" in both sequences (ID# 39, 41, and 44, in condition 1; ID# 3, and 13, in condition 2), and only one subject behaved consistently with a pattern of inequality aversion (ID# 32, in condition 1). Table 6 reports relative frequencies of patterns of behavior in the sequence of punishment and trust games, respectively. The patterns of behavior motivated by efficiency or inequality aversion are never relatively more important than other behaviors.[13]

Table 6.
Relative frequencies of different patterns of behavior

Sequence	position		Patterns of behavior	Condition 1 (N=48)	Condition 2 (N=30)	Total
Punishment games	player 1	E.	R L L	0/24=0	2/15=0.13	2/24=0.08
		I.A.	L R R	2/24=0.08	1/15=0.07	3/39=0.08
	player 2	E.	l l l, l l X, l X l, X l l	13/19=0.68	5/13=0.38	18/32=0.56
		I.A.	r r r, r r X, r X r, X r r	1/19=0.05	0/13=0	1/32=0.03
		other	r l l, l r l, l l r, r l X, r X l, X r l, l r X, l X r, X l r	5/19=0.26	8/13=0.62	13/32=0.41
Trust games	player 1	E.	R R R	6/24=0.25	5/15=0.33	11/39=0.28
		I.A.	L L L	11/24=0.46	0/15=0	11/39=0.28
	player 2	E. or I.A.	l l l, l l X, l X l, X l l	2/7=0.29	3/10=30	5/17=0.29
		other	r l l, l r l, l l r, r l X, r X l, X r l, l r X, l X r, X l r, r r r, r r X, r X r, X r r	5/7=0.71	7/10= 0.7	12/17=0.71

Note: E.= Efficiency, I.A.= Inequality Aversion. L=left, R=right, player 1. l=left, r=right, player 2. X=no observation.

[11] See Charness and Rabin (2002).
[12] We have the same pattern for the two theories on inequality aversion: Bolton-Ockenfels (2000), and Fehr-Schmidt (1999).
[13] With the exception of the case of second players in the punishment games in condition 1, in which the relative frequency of efficiency based behavior is 0.68.

5 Discussion

In our experiment subjects played two sequences of two-person extensive form games with anonymous and different counterparts. The payoffs structure of the games determined underlying reciprocity responses. The anonymity and the changing of counterparts avoided reputation effects. This procedure should have induced the subjects to play without changing their behaviors through the games, as if they were playing one-shot game each time. This was not the case. Results show subjects' sensitivity to the change (introduced by the experimental design) of the levels of the determinants of the "reciprocal interaction." In the sequence of punishment games we increased through the three games the level of the potential "pain" inflicted by the first movers to the second movers. In the sequence of trust games we increased through the three games the second movers' benefits determined by the first movers' trusting behavior. Results show subjects' sensitivity to these changes. The second players in the punishment games punished more through the sequence; and the second players in the trust games defected less through the sequence. This corroborates previous findings by Cooper and Kraker Stockman (2002) that: subjects learn to punish over time. Subjects do not show fixed preferences through the sequences of our experiment. Indeed, subjects' preferences depend on other players' behavior, and on the interpretation of these behaviors in terms of their intentions (in analogy to the findings of Brandts and Solà, 2001; and Falk et al., 1999.)

The second part of our analysis is dedicated to the study of different behaviors between the two conditions of our experimental design. The only difference between condition 1 and condition 2 is the order of the sequences played, i.e., in condition 1 the subjects played first the three punishment games and then the three trust games, whereas, in condition 2 the subjects played first the three trust games and then the three punishment games. The order and the payoffs structure of the games inside each sequence is the same in both conditions. Results show that the subjects (second players) in condition 2 punished more than the subjects in condition 1, and that the subjects in condition 2 trusted more than the subjects in condition 1. We conclude that different experiences determine different behaviors. The second players in condition 1 experienced a sequence of increased pain (inflicted by the first players), before they became first players in the sequence of trust games. They trusted less than the first players in condition 2. The first players that started with the sequence of trust games experienced a high level of defection before they became second players in the sequence of punishment games. They punished more than the second players in the punishment games in condition 1. The negative emotion caused by having experienced defection (from others) is carried over from one sequence to another. Subjects compare their obtained outcomes with the outcome they would have obtained if their counterpart behaved non-opportunistically. This comparison, based on upward counterfactual, determines "negative" emotional response that "influences on subsequent judgments, making salient the role of other people in causing negative events and perhaps events in general" (cf. Keltner et al., 1993, pp.751). We named this response as reciprocity-based anger. Experiencing anger determines punishment and non-trusting behaviors.

Negative reciprocity is triggered by emotional response to others' behavior. In this sense emotion contributes to the enforcement of social norms (see Gintis, 2000; Bowles and Gintis, 2001; Fehr and Gächter, 2002; and Carpenter et al., 2002.)[14] Our results show that subjects are willing to enforce the social norm of cooperation even if this process is based on costly punishment.

We conclude that different behaviors are determined by different emotional experience. Indeed, experienced emotions matter in reciprocal interactions. As stated by Bowles and Gintis (2001): “Adherence to social norms is underwritten by emotions, not only by the expectation of future reciprocation.”

[14] We adopt the definition of social norm given by Fehr and Gächter (2000 b): “It is: 1) a behavioral regularity; that is 2) based on a socially shared belief how one ought to behave; which triggers 3) the enforcement of the prescribed behavior by informal social sanctions.” (cf. pp.9)

References

Abbink, K., Bolton, G.E., Sadrieh, A., and Tang, F-F (2001). Adaptive Learning versus Punishment in Ultimatum Bargaining. *Games and Economic Behavior*, 37, 1-25.

Andreoni, J. (1987). Why Free Ride? Strategies and Learning in Public Goods Experiments. Discussion Paper University of Wisconsin, Department of Economics.

Axelrod, R. (1984). The Evolution of Cooperation. New York, Basic Books.

Axelrod, R., and Hamilton, W.D. (1981). The Evolution of Cooperation in Biological Systems. *Science*, 211, 1390-1396.

Berg, J., Dickhaut, J., and McCabe, K.A. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10, 122-142.

Bolton, G. (1997). The rationality of splitting equally. *Journal of Economic Behavior and Organization*, 32, 365-381.

Bolton, G., Brandt, J., and Ockenfels, A. (1998). Measuring Motivation for the Reciprocal Responses Observed in a Simple Dilemma Game. *Experimental Economics*, 1, 207-219.

Bolton, G.E., and Ockenfels, A. (2000). ERC-A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90, 166-93.

Bowles, S., and Gintis, H. (1999). The Evolution of Strong Reciprocity. Mimeo, University of Michigan.

Bowles, S., and Gintis, H. (2001). The economics of shame and punishment. Working papers.

Brandts, J., and Solà, C. (2001). Reference Points and Negative Reciprocity in Simple Sequential Games. *Games and Economic Behavior*, 36, 138-157.

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

Carpenter, J.P., Matthews, P.H., Ong'ong'a, O. (2002). Why Punish? Social Reciprocity and the Enforcement of Prosocial Norms. Mimeo.

Connolly, T., Ordoñez, L.D., and Coughlan, R. (1997). Regret and responsibility in the evaluation of decision outcomes. *Organizational Behavior and Human Decision Processes*, 70, 73-85.

Cooper, D.J., and Kraker Stockman, C. (2002). Learning to Punish: Experimental Evidence from a Sequential Step-Level Public Good Game. *Experimental Economics*, 5, 39-51.

Coricelli, G., McCabe, K.A., and Smith, V.L. (2000). Theory-of-mind mechanism in personal exchange. In Hatano, et al. (Eds.), Affective Mind. Elsevier Science Publisher, 250-259.

Dawes, R.M., and Thaler, R. (1988). Anomalies Cooperation. *Journal of Economic Perspectives*, 2, 187-197.

Falk, A., Fehr, E., and Fischbacher, U. (2000). Testing Theories of Fairness – Intentions Matter. Working Paper, University of Zurich.

Falk, A., Fehr, E., and Fischbacher, U. (1999). On the Nature of Fair Behavior. WP Institute for Empirical Research in Economics, University of Zurich.

Fehr, E., and Gächter, S. (2000 a). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, 14, 159-181.

Fehr, E., and Gächter, S. (2000 b). Cooperation and Punishment in Public Goods Experiments. *American Economic Review*, 980-994.

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans. *Nature*. 415, 137-140.

Fehr, E., and Schmidt, K. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114, 817-68.

Frijda, N.H. (1986). The Emotions. Cambridge, England: Cambridge University Press.

Frijda, N.H., Kuipers, P., and ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*. 57, 212-228.

Gale, J., Binmore, K.G., and Samuelson, L. (1995). Learning to be imperfect: The ultimatum game. *Games and Economic Behavior*, 8, 56-90.

Gintis, H. (2000). Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology*, 206, 169-179.

Güth, W., and Yaari, M. (1992). An evolutionary approach to explain reciprocal behavior in a simple strategic game. In Witt, U. (Eds.), Explaining Process and Change-Approaches to Evolutionary Economics, Ann Arbor, MI: University of Michigan Press, 23-34.

Kahneman, D., and Miller, D. (1986). Norm Theory: Comparing reality to its alternatives. *Psychological Review*. 93, 136-153.

Levine, D. (1998). Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics*, 1, 593-622.

Lewis, D. (1969). Convention: A Philosophical Study. Cambridge, Mass.: Rutgers University Press.

Lewis, D. (1973). Counterfactuals Harvard. University Press. Cambridge, Massachusetts

Loomes, G., and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92, 805-824.

Loomes, G., and Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *Review of Economic Studies*, 53, 271-282.

McCabe, K.A., and Smith, V.L. (2000). A comparison of naïve and sophisticated subject behavior with game theoretic predictions. *Proceedings National Academy of Sciences*, 97, 3777-3781.

Metcalfe, J., and Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological Review*, 106, 3-19.

Niedenthal, P.M., Tangnez, J.P., and Gavanski, I. (1994). "If only I weren't" versus "if only I hadn't": Distinguishing shame and guilt in counterfactual thinking. *Journal of Personality and Social Psychology*. 67, 585-595.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review*, 83, 1281-1302.

Roese, N.J., and Olson, J.M. (1995). What might have been: The social psychology of counterfactual thinking. Mahwah, NJ: Erlbaum .

Roth, A.E., Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8, 164-212.

Schelling, T. (1984). Self-Command in Practice, in Policy, and in a Theory of Rational Choice. *American Economic Review*, 1-11.

Sugden, R. (1986). The Evolution of Rights, Cooperation and Welfare. New York: Basil Blackwell.

Trivers, R.L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.

Young, H.P. (1993 a). The evolution of conventions. *Econometrica*, 61, 57-84.

Young, H.P. (1993 b). An Evolutionary Model of Bargaining. *Journal of Economic Theory*, 59, 145-168.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*. 12, 93-106.

Zeelenberg, M. van Dijk, W.W., and Manstead, A.S.R. (2000). Regret and Responsibility Resolved? Evaluating Ordoñez and Connolly's (2000) Conclusions. *Organizational Behavior and Human Decision Processes*, 81, 143-154.

Appendix A1

Individual data. Condition 1

	session 1				session 2			
ID	position	P1	P2	P3	position	T1	T2	T3
1	1	R	R	R	2	r	X	r
2	1	L	L	L	2	X	l	l
3	1	R	R	R	2	l	X	X
4	1	L	R	R	2	X	X	r
5	1	R	R	R	2	r	X	X
6	1	L	R	R	2	X	X	X
7	1	L	R	R	2	r	l	X
8	1	R	R	L	2	X	X	X
9	1	L	L	L	2	r	r	l
10	1	L	R	R	2	l	l	X
11	1	L	R	L	2	X	X	X
12	1	R	R	R	2	X	r	X
13	1	L	L	L	2	X	X	l
14	1	R	R	R	2	X	X	X
15	1	R	R	R	2	r	X	X
16	1	R	R	L	2	r	X	X
17	1	R	R	L	2	l	X	r
18	1	L	L	R	2	r	r	r
19	1	R	R	R	2	r	X	X
20	1	R	R	R	2	X	r	X
21	1	R	R	R	2	X	X	X
22	1	R	R	R	2	X	X	l
23	1	L	L	L	2	r	X	X
24	1	R	R	R	2	X	X	X
25	2	l	r	X	1	R	L	R
26	2	X	X	r	1	L	L	L
27	2	l	r	X	1	R	R	L
28	2	X	l	l	1	L	L	L
29	2	l	l	X	1	R	R	R
30	2	X	l	l	1	L	L	L
31	2	X	X	l	1	R	L	L
32	2	r	r	X	1	L	L	L
33	2	X	l	X	1	R	L	L
34	2	X	X	l	1	R	R	R
35	2	X	l	l	1	L	L	R
36	2	l	l	l	1	L	L	L
37	2	X	l	r	1	L	L	L
38	2	l	l	l	1	L	L	L
39	2	l	l	X	1	R	R	R
40	2	l	l	l	1	L	L	L
41	2	l	X	l	1	R	R	R
42	2	X	l	X	1	R	R	R
43	2	l	l	r	1	R	L	L
44	2	l	l	l	1	R	R	R
45	2	l	X	l	1	L	L	L
46	2	l	l	l	1	L	L	L
47	2	X	l	l	1	L	L	L
48	2	l	r	X	1	R	L	L

Note: L=left, R=right, player 1. l=left, r=right, player 2. X=no observation.

Appendix A2
Individual data. Condition 2

	sequence 1				sequence 2			
ID	position	T1	T2	T3	position	P1	P2	P3
1	1	R	R	R	2	X	X	l
2	1	R	L	L	2	r	r	l
3	1	R	R	R	2	l	X	l
4	1	L	L	R	2	l	X	l
5	1	R	R	R	2	l	r	X
6	1	R	L	L	2	X	l	r
7	1	R	R	L	2	l	r	X
8	1	R	R	R	2	X	l	r
9	1	L	L	L	2	r	l	r
10	1	R	R	L	2	l	r	r
11	1	R	L	L	2	l	r	X
12	1	R	L	L	2	l	X	l
13	1	R	R	R	2	X	l	l
14	1	R	R	L	2	l	l	X
15	1	L	R	R	2	X	X	X
16	2	l	X	X	1	L	R	R
17	2	r	X	X	1	R	L	L
18	2	l	l	l	1	R	R	R
19	2	X	r	X	1	R	R	R
20	2	r	l	X	1	R	L	R
21	2	r	l	X	1	L	L	L
22	2	l	X	X	1	R	R	R
23	2	l	l	l	1	L	L	R
24	2	X	X	X	1	R	L	L
25	2	r	l	r	1	R	R	L
26	2	r	X	l	1	R	R	R
27	2	r	r	X	1	R	R	R
28	2	l	l	r	1	L	R	R
29	2	r	X	r	1	R	R	R
30	2	X	l	l	1	L	R	L

Note: L=left, R=right, player 1. l=left, r=right, player 2. X=no observation.