The Journal of Socio-Economics 38 (2009) 871–877

Contents lists available at ScienceDirect

The Journal of Socio-Economics

journal homepage: www.elsevier.com/locate/soceco

Reputational cues in repeated trust games

Riccardo Boero[a], Giangiacomo Bravo[b,*], Marco Castellani[c], Flaminio Squazzoni[c]

[a] *Dipartimento di Scienze Economiche e Finanziarie "G. Prato", Università di Torino and GECS - Research Group in Experimental and Computational Sociology, Italy*
[b] *Dipartimento di Scienze Sociali, Universita di Torino and GECS - Research Group in Experimental and Computational Sociology, Italy*
[c] *Dipartimento di Studi Sociali, Università di Brescia and GECS - Research Group in Experimental and Computational Sociology, Italy*

ARTICLE INFO

Article history:
Received 17 September 2008
Received in revised form 21 April 2009
Accepted 3 May 2009

JEL classification:
A14
C72
C92

Keywords:
Reputation
Trust game
Cognitive mechanisms
Rating mechanism

ABSTRACT

The importance of reputation in human societies is highlighted both by theoretical models and empirical studies. In this paper, we have extended the scope of previous experimental studies based on trust games by creating treatments where players can rate their opponents' behavior and know their past ratings. Our results showed that being rated by other players and letting this rating be known are factors that increase cooperation levels even when rational reputational investment motives are ruled out. More generally, subjects tended to respond to reputational opportunities even when this was neither rational nor explainable by reciprocity.

1. Introduction

Reputation is of paramount importance in creating and maintaining prosocial behaviors in large human groups. Evidence of this can be found both in theoretical models (e.g. Bravo and Tamburino, 2008; Nowak and Sigmund, 1998a,b; Pollock and Dugatkin, 1992; Raub and Weesie, 1990) and in empirical studies (see below). From an evolutionary perspective, the importance of building and maintaining good reputational standing is usually linked either to indirect reciprocity strategies (Boyd and Richerson, 1989; Leimar and Hammerstein, 2001; Nowak and Sigmund, 1998b, 2005; Panchanathan and Boyd, 2004) or with costly signaling motives (Gintis et al., 2001; Henrich et al., 2006; Zahavi and Zahavi, 1997). In both cases, reputation-based behavior tends to increase altruism and cooperation in large groups of unrelated individuals where other mechanisms, such as kin selection or direct reciprocity, are unlikely to succeed.

A wide range of empirical work highlights the key role of reputation in influencing human behavior. In an experiment based on Nowak and Sigmund (1998b) indirect reciprocity model, Seinen and Schram (2006) found that the simple possibility of knowing the past behavior of other individuals has a dramatic effect on cooperation. In their experiment, subjects interacted either in the position of "donors" or of "recipients". Each donor faced the decision of whether to pay or not a cost b in order to provide a benefit $c > b$ for a recipient.

The experiment showed that the donors' decisions are strongly influenced by the possibility of knowing the recipients' past behavior, with a helping rate of 70% when this was allowed *versus* 22% when this was not possible. The authors explained the outcome as an effect of the sensitivity of donors to the recipients' "social status" (i.e. reputation). A comparable effect of reputation has been found also in experiments based on the ultimatum game (Fehr and Fischbacher, 2003), on trust games (Keser, 2003), on various combinations of public good and indirect reciprocity games (Milinski et al., 2002a,b, 2006) and on a combination of public good and trust games (Barclay, 2004).

It is worth noting that reputation-based behavior does not necessarily need to be grounded on the direct knowledge of past behavior of other individuals, since humans can exploit the many sources of information available in social life. For instance, gossip plays an important role in transmitting reputational information in human societies (see Dunbar, 1996). A recent study showed that gossip has a strong influence on the behavior of experimental subjects even when they are also able to use other sources of information, including direct observation (Sommerfeld et al., 2007).

Another experiment showed that individuals react to the possibility of being the subject of gossip by increasing their

[*] Corresponding author. Tel.: +39 011 670 26 35 fax: +39 011 670 26 12.
E-mail address: giangiacomo.bravo@unito.it (G. Bravo).

1053-5357/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.socec.2009.05.004

contributions in a dictator game (Piazza and Bering, 2008). The relevance of direct observations and gossip for the sustainability of cooperation in large social groups has been also emphasized in simulation models (e.g. Conte and Paolucci, 2002).

If reputation (*via* gossip or direct observation) plays such an important role, it is likely that humans possessed a high sensitivity for anything that could lead to a change in their reputational status. This sensitivity is probably an adaptation to the life of hunter-gatherer groups that represented the original adaptive environment of our species, where the actions of each individual were easily observable and deeply influenced the behavior of the other group members and, more generally, cooperation inside each group (Alexander, 1987; Barkow et al., 1992; Fehr and Fischbacher, 2003; Milinski and Rockenbach, 2007; Nowak and Sigmund, 2005; Richerson et al., 2003). Coherently with this approach, our general hypothesis is that human beings should react to reputational opportunities even when this is barely rational in terms of future expected utility. More specifically, it is necessary to distinguish between a strategic reputation building (the one that clearly occurs in the experiments quoted above) and a deeper and more cognitive response to situation where other people are in the condition to observe and judge one's actions. While the former, based on rational calculus, is undoubtedly important in many situations, we argue that the latter is deeply rooted in social interaction mechanisms and produces non-trivial effects on human behavior.

The interplay of the two mechanisms is probably akin to that observed in a few studies where the mental processes of the subjects were monitored through functional magnetic resonance imaging. Especially interesting is a recent experiment realized by Hsu et al. (2008) that shows the activation of different brain regions when subjects face a difficult trade-off between rational considerations based on efficiency motives and a widespread social norm as equity. More specifically, the authors found that a specific brain region (the putamen) responds to efficiency while a second one (the insula) responds to equity. A third region (the caudate/septal subgenual) encodes a unified measure of the two motives above and is probably linked with the resolution of the trade-off. Moreover, a behavioral measure of individual differences in inequity aversion correlates with the activity measured in the equity encoding regions. Following this example, we can imagine similar psychological mechanisms acting on reputation and leading to rational reputation-building actions that are, at least partially, separated from more social cognition driven behaviors.

Indeed, recent experimental work shows the existence of subtle reputation-related cues able to significantly modify individuals' behavior. Those cues are especially linked with the possibility of being observed. For instance, in Haley and Fessler (2005) study, the presence of stylized eyespots on the computer desktops used for the experimental sessions significantly increases the generosity of players in a dictator game despite no differences in actual anonymity. In another work, conducted in a real-world setting, Bateson et al. (2006) found a similar effect of apparently unimportant cues of being watched. Their results show that people put nearly three times as much money in a "honesty box", used to collect money for drinks in a university coffee room, when the cost of the drinks was displayed on a board along with a picture of eyes staring at the consumer than when the notice included a flower control picture.

The effect of being watched is so striking that subjects react also when it is evident that the "observer" is not human: the participants in another experiment contributed significantly more to a public good when a robot picture – that obviously represented a machine but endowed with two large eyes – was placed on their computer desktops (Burnham and Hare, 2007). Overall, the results of those experiments suggest the existence of a cognitive mechanism that enhances cooperation in response to the (actually false) possibility that somebody else is watching the subject's actions, and hence to the possibility of a modification in their reputational status.

Note that individuals not only care about their own reputational status but also react promptly to other reputations. As before, in some situations this may be explained by rational calculus, e.g. when subjects face the decision of trusting another person whose decisions can significantly change their own payoffs. However, the reaction can also be driven from different psychological mechanisms. From this point of view, reputation is especially important in trust situations. An elegant formalization of a trust situation is the trust game (originally called investment game by Berg et al., 1995). In a trust game, Player *A* decides how much of his/her endowment to send (or "invest") to Player *B*, who receives the amount sent multiplied by a given factor greater than one (usually three). Subsequently, *B* decides the proportion of the received amount to return to *A*. Since rationally *B* has no incentive to return anything, the only rational strategy is also for *A* to keep the entire endowment. Empirically, the amounts exchanged by real subjects in a one shot game with no reputation possibilities are usually higher than the amounts predicted by a rational choice perspective. This result is usually explained using direct reciprocity motives (e.g. Berg et al., 1995; Fehr et al., 1998; McCabe and Smith, 2000).

Those experiments involved no reputational motives, but Keser (2003) introduced both repetition and reputation in the trust game. In her design, subjects interacted for a fixed number of rounds and *A* players where allowed to rate the behavior of their opponents. In the next round, the ratings were presented to the new *A* players before they took their decision. The main Keser's result is that reputation increases significantly the overall cooperation levels of the game.

It is worth noting that the introduction of the reputational opportunity in the game has a stronger effect on the proportion of the received amount returned by *B* players (+41.5% in the "short run reputation" treatment in comparison with the baseline game) than on the proportion of the endowment invested by the *A* ones (+31.7%). In other words, the investment in reputation among *B* players plays a stronger effect than the reduction of uncertainty for the *A* ones.

Although there is little doubt that a rational investment in reputation plays an important role in explaining Keser's results, we could suspect that it is not the whole story and that part of the enhancement of *B*'s contributions is due to a different cognitive reaction to a situation where the subject is "under judgment". Unfortunately, Keser's experimental protocol does not permit us to separate the two mechanisms. We therefore designed two experiments to disentangle them by adding a treatment where also *A* players are under judgment and by selectively remove all the rational opportunities of reputational investment.

Our results were overall consistent with the predictions that a significant part of the subjects' behavior is not explicable only as rational investment in reputation. More specifically, besides the obviously strong effect of rational reputation seeking, we found that reputation matters also when the observed behavior can only be based on cognitive mechanisms akin to the ones that become apparent in the "eyespot" experiments. The effects of those mechanisms are less uniformly distributed among subjects, but are still noticeable at the aggregate level. The paper is organized as follows: Section 2 describes the first experimental setting and presents its results; Section 3 presents the setting and the results of the second experiment; Section 4 is devoted to a general discussion of the two experiments.

2. Experiment 1

2.1. Methods

One hundred and twenty subjects participated in the first experiment in three groups of 40 individuals. Subjects were students of the University of Brescia – 58 females and 62 males – recruited through public announcements within different faculties. The experiment took place on a single day in the computer laboratory of the Faculty of Economics, which is equipped with the experimental software z-Tree (Fischbacher, 2007). All interactions took place through the computer network and the subjects had no possibility of identifying their counterparts. The experiment used a within-subject design: participants played a trust game for a total of 35 rounds, with the first 10 consisting of a baseline treatment, based on Berg et al. (1995) game, and the last 25 of a different treatment for each group. The subjects were informed in advance of the duration of the game (i.e. the number of periods) of each of the two stages of the experiment. Each group played for about one hour and a half and the average earnings were 14.35 Euros that were paid immediately after the experiment.

The experiment used a "stranger" matching protocol (i.e. random re-coupling in each period) and the players' roles were randomly assigned at the beginning of each period. The sequence of the players' moves during both the baseline (10 periods) and subsequent treatments (25 periods) was as follows: (i) both player *A* (the trustor) and player *B* (the trustee) received an initial endowment of 10 experimental currency units (ECU), with an exchange rate of 1 ECU = 1.5 Euro cents; (ii) player *A* decided his/her investment and the invested amount was tripled and sent to player *B* in addition to his/her own endowment; (iii) *B* chose the amount to return to *A*; (iv) the sums earned by both players in the current period were displayed to the subjects.

After the common baseline, each of the three groups of 40 subjects played a different treatment. The sequence of players' moves in all treatments was as in the baseline except for the introduction of a "rating" stage at the end of each period. In the first treatment (hereafter indicated as B-Rep), the focus was on *B* reputation. This was achieved by allowing *A* to rate *B* behavior as "negative", "neutral", or "positive". Note that, at the rating stage, *A* already knew the sum returned by *B* and that *A* had the possibility to rate *B* only when his/her investment was greater than zero. The subsequent *A* players interacting with *B* were informed of the last rating received by the latter *before* making their investment decision. When a *B* player had not already been rated, e.g. in the first period of the treatment or because he/she not yet played as *B*, an "unknown" rating appeared on *A* decision screen.

The second treatment (hereafter A-Rep) was exactly like the first one except for the fact that *B* was allowed to rate *A*. This information was available in the next period for the subject playing as *B* with *A* players who were already rated. As in previous case, the rating possibilities were "negative", "neutral" and "positive" and this information was presented to the *B* players before their return decision.

In the third treatment (hereafter Both-Rep), both *A* and *B* players were allowed to rate each other and knew this information in the following period before their investment/return decision. The main purpose behind the design of this treatment was to investigate whether the introduction of a two-way reputation system could lead to higher cooperation levels than the sum of the two one-way rating schemes.

2.2. Experiment 1 results

Table 1 presents the overall results of the first experiment. All analysis was conducted on the *R 2.8.1* statistical platform (R Development Core Team, 2008), using the *plm 1.1-1* package for panel models. In all our treatments, the baseline periods produced results that were fully consistent with previous experiments based on the standard repeated trust game (e.g. Berg et al., 1995; Ortmann et al., 2000). This means that the overall experimental setting of our baseline treatment (e.g., the spatial setting of the laboratory, the computer interface, and the actual values of the amounts exchanged) was capable of reproducing the results of a standard trust game and can hence fruitfully be compared with the further treatments that we tested.

Table 1
Average investment and returns by treatment in the first experiment. Standard deviations are in parenthesis.

	A investment (ECU)	*B* return (ECU)
Baseline rounds (all groups)	3.91 (2.67)	3.88 (4.46)
B-Rep	4.39 (2.84)	6.30 (5.23)
A-Rep	5.61 (2.96)	5.28 (5.96)
Both-Rep	5.42 (3.17)	7.25 (6.46)

In order to investigate the effect of the introduction of a reputation system in the game, taking at the same time into account the panel structure of the data, we estimated a fixed effects (or "within-groups") panel regression model for each treatment (see Baltagi, 2001). This is a strategy similar to that used in Barrera and Buskens (in press), although we used a fixed effect model instead of random effect one, which is more appropriate in case of in-group comparisons.

Table 2 presents a summary of the coefficients estimates for the B-Rep treatment model. The treatment effect is significant and positive for both *A* investments and *B* returns, when controlling for the amounts received by players in the previous rounds, for end effects of both the baseline and the treatment periods, and, in the case of *B* returns, for the amount invested by *A* in the current round. The treatment effect for *B* returns is more pronounced than the one for *A* investment, a fact that fits with the idea that being under rating enhances *B* willingness to return high amounts to *A*.

It is worth noting that, while *A* investments are not significantly influenced by any of the control variables included in the model, many of them affect *B* returns. More specifically, Table 2 shows a positive effect of an increase of *A* investment and a strong negative effect of the last period of the treatment. The relation between *A* investments and *B* returns is explicable in terms of reciprocity and has been widely observed in trust games (e.g. Berg et al., 1995). The end effect is however understandable by considering that any further investment of reputation will be lost with the end of the game. Overall, *B* actions appears to be more easily influenced by many details of the game than the *A* ones (notice that the coefficients related to the past amounts received by the current *B* when he/she played in both *A* and *B* positions are also significant). The enhancement in *A* investment compared with the baseline depends instead only on the possibility of discriminating between trustworthy and untrustworthy *B*. A new panel model (excluding the baseline periods) shows that *A* players actively discriminates among the *B* ones depending on their rating (Table 3). A "positive" judgment leads to a significant higher investment relatively to the "unknown" case, while a "negative" one to a significant lower investment. On the other hand, there is no significant difference in investments when *B* has a "neutral" judgment relative to the case where he/she has not yet been rated. Overall, the results of the B-Rep treatment show a strong effect of reputation on player's action and are consistent

Table 2
Coefficient estimations for the B-Rep treatment model.

	Dependent: *A* investment		Dependent: *B* return	
Treatment effect	0.740	(0.217)***	2.141	(0.281)***
Previous return received when playing as *A*	0.013	(0.020)	0.071	(0.026)**
Previous investment received when playing as *B*	0.017	(0.037)	0.102	(0.043)*
Last treatment period	−0.091	(0.539)	−5.491	(0.687)***
Last baseline period	−0.451	(0.557)	−0.945	(0.706)
Current A investment	–	–	1.048	(0.043)***
F	3.855**		143.055***	

Standard errors are in parenthesis. Significance codes:
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table 3
Coefficient estimations for rating effects on *A* investments in B-Rep and Both-Rep treatments.

	B-Rep treatment		Both-Rep treatment	
B negative rating	−1.611	(0.333)***	−1.132	(0.356)**
B neutral rating	0.106	(0.343)	0.246	(0.394)
B positive rating	2.489	(0.329)***	1.764	(0.356)***
Previous return received when playing as *A*	0.008	(0.018)	0.039	(0.017)*
Previous investment received when playing as *B*	0.028	(0.034)	0.037	(0.032)
Last treatment period	−0.104	(0.423)	−1.798	(0.471)***
F	66.423***		35.334***	

Standard errors are in parenthesis. Significance codes:
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

with the ones reported from the Keser (2003) experiment presented above.

Treatment A-Rep, where *B* was allowed to rate *A* behavior, produced an increase in the average investment compared to the baseline periods and a decrease in the average return when controlled for *A* investments (Table 4). *B* players are no longer under rating and the increase in absolute terms of their returns is only due to the large increase of *A* investments (they actually reduced their returns, when proportioned to the received amount). The amount received in the last period when they played in the *A* role has a weak but significant effect for both *A* and *B* players, while none of the other control variables had a significant effect.

The large increase in *A* investments is hard to explain following rational expectations. *B* players are not in a risk situation and possess the information regarding the actual amount received in the current round. So why should they consider the rating of their counterparts? Nevertheless, statistical evidence shows that *B* players also used *A*s' past ratings in their decisions. In order to analyze this point, we estimated a further fixed effects model, where the panel included only the treatment periods. The resulting model was highly significant [$F(7,492) = 48.002$, $p < 0.001$] and had a negative and significant effect for both a past negative and a past "neutral" judgment compared with an unknown judgment ($t = -2.079$, $p = 0.038$ and $t = -0.705$, $p = 0.007$, respectively), while the effect for a "positive" judgment was not significant ($t = -1.5707$, $p = 0.116$). As expected, there is a positive and significant effect for the amount invested by *A* ($t = 15.135$, $p < 0.001$), and a positive effect for the amount received in the last period when the current *B* played as *A* ($t = 2.235$, $p < 0.025$), while the other control variables showed non-significant effects.

Treatment Both-Rep led to a significant increase in both investments and returns in comparison with the baseline periods. Both effects were however weaker than the corresponding ones in the B-Rep and A-Rep treatments. There is a significant end effect for both *A* and *B* players, coherently with the prediction that the incentive for investing in reputation during the last period of the game should decrease, while only *A* players respond significantly to an increase in the amounts received in the previous period (Table 5). The rating effects are similar to the ones registered in the B-Rep treatment (Table 3).

2.3. Discussion of experiment 1 results

The first experiment showed that the amounts invested/returned were systematically higher when subjects were under rating. This result is robust across treatments and independent from

Table 4
Coefficient estimations for the A-Rep treatment model.

	Dependent: *A* investment		Dependent: *B* return	
Treatment effect	1.368	(0.169)***	−1.226	(0.339)***
Previous return received when playing as *A*	0.102	(0.014)***	0.070	(0.031)*
Previous investment received when playing as *B*	0.039	(0.027)	−0.030	(0.049)
Last treatment period	−0.720	(0.434)	−0.581	(0.829)
Last baseline period	−0.510	(0.449)	−0.840	(0.858)
Current A investment	–	–	1.035	(0.050)***
F	28.854***		77.120***	

Standard errors are in parenthesis. Significance codes: $^{**}p < 0.01$.
* $p < 0.05$.
*** $p < 0.001$.

Table 5
Coefficient estimations for the Both-Rep treatment model.

	Dependent: *A* investment		Dependent: *B* return	
Treatment effect	0.979	(0.210)***	1.679	(0.332)***
Previous return received when playing as *A*	0.053	(0.016)**	0.004	(0.026)
Previous investment received when playing as *B*	0.050	(0.030)	0.059	(0.045)
Last treatment period	−2.140	(0.537)***	−2.409	(0.830)**
Last baseline period	−0.242	(0.550)	−1.174	(0.848)
Current A investment	–	–	1.215	(0.046)***
F	11.853***		143.580	***

Standard errors are in parenthesis. Significance codes: $^{*}p<0.05$.
** $p<0.01$.
*** $p<0.001$.

the position held by the subjects in the game. The B-Rep treatment results confirm the earlier findings of Keser (2003). The following two treatments extend her findings by showing that also *A* players are sensitive to the possibility of invest in their reputation and that the *B* ones responds to *A*s' reputation. The latter result is especially interesting, since in the A-Rep treatment *B* players are neither in a risk position nor under rating and therefore have no rational incentives to modify their behavior on the bases of the judgments received by *A* players.

The Both-Rep treatment confirms the findings of the previous ones, even if its effect on *A* an *B* behavior is somewhat weaker. This suggest that a system implementing a bidirectional reputation scheme does not perform necessarily better than one using only one-way ratings.

Overall, while most of the investment/return increase is probably motivated mainly by rational reputation-seeking behaviors, there are some hints that also other cognitive mechanisms are at work. First, *A* players enjoy an actual benefit from knowing *B* ratings, since this information can help to reduce uncertainty on *B* behaviors, but the opposite is not true. Therefore, at least for *A* players it is difficult to completely reduce the motivation behind this "rating effect" to a simple rational investment in reputation. The decision for *B* players is actually analogous to that faced by subjects playing a dictator game, and the relative returns found in the baseline periods are indeed similar to the proportion of the endowment offered by first players in standard dictator games. It is well known that the results of dictator games are highly sensitive to many details of the experimental setting. However, most of the time this has little to do with rational reasoning and depends instead from the functioning of other mental mechanisms (for a review of dictator experiments, see Camerer, 2003, 48–63).

Secondly, if both *A* and *B* players were only motivated by a rational investment in reputation, they should increase their investment/return levels as soon as the reputational opportunity arises, i.e., from the first period of the treatment following the baseline. Nevertheless, neither the returns in the first period allowing ranking of the B-Rep treatment, nor the investments in the first period of the A-Rep treatment are significantly different from those in the opening period of the experiment [Wilcoxon signed rank test: $V = 13$, $p = 0.152$ (two tailed) and $V = 2.5$, $p = 0.424$ (two tailed), respectively]. From the second treatment round on, players start instead to increase their investments/returns when receiving bad rankings: a behavior that leads to the overall higher cooperation level highlighted in Tables 2 and 4. This result suggests that subjects do not rationally plan to invest in their own reputation: they react instead to the judgments they receive during the game in order to maintain a sufficient reputational status.

Finally, it is worth noting that the "rating effect" apparently crowds out even the effect of incertitude reduction for *A* players. The amount invested by *A* players in the Both-Rep treatment is indeed very close to the amount invested in the A-Rep one (actually somewhat lower) where only *B* players were allowed to rate the *A* ones (see Table 1). This implies that the difference in *A* investments between the baseline and the treatment periods in the B-Rep treatment – due to the fact that *A* players could rationally trust the *B* ones having a "positive" judgment – disappears in the Both-Rep treatment. In the latter case, only the two-way effect of rating still produces significant effects on the observed behaviors, which are similar to those found in the A-Rep treatment for *A* players and in the B-Rep one for *B* players.

3. Experiment 2

3.1. Methods

We designed a second experiment to verify some of the hypotheses arising from the first one, with a specific focus on non-rational investments in reputation. A total of 84 students, 34 females and 50 males, participated to the new experiment, with an average earning of 12.88 Euros. We ran two new treatments with 42 students each. The first one (hereafter B-Rep-NR, where NR stands for "non-rational") was a modification of the B-Rep treatment, while the second one (A-Rep-NR) was a modification of the A-Rep treatment. In order to keep the situation as simple as possible, the second experiment used a between-subject design, which implies that no baseline period was played. This meant that in order to understand the treatment effect, we compared the subjects under rating (i.e. *B* players in the B-Rep-NR treatment and *A* players in the A-Rep-NR one) with those that, playing in the corresponding role in the other treatment, were not under rating (i.e. *B* players in the A-Rep-NR treatment and *A* players in the B-Rep-NR one).

Besides the fact that subjects did not play the baseline periods, A-Rep-NR and B-Rep-NR treatments were exactly like the corresponding ones of the first experiment, except that the players were able to know their opponents' rating only *after* they took the decision regarding the amount to invest or to return. This change was designed with the explicit purpose of eliminating any rational incentive for reputation investment. The fact that the ratings were revealed to the subjects only after they took the decision about the amount to invest/return implies that the rating had no possibility of influencing their decisions and, consequently, that no player could rationally increase his/her future earnings by investing in reputation. Despite the lack of rational incentives, we still expect that subjects should react to the fact of being under rating because of the working of different cognitive mechanisms. The findings of the first experiment suggested that especially *B* players, who appeared to be more sensitive to the details of the game, should significantly improve their returns in the B-Rep-NR treatment comparing to the A-Rep-NR one.

3.2. Experiment 2 results

Table 6 presents experiment 2 average investments and returns. It is immediately clear that the differences between

Table 6
Average investment and returns by treatment in the second experiment. Standard deviations are in parenthesis.

	A investment (ECU)	*B* return (ECU)
B-Rep-NR	4.30 (3.20)	4.60 (5.61)
A-Rep-NR	4.47 (3.16)	3.28 (4.83)

the two treatments influence much more *B* returns than *A* investments.

In order to determine the significance of this effect, we estimated a random effects panel model (Baltagi, 2001; for an application of random effect models to the analysis of experimental data, see Barrera and Buskens, in press), which included a dummy variable indicating the treatment and all the control variables used in the first experiment. The coefficient estimates are reported in Table 7. As expected, there is a positive and significant effect on *B* returns for the B-Rep-NR dummy, controlling for *A* investments, which implies that *B* players returned significantly more when under rating. Also the sign of the coefficient for *A* players' investments goes in the expected direction, but the coefficient itself is not significant. Both *A* and *B* players are also significantly affected by the amount received in the last period when they played as *A*, but there was no significant end effect.

An interesting result is that, despite the fact that the ratings have no practical effect, the subjects playing in both *B* position in the B-Rep-NR treatment and *A* position in the A-Rep-NR treatment tend to increase their returns after receiving a "negative" judgment while they tend to decrease their returns when receiving a "neutral" or "positive" judgment (Table 8).

3.3. *Discussion of experiment 2 results*

The results of the second experiment reinforced the idea of a specific cognitive reputational mechanism. The average amounts returned by subjects are significantly higher when under rating. Also the amounts invested by *A* players tended to increase, even if this effect is not statistically significant. Unlike the previous experiment, this result can no longer be explained under a rational reputation investment framework, since the experiment design did not allow the future opponents of a given player to know his/her reputation before their investment/return decisions. The fact that a non-rational mechanism is now at work is also highlighted by the lack of any end effect.

Especially interesting is the fact that both *A* and *B* players reacted strongly to the judgments received. This suggests a reputation effect also on *A* behaviors, despite the non-significance of the first coefficient of the panel model. The result that, on average, *A* investments did not significantly increase in the A-Rep-NR treatment may indeed be due to the higher tolerance of *B* players, who less frequently than the *A* ones rated their opponents' behavior as "negative" (Table 8). This may have created less incentives for *A* players than for the *B* ones to change their behavior, resulting in lower overall cooperation levels than in the B-Rep-NR treatment.

4. General discussion

Our experiments show that humans are highly sensitive to their own reputational status and react promptly to other people's judgment. A large part of this effect is undoubtedly due to a rational investment in reputation, but the second experiment suggests that reputation-seeking behavior still matters even when any practical effect of the players' judgments is ruled out. The non-rational reputation effect appears to be weaker than the rational one, but the two mechanisms are probably not mutually exclusive, working instead often side by side. In the first experiment, the increase of the amounts invested/returned is hence due to *both* the rational reputation building behavior and the cognitive response to the rating scheme. Only the latter mechanism is at work in the second experiment, a fact that explains the weaker reputation effect.

A related point is the significant end effect found in the first experiment, but not in the second one (Tables 5 and 7). This is consistent with the idea of a rational investment in reputation working in the former case, while a different mechanism is at work when judgments cannot influence opponents' behavior.

Finally, it is interesting to note that the effects of the cognitive reputation mechanism are less uniformly distributed among subjects than rational reputation building. This is shown by the variance of *B* players' returns in the B-Rep-NR treatment, which is significantly (at the 10% level) higher than the one in the B-Rep treatment [$F(499{,}524) = 0.868$, $p = 0.055$ (one tailed)]. Similarly, the variance of *A* players' investments in the A-Rep-NR treatment is higher than the one in the A-Rep treatment [$F(499{,}524) = 0.879$, $p = 0.072$ (one tailed)]. This opens an interesting analogy with the Hsu et al. (2008) findings showing that the inequity aversion varies significantly among individuals, both as behavioral measure and as activity of the related brain regions. Unfortunately, we were not able to perform brain imaging on our subjects, but in our experiments the behavior of subjects also appears to be less uniform when only cognitive reputation mechanisms are at work (experiment 2) than when they also rationally invest in their reputation (experiment 1).

Summarizing, our results are consistent with the hypothesis that reputation motives significantly modifies human behavior towards greater cooperation not only when rational incentives in reputation building exist, but also when the judgments expressed have no practical consequences. From an evolutionary point of view, the sensitivity towards reputational status is probably part of the human "tribal social instinct", i.e. the set of emotions and cognitive mechanisms that represent one of the key adaptations for social life of our species (Richerson and Boyd, 2001; Richerson et al., 2003). The strong concern for reputational status, even when it has no actual effect on individuals' payoffs, is a cognitive mechanism that can significantly explain a part of human behavior and has both evolutionary meaning and empirical plausibility.

Table 7
Coefficient estimations for the A-Rep-NR vs. B-Rep-NR comparison model.

	Dependent: *A* investment		Dependent: *B* return	
B-Rep-NR treatment effect	−0.330	(0.331)	1.411	(0.702)*
Previous return received when playing as *A*	0.111	(0.014)***	0.042	(0.020)*
Previous investment received when playing as *B*	0.015	(0.024)	0.027	(0.033)
Last treatment period	−0.264	(0.342)	−0.938	(0.530)
Current A investment	–	–	0.765	(0.033)***
F	18.044***		110.183***	

Standard errors are in parenthesis. Significance codes: ** $p < 0.01$.

* $p < 0.05$.

*** $p < 0.001$.

Table 8
A investment and *B* return change by received rating in the second experiment.

	Rating proportion	*A* investment change (ECU)	Rating proportion	*B* return change (ECU)
Negative	0.389	0.511	0.455	0.470
Neutral	0.228	−0.209	0.188	−0.382
Positive	0.383	−0.816	0.356	−0.703

Acknowledgments

Financial support was provided by a FIRB 2003 grant from the Italian Ministry for the University and the Scientific Research (SOCRATE Research Project, coordinated by Rosaria Conte, Protocol: RBNE03Y338_002). For remarks on a preliminary version of this paper, we would like to thank Ralf Sommerfeld, Rosaria Conte, Matteo Gallizzi and all the participants of a seminar at the Department of Social Sciences, University of Brescia. We would like to express our gratitude to Beatrice Marelli and to all the participants of a seminar at the Indiana University Workshop in Political Theory and Policy Analysis for remarks and suggestions. We are especially grateful for the thoughtful remarks of an anonymous reviewer. We would also like to thank Robert Coates, Centro Lingusitico, Università Commerciale Luigi Bocconi, for his lingusitic revisions. Usual disclaimers apply.

References

Alexander, R.D., 1987. The Biology of Moral Systems. Basic Books, New York.

Baltagi, B.H., 2001. The Econometrics of Panel Data, Second ed. John Wiley & Sons, London.

Barclay, P., 2004. Trustworthiness and competitive altruism can also solve the "tragedy of the commons". Evolution and Human Behavior 25, 209–220.

Barkow, J.H., Cosmides, L., Tooby, J. (Eds.), 1992. The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford University Press, Oxford.

Barrera, D., Buskens, V., in press. Third-Party Effects on Trust in an Embedded Investment Game. In: Cook, K., Snijders, C., Buskens, V., Cheshire, C. (Eds.). Trust and Reputation. New York, Russell Sage.

Bateson, M., Nettle, D., Roberts, G., 2006. Cues of being watched enhance cooperation in a real-world setting. Biology Letters 2, 412–414.

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

Boyd, R., Richerson, P.J., 1989. The evolution of indirect reciprocity. Social Networks 11, 213–236.

Bravo, G., Tamburino, L., 2008. The evolution of trust in non-simultaneous exchange situations. Rationality and Society 20 (1), 85–113.

Burnham, T.C., Hare, B., 2007. Engineering human cooperation: does involuntary neural activation increase public goods contributions? Human Nature 18, 88–108.

Camerer, C.F., 2003. Behavioral Game Theory. Experiments in Strategic Interaction. Russel Sage Foundation/Princeton University Press, New York/Princeton.

Conte, R., Paolucci, M., 2002. Reputation in Artificial Societies: Social Beliefs for Social Order. Kluwer Academic Publishers, Dordrecht.

Dunbar, R.I.M., 1996. Grooming, Gossip and the Evolution of Language. Harvard Univerity Press, Cambridge, MA.

Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 525, 785–791.

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

Fischbacher, U., 2007. z-Tree. Zurich toolbox for readymade economic experiments. Experimental Economics 10, 171–178.

Gintis, H., Smith, E.A., Bowles, S., 2001. Costly signaling and cooperation. Journal of Theoretical Biology 213, 103–119.

Haley, K.J., Fessler, D.M., 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior 26, 245–256.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., Ziker, J., 2006. Costly punishment across human societies. Science 312, 1767–1770.

Hsu, M., Anen, C., Quartz, S.R., 2008. The right and the good: distributive justice and neural encoding of equity and efficiency. Science 320, 1092–1095.

Keser, C., 2003. Experimental games for the design of reputation management systems. IBM Systems Journal 42, 498–506.

Leimar, O., Hammerstein, P., 2001. Evolution of cooperation through indirect reciprocity. Proceedings of the Royal Society London 268, 745–753.

McCabe, K.A., Smith, V.L., 2000. A comparison of naïve and sophisticated subject behavior with game theoretic predictions. Proceedings of the National Academy of Sciences of the United States of America 97, 3777–3781.

Milinski, M., Rockenbach, B., 2007. Spying on others evolves. Science 317, 464–465.

Milinski, M., Semmann, D., Krambeck, H.-J., 2002a. Donors to charity gain both indirect reciprocity and political reputation. Proceedings of the Royal Society 269, 881–883.

Milinski, M., Semmann, D., Krambeck, H.-J., 2002b. Reputation helps solve the tragedy of the commons. Nature 415, 424–426.

Milinski, M., Semmann, D., Krambeck, H.-J., Marotzke, J., 2006. Stabilizing the earth's climate is not a losing game: supporting evidence from public goods experiments. Proceedings of the National Academy of Sciences of the United States of America 103 (11), 3994–3998.

Nowak, M.A., Sigmund, K., 1998a. The dynamics of indirect reciprocity. Journal of Theoretical Biology 194, 561–574.

Nowak, M.A., Sigmund, K., 1998b. Evolution of indirect reciprocity by image scoring. Nature 393, 573–577.

Nowak, M.A., Sigmund, K., 2005. Evolution of indirect reciprocity. Nature 437, 1291–1298.

Ortmann, A., Fitzgerald, J., Boeing, C., 2000. Trust, reciprocity, and social history: a re-examination. Experimental Economics 3, 81–100.

Panchanathan, K., Boyd, R., 2004. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature 432, 499–502.

Piazza, J., Bering, J.M., 2008. Concerns about reputation via gossip promote generous allocations in an economic game. Evolution and Human Behavior 29, 172–178.

Pollock, G.B., Dugatkin, L.A., 1992. Reciprocity and the evolution of reputation. Journal of Theoretical Biology 48, 25–37.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing.

Raub, W., Weesie, J., 1990. Reputation and efficiency in social interactions: an example of network effects. The American Journal of Sociology 96 (3), 626–654.

Richerson, P.J., Boyd, R., 2001. The biology of commitment to groups: a tribal instincts hypothesis. In: Nesse, R. (Ed.), Evolution and the Capacity for Commitment. Russell Sage Foundation, New York, pp. 186–220.

Richerson, P.J., Boyd, R.T., Henrich, J., 2003. Cultural evolution of human cooperation. In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. The MIT Press, Cambridge, pp. 357–388.

Seinen, I., Schram, A., 2006. Status and group norms: indirect reciprocity in a helping experiment. European Economic Review 50, 581–602.

Sommerfeld, R.D., Krambeck, H.-J., Semmann, D., Milinski, M., 2007. Gossip as an alternative for direct observation in games of indirect reciprocity. Proceedings of the National Academy of Sciences of the United States of America 104 (44), 17435–17440.

Zahavi, A., Zahavi, A., 1997. The Handicap Principle: A Missing piece of Darwin's Puzzle. Oxford University Press, Oxford.