

Third Party Reputation in Repeated Trust Games

Riccardo Boero^{*}, Giangiacomo Bravo[°],
Marco Castellani[†], Francesco Laganà[†],
Flaminio Squazzoni[†]

^{*} Dipartimento di Scienze Economiche e Finanziarie “G. Prato”, Università di Torino and GECS -
Research Group in Experimental and Computational Sociology, Università di Brescia,
boero@econ.unito.it

[°] Dipartimento di Scienze Sociali, Università di Torino and GECS -Research Group in
Experimental and Computational Sociology, Università di Brescia, giangiaco.bravo@unito.it

[†] Dipartimento di Studi Sociali and GECS — Research Group in Experimental and Computational
Sociology, Università di Brescia {gbravo, castella, squazzon}@eco.unibs.it

[†] Dipartimento di Studi Sociali, Università di Brescia, lagana@eco.unibs.it

Abstract. This paper presents the results of some laboratory experiments on the relevance of reputation for the emergence of cooperation. We have extended some previous repeated trust games adding new treatments where reputation is taken more explicitly into account than in the present literature. We compare treatments where the investor and the trustee rate each other and treatments where the investor and the trustee are rated by a third party. The results show that, although reputation positively affects trust and cooperation, some differences in the reputation mechanism can generate different cooperative outcomes. These results have implications on the recent sociological debate on markets and on normative pillars of economic phenomena.

Keywords: Reputation; Trust; Game Theory; Social Norms; Strategic Interactions.

JEL: C91; C73; Z13.

1 Introduction

Trust and reputation are a key issue in economics and sociology (e.g., Granovetter 2002; Cook, Hardin and Levi 2005; Gui and Sudgen 2005; Barrera 2008), not to mention organization sciences and the growing literature on e-markets and communities (e.g., Zacharia and Maes 2000; Jackson 2004; Sabater and Sierra 2005; Change *et al.* 2006). This interest is due to a twofold reason. Firstly, globalization involves that information evaluation processes, e.g., on goods, partners or services, are an increasingly difficult issue today for any economic actor, because of uncertainty, ambiguity, information asymmetries and the increasing knowledge-based source of goods and services that are exchanged on markets. As shown by many economic sociology studies, rather than solely on standard information signals, e.g., price, economic actors tend more and more to rely upon social norms and networks to reduce strategic and operative uncertainty, get sensitive market information and establish informal agreements with partners (Raub and Weesie 1990; Granovetter 2002; Buskens and Raub 2002; Burt 2005). Secondly, the Internet and e-Bay age has caused that also the market arena has gone through significant changes, as many transactions now are undertaken by dispersed actors interacting directly through on-line reputation systems that embody particular incentives and social norms. In such context, where uncertainty, transformations, chaos and societal change are the rules, it is quite natural to expect that scientists pay more attention to the study of self-regulating enforcement mechanisms and of basic informal, unplanned and spontaneous institutions that can eventually promote higher social order from the bottom-up, with the aim to grab

both the positive or negative sides of this institutional innovations as well (e.g., Guennif and Revest 2005).

In recent years, the most significant progresses in understanding these issues have been made by the so-called “experimental behavioural sciences”, where cooperation and social norms and patterns are studied in the laboratory (e.g., Henrich *et al.* 2005; Gintis *et al.* 2005; Fehr and Gintis 2007). One of the most important idea behind this approach, the same suggested in this paper, is to apply the experimental method to the analysis of socio-economic phenomena so as to capture in clear terms and dissect the micro-foundations of human behaviour, and in particular the interplay of material interests and normative behaviour in market-like interactions (Bowles 2004). As Fehr and Gintis (2007, 44) have correctly argued in some sentences that are worth reporting here, the experimental approach can be beneficial for the development of sociology:

“sociological theory has not developed a coherent, broadly accepted framework that facilitates cumulative scientific progress and explains the emergent aggregate patterns of social behaviour in terms of individuals’ preferences, their beliefs, and the social and economic constraints they face. Nor has sociological research developed a parsimonious, empirically grounded view of the basic motivational driving forces of human behaviour, which may be due to the role that controlled experiments played in the development of the discipline. [...] Most empirical work in sociology is based on non-experimental methods [...]. This is unfortunate because the experimental method permits cumulative empirical and theoretical progress through rigorous testing of alternative theories, the establishment of causal relationships between key variables, and the replicability of results. In contrast, non-experimental field data rarely permit a clean discrimination between theories, the inference of causality is typically not possible from these data, and replicability often proves difficult or impossible”.

A good example of this approach in the field of trust and reputation is Keser (2003), where reputation is introduced into repeated trust games. In a trust game, the investor (*A* player) decides how much of his/her endowment to send to the trustee (*B* player), 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* (e.g., Berg *et al.* 1995; Camerer 2003). In Keser’s design, subjects interact for a fixed number of rounds and *A* players are allowed to rate the behaviour of their opponents. In the following round, the reputation scores are presented to the new *A* players before they took their decision. The Keser’s main result is that the possibility of relying on others’ experiences significantly increases cooperation in the game, by providing incentives for *A* investments and promoting *B* returns.

Experimental studies and analytical and simulation models on reciprocity have confirmed that the possibility of knowing others’ past behaviour significantly increases cooperation (e.g., Novak and Sigmund 1998a; 1998b; 2005; Fehr and Fischbacher 2003; Seinen and Schram 2006; Bravo and Tamburino 2008). The point is that recent experimental finding suggest that reputation does not need to be necessarily grounded on direct knowledge of past behaviour of other individuals, since humans can exploit the many sources of information available in social life, nor it is just closely related to reciprocity mechanisms. For instance, gossip plays an important role in transmitting reputational information in human societies (e.g., Burt and Knez 1995; Dunbar 1996; Kurubawa 2005; Ferrin, Dirks and Shah 2006). Some recent works show that gossip has a strong influence on the behaviour of experimental subjects even when subjects can use other information sources, e.g. direct observation (Sommerfeld *et al.* 2007). Another experiment show that individuals react to the possibility of being the subject of gossip by increasing their contributions in a dictator game (Piazza and Bering 2008). The normative nature of reputation mechanisms has also been emphasized in many interesting computer simulation models (e.g., Conte & Paolucci 2002; Janssen 2006; Hahn *et al.* 2007).

This paper aims to show some experimental results where the role of third parties in reputation mechanisms is taken more explicitly into account than at the present. Previous evidence shown that human subjects are highly sensitive to their own reputation and react

promptly to other people's judgments. Although a large part of the reputation effect is due to a strategic investment in reputation made by players (e.g., Bolton, Katok and Ockenfels 2004; Barrera 2008), Keser (2003) and Boero *et al.* (2008) have shown that reputation matters even when any practical effect of reputation scores on subjects' payoffs is ruled out. Such results help corroborate the hypothesis that subjects attach to reputation a normative flavour.

In this paper we present four new treatments on a repeated trust game where reputation scores are not directly expressed by the same players (*A* and *B*) who are engaged in interaction, but by *C* players who do not take part in interaction and are called to observe and rate *A*'s and *B*'s behaviour. We aim to focus on two reputation mechanisms. In the first one, reputation scores are assigned and transmitted by *C* players. In the second one, reputation scores of *B* players are assigned by *A* players and transmitted to *C* players who decide whether transmit the scores to the other players. In this case, *C* players do not have full responsibility on scores' assignment, so that reputation mechanism looks more like gossip (e.g., Conte and Paolucci 2002). The research questions are as follows: Does a third party reputation mechanism generate the same cooperative outcomes of dyadic reputation? What is the consequence of a gossip-like third party reputation mechanism on trust and cooperative outcomes?

The paper is organized as follows. In the second section we describe the experiment. The third one presents the results. The fourth one suggests concluding remarks and future developments.

2. The Experiment

The experiment took place on two days in the computer laboratory of the Faculty of Economics of the University of Brescia, which is equipped with the experimental software z-Tree (Fischbacher 2007). The subjects were 126 students of the University of Brescia recruited through public announcements in the various faculties. All interactions took place through the computer network and the subjects were unable to identify their counterparts. They played 25 rounds of our reputation-based repeated trust game. The subjects were informed in advance of the duration of the game. Each subject played for nearly one hour and the average earning was about 15 Euros, which was paid immediately after the experiment.

The experiment followed four previous treatments where Boero *et al.* (2008) extended the Keser's repeated trust game to allow *A* and *B* players to rate their opponents. First, we replicated the repeated trust game baseline, which is as follows: (i) subjects are randomly assigned to the role of *A* and *B* players each round; (ii) both player *A* (the investor) and player *B* (the trustee) received an initial endowment of 10 experimental currency units (ECU), having an exchange rate of 1 ECU = 1.5 Euro cents; (iii) player *A* decided his/her investment and the invested amount was tripled and sent to player *B* in addition to his/her own endowment; (iv) *B* chose the amount to return to *A*; (v) the sums earned by both players in the current period were displayed to both subjects.

Upon the baseline, we add three new treatments. In the first treatment (indicated as *Treatment 1*), *A* was allowed to rate *B*'s behaviour as "negative", "neutral", or "positive", as in Keser's repeated trust game. The subsequent *A* players interacting with *B* were informed of the last reputation score received by the latter *before* making their investment decision. The second treatment (*Treatment 2*) was exactly like the first except for the fact that only *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 cases, the possible scores were "negative", "neutral" and "positive" and this information was presented to the *B* players before their return decision. In the third treatment (*Treatment 3*), both *A* and *B* players were allowed to rate each other and knew the scores in the following period before their investment/return decision. The results of this first experiment are used in this paper as baselines against which new treatments are compared.

The new experiment consisted of four new treatments (called Treatment *C1*, *C2*, *C3* and *Inf*). Subjects are randomly assigned to the roles of *A*, *B* and *C* players each round, playing on average the same number of round in each role. *Treatment C1* worked as *treatment 1* with the exception that *B*'s behaviour is rated as "negative", "neutral", or "positive" by *C* players who are called to observe the interaction. Reputation scores of *B* players are transmitted by *C* players to the subsequent *A* players. In *treatment C2*, *C* players observed and rated *A*'s behaviour. Reputation scores are transmitted to the subsequent *B* players. In *treatment C3*, *C* players observed and rated both *A*'s and *B*'s behaviour and reputation scores are transmitted to both. *Treatment Inf* worked as *treatment C1* with the exception that *B*'s reputation is not assigned by *C* players but by *A* players. Scores are then transmitted to *C* players who decide whether to transmit the scores to subsequent *B* players. Before to decide, *C* players know the past reputation score of the *A* players the last time they played in the *B*'s role. While the difference between *treatment 1*, *2* and *3* and *treatment C1*, *C2* and *C3* is that in the second case reputation scores originated from a player that was not directly involved in the interaction (third party), the difference between *C1* and *Inf* is that, in the second case, *C* players are simply *informers* who do not have any direct responsibility on the quality of reputational information they transmit. As a matter of fact, in real social life, it is likely that reputation takes place also when people relate to reputational information at second hand, such as in case of gossip (e.g., Conte and Paolucci 2002). *Treatment Inf* allows us to observe the impact of a gossip-like reputation mechanism on trust and cooperation.

3. Results

Let us begin with the results of treatments *1*, *2*, *3* and *C1*, *C2*, and *C3*. To get a reasonable comparison, treatments should be matched as follows: *1* with *C1* (*B* players are under rating), *2* with *C2* (*A* players are under rating), *3* with *C3* (both *A* and *B* players are under rating). While in *treatment 1*, *2* and *3* reputation took place within the groups of players, in *C1*, *C2* and *C3* reputation was due to a third party that was external to the groups of players. Table 1 show mean and standard deviation of *A* investments, *B* returns and *B* return proportional to the *B*'s endowment.

In *treatment 1* and *C1*, before to decide whether and how much to invest, *A* can use the *B*'s reputation score that another *A* has assigned to *B* at the end of the previous round (*treatment 1*), or the *B*'s reputation score assigned by *C*, that is, a third party (*treatment C1*). *B* returns are important to understand whether *B* is interested in building a good reputation to *A*'s eyes. Results show that the introduction of a third party basically replicates the high level of cooperation of *treatment 1*. *A* investments and *B* returns are even higher in absolute terms, although *B* returns in proportion to the endowment are essentially the same. Some statistical analyses on the ratio between proportional returns and rating in *treatment 1* and *C1* show that there is a co-graduation between the treatments. *Treatment 1* shows a value of $\gamma=0,591$ and Kendall's $\tau=0,481$ and *treatment C1* shows corresponding values of $\gamma=0,589$ and Kendall's $\tau=0,464$. This means that in the two treatments subjects essentially behaved in the same way. Moreover, the non significance of Mann-Whitney test shows that the distribution of *B* returns are not significantly different between the two treatments ($p=0.709$).

In *treatment 2* and *C2*, before to decide whether and how much to return to *A*, *B* can use the *A*'s reputation score. In this case, *B* can make use of his/her direct evidence (how much *A* has decided to invest) and the past *A*'s reputation as well, with this second information source that should not make great sense if *B* players would behave as 'pure' rational agents. The results show that *C* players essentially do not make a real difference as regards to *treatment 2*. *A* players invested slightly more in *treatment 2* than in *treatment C2* (see Table 1), but the difference is not statistically relevant (see Table 2). The same holds true when we look at the absolute ($p=0.086$) and proportional ($p=0.572$) returns in *treatment 3* and *C3* (see Table 2). Although some small difference can be pointed out, such as a higher

investment of *A* players in *treatment 3* ($p=0.01$), the *treatment 3* and *C3* show the same results, with the same *B* proportional returns. This is confirmed when we look at the Mann-Whitney test comparison among these six treatments in Table 2. The same trend holds true when we look at the average of *B*'s returns and the consequent reputation score assigned by *A* at the end of each round (see table 4), that is, when reputation score's assignment takes place.

Table 1. Mean and standard deviation of *A* investments, *B* returns and *B* returns proportional to the endowment.

<i>Treatments</i>	<i>Statistics</i>	<i>A investment</i>	<i>B return</i>	<i>B return</i>
1 (<i>A</i> → <i>B</i>)	Mean	4.29	6.40	0.25
	Std. Dev.	2.70	4.84	0.14
2 (<i>B</i> → <i>A</i>)	Mean	5.61	5.83	0.19
	Std. Dev.	2.77	5.80	0.17
3 (<i>A</i> ↔ <i>B</i>)	Mean	5.27	7.08	0.24
	Std. Dev.	3.02	5.98	0.16
C1 (<i>C</i> → <i>B</i>)	Mean	5.00	6.70	0.24
	Std. Dev.	3.39	6.16	0.18
C2 (<i>C</i> → <i>A</i>)	Mean	5.46	4.86	0.16
	Std. Dev.	3.37	5.88	0.17
C3 (<i>C</i> → <i>A</i> & <i>B</i>)	Mean	4.64	6.59	0.24
	Std. Dev.	3.08	6.59	0.20

Table 2. Mann-Whitney test comparison of *treatment 1-C1*, *2-C2*, and *3-C3* (one-tailed).

<i>Treatments</i>	<i>A investment</i>	<i>B return</i>	<i>B proportional return</i>
1-C1	W=27742 $p=0.010$	W=32108.5 $p=0.6455$	W=32562.5 $p=0.2578$
2-C2	W=32099 $p=0.3563$	W=35357.5 $p=0.0082$	W=35265.5 $p=0.0097$
3-C3	W=35017.5 $p=0.01$	W=33719 $p=0.08653$	W=32421.5 $P=0.286$

Table 3. Average of *A* investments, *B* returns and *B* proportional in *treatment 1* and *C1*, *2* and *C2*, and *3* and *C3* for past reputation scores of players.

<i>B</i> 's score	<i>Treatment 1</i>			<i>Treatment C1</i>		
	<i>A investment</i>	<i>B return</i>	<i>B proportional return</i>	<i>A investment</i>	<i>B return</i>	<i>B proportional return</i>
Unknown	4,12	5,66	0,24	4,19	5,81	0,22
Negative	2,57	3,34	0,17	4,41	4,70	0,18
Neutral	4,03	6,21	0,27	4,32	5,52	0,22
Positive	6,33	10,00	0,33	6,16	9,41	0,31
<i>A</i> 's score	<i>Treatment 2</i>			<i>Treatment C2</i>		
Unknown	6,18	5,15	0,52	3,19	4,55	0,45

Negative	3,25	3,91	0,39	3,38	4,05	0,41
Neutral	4,44	4,91	0,49	5,10	5,15	0,52
Positive	7,79	6,99	0,70	6,49	7,03	0,70
<i>B's score</i>						
	<i>Treatment 3</i>			<i>Treatment C3</i>		
Unknown	5,46	6,46	0,22	4,29	5,71	0,22
Negative	4,32	4,07	0,15	3,76	3,11	0,13
Neutral	5,67	8,33	0,28	4,49	5,89	0,22
Positive	5,82	9,30	0,31	5,45	9,55	0,33
<i>A's score</i>						
Unknown	8,55	5,37	0,54	5,66	4,14	0,41
Negative	6,84	4,65	0,47	5,24	3,38	0,34
Neutral	6,04	4,23	0,42	5,93	4,82	0,48
Positive	7,38	6,14	0,61	8,51	5,78	0,58

Table 4. Average *B*'s return and reputation score assigned by *A/C* players after the *B*'s return at the end of each round in *treatment 1* and *C1*.

<i>Score assigned to B</i>	<i>Statistics</i>	<i>Treatment 1</i>		<i>Treatment C1</i>	
		B returns	B proportional returns	B returns	B proportional returns
Negative	Mean	3.47	.1289	3.48	.1226
	Std. Dev.	3.85	.1387	4.45	.1333
Neutral	Mean	6.57	.2743	5.55	.2088
	Std. Dev.	4.19	.0887	4.89	.1523
Positive	Mean	9.21	.3645	9.27	.3588
	Std. Dev.	4.51	.1087	6.12	.1693

Figure 1 and 2 show the dynamics of *A* investments and *B* returns over time. Results indicate that the third party reputation mechanism guaranteed on average a higher level of cooperation at the beginning of the game (see Figure 1 and 2 on the right). Cooperation does not increase or tends even to decline as the rounds follow. Looking at the average *A* investments, *B* returns and *B* proportional in *treatment 1* and *C1* (see Table 3), where the target of reputation scores were *B* players, we can argue that investments/returns and reputation scores have had a strong relation. This is confirmed when looking at values of Kendall's $\tau=0,481$ (*treatment 1*), $\tau=0.5962$ (*treatment 2*), $\tau=0.5984$ (*treatment 3*, proportional investment), $\tau=0.5670$ (*treatment 3*, proportional return), $\tau=0.4647$ (*treatment C1*), $\tau=0.5317$ (*treatment C2*), $\tau=0.4560$ (*treatment C3*, proportional investment), $\tau=0.4176$ (*treatment C3*, proportional return), $\tau=0.5009$ (*treatment Inf*).

Fig. 1. *A* investments over time. From left top, *treatment 1* and *C1*, 2 and *C2*, and 3 and *C3*.

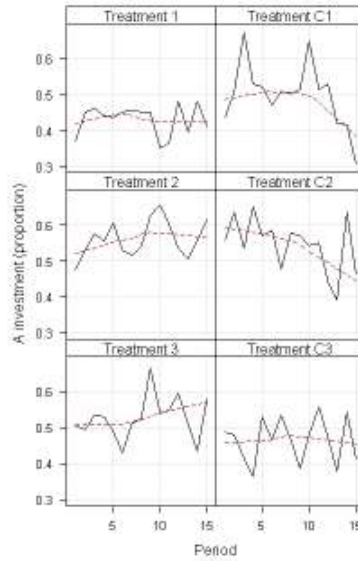
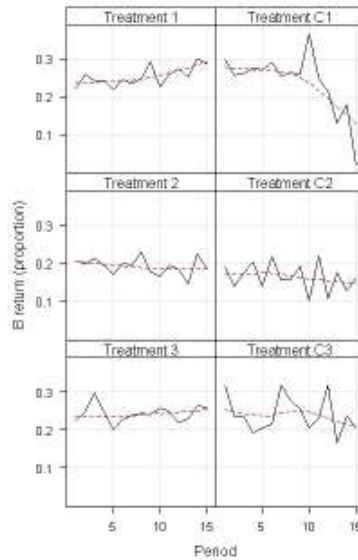


Fig. 2. *B* returns over time. From left top, *treatment 1* and *C1*, 2 and *C2*, and 3 and *C3*.



The result of *treatment Inf* shown that *A* invested on average 4.16 ECU and *B* returned on average 4.36 ECU, and 0.17 in proportion to his/her endowment, that is, less than in *treatment 1* and *C1*. The same evidence holds true when we look at Mann-Whitney test that compares *treatment Inf* with *treatment 1* and *C1*, and *treatment 1* with *treatment Inf* (see Table 5). *C* players decided to send to *A* players the *B*'s reputation score in 67% of the cases. When *B* players had a negative reputation score, *C* players decided to pass it along in 62% of the cases, when a good one, in 72% of the cases, when a neutral one, in 60% of the case, when unknown, in 81% of the cases. The reputation score that *A* players assigned to *B* players followed these *B*'s returns and proportional returns: *A* assigned a negative score when *B* returned on average 2.26 ECU (proportional return is .089), a neutral score when *B*

returned on average 4.44 ECU (proportional return is .198), and a positive when *B* returned on average 9.27 ECU (proportional return is .320).

No significant effect has been played also by past reputation scores of *A* players when they played in a *B*'s role. *C* players decided to send the *B*'s reputation score when *A* had a negative reputation in 48% of the cases, a neutral in 21% of the cases and a positive in 31% of the cases. They decided not to pass the *B*'s reputation score when *A* had a negative reputation in 52% of the cases, a neutral in 24% of the cases and a positive in 34% of the cases. The past reputation score of *A* players did not have great meaning to *C* players.

These last evidences allow us to conclude that the decision of the *informers* as such did not significantly impact the game. But, if this is true, what else can explain the relatively worst performance of this last reputation mechanism as regards to what happened in *treatment 1* (direct reputation) and *C1* (third party reputation)?

Table 5. Mann-Whitney test comparison of *treatment 1-C1*, *2-C2*, and *3-C3* (one-tailed).

<i>Treatments</i>	<i>A investment</i>	<i>B return</i>	<i>B return</i> (proportion to <i>B</i> 's endowment)
1- <i>Inf</i>	W=32769 <i>p</i> =0.2177	W=40901 <i>p</i> =0.000	W=41295 <i>p</i> =0.000
<i>C1- Inf</i>	W=18958.5 <i>p</i> =0.006	W=16862.5 <i>p</i> =0.000	W=17006.5 <i>p</i> =0.000

Figure 3 shows the dynamics of *B*'s returns over time in *treatment Inf*. The evidence is that, at the beginning of the game, players started by cooperating but soon in a few rounds cooperation started to decline according to a clear pattern. The fact that players started by cooperating means that they understood the game and did not approach it following a negative flavour. Patterns of players' behaviour are relatively stable and not affected by random factors.

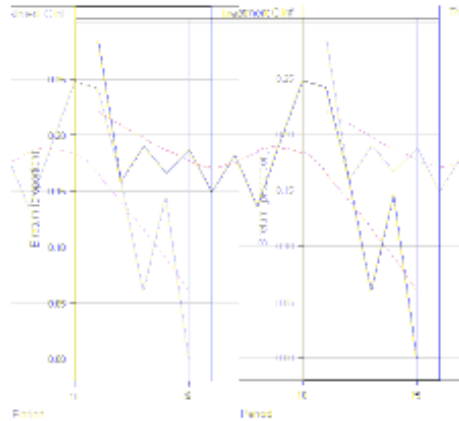


Fig. 3. *B*'s returns over time in *treatment Inf*.

To capture this pattern, we have dissected the impact of reputation scores in *treatment 1*, *C1* and *Inf* (see Figure 4, 5 and 6). The comparison shows that *treatment Inf* is characterized by the predominance and the persistence of negative scores. At the end of the round 15, 69% of *B* players have had negative reputation scores, 51% of whom kept a negative score nearly along all the game (see Figure 6). A possible explanation is that *A* players started to cooperate and to invest, but *B* players returned less than in *treatment 1* and *C1*. This, in turn, brought about negative reputation scores for *B* players, reduction of investments of *A* players and weak incentives for the emergence of trust and cooperation.

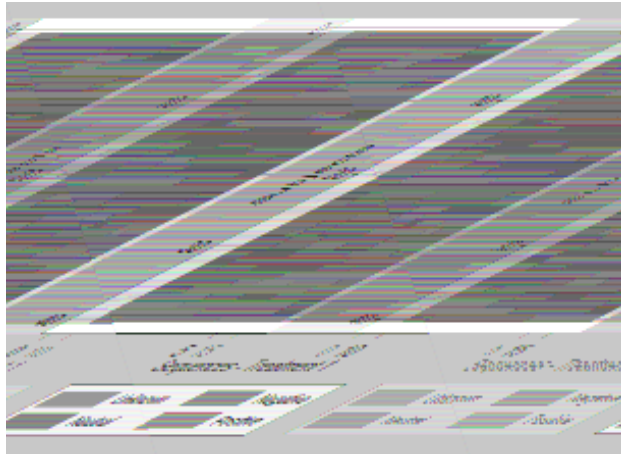


Fig. 4. Sequence of B 's ratings in *treatment 1*.

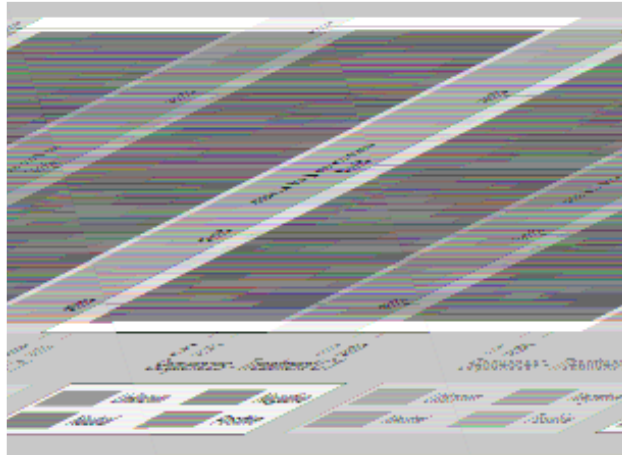


Fig. 5. Sequence of B 's ratings in *treatment C1*.

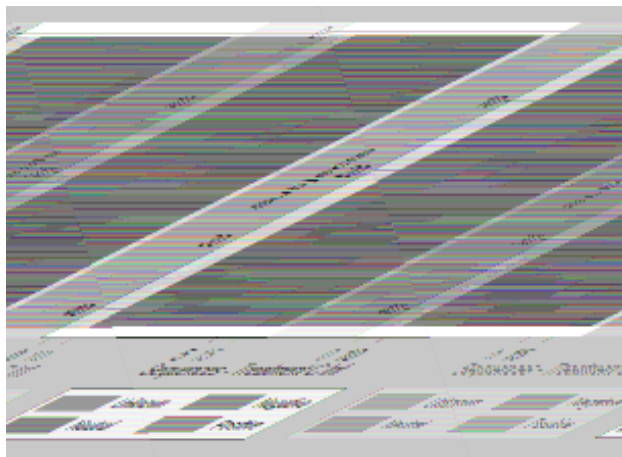


Fig. 6. Sequence of B 's ratings in *treatment Inf*.

There are two possible explanatory mechanisms of B 's behaviour, which could be at work simultaneously in a complementary way. First, B players decided to return little because they could suspect to be cheated by C informers, who could decide not to pass along their possible positive reputation to other A players. Secondly, B players decided to

return little because they thought that *C* informers did not have any concrete incentive to pass to *A* players their possible negative reputation. The point is that *C* informers have been viewed by players as a source of uncertainty in the reputation process. Players thought that *C* informers were not responsible for the quality of the reputational information they were called to spread, because scores were expressed by the players themselves. On the contrary, the introduction of the third party reputation mechanism in *treatment C1* was viewed as a stable, largely predictable reputation mechanism, exactly because in that case *C* players were responsible for the score they were called to pass along. The evidence is that our gossip-like reputation mechanism did not guarantee the same level of trust and cooperative regimes that was guaranteed by the third party reputation mechanism, where *C* players called to assign reputation had full responsibility of the reputation they handed out.

4. Concluding Remarks and Future Developments

The experiments presented in this paper have shown three main findings that can help extend the explanatory power of reputation-based trust for the emergence of cooperation among actors, as well as refine standard views on this.

First, our experiments have confirmed that individuals tend to be really sensitive to their reputation even when this would not affect their material payoffs. This is an interesting finding when considered that, in our experiments: (a) actors were interacting anonymously with random unknown partners, (b) there was no social structure effect, as actors were randomly assigned to roles (*A*, *B* and *C*) each round, and (c) they were subject to economic incentives, which are expected to enforce self-interest motivation. As a matter of fact, if the conditions above are nullified, i.e., if there is no economic incentive, actors are embedded in a dense social structure, and they interact with identifiable partners, it is expected that trust and cooperation can easily result. Vice versa, by assuming (a), (b) and (c), our results show that: (i) against the prediction of the rational choice theory, actors did not take into account only material self-interest consequences of their behaviour while taking economic decisions; (ii) they mostly referred to social norms, such as reputation, fairness and trust responsiveness (e.g., Rabin 1993; Pelligra 2005), even when this reference could have penalised their expected pay-off and even if information, i.e., reputational information, was undoubtedly ambiguous, uncertain, and subjectively biased; and (iii) they complied with social norms although they were not embedded in a dense social structure and social sanctions were absent.

A second finding helps overcome some limitations of the standard trust game literature, by paying more attention to sociological features (Swedberg 2001). In the standard literature, one of the strongest and unrealistic assumption is that information on past behaviour of a player is always available and always true for the counterpart. As a consequence, information asymmetries and the formation of the reputation are taken for granted (Guennif and Revest 2005). On the contrary, sociological findings have shown exactly the relevance of the third parties as the “intermediaries of trust” (e.g., Burt 2005), in particular in large and dynamic networks where it is expected to find high level of information asymmetries and high uncertainty on partners’ trustworthiness (Corten and Cook 2008). In this respect, our experimental evidence suggests that trust can be promoted by third party reputation also in random interactions.

Finally, a third finding is that, though it is evident that in our experiment reputation mechanisms “encapsulated” trust, to paraphrase Cook, Hardin and Levi (2005), the effect of third parties can be positive on this provided that they are perceived by other players to be responsible of the reputational information that they transmit. This can explain the bad performance of treatment *Inf* when compared with *treatment 1* and *C1*, where reputation was based on reputation scores expressed directly by the same players or by *C* players who had the full responsibility of assigning the score. This evidence can have an interesting implication on the debate on reputation on markets, and in particular on auditing (e.g.,

Power 2007). Our experimental results could suggest that, to promote cooperation in triadic relations between investors, trustees and third parties, these last should be perceived by the others as fully committed to the evaluation they elaborate and transmit.

Of course, this last point requires further elaborations and some new experimental treatments. Further treatments should integrate the previous ones by: (i) adding economic incentives to the third party, respectively aligned to investors and trustees pay-offs; (ii) providing a fixed structure of the roles. These two modifications will allow us to verify whether economic incentives can strengthen or blunt the reputation-based trust mechanisms described in this paper, and whether a perceived fixed structure of roles, i.e., a functional specialisation of the figures involved in the game, can provide more robust social order-like outcomes.

Acknowledgements

Financial support was provided by a FIRB 2003 grant (Strategic Program on Human, Economic and Social Sciences) from the Italian Ministry for the University and the Scientific Research (SOCRATE Research Project, coordinated by Rosaria Conte, Protocol: RBNE03Y338_002). A first version of this paper has been presented at a seminar held at the Department of Sociology, Utrecht University in February 2009 and at the ICORE-International Conference on Reputation: Theory and Technology, held in Italy, in March 2009. We would like to thank Vincent Buskens, Rense Corten and three anonymous conference referees for very helpful remarks and comments. Usual disclaimers apply.

References

- Barrera, D.: The social mechanisms of trust. *Sociologica*, 2, doi: 10.2383/27728 (2008)
- Berg, J., Dickhaut, J., & McCabe, K. A.: Trust, reciprocity and social history. *Games and Economic Behaviour*, 10, 122-142 (1995)
- Boero, R., Bravo, G., Castellani, M., Squazzoni F.: Reputational clues in repeated trust games. University of Brescia Department of Social Sciences Working Paper SOC 01-08: <http://www.unibs.it/on-line/dss/Home/Inevidenza/PaperdelDipartimento/documento8952.html> (2008)
- Bolton, G. E., Katok, E., and Ockenfels, A.: How effective are electronic reputation mechanisms? An Experimental investigation. *Management Science*, 50, 1587-1602 (2004)
- Bowles, S.: *Microeconomics: behavior, institutions, and evolution*. Princeton: Princeton University Press (2004)
- Bravo, G. and Tamburino, L.: The evolution of trust in non-simultaneous exchange situations. *Rationality and Society*, 20, 1, 85-113 (2008)
- Burt, R.: *Brokerage and closure*. Oxford: Oxford University Press (2005)
- Burt, R. and Knez, M.: Kinds of third-party effects on trust. *Rationality and Society*, 7, 3, 255-292 (1995)
- Buskens, V. and Raub, W.: Embedded trust: control and learning. In Thye S. R. and Lawler E. J. (Eds.), *Group cohesion, trust and solidarity*, Amsterdam: Elsevier Science (2002)
- Camerer, C. F.: *Behavioural game theory. Experiments in strategic interaction*. New York/Princeton: Russell Sage Foundation/Princeton University Press (2003)
- Chang, E., Dillon, T. and Hussain, F. K.: *Trust and reputation for service-oriented environments*. Chichester: Wiley (2006)
- Conte, R. and Paolucci, M.: *Reputation in artificial societies: Social beliefs for social order*. Dordrecht: Kluwer Academic Publishers (2002)
- Cook, C., Hardin, R., and Levi, M.: *Cooperation without trust?*. New York: Russell Sage Foundation (2005)
- Corten, R., and Cook, C.: *Cooperation and Reputation in Dynamic Networks*. Working paper, Utrecht University. http://www.uu.nl/content/cortencooc2008_v090908b.pdf (2008)
- Dunbar, R. I. M.: *Grooming, Gossip and the evolution of language*. Cambridge, MA: Harvard University Press (1996)

- Fehr, E. and Fischbacher, U.: The nature of human altruism. *Nature*, 525, 785-791 (2003)
- Fehr, E. and Gintis, H.: Human motivation and social cooperation: experimental and analytical foundations. *Annual Review of Sociology*, 33, 43-64 (2007)
- Ferrin, D. L., Dirks, K. T. and Shah, P. P.: Direct and indirect effects of third-party relationships on interpersonal trust. *Journal of Applied Psychology*, 91, 4, 970-883 (2006)
- Fischbacher, U.: z-Tree. Zurich toolbox for readymade economic experiments. *Experimental Economics*, 10(2), 171-178 (2007)
- Gintis, H., Bowles, S., Boyd R., Fehr R.: *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge, MA: The MIT Press (2005)
- Granovetter, M.: A theoretical agenda for economic sociology. In Guillen M., Collins R., England P. and Meyer M. (Eds.) *The New Economic Sociology: Developments in an Emerging Field*. New York: Russell Sage Foundation, 35-59 (2002)
- Guennif, S. and Revest, V.: Social structure and reputation: The NASDAQ case study. *Socio-Economic Review*, 3, 417-436 (2005)
- Gui, B. and Sugden, R. (Eds.): *Economics and social interaction. Accounting for interpersonal relations*, Cambridge, MA: Cambridge University Press (2005)
- Hahn, C., Fley, B., Florian, M., Spresny, D., and Fischer, K.: Social reputation: A mechanism for flexible self-regulation in multiagent systems. *JASSS*. 10(1): <<http://jasss.soc.surrey.ac.uk/10/1/2.html>> (2007)
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E.: «Economic man» in cross-cultural perspective: behavioural experiments in 15 small-scale societies. *Behavioural and Brain Sciences*, 28, 795-815 (2005)
- Jackson, K. T.: *Building reputational capital*. Oxford: Oxford University Press (2004)
- Janssen, M.: Evolution of cooperation when feedback to reputation scores is voluntary. *JASSS*. 9(1): <<http://jasss.soc.surrey.ac.uk/9/1/17.html>> (2006)
- Keser, C.: Experimental games for the design of reputation management systems. *IBM Systems Journal*, 42, 498-506 (2003)
- Kuwabara, K.: Affective attachment in electronic markets. A sociological study of eBay. In Nee V. and Swedberg R. (Eds.), *The economic sociology of capitalism*, Princeton, Princeton University Press, 268-288 (2005)
- Nowak, M. A. and Sigmund, K.: The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194, 561-574 (1998a)
- Nowak, M. A. and Sigmund, K.: Evolution of indirect reciprocity by image scoring. *Nature*, 393, 573-577 (1998b)
- Nowak, M. A. and Sigmund, K.: Evolution of indirect reciprocity. *Nature*, 437, 1291-1298 (2005)
- Pelligra, V.: Under trusting eyes: the responsive nature of trust. In Gui and Sugden, 105-124 (2005)
- Piazza, J. and Bering, J. M.: Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behaviour*, 29, 172-178 (2008)
- Rabin, M.: Incorporating fairness into game theory. *The American Economic Review*, 83, 1281-1302 (1993)
- Raub, W. and Weesie, J.: Reputation and efficiency in social interaction: An example of network effect. *American Journal of Sociology*, 3, 626-654 (1990)
- Sabater, J. and Sierra C.: Review on computational trust and reputation models. *Artificial Intelligence Review*, 24, 33-60 (2005)
- Seinen, I. and Schram, A.: Status and group norms: Indirect reciprocity in a helping experiment. *European Economic Review*, 50, 581-602 (2006)
- Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., and Milinski, M.: Gossip as an alternative for direct observation in games of indirect reciprocity. *PNAS* 104(44), 17435-17440 (2007)
- Swedberg, R.: *Sociology and game theory: contemporary and historical perspectives*. *Theory and Society*, 30, 301-315 (2001)
- Zacharia, G. and Maes, P.: Trust management through reputation mechanisms. *Applied Artificial Intelligence*, 14, 881-907 (2000)