***Riccardo Boero*[*], *Giangiacomo Bravo*[†,‡], *Marco Castellani*[†] *and Flaminio Squazzoni*[†]**

Reputation and judgment effects in repeated trust games

DSS PAPERS SOC 01–08

* *Dipartimento di Scienze Economiche e Finanziarie "G. Prato", Università di Torino and GECS — Research Group in Experimental and Computational Sociology*

† *Dipartimento di Studi Sociali, Università di Brescia and GECS — Research Group in Experimental and Computational Sociology*

‡ *Contacting author. Mail address: Dipartimento di Studi Sociali, Via San Faustino 74B, 25122 Brescia, Italy. Email: gbravo@eco.unibs.it*

Contents

Abstract

Reputation is of paramount importance in creating and maintaining prosocial behaviors in large human groups. Evidence of this can be found in both theoretical models and empirical studies. Reputation is especially relevant in trust-based interactions, which are typically formalized in experimental studies using the trust game (also known as investment game). In this paper, we extend the scope of previous experimental studies based on the trust game by creating new treatments where both first and second players have the possibility to rate their opponents' behavior and to know their past ratings. Our results show that being rated by other players and letting this rating be known represent factors that significantly increases cooperation levels. More generally, the importance of reputation is so deeply rooted in human psychological mechanisms that subjects tend to build and maintain a high reputational status even when this is neither rational nor explainable by reciprocity motives.

1. Introduction

Reputation is of paramount importance in creating and maintaining prosocial behaviors in large human groups. Evidence of this can be found in both theoretical models (e.g. Nowak and Sigmund, 1998a,b; Pollock and Dugatkin, 1992; Raub and Weesie, 1990) and empirical studies (e.g. Barclay, 2004; Bolton *et al.*, 2005; Engelmann and Fischbacher, 2002; Keser, 2003; Milinski *et al.*, 2002; Seinen and Schram, 2006). From an evolutionary perspective, reputation is often linked with the dynamics of indirect reciprocity (Nowak and Sigmund, 2005). The indirect reciprocity concept, originally introduced by Robert Alexander (1987), plays an important role in explaining cooperation in large groups, where the probability of diadic repeated encounters is low. In this context, strategies based on direct reciprocal altruism (Axelrod, 1984; Trivers, 1971) are often evolutionary unstable (especially when the possibility of making error is taken into account) and cannot explain a significant part of the cooperation empirically found in nature (Nowak and Sigmund, 1998a; Pollock and Dugatkin, 1992). This problem is especially relevant for human beings, that usually cooperate in groups of unrelated individuals larger than other primates (Fehr and Fischbacher, 2003; Gintis *et al.*, 2003). On the other hand, both analytical and simulation models show that indirect reciprocity can explain the development of cooperation in fairly large populations even under random matching conditions (Nowak and Sigmund 1998a,b; but Leimar and Hammerstein 2001).

Nowak and Sigmund (1998b) influential model showed that a fundamental requisite to exploit indirect reciprocity strategies is a reliable knowledge of the past behavior of any potential partner. In other words, individuals using indirect reciprocity should base their actions on the *reputation* of others (Nowak and Sigmund, 2005). Reputation can be based on the direct observation of others' actions, but also on the exploitation of any other source of information that is available in social life. For instance, gossip plays an important role in transmitting reputational information among individuals in human societies (see Dunbar, 1996). A recent work shows indeed that gossip has a strong influence on the behavior of experimental subjects even when they are able to use also other sources of information, including direct observation (Sommerfeld *et al.*, 2007). The relevance of direct observations and gossip for the sustainability of cooperation in large social groups has been also emphasized in many simulation models (e.g. Conte and Paolucci, 2002).

If reputation is so important in maintaining cooperation in large groups, it is likely that human 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 in 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 other's behaviors and, more generally, the cooperation inside each group (Fehr and Fischbacher, 2003; Gintis *et al.*, 2003; Richerson *et al.*, 2003).

Given this sensitivity, our general hypothesis is that human beings should react to reputational opportunities even when this is little rational in terms of future expected utility. For instance, a previous experimental work showed that even subtle cues suggesting the possibility of being observed, like the presence of stylized eyespots on the computer desktops used for the experimental sessions, are able to significantly increase the generosity of players in a dictator game despite no differences in actual anonymity (Haley and Fessler, 2005). This result can easily be interpreted as the reaction of an unconscious psychological mechanism to the (actually false) possibility that somebody else was watching the subject's actions and, hence, to the possibility of a modification in the reputational status of the players. A corollary of this hypothesis is that subjects should also react promptly to others' reputations: both when 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, *and* when there are no rational self-interested motives, e.g. in the case of the so called altruistic "punishment" (Barclay, 2006; Boyd *et al.*, 2003; Fehr and Fischbacher, 2004; Fehr and Gächter, 2002). From this point of view, it is especially interesting to notice that recent studies showed the existence of neural basis for the altruistic punishment of defectors in economic exchanges (de Quervain *et al.*, 2004).

A number of past experimental results are coherent with our hypothesis. In an experiment based on Nowak and Sigmund (1998b) 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 this study, 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 $c > b$ to a recipient. The experiment showed that the donors' decisions is strongly influenced by the possibility of knowing the recipients' past behaviors, with a helping rate of 70% when this was allowed *versus* a 22% one 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 a combination of public good and helping games (Milinski *et al.*, 2002) and on a combination of public good and trust games (Barclay, 2004).

Reputation is especially important in trust situations (Dasgupta, 1999; Ostrom, 1998). 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 send (or "invest") to Player B, who receive the amount sent multiplied for 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 incentives 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).

In repeated interaction situations where players are re-coupled in each period and know the past behavior of their partners, it is worth noting that the trust game presents some analogies with the helping game described above. More specifically, if A players are able to distinguish between trustworthy and untrustworthy Bs, they can use discriminating strategies by sending a significant proportion of their endowments only to the B players that are in a good standing. In other words, besides the direct reciprocity strategies typical of the trust game — where A trusts B by investing a significant proportion of the endowment and B repays A's trust by returning a fair quota of what he/she received — in a repeated interaction setting players can also employ indirect reciprocity strategies, where A trusts B because B played fair with another subject. Notice that this is exactly the discriminating strategy highlighted by Nowak and Sigmund (1998b) model. Direct and indirect reciprocity strategies are actually more complementary that alternative in this setting. Using an evolutionary model based on a binary version of the repeated trust game, Bravo and Tamburino (2008) showed indeed that the interplay between direct and indirect strategies can lead to the evolution of cooperative agents. More specifically, direct reciprocity is able to support indirect reciprocity when the available information regarding the past behavior of other agents is imperfect, while indirect reciprocity works well also in conditions of continuous re-matching of the interacting partners, where direct reciprocity is difficult to use.

In an experimental work based on a repeated trust game, Claudia Keser (2003) showed that allowing B players to build a reputation increases the overall cooperation levels of the game. It is worth noting that the introduction of a reputation system in the game had 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 has a stronger effect than the reduction of uncertainty for the A ones. From a rational perspective, this result can be explained with the strong incentive in increasing their reputation that B players faced due to their expectation of receiving significantly higher amounts from A players in case of good standing. Although rational calculus may also play a role, we derive from our general hypothesis the prediction that a significant part of B player behavior should not be explicable as a rational investment in reputation, but only in the light of human sensitivity for reputation. Moreover, we predict that subjects should react to others' reputational status even when this has no rational self-interest meaning at all.

In order to test those predictions, we extended Keser's experiment by adding two new treatments: one where B is allowed to rate A's behavior and one where both A and B can rate their opponents. We expect that the subjects in both A and B positions react to the simple fact of being under rating and that they punish who is in a bad standing by sending him/her significantly lower amounts. Our result are consistent with those predictions. More generally, the importance of reputation is so deeply rooted in human psychological mechanisms that subjects tend to build and maintain a high reputational status even when this is neither rational nor explainable by reciprocity motives. The paper is organized as follows: section 2. describes the experimental setting, section 3. presents the results, while section 4. discusses them.

2. Methods

One hundred and twenty subjects participated to the experiment in three groups of 40 individuals. Subjects were students, 58 females and 62 males, recruited through public announcements within different faculties of the University of Brescia. The experiment took place in a single day in the computer laboratory of the Faculty of Economics, which is equipped with the experimental software z-Tree (Fischbacher, 1999). All interactions took place through the computer network and the subjects had no possibility to identify their partners. The 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). Each group played for nearly two hours and the average earning was 14.35 Euros, including the show-up fee, 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 the 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), having 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 the treatments was as in the baseline one except for the introduction of a "rating" stage at the end of each period. In the first treatment (hereafter indicated as $A \rightarrow B$), A was allowed to rate B behavior as "negative", "neutral", or "positive". Notice 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 has 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 $B \rightarrow A$) was exactly as the first one 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 rating possibilities were "negative", "neutral" and "positive".

In the third treatment (hereafter $A \leftrightarrow B$), both A an B players were allowed to rate each other and knew this information in the following period. The main purpose behind the design of this further treatment is to investigate whether the introduction of a two-ways reputation system can lead to higher cooperation levels than the sum of the two one-way rating schemes.

3. Results

Table 1 presents the overall results of the experiment. Both the baseline and the first treatment ($A \rightarrow B$) produced results that are consistent with previous experiments based on the trust game (e.g. Berg *et al.*, 1995; Ortmann *et al.*, 2000), including the Keser (2003) one presented above. Both A absolute average investment and B absolute average return in the first treatment are significantly higher than in the baseline (Mann Whitney U test, $z = -2.766$, $p = 0.006$ two tailed, and $z = -8.346$, $p < 0.001$ two tailed, respectively). Also considering B relative returns (i.e. the proportion of the total amount received by B that he/she returned to A) leads to a significant difference with the baseline (Mann Whitney U test, $z = -9.367$, $p < 0.001$ two tailed). More specifically, A average investment increased by 12.2% while B average return increased by 62.4% in absolute terms and by 41.7% relatively to the received amount.

It is worth noting that A players discriminated among the B ones depending on their reputation. Table 2 shows the average A investment as a function of B ratings. A positive rating leads to a significant higher investment relatively to the "unknown" case (Mann Whitney U test, $z = -5.674$, $p < 0.001$ two tailed), while a negative one to a significant lower investment (Mann Whitney U test, $z = -4.730$, $p < 0.001$ two tailed). On the other hand, there is no significant difference in investments when B has a neutral rating relatively to the case where he/she has not yet been rated (Mann Whitney U test, $z = -0.654$, $p = 0.513$ two tailed).

Finally, B returns tend to increase for the first 15-20 periods and subsequently declines, with a pronounced end effect in the last three periods (Fig. 1b), a fact that is coherent with the idea that reputation has little impact when the probability of

	A investment (ECU)	B return (ECU)	B return (proportion)
Baseline	3.91	3.88	0.16
	(2.67)	(4.46)	(0.16)
Treatment 1 ($A \rightarrow B$)	4.39	6.30	0.25
	(2.84)	(5.23)	(0.16)
Treatment 2 ($B \rightarrow A$)	5.61	5.28	0.17
	(2.96)	(5.96)	(0.17)
Treatment 3 ($A \leftrightarrow B$)	5.42	7.25	0.24
	(3.17)	(6.46)	(0.18)

Table 1: Average investments and returns by treatment. Standard deviations are in parenthesis.

B rating	A investment (ECU)	Std. dev.
Unknown	4.12	2.24
Negative	2.44	2.31
Neutral	4.20	2.04
Positive	6.61	2.36

Table 2: A average investment as a function of B rating. An "unknown" rating corresponds to the early periods of the game.

future interaction declines. On the other hand, A investments are approximatively constant during the whole treatment (Fig. 1a).

The second treatment ($B \rightarrow A$), where B was allowed to rate A behavior, produced an *increase* in the average investment and a *decrease* in the average return in comparison with treatment one (Mann Whitney U test, $z = -6.846$, $p < 0.001$ two tailed, and $z = -4.408$, $p < 0.001$ two tailed, respectively) (Tab. 1). Notice also that B relative returns in treatment two are no longer significantly different from the baseline ones (Mann Whitney U test, $z = -0.870$, $p = 0.384$ two tailed). This means that the significant, although small, increase in the absolute amounts returned by B players in comparison with the baseline (Mann Whitney U test, $z = -2.321$, $p = 0.020$ two tailed) is simply a response to the strong increase of A players' investments. This interesting results will be further discussed below.

The behavior of A players appears hard to explain following rational expectations. Why should B players consider the past rating of their partners even if they

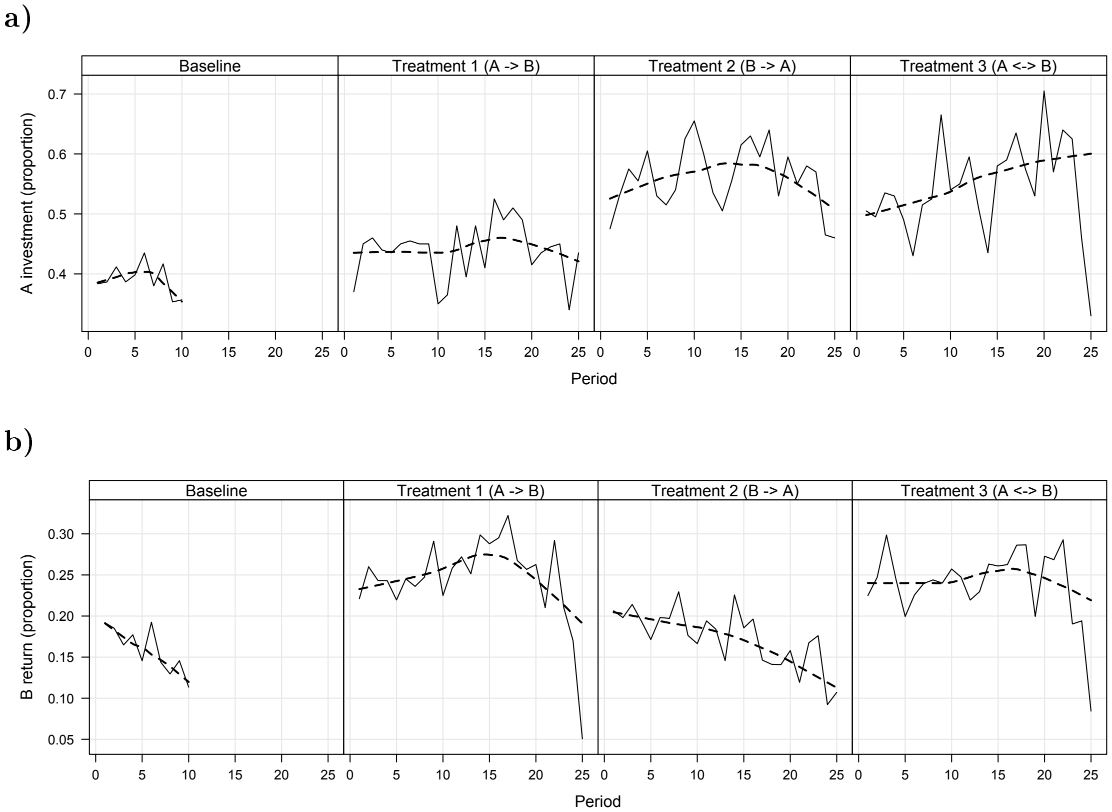

Figure 1: Game dynamic for A investments (a) and B returns (b) in the different treatments. Straight lines represent the actual data, dashed lines their smoothing (loess).

posses the information regarding the actual amount they received? Nevertheless, statistical evidence shows that B players used also A past ratings in their decisions. In order to analyze this point, we employed a regression model for panel data (Baltagi, 2001), where the panel is formed by the observations related to the subjects playing as B for each of the 25 periods of the treatment, the dependent is the amount returned by B, and the independents are A investment and A past rating. We estimated both the fixed and the random effect model and subsequently choose the latter following a Hausman test ($\chi^2_4 = 0.655$, $p = 0.957$). The resulting model is highly significant ($F_{39,495} = 82.624$, $p < 0.001$) and shows a positive and significant effect for the amount invested by A ($z = 15.255$, $p < 0.001$), a negative and significant effect for both a past negative and a past neutral ranking in comparison with an unknown ranking ($z = -2.031$, $p = 0.042$ and $z = -2.640$, $p = 0.008$, respectively), while the effect for a positive ranking is not significant ($z = -1.432$, $p = 0.152$).

The dynamics of this treatment show a slowly decline for B returns during the whole treatment (Fig. 1b), while A investments follow an inverted-U trend similar to the one for B returns in treatment one, although with a less pronounced end effect (Fig. 1a).

The third treatment, where both A and B were allowed to rank the behavior of their partners ($A \leftrightarrow B$), led to a significant increase in investments in comparison with the baseline one (Mann Whitney U test, $z = -8.089$, $p < 0.001$ two tailed). Also the returns are significantly higher than the baseline ones both in absolute and in relative terms (Mann Whitney U test, $z = -8.339$, $p < 0.001$ two tailed, and $z = -7.627$, $p < 0.001$ two tailed, respectively) (Tab. 1). However, while the difference in the invested amounts between the first and the third treatments is also significant (Mann Whitney U test, $z = -5.281$, $p < 0.001$ two tailed), this is not the case for the difference in the returns (Mann Whitney U test, $z = -1.468$, $p = 0.142$ two tailed for the absolute amounts returned and $z = -0.080$, $p = 0.936$ two tailed for the relative ones). Symmetrically, the difference between the second and the third treatment is significant for the returned amounts, both in absolute and in relative terms (Mann Whitney U test, $z = -5.158$, $p < 0.001$ two tailed, and $z = -5.946$, $p < 0.001$ two tailed, respectively), but not for the investments (Mann Whitney U test, $z = -1.263$, $p = 0.207$ two tailed).

The dynamics of this treatment show essentially a slow increase in the investments and a substantial stability of the returns. In both cases, a pronounced end effect is observable, coherently with the idea of a reduction of the interest of a reputation investment in the last periods of interaction (Fig. 1).

4. Discussion

Coherently with our hypothesis, the experiment shows that humans are highly sensitive to their own reputational status and react promptly to others' judgments even when this should not have any actual effect on their payoffs. The amounts sent are indeed systematically higher when subjects are under rating. This result is robust across treatments and independent from the position held by the subjects in the game. While for B players it may be explained as a rational investment in reputation (e.g. in the first treatment and in Keser 2003 experiment), the same motive cannot be easily advanced for A players (e.g. in the second treatment). The trust game is indeed asymmetric: while A players enjoy an actual benefit from knowing B ratings, since this information can help in reducing uncertainty on B behaviors, the opposite is not true. The decision for B players is actually analogous to the one faced by subjects playing a dictator game. The relative returns found in the baseline trust game are indeed similar to the proportion of the endowment offered by first players in standard dictator games (for a review of dictator experiments, see Camerer, 2003, 48–63). The motivations behind this “judgment effect” should therefore be others than a simple investment in reputation, at least for A players. Moreover, if both A and B players were only motivated by a rational investment in their own reputations, they should increase their investment/return levels as soon as the reputational opportunity arises (i.e. from the first period of the treatment that followed the baseline). However, neither the returns in the first period of the first treatment nor the investments in the first period of the second treatment are significantly different from the baseline first period ones (Mann Whitney U test, $z = -0.748$, $p = 0.454$ two tailed, and $z = -1.094$, $p = 0.274$ two tailed, respectively). The second period returns of the first treatment and the second period investments of the second treatment are instead significantly higher than the corresponding baseline ones (Mann Whitney U test, $z = -2.826$, $p = 0.005$ two tailed, and $z = -2.162$, $p = 0.031$ two tailed, respectively). This result suggests that the subjects do not rationally plan to invest in their own reputation. They react instead promptly to others' judgments trying to maintain a sufficient reputational status.

It can be argued that also A players have some rational reasons to invest in their own reputation, since B players actually return slightly lower amounts to subjects in bad standing. However, this just shifts the question to why B players should give any weight to A players' reputations. From our point of view this is a form of “altruistic punishment” that B players impose on the A ones without a positive reputation. This result is consistent both with our predictions and with past works on altruistic punishment (Barclay, 2006; Fehr and Fischbacher, 2004; Fehr and Gäcther, 2002; de Quervain *et al.*, 2004). However, an alternative hypothesis able to explain the observed behavior is that subjects simply took advantage of As' bad standing in order to keep a larger share of their endowment, without being really concerned by punishing them. Unfortunately, the present data do not permit to discriminate between the two hypotheses, but the design of new experiments able to shed light on this may represent an interesting extension to the present work.

Finally, it is worth noting that the “judgment effect” apparently crowds out even the effect of incertitude reduction for A players. The amount invested by A players in the third treatment is indeed very close to the amount invested in the second treatment (actually somewhat lower, but the difference is not significant) where only B players were allowed to rate the A ones (see Table 1). This implies that the difference in A investments existing between the baseline and the first treatment — due to the fact that A players could rationally trust the B ones having a “positive” rating — disappears in the third treatment. In the latter case, only the two-way effect of rating still produces significant effects on the observed behaviors, which are indeed similar to the ones found in the second treatment for A players and in the first one for B players.

Summarizing, our results are consistent with the hypothesis that reputation motives represent themselves a significant force able to explain human behavior and with the related idea that human beings use forms of altruistic punishment against subject having bad reputational status (even if we were not able to completely rule out the possibility of the presence of self-interested motives in this second case). From an evolutionary point of view, the sensitivity for the reputational status is probably part of the human “tribal social instincts” (Richerson and Boyd, 2001), i.e. the set of emotions and cognitive mechanisms, tuned to rise group cohesion and collective action, that are already present in other primates (e.g. Manson and Wrangham, 1991) but represents one of the key adaptations for social life of our species (Richerson and Boyd, 2001; Richerson *et al.*, 2003). The link existing between reputation and cooperation flows clearly through the indirect reciprocity mechanism (Alexander, 1987; Nowak and Sigmund, 1998a,b), which permits high levels of cooperation even in large groups where direct reciprocity suffers from significant shortcuts (Bravo and Tamburino, 2008; Nowak and Sigmund, 2005). The strong concern for the reputational status, even when it has no actual effect on individuals’ payoffs, represents indeed a mechanism that is able to explain a significant part of human behavior and possesses both evolutionary meaning and empirical plausibility.

Acknowledgments Financial support was provided by a FIRB (Strategic Program on Human, Economic and Social Sciences) grant from the Italian Ministry for the University and the Scientific Research. We would like to thank our partners in the SOCRATE research project: ISTC-CNR Rome, LIUC “Carlo Cattaneo” University, University of L’Aquila, ThinkinGolem, and Democenter. Finally, we would like to thank Rosaria Conte (CNR, Rome). Our interest in reputation issues has been strongly influenced by her insights and suggestions. Usual disclaimers apply.

References

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

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

Baltagi, B. H. (2001). *The Econometrics of Panel Data (Second Edition).* 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.

Barclay, P. (2006). Reputational Benefits for Altruistic Punishment. *Evolution and Human Behavior*, 27: 325–344.

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

Bolton, G. E., Katok, E. and Ockenfels, A. (2005). Cooperation Among Strangers With Limited Information About Reputation. *Journal of Public Economics*, 89: 1457–1468.

Boyd, R., Gintis, H., Bowles, S. and Richerson, P. J. (2003). The Evolution of Altruistic Punishment. *Proceedings of the National Academy of Science USA*, 100: 3531–3535.

Bravo, G. and Tamburino, L. (2008). The Evolution of Trust in Non-Simultaneous Exchange Situations. *Rationality and Society*, 20: 85–113.

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

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

Dasgupta, P. (1999). Economic Progress and the Idea of Social capital. In Dasgupta, P. and Serageldin, I., eds., *Social Capital. A Multifaceted Perspective*, pp. 325–424. The World Bank, Washington DC.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. and Fehr, E. (2004). The Neural Basis of Altruistic Punishment. *Science*, 305: 1254–1258.

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

Engelmann, D. and Fischbacher, U. (2002). Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game. University of Zurich, Institute for Empirical Research in Economics, Working Paper 132.

Fehr, E. and Fischbacher, U. (2003). The Nature of Human Altruism. *Nature*, 525: 785–791.

Fehr, E. and Fischbacher, U. (2004). Third-Party Punishment and Social Norms. *Evolution and Human Behavior*, 25: 63–87.

Fehr, E. and Gächter, S. (2002). Altruistic Punishment in Humans. *Nature*, 415: 137–140.

Fehr, E., Kirchsteiger, G. and Riedl, A. (1998). Gift Exchange and Reciprocity in Competitive Experimental Markets. *European Economic Review*, 42: 1–34.

Fischbacher, U. (1999). z-Tree. Zurich toolbox for readymade economic experiments. University of Zurich, Working paper no. 21.

Gintis, H., Bowles, S., Boyd, R. and Fehr, E. (2003). Explaining Altruistic Behavior in Humans. *Evolution and Human Behavior*, 24: 153–172.

Haley, K. J. and Fessler, D. M. (2005). Nobody's watching? Subtle Cues Affect Generosity in an Anonymous Economic Game. *Evolution and Human Behavior*, 26: 245–256.

Keser, C. (2003). Experimental Games for the Design of Reputation Management Systems. *IBM Systems Journal*, 42: 498–506.

Leimar, O. and Hammerstein, P. (2001). Evolution of Cooperation Through Indirect Reciprocity. *Proceedings of the Royal Society London*, 268: 745–753.

Manson, J. H. and Wrangham, R. W. (1991). Intergroup aggression in chimpanzees. *Current Anthropology*, 32: 369–390.

McCabe, K. A. and Smith, V. L. (2000). A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Predictions. *Proceedings of the National Academy of Sciences USA*, 97: 3777–3781.

Milinski, M., Semmann, D. and Krambeck, H.-J. (2002). Reputation Helps Solve The Tragedy of the Commons. *Nature*, 415: 424–426.

Nowak, M. A. and Sigmund, K. (1998a). The Dynamics of Indirect Reciprocity. *Journal of Theoretical Biology*, 194: 561–574.

Nowak, M. A. and Sigmund, K. (1998b). Evolution of Indirect Reciprocity by Image Scoring. *Nature*, 393: 573–577.

Nowak, M. A. and Sigmund, K. (2005). Evolution of Indirect Reciprocity. *Nature*, 437: 1291–1298.

Ortmann, A., Fitzgerald, J. and Boeing, C. (2000). Trust, Reciprocity, and Social History: A Re-examination. *Experimental Economics*, 3: 81–100.

Ostrom, E. (1998). A Behavioral Approach to the Rational Choice Theory of Collective Action. *American Political Science Review*, 92: 1–22.

Pollock, G. B. and Dugatkin, L. A. (1992). Reciprocity and the Evolution of Reputation. *Journal of Theoretical Biology*, 48: 25–37.

Raub, W. and Weesie, J. (1990). Reputation and Efficiency in Social Interactions: An Example of Network Effects. *The American Journal of Sociology*, 96: 626–654.

Richerson, P. J. and Boyd, R. (2001). The Biology of Commitment to Groups: A Tribal Instincts Hypothesis. In Nesse, R., ed., *Evolution and the Capacity for Commitment*, pp. 186–220. Russell Sage Foundation, New York.

Richerson, P. J., Boyd, R. T. and Henrich, J. (2003). Cultural Evolution of Human Cooperation. In Hammerstein, P., ed., *Genetic and Cultural Evolution of Cooperation*, pp. 357–388. The MIT Press, Cambridge.

Seinen, I. and 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., and Milinski, M. (2007). Gossip as an Alternative for Direct Observation in Games of Indirect Reciprocity. *Proceedings of the National Academy of Science USA*, 104: 17435–17440.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *Quarterly Review of Biology*, 46: 35–57.