Centre de Recherche en Économie et Management
Center for Research in Economics and Management

University of Caen

University of Rennes 1

Working Paper

Reputation and Social (Dis)approval in Feedback Mechanisms: An Experimental study

Marianne Lumeau
University of Paris 13, CEPN and Labex ICCA, France

David Masclet
CREM UMR CNRS 6211, Université de Rennes 1, France

Thierry Pénard
Université de Rennes 1, CREM UMR CNRS 6211, France

December 2013 - WP 2013-43

Reputation and Social (Dis)approval in Feedback Mechanisms: An Experimental study

Marianne Lumeau, David Masclet and Thierry Pénard

December 2013

Abstract Several studies have highlighted the role of feedback mechanisms in the success of electronic marketplaces. In this current experiment, we attempt to isolate experimentally the role of reputation and social (dis)approval associated to ratings using a trust game experiment with the opportunity to rate one's partner (Keser, 2003; Masclet and Penard, 2012). For this purpose we compare two experimental feedback systems that differ in the information that is publically available to participants. In a first feedback system, individuals' rating profiles are public whereas in the second feedback system this information is private. Our findings indicate that both private and public ratings improve cooperation. However, we observe that private feedbacks are less efficient in enhancing trust and trustworthiness than public systems. This is mainly due to fact that fewer ratings are assigned in the private feedback system than in the public system. Altogether these findings suggest that, even if social (dis)approval matters, publicly observed feedback remains crucial to induce honest behaviors and improve efficiency on markets characterized by imperfect information.

Keywords: Reputation, (dis)approval, Experiment, Trust Game

Contact: Marianne Lumeau – University of Paris 13/CEPN and Labex ICCA (MSH Paris Nord – 4 rue de la Croix Faron – 93 579 La plaine Saint-Denis – France). marianne.lumeau@univ-paris13.fr

David Masclet – University of Rennes1/CREM, CNRS (7 place Hoche – 35 065 Rennes Cedex – France). david.masclet@univ-rennes1.fr

Thierry Penard – University of Rennes1/CREM (7 place Hoche – 35 065 Rennes Cedex – France). thierry.penard@univ-rennes1.fr

Acknowledgments: The authors thank the participants of the Workshop on Conflicts in Rennes, of the ASFEE in Fort de France and of the ESA in New-York for helpful comments. We also thank Elven Priour for programming the experiment. Financial support from the *MSHB* project is gratefully acknowledge

1. Introduction

Several studies have highlighted the role of feedback mechanisms in the success of electronic marketplaces such as eBay or AmazonMarketPlace (Cabral and Hortaçsu, 2010; Dellarocas, 2003, 2006; Resnick and al., 2006). Feedback systems would promote trust among anonymous traders by offering the opportunity for traders to rate each other and providing information about each partner's past transactions. For instance, after a transaction on eBay, the buyer and the seller have the possibility to evaluate their partner and all received ratings are recorded in a feedback profile that is publicly available.

In this current paper, we conjecture that the effectiveness of online feedback systems is due to the presence of two components in the rating system: a "reputational" component and a "(dis)approval" component. The first component refers to the idea that received ratings will affect the trader's reputation and therefore her future *monetary* benefits. Consequently, each participant may anticipate that adopting an opportunistic (cooperative) behavior may lead to a negative (positive) rating that will be publicly available on her profile, hence inducing a reputational loss (gain). Indeed, many studies have shown that for a seller a high number of positive ratings increases her probability of selling items at a higher price (Bajari and Hortacsu, 2004; Lucking-Reiley and al., 2007; Houser and Wooders, 2006). This may strongly incite the traders to adopt a cooperative behavior on these online marketplaces. We will refer to this effect, as the *Reputational Response Hypothesis* (RRH).

In addition to this "reputational" component, we conjecture that receiving a negative (positive) rating may induce a *non-monetary* (dis)utility. This may be the case if the individual is sensitive to (dis)approval expressed by others through their ratings. There are several reasons to believe that this might be the case. The fact that expressions of approval and disapproval are commonly observed in human interaction suggests that they must influence the behavior of at least some individuals. In recognition of the importance of informal sanctions, economists have integrated phenomena such as peer pressure (Kandel and Lazear, 1992; Barron and Paulson-Gjerde, 1997) and the avoidance of social disapproval (Hollaender, 1990; Akerlof, 1980; Lindbeck et al., 1999) into theoretical models. More recently, several experiments have shown that, in various contexts, individuals do not only respond to the receipt of monetary sanctions (or rewards) but are also sensitive to social (dis)approval of others (e.g. Gaechter and Fehr, 1999; Masclet and al., 2003; Rege and Telle, 2004; Sefton and al., 2007; Dugar, 2010). We will refer to this explanation as the *(dis)approval Response Hypothesis* (ARH).

In this current experiment, we attempt to isolate experimentally the role of reputation and social (dis)approval associated to ratings using a trust game experiment with the opportunity to rate one's partner (Keser, 2003; Masclet and Penard, 2012). For this purpose we compare two experimental feedback systems that differ in the information that is publically available to participants. In a first feedback system, individuals' rating profiles are public whereas in the second feedback system this information is private. The comparison of the two systems should allow us to isolate the role played by reputation and social (dis)approval in ratings.

Our experiment consists of three treatments. Our baseline treatment, called *No Feedback,* is a finitely repeated trust game inspired from the trust game designed by Berg and al. (1995).[1] The second treatment, called *Public Feedback,* consists of a two-stage game. After the trust-game similar to the baseline, participants observe the decision of their partner and have the opportunity to assign either a negative (-1) or a positive (+1) rating point to their counterpart. In this treatment it is common knowledge that each rating point is recorded on the partner's feedback profile. Moreover, at the beginning of each period, participants can observe their counterpart's profile, so that each participant is aware of his or her partner's reputation. The third treatment, called *Private Feedback,* is similar to the previous treatment with the notable difference that assigned ratings are private information. In other words, feedback profiles are never available to the subsequent partners.

To anticipate our findings, we observe that both private and public feedbacks improve trust and trustworthiness compared to our baseline which suggests that not only the reputational component but also the approval and disapproval component of ratings matter. However, we observe that private feedbacks are less efficient in enhancing cooperation than public systems. This is mainly due to fact that fewer ratings are assigned in the private feedback system than in the public system. Altogether these findings suggest that, even if social (dis)approval matters, publicly observed feedback remains crucial to induce honest behaviors and improve efficiency on markets characterized by imperfect information.

This paper is organized as follows. The next section reviews previous experiments related to the role of reputation and the role of social (dis)approval. Section 3 describes the experiment. Section 4 presents our main findings. Section 5 discusses and concludes.

2. Previous literature

Our paper draws upon two distinct strands of the economic literature. The first strand of literature consists of experimental papers that analyze the effectiveness of decentralized feedback systems in the context of trust game or buyer-seller games. The second strand of literature concerns papers that investigate the role played by reputation or social (dis)approval considerations on trust and cooperation.

[1] This game provides a good abstraction of the context in which transactions occur in electronic marketplaces like eBay. In such an environment, the buyer makes payment to the seller, in return for the promise of receiving the purchased item. The buyer is therefore required to trust the seller, who in turn can elect to be honest or, conversely, opportunistic by not delivering the item or by sending an item that does not correspond to that listed in the auction description.

Keser (2003) implemented a one-sided feedback mechanism where Player A has the possibility to rate Player B at the end of the trust game. Her results indicate that such a feedback mechanism improves trust and trustworthiness. These findings were replicated by Chen and Hogg (2005) and Masclet and Penard (2012) with two-sided feedback mechanisms in which both traders can evaluate their counterpart. Masclet and Penard also tested the robustness of these findings to the introduction of a cost for posting a rating and showed that individuals do not hesitate to rate their partner even with this cost. Using a buyer-seller game[2], Bolton and al. (2013), Li and Xiao (2010) or Gazzale and Khopkar (2011) also showed that introducing a feedback mechanism significantly improves transaction efficiency. Some experimental studies have also investigated the effect of reputation on trust. For instance, Di Cagno and Sciubba (2010) showed that individuals are more cooperative when they know that their behaviors will be publicly observable. By introducing a second stage of network formation after a trust game, the authors found higher trust and trustworthiness levels than in a standard trust game. Bolton and al. (2004) compared a buyer-seller game played repeatedly by strangers to a variant treatment in which buyers are perfectly informed about sellers' past behavior. They showed evidence of sellers' concerns for their reputation. Similar results are obtained by Bohnet and al. (2005), and Huck and Lünser (2010) who consider different informational settings.[3] However to our knowledge, none of these studies have attempted to disentangle the "reputational" and "(dis)approval" components of rating systems.

Other experiments have investigated the role played by social (dis)approval on trust, irrespective of reputation concerns. In the context of voluntary contribution mechanisms, Masclet and al. (2003) introduced the opportunity for players to express their disapproval by assigning negative points to their partners. Their results indicate that such a mechanism has a positive impact on the level of cooperation. However the authors also observe that long-term contributions are significantly lower compared to a monetary punishment system (see also Noussair and Tucker, 2005). Dugar (2010) extended the analysis of Masclet et al. (2003) to the case where individuals have also the opportunity to assign points of approval. The author replicates Masclet et al. (2003)'s findings and find that disapproval was more efficient than approval. These findings are generally interpreted in term of emotional response to received (dis)approval. (Coricelli and al., 2010; Hopfensitz and Reuben, 2009; Joffili and al., 2013). Using physiological measures, these studies show that negative emotions are generally associated to the fact of receiving sanctions that have positive influence on future cooperation.

---

[2] The buyer-seller game is a variant of the trust game in which the buyer decides whether to buy an item or not (i.e., to trust the seller) and the seller decides whether to ship or not the item to the buyer (i.e., to be trustworthy).

[3] Bohnet and al. (2005) consider three different treatments: only buyers have access to sellers' past history, other sellers have access to this information, and finally both buyers and sellers can consult sellers' past history. Huck and Lünser (2010) compare a situation in which buyers are informed on past behaviors of all sellers at the previous period with a situation in which buyers are only informed on the past behavior of his partner.

3. The experiment

3.1. Experimental design

Our experiment consists of three treatments based on a modified version of the trust game originated by Berg and al. (1995).

The No Feedback treatment

The *No Feedback* treatment is a simultaneous version of the trust game repeated during20 periods. Precisely, at the beginning of each period, a player A and a player B are randomly matched together and play a trust game. The trust game is played as follows. Let $I = \{1,2\}$ stand for a group of two players referred respectively as players of types A and B. At the beginning of the game, each player $i \in I$ is exogenously endowed with 10 experimental units. Player A decides how much of her/his endowment to send to player B. Any amount sent by player A – denoted s_A (with $s_A \in \{0,1,2,\ldots,10\}$) – is tripled by the experimenter, such that player B receives $3s_A$. Player B then decides how much to return to player A, with s_B the amount returned by player B (with $s_B \in \{0,1,\ldots,3s_A\}$) and r_B the proportion returned (with $r_B = \frac{s_B}{3s_A} \times 100$). The amount sent by player A and the proportion returned by player B can be seen as measures of trusty and trustworthy behavior, respectively.[4] Players A and B' payoffs are respectively given by:

$$\pi_A(s_A, s_B) = 10 - s_A + s_B \text{ (1a) and } \pi_B(s_A, s_B) = 10 + 3s_A - s_B \text{(1b)}$$

The theoretical prediction of the trust game can be easily found by backward induction. Assuming selfish preferences, the subgame perfect equilibrium is straightforward: anticipating that player B will never return a positive amount, a rational player A will always send nothing to player B. Players dominant strategies are thus $s_i = 0$, and the payoffs are therefore $\pi_A(0,0) = 10$ and $\pi_B(0,0) = 10$. Obviously, this game-theoretical solution is Pareto inefficient since a positive amount sent by player A could lead to a Pareto improvement and the social optimum would occur if and only if $s_A = 10$. Since the trust game is finitely repeated under a stranger matching protocol, the subgame perfect equilibrium applies in each period.[5]

[4] These measures are imperfect as individuals can have other motivations behind their decision (for instance, increase the total payoff, aversion to inequality …).

[5] The fact that the trust game is played simultaneously in our experiment does not change the theoretical solution (Brandts and Charness, 2011). A simultaneous trust game means that players A and B play at the same time: player A chooses the amount to send to player B, while, at the same time, player B determines the amount to return for all potential amounts received from A. Such a procedure is justified in treatments with a second stage of rating since players A and B are in a symmetrical position when they have to decide whether to rate or not their partner.

The Public Feedback treatment

In this treatment, a second stage is added in which both players can rate their partners, being informed that rating will be publicly available. During this second stage, each participant observes the amount sent or returned by her/his partner and can assign either a negative (-1) or positive $(+1)$ rating point. We denote the rating decision of each player i by e_i ($e_i \in \{-1, 0, +1\}$). Assigning point is costly for the rater who incurs a direct monetary cost of 1 ECU (corresponding to 1/10 of initial endowment). This cost may be interpreted as the opportunity cost of rating, measured by the amount of time and effort devoted to this task (see Masclet and Penard 2012 for a discussion on this point). In our experiment, the target does not incur any direct monetary cost or benefit from receiving ratings. However she/he will face an indirect reputational cost or gain since ratings are recorded on her/his feedback profile and are perfectly observable by future partners during subsequent periods. Precisely, at the beginning of each period, each player is aware of the feedback profile of her/his partner. This information can be used to make inference about her/his reputation: a large number of positive ratings can be a signal of good reputation. By sharp contrast, a partner that has received several negative ratings can be perceived as untrustworthy and may suffer from a bad reputation. In addition to this reputational gain (loss), the rated player may also incur another indirect gain (cost) in term of social (dis)approval since ratings are a way to express one's (dis)approval (Masclet et al. 2003; Dugar, 2010). The payoffs functions in this treatment are as follow:

$\pi_A(s_A, s_B, e_A) = 10 - s_A + s_B - c_A$ (2a) and $\pi_B(s_A, s_B, e_B) = 10 + 3s_A - s_B - c_B$ (2b)

where $c_i = \begin{cases} 1\ if & e_i \in \{-1,1\} \\ 0\ if & e_i = 0 \end{cases}$

By working backward, it can be easily seen that in the second stage, players should never rate their partner since posting a rating is costly. In the absence of rating, the *Public Feedback* treatment is therefore similar to the trust game described before (i.e., the trust game repeated a finite number of time) with $s_i = 0$ and $e_i = 0$ in each period. Players' payoffs are therefore $\pi_A(0,0,0) = 10$, $\pi_B(0,0,0) = 10$.[6]

The Private Feedback treatment

The *Private Feedback* treatment is similar to the *Public Feedback* treatment, except that the player's feedback profile is not observable by her/his partner. Consequently, there is no reputational cost or gain when players receive some ratings. Only the (dis)approval role of ratings should be effective in this treatment.

[6] Note that introducing a cost in the decision to rate allows us to compare our three treatments with identical theoretical prediction.

From a theoretical point of view, whether ratings are observable or not should not affect predictions since in both cases, players should never assign ratings in the second stage since evaluation is costly for the rater.

3.2. Behavioral Hypotheses

To illustrate what we expect to happen in our experiment, we present in this sub-section some behavioral hypotheses. Consider first our baseline, i.e. the *No Feedback* treatment. We may relax our hypothesis that individuals are only interested in maximizing their own payoff and consider that they have some social preferences but also beliefs regarding others' social preferences. For instance, player A may send a positive amount to player B because she/he has other-regarding social preferences such as altruism. However, since trust is mainly a matter of beliefs that one individual has about the behaviors of others, even if player A has self-regarding preferences, she/he may still be incited to send a positive amount if she/he believes that player B has herself other-regarding preferences or will reciprocate positively (Cox, 2004). Concerning player B, she/he will be incited to return a positive amount to player A if she/he has other regarding preferences or if she/he is motivated by positive reciprocity. Previous experiments have provided strong evidence that players A do not hesitate to send positive amounts and that players B respond by reciprocating positively, despite the absence of repetition with the same partner in the game (see for instance Anderhub et al. 2002; Cochard et al. 2004; Engle-Warnick and Slonim 2004). Based on these previous observations we can write the following hypothesis.

Hythothesis 1*: Players A should send positive amounts to players B while players B should return positive amounts in the Baseline*.

Let's now consider the *Public Feedback* treatment. We conjecture that introducing a rating mechanism – where ratings are publicly observable – should improve transfers from both players compared to the baseline treatment. Indeed feedback systems may affect decisions in two different ways. First, displaying reputation feedback profiles might help players infer their partners' intentions. Many studies in different experimental contexts have shown the influence of releasing the past history of individual players' decisions. For instance, Berg and al. (1995) find that the provision of social history (i.e., information on the amounts sent and returned in previous experimental sessions) significantly increases amounts sent and returned. In related studies, Keser (2003) and Masclet and Penard (2012) find that the introduction of a reputation feedback system increases the overall efficiency by improving both the levels of trust and trustworthiness. Similar findings are obtained by Bolton and al. (2004), Bohnet and al. (2005) and Huck and Lünser (2010) with perfect reputation mechanisms. They also showed that players condition their choice on their partner's reputation that provides incentives to cooperate in order to avoid (benefit from) the future monetary consequences of having a bad (good) reputation. Based on these previous studies we conjecture that:

Hythothesis 2 *(Reputational Response Hypothesis): Both players A and B may be incited to cooperate to avoid (get) a bad (good) reputation, that may reduce (increase) their future trade opportunities. As a consequence both players A and B' amounts sent or returned should be higher in the Public Feedback treatment than in the Baseline.*

Second, the attribution of positive (negative) ratings may play a rewarding (disciplining) role as they vehicle social (dis)approval (Gaechter and Fehr, 1999; Masclet and al., 2003; Rege and Telle, 2004; Sefton and al., 2007; Dugar, 2010). Based on these findings we conjecture that the opportunity of evaluating one's partner in the public feedback treatment may have positive and significant effects on trust and trustworthiness. This is summarized in the third hypothesis.

Hythothesis 3 *((Dis)Approval Response Hypothesis): Both players A and B may be incited to cooperate to receive (avoid) social (dis)approval associated to positive (negative) rating. This should lead to higher amounts sent or returned by both players in the Public Feedback treatment than in the Baseline.*

Our fourth hypothesis concerns the *Private Feedback* treatment. We conjecture that, in absence of observability, only the second effect of ratings, i.e. the social (dis)approval effect may lead people to cooperate. Consequently one should observe lower trust and trustworthiness in the private feedback treatment than in the public feedback treatment. In addition, players may also anticipate that the impact of their ratings will be lower and they should be less incited to demand rating points (see Anderson and Putterman, 2006; Carpenter, 2007; Nikiforakis and Normann, 2008). Consequently one should reasonably expect to observe fewer ratings in the Private Feedback treatment than in the Public Feedback treatment. Altogether, this implies that cooperation should be lower in the Private Feedback treatment than in the Public Feedback treatment. This conjecture is summarized as follow:

Hythothesis 4*: One should observe lower amounts sent or returned in the Private Feedback treatment than in the Public Feedback treatment. This is due to the fact that i) fewer ratings are assigned in the Private Feedback treatment and ii) ratings have a lower impact on future trust and trustworthiness levels.*

3.3. Procedure

All sessions were held at the Center for Research in Economics and Management (CREM), University Rennes 1, France. The experiment was computerized using the Z-tree program (Fischbacher, 2007) and consisted in 19 sessions, summarized in Table 1. We used a between-subject design. In total, 182 subjects – between eight and ten per session – were recruited among a population of undergraduate students from a variety of majors. All of them have never played a trust game. On average a session lasted 110 minutes, including initial instructions and subject payment.

At the beginning of the experiment, the instructions were distributed and read to the subjects.[7] Each session consisted of 20 periods. Each period within a session proceeded under identical rules. At the beginning of the experiment, each subject was assigned the role of Player A or Player B. They kept this role during the entire session. The computer network then matched subjects into pairs of players, with one player A and one player B. At the end of each period, the composition of the pairs changed under a stranger matching protocol so that subjects were rematched with another partner on a random basis.

TABLE 1 – Characteristics of experimental sessions

Sessions	Treatment	Number of subjects per session
1	Public Feedback	10
2	Public Feedback	10
3	Public Feedback	10
4	Public Feedback	10
5	Public Feedback	10
6	Public Feedback	10
7	Private Feedback	10
8	Private Feedback	10
9	Private Feedback	10
10	Private Feedback	10
11	Private Feedback	10
12	Private Feedback	8
13	No Feedback	8
14	No Feedback	10
15	No Feedback	10
16	No Feedback	8
17	No Feedback	10
18	No Feedback	10
19	No Feedback	8
Total		182

[7] Game instructions are available upon request.

4. Results

4.1. Amounts sent and returned

Table 2 provides summary statistics of average behavior of both players across the twenty periods in each treatment. The amount sent by Player A is a good proxy of his or her level of trust while Player B's return rate approximates his or her trustworthiness. We will first present Players A' amounts sent (subsection 4.1.1) and then Players B' return rate (subsection 4.1.2.).

TABLE 2 – Average transfers per treatment

	No Feedback	Public Feedback	Private Feedback
Amount sent by Player A (s_A)	2.24	4.57	3.60
	(2.91)	(3.69)	(3.53)
Amount returned by Player B (s_B)	1.46	4.88	2.77
	(3.17)	(5.82)	(4.12)
Return rate of Player B (r_B)	11.87	24.29	15.93
	(22.03)	(25.26)	(20.45)

Notes: Amount in experimental units and return rate in %; Standard errors in parentheses

4.1.1. Player A's amount sent

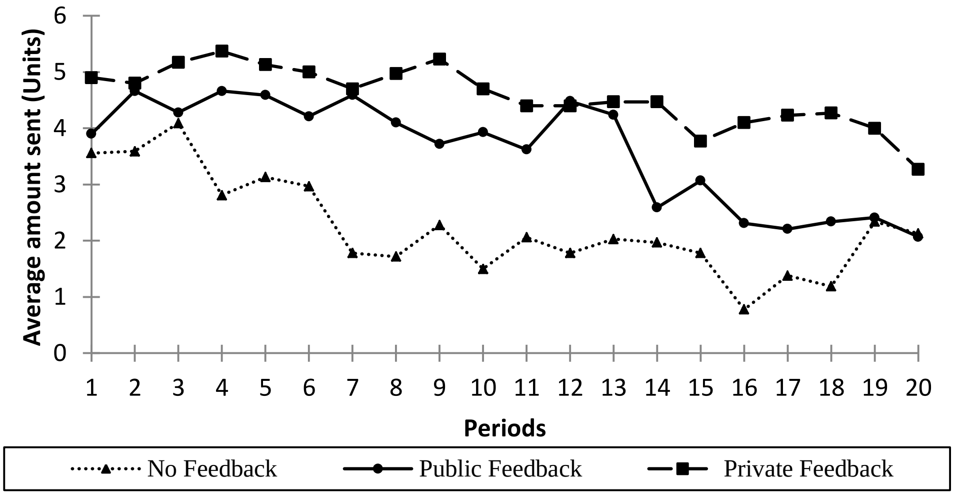
Figure 1 illustrates the time series of player A's average amount sent by period for each of the three treatments. It shows that introducing a feedback mechanism increases the amount sent by Player A. Table 2 indicates that amounts sent are 2.24 units in the No Feedback treatment, and respectively 4.57 units and 3.6 units in the Public and Private Feedback treatments. A Mann-Whitney test[8] indicates that the level of amount sent is significantly higher in the Public Feedback treatment than in the No Feedback treatment (z = -2.714; p = 0.0066). Similar findings are obtained if one considers the first ten periods of the game (z = -2.429; p = 0.0152) or the last ten periods in isolation (z = 2.571; p = 0.0101). The difference is also significant between the No Feedback and the Private Feedback treatments during all the periods (z = 2.000; p = 0.0455), during the first ten periods (z = 0.155; p = 0.0312) and during the last ten period (z = 1.714; p = 0.0865). Finally, the comparison between amounts sent in the Private Feedback treatment (3.6 units) and in the Public Feedback treatment (4.47 units) shows that average levels are lower in the Private Feedback treatment although the difference is not statistically significant (z = 0.882; p = 0.3776). Decomposing the twenty periods, we observe no significant difference in periods 1-10 (z = -0.802; p = 0.4225) but a borderline

[8] All statistical tests reported in this section are two-tailed Mann-Whitney tests using the average outcome of each session as one independent observation.

significant difference in periods 11-20 (z = -1.601; p = 0.1093).[9] The findings on player A's decisions are stated in result 1.

RESULT 1 *a) The average amount sent by Player A is significantly higher in the Public and Private Feedback treatments than in the No Feedback treatment. b) The average amount sent by Player A is lower in the Private Feedback treatment than in the Public Feedback treatment, although the difference is not statistically significant.*

FIGURE 1 – Evolution of player A's average amount sent



Support for result 1. Table 3 provides additional support to result 1. Table 3 consists of two panels. The left panel displays GLS estimates on the determinants of the amount sent by Player A (s_A). The right panel – that we will discuss later – displays estimates on the determinants of Player B's return rate (r_B). The independent variables include dummies for each treatment. A variable *period* is also included to study the evolution of amounts sent or returned over time.

Column 1 of Table 3 indicates that the amount sent by Player A increases by 2.323 units in the Public Feedback treatment and by 1.355 units in the Private Feedback treatment, as compared to the No Feedback treatment. Column 2 of Table 3 indicates that the amount sent by Player A is lower under a private feedback mechanism than under a public feedback mechanism, but the difference is not significant.

9 Decomposing the twenty periods of interaction into quarters, we observe a significant difference in periods 16-20 (z = 1.925; p = 0.0542), but not in periods 11-15 (z = 0.614; p = 0.5218).

TABLE 3 – Determinants of s_A and r_B

	Player A		Player B	
	(1) All Treatments	(2) Feedback Treatments	(3) All Treatments	(4) Feedback Treatments
No Feedback treatment	Ref.		Ref.	
Public Feedback treatment	2.323***	Ref.	12.42***	Ref.
	(0.588)		(3.203)	
Private Feedback treatment	1.355**	-0.968	4.059	-8.365***
	(0.546)	(0.641)	(2.972)	(3.191)
Periods	-0.107***	-0.106***	-0.947***	-1.050***
	(0.011)	(0.014)	(0.0742)	(0.0924)
Constant	3.367***	5.677***	21.81***	35.31***
	(0.377)	(0.467)	(2.328)	(2.692)
Observations	1820	1180	1820	1180
R^2	0.1060	0.0457	0.1053	0.0990
Wald χ^2	361.64	276.22	314.58	257.96
Prob > χ^2	0.000	0.000	0.000	0.000

*Notes: GLS model with individual random effect. Robust standard errors in parentheses. *** p<0.01; ** p<0.05; * p<0.1.*

4.1.2. Player B's return rate

Figure 2 displays Player B's average return rate over the 20 periods for the three treatments. It shows that the return rate is higher in the presence of a public feedback mechanism (24.29%) than without feedback mechanism (11.87%). The difference is significant if one considers all periods (z = 2.286; p = 0.0233), the first ten periods only (z = 2.143; p = 0.0321) and the last ten periods only (z = 2.249; p = 0.0152).

FIGURE 2 – Evolution of player B's average return rate

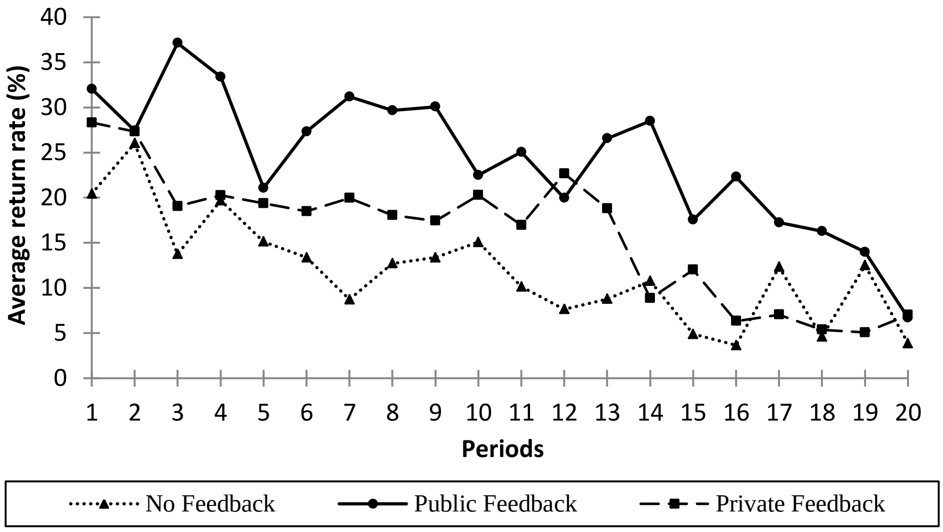


Figure 2 also suggests that the return rate under the Private Feedback treatment (15.93%) is higher than under the No Feedback treatment. However, Mann-Whitney tests indicate no significant difference between these both treatments during all the periods (z = 1.286; p = 0.1985), during the first ten periods (z = 1.286; p = 0.1985) and during the last ten periods (z = 1.000; p = 0.3173). Finally, the difference in the return rate between the Private Feedback treatment (15.93%) and the Public Feedback treatment (24.29%) is significant during all the periods (z = -1.922; p = 0.0547), the first ten periods (z = -1.761; p = 0.0782) and the last ten periods (z = -1.761; p = 0.0782). These findings are summarized in result 2.

RESULT 2 *a) Player B's return rate is higher with a Public Feedback mechanism than without.b) No significant difference is found between the Private Feedback treatment and the No Feedback treatment. c) The average return rate is significantly lower in the Private Feedback treatment than in the Public Feedback treatment.*

Support for result 2. The right panel of Table 3 confirms our findings above. Column 3 of Table 3 indicates that a public feedback mechanism has a positive and significant impact on Player B's return rate. Compared to the No Feedback treatment the return rate increases by 12.42 units when players can assign public ratings. The coefficient associated to the Private Feedback treatment is also positive but smaller than the coefficient for the public feedback treatment and is not significant. Column 4 of Table 3 confirms that the average return rate is significantly lower in the Private Feedback treatment than in the Public Feedback treatment.

4.2. Received rating

Table 4 shows the structure of ratings received by Players A and Players B in both feedback treatments. First, we observe that, in both treatments, a significant proportion of players are evaluated by their partners. That means that players do not hesitate to use the feedback system even if it is costly and ratings are not publicly observable. However the feedback frequencies are lower in the Private Feedback treatment than in the Public Feedback treatment for both players.

Player A receives ratings from Player B in 31% of the case in the Public Feedback treatment and in 24.66% of the case in the Private Feedback treatment. A Mann-Whitney test indicates that this difference is statistically significant (z = 1.764; p = 0.0776). The frequencies of ratings received by Player B also differ significantly across feedback treatments, with respectively 29.5% and 12.07% in the Public and Private Feedback treatments (z = 2.887; p = 0.0039). On average, Player A receive twice more ratings than Player B in the Private Feedback treatment (z = 1.858; p = 0.0632), while there is no significant difference in the Public Feedback treatment (z = 0.241; p = 0.8095). These findings are stated in result 3.

RESULT 3 *a) Both players A and B receive fewer ratings in Private Feedback than in Public Feedback treatments. b) Player A receives twice more ratings than Player B in Private Feedback treatment.*

TABLE 4 – Feedback structure per treatment

		Received by A from B	*A's amount sent (s_A)*	Received by B from A	*B's return rate (r_B)*
Public Feedback	Feedback frequency	31.00%		29.50%	
	Negative feedback	41.94%	*2.9*	68.36%	*17.65*
	Positive feedback	58.06%	*7.54*	31.64%	*55.27*
Private Feedback	Feedback frequency	24.66%		12.07%	
	Negative feedback	41.26%	*0.58*	81.43%	*13.28*
	Positive feedback	58.74%	*7.7*	18.57%	*50.66*

Table 4 also shows the nature of received feedbacks. In both treatments, negative (positive) ratings are associated to low (high) amount sent by Player A or low (high) return rate of Player B. Table 4 further indicates that, in both treatments, Player B receives more negative feedbacks than positive feedbacks, with a relatively higher frequency in Private Feedback treatment (81.43%) than in Public Feedback treatment (68.36%). Player A receives relatively less negative feedback than player B[10]; and the frequency of negative ratings received by Player A does not differ between the Public Feedback (41.94%) and the Private Feedback (41.26%) treatments.

4.3. Influence of Feedback on the amount sent and the return rate

To sum up the results of the previous sub-sections, Players B (and to a lesser extent, Players A) transfer significantly less in the Private Feedback treatment than in the Public Feedback treatment. Another result is that players (particularly Players of type B) receive fewer ratings in the Private Feedback treatment than in the Public Feedback treatment. Can we explain lower transfers in the private feedback system by the smaller frequency of ratings? According to our fourth hypothesis, players should be also less concerned by private ratings than by public ratings. Indeed, when feedbacks are observed by partners, some players can behave honestly to improve their reputation. Therefore players who

[10] This result can be explained by the nature of the trust game that gives the opportunity for player B to sanction player A either by a negative rating or by returning a small amount.

have only a reputation-building strategy should send or return less when ratings are kept private.

To check this, we estimated the effects of receiving ratings. Precisely we measure the impact of ratings received in *t-1* on the change in individual's amount sent or return rate between period *t-1* and *t*. The results are shown in Table 5. The first two columns display estimates in which the dependent variable is the change in Player A's amount sent between period *t-1* and *t* ($s_{A,t} - s_{A,t-1}$). Columns (3) and (4) show the results of estimates in which the dependent variable is the change in Player B's return rate between period *t-1* and *t* ($r_{B,t} - r_{B,t-1}$). The independent variables include dummies for having received positive (negative) rating in the previous period (*t-1*) and interaction variables *Positive (negative) received rating in t-1 × Private Feedback treatment*. Variables *Player A's amount sent in t-1* and *Player B's amount returned in t-1* are also included as control variables, as well as a trend variable. In estimates (2) and (4) we introduce individual's amount sent or return rate in period *t-1* (i.e., $s_{A,t-1}$ or $r_{B,t-1}$) in order to control for potential "regression toward the mean" effects" (Kahneman and Tversky, 1973) . These variables are standardized to have a mean of zero and a standard deviation of one (see Lacomba et al., 2013).[11]

Column 1 of Table 5 shows that receiving a negative rating induces an increase in the amount sent by Players A in the next period. The impact of a negative feedback is not significantly different when the feedback is public or private (as shown by the insignificant coefficient associated to the interaction variable *Negative received rating in t-1 × Private Feedback treatment*). Players A who receives a positive rating tend to reduce their amount sent in the next period. To check whether this effect may partly reflect a regression toward the mean effect, we controlled for player A's amount sent in t-1 in column (2). This seems to be indeed the case as shown by highly significant coefficient associated the variable $s_{A,t-1}$ *standardized*.

The interaction variable *Positive received rating in t-1 × Private Feedback treatment* is significant at 10% in column (2), indicating that receiving a positive feedback has a lower influence in the Private Feedback treatment. Precisely, the impact of receiving a positive rating in the previous period tends to disappear in the private Feedback treatment, as compared to the Public Feedback treatment. Finally, the negative and significant coefficient of the variable $s_{A,t-1}$ *standardized* indicates that there is a tendency to regress to the mean.

[11] The variable $s_{A,t-1}$ *standardized* is introduced to capture a pure "regression toward the mean" effect. It is constructed as follow : $\frac{s_{i,t-1}-s_{-i,t-1}}{\sigma_{-i,t-1}}$. By symmetry, the variable $r_{B,t-1}$ *standardized* is constructed as follow: $\frac{r_{i,t-1}-r_{-i,t-1}}{\sigma_{-i,t-1}}$.

TABLE 5 – Determinants of amount sent and return rate difference

	Player A		Player B	
	(1)	(2)	(3)	(4)
	$s_{A,t} - s_{A,t-1}$	$s_{A,t} - s_{A,t-1}$	$r_{B,t} - r_{B,t-1}$	$r_{B,t} - r_{B,t-1}$
Player A's amount sent in t-1 ($s_{A,t-1}$)			-2.802*** (0.209)	-1.660*** (0.292)
Player B's amount returned in t-1 ($s_{B,t-1}$)	-0.113*** (0.0208)	-0.0215 (0.0275)		
No received rating in t-1	Ref.	Ref.	Ref.	Ref.
Positive received rating in t-1	-1.322*** (0.385)	-1.097** (0.454)	-10.95** (4.323)	-5.190 (3.570)
Negative received rating in t-1	1.203*** (0.430)	0.858* (0.483)	12.30*** (2.368)	5.476*** (2.290)
Positive received rating in t-1 × Private Feedback Treatment	0.689 (0.498)	1.184* (0.683)	-1.544 (8.719)	-0.0702 (7.422)
Negative received rating in t-1 × Private Feedback Treatment	-0.629 (0.491)	-0.772 (0.676)	-3.860 (3.514)	-0.354 (3.715)
$s_{A,t-1}$ standardized		-0.919*** (0.170)		
$r_{B,t-1}$ standardized				-7.741*** (0.861)
Periods	-0.039** (0.0186)	-0.0175 (0.0133)	-0.248** (0.123)	-0.0763 (0.0932)
Constant	0.850*** (0.237)	0.266 (0.238)	12.13*** (1.960)	6.759*** (1.963)
Observations	1121	1121	1121	1121
R^2	0.1126	0.1633	0.1786	0.2931
Wald χ^2	96.78	43.88	242.54	205.37
Prob > χ^2	0.000	0.000	0.000	0.000

*Notes: GLS model with individual re. Robust standard errors in parentheses. *** p<0.01; ** p<0.05; * p<0.1.*

Columns (3) and (4) of Table 5 provides findings for Players B. They show that players B tend to increase their return rate if they received a negative feedback in the previous period. In contrast, receiving a positive rating has not any impact on Players B' behavior. Moreover Players B' responses are not statistically different when the feedback is private or public. Finally, there is a tendency to regress to the mean, as shown by the negative and highly significant coefficient associated to the variable $r_{B,t-1}$ *standardized*. Altogether these findings, that are inconsistent with the hypothesis 4, indicate that the reduction in the amount sent and the return rate in the Private Feedback treatment is mainly due to a decline in the number of ratings rather than a change in the reactions of players to a private or public feedback. This is summarized in result 4.

RESULT 4 *a) Receiving a negative rating in t-1 induces an increase in the amount sent and the return rate in the next period. b) The effect of receiving ratings is similar in both Feedback treatments.*

5. Discussion and conclusions

In this paper we have compared the effectiveness of private and public feedback mechanism on trust and trustworthiness in the context of a trust game. The aim was to isolate the social (dis)approval effects and the reputational effects of feedback mechanisms. We have three main findings.

First, we provide new evidence of the benefits of feedback mechanisms like those introduced on *eBay*, Amazon or AirBnB. Our data indicate that in the presence of feedback mechanisms, individuals do not hesitate to evaluate their counterpart even if it is costly. Furthermore we find that both trust and trustworthiness are significantly improved in presence of rating systems.

Second we find that both trust and trustworthiness are lower with a private feedback mechanism compared to a public feedback mechanism, indicating the important role of rating observability in the effectiveness of feedback mechanisms. Private feedbacks alone do not seem sufficient to improve cooperation in the long run, suggesting that reputation seems to play a critical role in the success of feedback mechanisms.

Third, in both treatments, ratings are strongly correlated with the amount sent or returned by the partner, which shows that feedback mechanisms play also their role of rewarding cooperative behaviors and punishing opportunistic behaviors. Individuals assign positive (negative) ratings when transfers received are high (low). However, fewer ratings are assigned with a private feedback mechanism compared to a public feedback mechanism. We also observe that receiving a negative rating induces an increase in the amount sent or returned in the next period. This effect is similar with both feedback mechanisms. This seems to indicate that individuals are more sensitive to others' approval or disapproval than to their own reputation and future gains.

Altogether these findings suggest that the lower levels of amount sent and return rate in the Private Feedback treatment are mainly due to a reduction in the volume of received ratings rather than a decline of their influence.

This study has thus paved the way to the examination of the role of emotions in environments characterized by anonymity and information asymmetry. Even in absence of direct contact between individuals, emotions can influence decisions, which highlight the need of incorporation of this aspect on the study of behavior.

References

Akerlof, G., 1980. A theory of social custom, of which unemployment may be one consequence, Quarterly Journal of Economics 94, 749-775.

Anderhub, V., Engelmann, D., Güth, W., 2002. An Experimental Study of the Repeated Trust Game with Incomplete Information. Journal of Economic Behavior and Organization 48, 197–216.

Anderson, C., Putterman, L., 2006. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. Games and Economic Behavior 54, 1–24.

Bajari, P., Hortaçsu, A., 2004. Economic Insights from Internet Auctions. Journal of Economic Literature 42(2), 457-486.

Barron, J., Paulson-Gjerde, K., 1997. Peer pressure in an agency relationship. Journal of Labor Economics 15, 235-254.

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

Bohnet, I., Harmgart, H., Huck, S., Tyran, J. R., 2005. Learning trust. Journal of the European Economic Association 3(2-3), 322-329.

Bolton, G., Katok, E., Ockenfels, A., 2004. How effective are online reputation mechanisms? An experimental investigation. Management Science 90 (11), 1587-1602.

Bolton, G., Greiner B., Ockenfels, A., 2013. Engineering Trust: Reciprocity in the Production of Reputation Information. Management Science 59(2), 265-285.

Brandts, J., Charness, G., 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. Experimental Economics 14, 375–398.

Cabral, L., Hortaçsu, A., 2010. The dynamic of seller reputation: evidence from eBay. Journal of Industrial Economics 58(1), 54-78.

Carpenter, J., 2007. The demand for punishment. Journal of Economic Behavior and Organization 62, 522–542.

Chen, K. Y., Hogg, T., 2005. Experimental evaluation of an eBay-Style self-reporting reputation mechanism. Internet and Network Economics. Lecture Notes in Computer Science 3828, 434-443.

Cochard, F., Nguyen Van, P., Willinger, M., 2004. Trusting behavior in repeated investment game. Journal of Economic Behavior and Organization 55, 31-44.

Coleman, J., 1990. Foundations of social theory. Harvard University Press, Cambridge, MA, USA.

Coricelli, G., Joffily, M., Montmarquette, C., Villeval, M.-C., 2010. Cheating, emotions, and rationality: an experiment on tax evasion. Experimental Economics 13 (2), 226-247.

Cox, J.C., 2004. How to identify trust and reciprocity. Games and Economic Behavior 46, 260-281.

Dellarocas, C., 2003. The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms. Management Science 49(10), 1407-1424.

Di Cagno, D., Sciubba, E., 2010. Trust, trustworthiness and social networks: playing a trust game when networks are formed in the lab. Journal of Economic Behavior & Organization 75, 156-167.

Dugar, S., 2010. Nonmonetary sanctions and rewards in an experimental coordination game. Journal of Economic Behavior and Organization 73(3), 377–386.

Engle-Warnick, J., Slonim, R. L., 2004. The evolution of strategies in repeated trust game. Journal of Economic Behavior & Organization 55 (4), 553-573.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171-178.

Gazzale, R. S., Khopkar, T., 2011. Remain Silent and Ye Shall Suffer: Seller Exploitation of Reticent Buyers in an Experimental Reputation System. Experimental Economics 14(2), 273-285.

Gaechter, S., Fehr, E., 1999. Collective action as a social exchange. Journal of Economic Behavior and Organization 39, 341-369.

Hollaender, H., 1990. A social exchange approach to voluntary cooperation. American Economic Review 80(5), 1157-1167.

Hopfensitz, A., Reuben, E., 2009. The importance of emotions for the effectiveness of social punishment. Economic Journal 119, 1534–1559.

Houser, D., Wooders, J., 2006. Reputation in auctions: theory and evidence from eBay. Journal of Economics and Management Strategy 15(2), 353–369.

Huck, S., Lünser, G. K., 2010. Group reputations: An experimental foray. Journal of Economic Behavior & Organization 73, 153-157.

Joffily, M., Masclet, D., Noussair, C., Villeval, M.-C., 2013. Emotions, Sanctions and Cooperation. Southern Economic Journal In-Press.

Kandel, E., Lazear, E., 1992. Peer pressure and partnership. Journal of Political Economy 100, 801-817.

Kahneman, D., Tversky, A., 1973. On the psychology of prediction. Psychological Review 80, 237-251.

Keser, C., 2003. Experimental games for the design of reputation management. IBM Systems Journal 42(3), 498-506.

Lacomba, J., Lagos, F., Reuben, E., Van Winden, F., 2013. On the Escalation and De-escalation of Conflict. working paper University of Amsterdam

Li, L., Xiao, E., 2010. Money Talks? An Experimental Study of Rebate in Reputation System Design. MPRA Working paper.

Lindbeck, A., Nyberg, S., Weibull, J., 1999. Social norms and economic incentives in the welfare state. Quarterly Journal of Economics 116(1), 1-35.

Lucking-Reiley, D., Bryan, D., Prasad, N., Reeves, D., 2007. Pennies from eBay: the Determinants of Price in Online Auctions. Journal of Industrial Economics 55(2), 223-233.

Masclet, D., Noussair, C., Tucker, S., Villeval, M.-C., 2003. Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism. American Economic Review 93(1), 366–80.

Masclet, D., Pénard, T., 2012. Do reputation feedback systems really improve trust among anonymous traders? An experimental study. Applied Economics 44 (35), 4553-4573.

Nikiforakis, N., Normann, H.T., 2008. A comparative statics analysis of punishment in public-good experiments. Experimental Economics 11, 358–369.

Noussair, C., Tucker, S., 2005. Combining monetary and social sanctions to promote cooperation. Economic Inquiry 43(3), 649–660.

Rege, M., Telle, K., 2004. The Impact of Social Approval and Framing. Journal of Public Economics 88(7–8), 1625–44.

Resnick, P., Zeckhauser, R., Swanson, S., Lockwood, K., 2006. The value of reputation on eBay: A controlled experiment. Experimental Economics 9(2), 79-101.

Sefton, M., Shupp, R., Walker, J., 2007. The Effect of Rewards and Sanctions in the Provision of Public Goods. Economic Inquiry 45 (4), 671-690.