ACCEPTED MANUSCRIPT

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

Marianne Lumeau, David Masclet and Thierry Penard

December 2014

Abstract: Several previous studies have highlighted the role of feedback mechanisms in the success of electronic marketplaces. This paper contends that the effectiveness of online feedback mechanisms passes through two channels, namely a 'reputational' effect that has been largely documented in the literature, but also a '(dis)approval' effect that has received less attention. We attempt to isolate these two effects using an experimental approach. For this purpose, we compare two experimental feedback systems that differ in the set of information available to participants. In the first feedback system, each player can observe the feedback profile of the other party, whereas in the second feedback system, this information is private. Our findings indicate that both systems improve cooperation. However, we observe that private feedback is less efficient in enhancing trust and trustworthiness than systems in which rating profiles are observed by partners. This finding is due to both a reduction of the number of assigned ratings and a lower impact of private ratings on subsequent decisions. All these results suggest that even if social (dis)approval matters, rating observability – and thus reputation – remains critical to induce honest behavior and improve efficiency in markets characterized by imperfect information.

Keywords: Reputation, (dis)approbation, Experiment, Trust Game, Feedback mechanism.

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) and Cirano, Montreal-Canada. david.masclet@univ-rennes1.fr

Accepted Manuscript

ACCEPTED MANUSCRIPT

Thierry Penard – University of Rennes1/CREM. thierry.penard@univ-rennes1.fr

1. Introduction

Several previous studies have highlighted the role of feedback systems in the success of electronic marketplaces, such as eBay and AmazonMarketPlace (Bajari and Hortaçsu, 2004; Cabral, 2012; Dellarocas, 2003; Dellarocas *et al.*, 2006). After the completion of a transaction, the buyer and the seller have the opportunity to rate each other. All received ratings are aggregated and displayed in each member's feedback profile that is made publicly available. Similar feedback systems exist on AirBnB, a marketplace for vacation rentals that connects hosts and travelers. By reducing information asymmetry, feedback systems help to build trust and foster cooperation among strangers. Such feedback systems are often called reputation mechanisms as they provide a measure of individuals' reputation.

In this current paper, we contend that the effectiveness of online feedback systems passes through two channels, namely a 'reputational' effect that has been largely documented, but also a '(dis)approval' effect that has received less attention in the literature. The former refers to the fact that the received ratings affect the traders' reputation and therefore their future *monetary* gains. Indeed, many studies have found that for a seller, a high number of positive ratings increases the probability of selling items at a higher price (Cabral and Hortaçsu, 2010; Houser and Wooders, 2006; Livingston, 2005; Lucking-Reiley *et al.*, 2007; Melnick and Alm, 2002; Resnick *et al.*, 2006). When feedback profiles are publicly observable, individuals are incited to adopt a cooperative behavior in order to accumulate positive ratings (and avoid negative ratings) and build their long-term reputation. We will refer to this effect as the *Reputational Response Hypothesis*.

In addition to this 'reputational' component, we hypothesize that receiving a positive (negative) rating may generate a *non-monetary* (dis)utility. This may be the case if the individual is sensitive to (dis)approval expressed by others through their ratings. Several reasons support this conjecture. The fact that expressions of approval and disapproval are commonly observed in human interactions suggests that they must influence the behavior of at least some individuals. In recognizing the importance of informal sanctions, economists have tried to integrate phenomena such as peer pressure (Kandel and Lazear, 1992; Barron and Paulson-Gjerde, 1997) and 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, not only do individuals respond to monetary sanctions (or rewards) imposed but are also sensitive to the social (dis)approval of others (e.g., Gachter and Fehr, 1999; Masclet *et al.*, 2003; Rege and Telle, 2004; Dugar, 2010). Therefore, we conjecture that

Accepted Manuscript

ACCEPTED MANUSCRIPT

individuals are incited to adopt a cooperative behavior in order to obtain approval rating (and avoid disapproval ratings). We will refer to this explanation as the *(Dis)Approval Response Hypothesis*.

In this paper, we attempt to experimentally isolate the role of reputation and social (dis)approval associated with ratings by using a trust game experiment in which players have the opportunity to rate each other (Keser, 2003; Masclet and Penard, 2012).[1] For this purpose, we designed three experimental treatments that vary in the set of information – or feedback – that participants have about their current partner. Our baseline treatment, called *No Feedback,* is a finitely repeated trust game inspired from the game designed by Berg *et al.* (1995). In this setting, players have no information about past behavior of their counterpart. The second treatment, called *Private Feedback*, involves a two-stage game. This treatment is similar to the baseline except that a new stage is added in which participants have the opportunity to assign either a negative (-1) or positive (+1) rating point to their counterpart. This information is private information for those who receive ratings. In other words, this information is never communicated to other participants in the current and future periods. The third treatment, called *Public Feedback*, is similar to the previous treatment except that at the beginning of each period, each participant can observe her or his counterpart's profile containing a historical record of received ratings, along with a score that represents the cumulative sum of positive and/or negative points obtained over all of the previous periods. This means that each participant is aware of her or his partner's reputation before deciding how much to send or return to her/his partner. This feedback system may convey both reputation and dis(approval) effects whereas the private feedback system can only rely on dis(approval) effects to foster cooperation.

A comparison of these three treatments should allow us to isolate the role played by reputation and social (dis)approval. If we observe differences in the levels of trust and trustworthiness between the Private Feedback and No Feedback treatments, this would suggest the presence of a (dis)approval component in ratings. Moreover, differences in the levels of trust and trustworthiness between the Private Feedback and Public Feedback treatments would provide evidence of a reputational component in ratings.

To anticipate our findings, we show that both private and public feedbacks improve trust and trustworthiness significantly compared to the No Feedback treatment. Yet we also observe that private feedback is less efficient in enhancing cooperation than public

[1] The trust game provides a good abstraction of the context in which transactions occur in electronic marketplaces like eBay. In such environments, 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 either not delivering the item or sending an item that does not correspond to the one listed in the auction description.

ACCEPTED MANUSCRIPT

feedback. This is due to the fact that i) fewer ratings are assigned in the Private Feedback treatment and that ii) those ratings have a lower effect on future behavior. All these findings suggest that even if social (dis)approval matters, publicly observed feedback, and therefore reputation, remain critical to promoting honest behavior and improving efficiency in markets characterized by imperfect information.

This paper is organized as follows. The next section will review previous experiments related to the roles of reputation and social (dis)approval. Section 3 describes the experiment and Section 4 then presents our main findings. A discussion and conclusion are provided in Section 5.

2. The previous literature

Our paper draws upon two distinct strands of the experimental economics literature. The first strand consists of papers that analyze the effectiveness of feedback systems in the context of trust games or buyer-seller games[2], while the second strand focuses on papers that investigate the influence of reputation or social (dis)approval considerations on trust and cooperation.

The first strand comprises the study of Keser (2003) who implemented a one-sided feedback system whereby 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 (2012) also tested the robustness of these findings when introducing a cost for posting a rating and showed that individuals do not hesitate to rate their partner despite the imposed cost. Using a buyer-seller game, Bolton *et al.* (2013), as well as Gazzale and Khopkar (2011) and Li and Xiao (2010), found 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) observed that individuals are more cooperative when they know that their behavior will be publicly observable. By introducing a second stage of network formation after a trust game, these authors found higher levels of trust and trustworthiness than in a standard trust game. Bolton *et al.* (2004) compared a buyer-seller game played repeatedly by strangers to a variant treatment in which

[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 or not to ship the item to the buyer (i.e., to be trustworthy).

Accepted Manuscript

ACCEPTED MANUSCRIPT

buyers are perfectly informed about sellers' past behavior. They showed evidence of sellers' concerns over their reputation. Similar results were obtained by Bohnet *et al.* (2005), and then Huck and Lünser (2010), who considered different informational settings.[3] To the best of our knowledge however, 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, irrespective of reputational concerns. In the context of voluntary contribution mechanisms, Masclet *et al.* (2003) introduced the opportunity for players to express their disapproval by assigning non-monetary negative points to their partners. Their results indicate that such a mechanism has a positive impact on the level of cooperation. These authors, however, observed that long-term contributions are significantly lower compared to a monetary sanction system (see also Noussair and Tucker, 2005). Dugar (2010) extended the analysis of Masclet *et al.* (2003) to the case where individuals also have the opportunity to assign approval points and concluded that disapproval was more efficient than approval. These results are generally interpreted in terms of an emotional response to received (dis)approval (Coricelli *et al.*, 2010; Hopfensitz and Reuben, 2009; Joffili *et al.*, 2014). Using physiological measurements, these studies reveal that such sanctions activate a 'virtuous emotional circle' that fosters future cooperation.

3. The experiment

3.1 Experimental design

Our experiment consists of three treatments that are built on a modified version of the trust game introduced by Berg *et al.* (1995).

The 'No Feedback' treatment

The *No Feedback* treatment is a simultaneous version of the trust game repeated during 20 periods. More precisely, at the beginning of each period, a Player A and a Player B are randomly matched and play a trust game with the following protocol: Let $I = \{A, B\}$ stand for a group of two players referred to respectively as players of types A and B. At

[3] Bohnet *et al.* (2005) considered 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 transactions. Huck and Lünser (2010) compared a situation in which buyers are informed of the past behavior of all sellers with a situation in which a buyer is only informed of his partner's past behavior.

Accepted Manuscript

ACCEPTED MANUSCRIPT

the beginning of the game, each player $i \in I$ is exogenously endowed with 10 experimental units. Player A decides how much of his/her 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 ($r_B = \left(\frac{s_B}{3s_A}\right) \times 100$). The amount sent by Player A and the proportion returned by Player B can be seen as measures of trust and trustworthy behavior, respectively.[4] Players A and B' payoffs are respectively given by:

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

The theoretical prediction of the trust game can be easily found by means of 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 $\pi_B(0,0) = 10$. This game-theoretical solution is obviously 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 $(s_A = 0, s_B = 0)$ applies in each period.[5]

The 'Private Feedback' treatment

The *Private Feedback* treatment is similar to the baseline treatment except that in each period of the trust game, a new stage is added in which both players can rate their partners. During this second stage, each participant observes the amount sent or returned by her/his partner and may decide to assign a rating point. The rating can be either positive (denoted +1) or negative (denoted -1). Assigning a positive or negative point is costless. Moreover this rating remains private information for the rated participant. In

[4] These measures are imperfect since individuals may have other motivations behind their decision (e.g., an increase in 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 simultaneously 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 or not to rate their partner.

ACCEPTED MANUSCRIPT

other words, this information is never displayed to other participants in the current period, as well as in the future periods.

The 'Public Feedback' treatment

The *Public Feedback* treatment is similar to the previous treatment with the notable exception that the players' feedback profile is now available to future partners. More precisely, at the beginning of each period, both players have access to her/his current partner's profile that contains a historical record of previously received ratings, along with a score equal to the cumulative sum of positive and negative rating points obtained over all of the previous periods. In this treatment, players are only informed about their current partner's profile, but meanwhile they cannot observe the profile of the other participants.

By backward induction, it can easily be shown that in both feedback treatments the subgame perfect equilibrium is again for player A to send nothing and for player B to return nothing. Under standard rationality hypotheses, the equilibrium is thus unchanged since players can't build a reputation.[6] Moreover, as assigning points is costless, all rating (or no-rating) strategies are compatible with this subgame perfect equilibrium. Players' payoffs are then expressed as: $\pi_A(0,0) = 10$, $\pi_B(0,0) = 10$.

However, results in experimental economics have shown that subjects do not generally behave as predicted by game theory. Based on previous studies, we expect to find different conducts in our three experimental treatments. The next subsection will present a series of behavioral hypotheses.

3.2 Behavioral Hypotheses

Let's first consider our baseline, i.e. the *No Feedback* treatment. It is well known that individuals are not solely interested in maximizing their own payoff and have some levels of social preferences, but also beliefs about others' social preferences. For instance, Players A may send a positive amount to Players B as a result of their other-regarding social preferences, namely altruism. However, since trust is mainly a matter of beliefs that an individual holds regarding others' behavior, even if Players A have

[6] Note that one may however relax some assumptions and consider some ad hoc assumptions specifying concerns for reputation such that there would exist a large number of reputation building Nash equilibria in which players A send positive amounts and players B return positive amounts as well in order to build up positive reputation (see Keser, 2003).

Accepted Manuscript

ACCEPTED MANUSCRIPT

self-regarding preferences, they may still be incited to send a positive amount if they believe that Players B have other-regarding preferences or will reciprocate positively (Cox, 2004). Concerning Players B, they will be incited to return a positive amount to Players A if they are motivated by altruism or positive reciprocity. Previous experiments have provided strong evidence that Players A don't hesitate to send positive amounts and Players B often return a fraction of the amount received, despite the absence of repetition with the same partner during the game (see, for example, Anderhub *et al.*, 2002; Cochard *et al.*, 2004; Engle-Warnick and Slonim, 2004). Note however that there is an important heterogeneity in the amounts sent or returned (and the return rate) and that these levels are affected by many factors, such as whether payment is random, whether subjects play both roles in the experiment or cultural concerns (see Johnson and Mislin (2011) for a meta-analysis).

Let's now consider the *Private Feedback* treatment. We conjecture herein that introducing a rating mechanism, despite the fact that ratings are private information, may improve transfers from both players compared to the baseline treatment. This may be the case if individuals are sensitive to social (dis)approbation and if the received ratings generate non-monetary (dis)utility. In such case, the assignment of positive (negative) ratings may play a rewarding (disciplining) role since they conveys social (dis)approval (Dugar, 2010; Gachter and Fehr, 1999; Masclet *et al.*, 2003; Rege and Telle, 2004). Based on these findings, we conjecture that the possibility of rating one's partner in the *Private Feedback* treatment may have positive and significant effects on trust and trustworthiness. This is summarized in the first hypothesis:

Hypothesis 1. *((Dis)Approval Response Hypothesis): Both Players A and B may be incited to cooperate in order to receive (avoid) social approval (disapproval) associated with positive (negative) ratings. This should lead to higher amounts sent and higher return rates by both players in the Private Feedback treatment than in the Baseline.*

Let's now consider the *Public Feedback* treatment. The fact that ratings are now publicly observable by the partner should increase both trust and trustworthiness, as compared to the *Private Feedback* treatment. The reason is that, in addition to the (dis)approval effect, this feedback system may lead individuals to cooperate to avoid (get) the reputational cost (gain) of receiving negative (positive) ratings. Many studies in different experimental contexts have shown the influence of releasing the past histories of individual players' decisions. For instance, Berg *et al.* (1995) found that providing information on the amounts sent and returned during previous experimental sessions significantly increases the amounts sent and returned. In related studies, Keser

Accepted Manuscript

ACCEPTED MANUSCRIPT

(2003) and Masclet and Penard (2012) found that the introduction of a reputation-based feedback system increases overall efficiency by improving both the levels of trust and trustworthiness. Similar findings were obtained by Bohnet *et al.* (2005), Bolton *et al.* (2004) and Huck and Lünser (2010) using perfect reputation mechanisms. These authors also showed that players base their choice on the partner's reputation, which provides an incentive to cooperate to benefit from (avoid) the future monetary consequences of having a good (bad) reputation. Based on the evidence of individuals' concerns over their reputation, we conjecture that the level of trust and trustworthiness should be higher with a public feedback system compared to a private feedback system. This is summarized as follow:

Hypothesis 2 *(Reputational Response Hypothesis): Both Players A and B may be incited to cooperate in order to avoid (get) a bad (good) reputation, which in turn may reduce (increase) their future gains. This should lead to higher amounts sent and higher return rates in the Public Feedback treatment than in the Private Feedback treatment.*

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 of 19 sessions, as summarized in Table 1. We opted for a between-subject design.[7] In all, 162 subjects - between eight and ten per session - were recruited among a population of undergraduate students. All subjects had never before played a trust game. On average, each session lasted 110 minutes, including initial instructions and subject payment.

At the beginning of the experiment, instructions were distributed and read to the subjects.[8] Each session was broken down into 20 periods, with each of these periods proceeding under identical rules. At the beginning of the experiment, each subject was assigned the role of Player A or Player B and kept this role throughout the 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 was changed under

[7] In a between-matching design, participants can be part of the treatment group or the control group, but cannot be part of both. The main advantage of a between-matching design is that measures are independent since each participant is only subjected to a single treatment. Moreover, we do not need to control for order effects. A possible problem is a selection bias since it is impossible to maintain homogeneity across the groups. We have tested this possible selection bias and founded no significant differences across the participants between treatments regarding age, gender or field of study.

[8] Game instructions are available upon request.

Accepted Manuscript

ACCEPTED MANUSCRIPT

a stranger matching protocol so that subjects were re-matched with another partner on a random basis.[9] At the end of the experiment, all gains accumulated by each subject during the experiment were converted in euros as follows: 30 units = €1. In addition, participants received a show-up fee of €5. Average earnings were €17.

Table 1 Characteristics of experimental sessions

Sessions	Treatment	Number of subjects per session
1	Public Feedback	8
2	Public Feedback	8
3	Public Feedback	8
4	Public Feedback	8
5	Public Feedback	8
6	Public Feedback	8
7	Private Feedback	8
8	Private Feedback	8
9	Private Feedback	10
10	Private Feedback	8
11	Private Feedback	8
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		162

4. Results

[9] Due to the number of sessions and subjects, they were re-matched with another partner under an imperfect stranger protocol (see Keser, 2003; Masclet and Penard, 2012).

Accepted Manuscript

ACCEPTED MANUSCRIPT

4.1 Amounts sent and returned

Table 2 provides summary statistics of the average behavior of both players over the 20 periods in each treatment. The amount sent by Player A is a good proxy of her or his level of trust, while Player B's return rate approximates his or her trustworthiness. We will first present Players A's amounts sent (Subsection 4.1.1) and then B's return rates (Subsection 4.1.2).

Table 2 Average transfers per treatment

	No Feedback	Private Feedback	Public Feedback
Amount sent by Player A (s_A)	2.24	4.25	5.65
	(2.91)	(3.77)	(3.10)
Amount returned by Player B (s_B)	1.46	3.33	7.24
	(3.17)	(5.08)	(6.44)
Return rate of Player B (r_B)	11.87	18.15	36.85
	(22.03)	(23.83)	(26.39)

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 per period for each of the three treatments; it reveals that introducing a feedback mechanism increases the amount sent by Player A. Table 2 indicates that amounts sent equal: 2.24 units in the No Feedback treatment, and respectively 5.65 units and 4.25 units in the Public and Private Feedback treatments. A Mann-Whitney test suggests that the level of amount sent is significantly higher in the Public Feedback treatment than in the No Feedback ($z = 3.000$; $p = 0.0027$).[10] The difference is also significant between the No Feedback and Private Feedback treatments ($z = -2.286$; $p = 0.0223$). Lastly, the comparison between

[10] All statistical tests reported in this section are two-tailed Mann-Whitney tests using the average outcome of each session as one independent observation. Separate tests were also conducted on the first ten periods and the last ten periods. As the results are similar, we have only reported the results of the Mann-Whitney tests for the twenty periods.

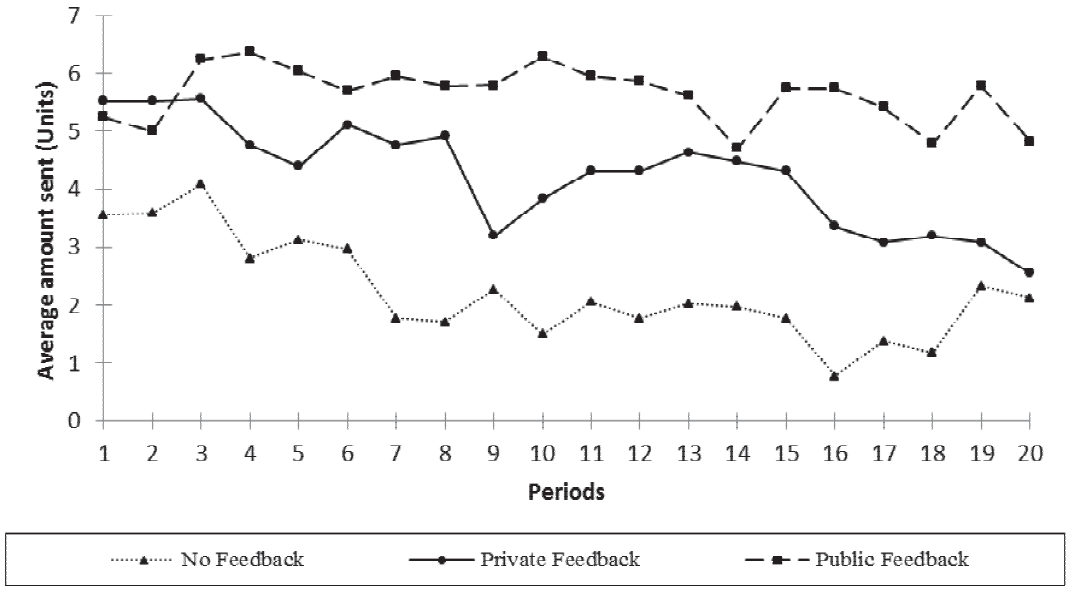
Accepted Manuscript

ACCEPTED MANUSCRIPT

amounts sent in the Private Feedback (4.25 units) and Public Feedback (5.65 units) treatment shows that average levels are lower in the Private Feedback, although the difference is borderline statistically significant (z = -1.601; p = 0.1093). The findings on Player A's decisions are summarized 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 baseline treatment; b) The average amount sent by Player A is lower in the Private Feedback than in the Public Feedback treatment, although this difference is not statistically significant.*

Figure 1 Evolution of player A's average amount sent



Support for result 1: Table 3 provides additional support for Result 1. Table 3 consists of two panels. The left-hand panel displays GLS estimates on the determinants of the amount sent by Player A ($s_{A,t}$) in period *t*. The right-hand panel, that will be discussed later, displays estimates on the determinants of Player B's return rate ($r_{B,t}$) in period *t*. The independent variables include dummies for each treatment. A variable *period* has also been include to take into account the evolution of amounts sent and returned over

ACCEPTED MANUSCRIPT

time, as well as a variable *amount received in period t-1* from the previous partner to control for possible indirect reciprocity. A set of variables describing the current partner's profile has also been included: the variable *positive (negative) rating in t-1* indicates whether the partner received a positive (negative) rating in the previous period; the *cumulative positive (negative) ratings* variable corresponds to the sum of positive (negative) ratings that the partner received since the beginning of the game.

Column (1) of Table 3 indicates that the amount sent by Player A increases significantly by 3.406 units in the Public Feedback treatment and by 2.004 units in the Private Feedback, as compared to the No Feedback treatment. Column (3) of Table 3 shows that the amount sent by player A is lower by 0.978 units under a Private Feedback mechanism than under a Public Feedback mechanism. Columns (2) and (3) of Table 3 indicate that the amount received in *t*-1 from one's partner is a good indicator of amount sent in period *t*. Player A is more strongly inclined to trust Player B (i.e., to send them a higher share of their endowment) if Player A received a greater sum from her/his previous partner. Moreover, column (4) of Table 3 reveals that with the public feedback system, the amount sent by Player A depends on her/his current partner's feedback profile (i.e., her/his past ratings). More precisely, players take into account the full history of past ratings and do not only focus on the most recent rating. The amount sent by Player A increases with the number of positive ratings displayed in the feedback profile of their current partner. This finding is consistent with previous results (see Masclet and Penard, 2012).

Table 3 Determinants of $s_{A,t}$ and $r_{B,t}$

	Player A				Player B			
	(1) All Treat.	(2) All Treat.	(3) Feedback Treat.	(4) Public Feedback Treat.	(5) All Treat.	(6) All Treat.	(7) Feedback Treat.	(8) Public Feedback Treat.
No Feedback Treat.	Ref.	Ref.			Ref.	Ref.		
Public Feedback Treat.	3.406***	2.527***	Ref.		24.986***	24.060***	Ref.	
	(0.473)	(0.471)			(3.893)	(3.894)		
Private Feedback Treat.	2.004***	1.691***	-0.978*		6.887*	5.390	-18.390***	
	(0.607)	(0.539)	(0.574)		(3.893)	(3.419)	(4.298)	
Amount received in t-1		0.163***	0.127***	0.067***		0.269	0.475**	0.392
		(0.032)	(0.038)	(0.025)		(0.183)	(0.209)	(0.324)
Partner's rating profile								
No rating in t-1				Ref				Ref
Positive rating in t-1				-0.358				-1.075
				(0.447)				(3.862)
Negative rating in t-1				-0.724				2.932
				(0.456)				(3.484)
Cumulative positive ratings				4.223***				27.643***
				(1.213)				(9.652)
Cumulative negative ratings				-0.724				2.932
				(0.456)				(3.484)

ACCEPTED MANUSCRIPT

Periods	-0.0926***	-0.073***	-0.069**	-0.017***	-0.955***	-0.904***	-1.022***	-1.175***
	(0.0206)	(0.019)	(0.029)	(0.026)	(0.103)	(0.117)	(0.172)	(0.233)
Constant	3.216***	2.735***	5.479***	4.563***	21.893***	20.756***	44.94***	41.668***
	(0.387)	(0.341)	(0.520)	(0.726)	(2.522)	(2.895)	(4.252)	(6.399)
Observations	1,620	1539	931	456	1620	1539	931	456
R^2	0.1836	0.298	0.170	0.311	0.205	0.210	0.195	0.258
Wald χ^2	77.37	151.01	29.91	78.78	124.08	123.64	83.95	220.42
Prob > χ^2	0.000	0.000	0.000	0.000	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

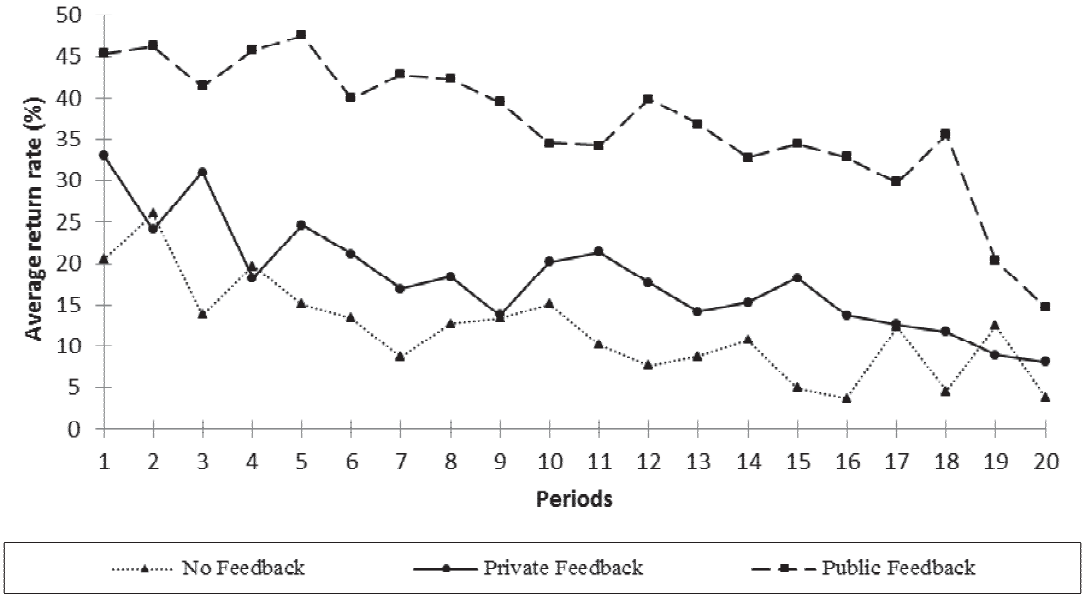
Figure 2 displays Player B's average return rate over the 20 periods for the three treatments; it shows a higher return rate in the presence of a public feedback mechanism (36.85%) than without any feedback (11.87%). The difference is significant (z = 3.000; p = 0.0027). Figure 2 also suggests that the return rate under the Private Feedback treatment (18.15%) is higher than under the baseline treatment. A Mann-Whitney test however indicates no significant difference between these two treatments (z = 1.429; p = 0.1531). Moreover, the difference in return rate between the Private Feedback (18.15%) and Public Feedback (36.85%) treatments is significant (z = -2.722; p = 0.0065). These findings are summarized below 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 than in the Public Feedback treatment.*

Figure 2 Evolution of player B's average return rate

Accepted Manuscript

ACCEPTED MANUSCRIPT



Support for Result 2: The right-hand panel of Table 3 displays GLS estimates on the determinants of the return rate of Player B ($r_{B,t}$). These estimates confirm our above findings. Column (5) of Table 3 indicates that a public feedback mechanism has a positive and significant impact on Player B's return rate. The return rate increases by 24.99 units in the Public Feedback treatment, compared to the No Feedback treatment. The coefficient associated with the Private Feedback treatment is also positive but smaller than the coefficient for the Public Feedback treatment and not robust to the introduction of the variable *Amount received in t-1* (see column 6). Column (7) of Table 3 confirms that the return rate is significantly lower in the Private Feedback treatment than in the Public Feedback Treatment.

Comparing columns (2) and (6) (or (4) and (8)) shows that the amount received in t-1 is of less importance to Player B than to Player A. A possible reason is that this information is less critical for Player B than for Player A since she can condition her decision on the amount received from Player A in the current period. Finally, the estimated coefficient for the *Cumulative positive ratings* is positive and significant, indicating that Player B conditions her decisions to her partner's feedback profile in the Public Feedback treatment (see column 8).

4.2 Received ratings

ACCEPTED MANUSCRIPT

Table 4 shows the breakdown of ratings received by Players A and B by both feedback treatments. We first observe that in both treatments, a significant proportion of players are evaluated by their partners, which means that some players do not hesitate to use the feedback system even if ratings are not publicly observable. Player A receives ratings from Player B in 88.12% of cases in the Public Feedback system and in 78.4% of the cases in the Private Feedback system. A Mann-Whitney test indicates that this difference is statistically significant (z = 1.690; p = 0.0910). The frequencies of ratings received by Player B also differ across feedback treatments, with respectively 89.58% and 82.60% in the Public and Private Feedback treatments but the difference is not statistically significant (z = 1.363; p = 0.1727).

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	88.12%		89.58%	
	Negative feedback	42.08%	*3.89*	51.63%	*20.47*
	Positive feedback	57.92%	*7.18*	48.37%	*54.53*
Private Feedback	Feedback frequency	78.4%		82.60%	
	Negative feedback	46.43%	*1.53*	68.04%	*8.89*
	Positive feedback	53.57%	*7.14*	31.96%	*45.18*

To provide a more formal proof of these findings, we ran estimates using Probit models on the determinants of the probability that Player A (B) receives a feedback from player B (A). These estimates are shown in Table 5. The independent variables include dummies for each treatment. A variable *period* has also been included to control for time trend in the probability of receiving a feedback. Columns (1) and (2) of Table 5 indicate that the probability that players A and B receive a feedback is significantly lower in the Private Feedback than in the Public Feedback treatments. These findings are stated in Result 3 below.

RESULT 3: *Both players A and B receive fewer ratings in the Private Feedback treatment than in the Public Feedback treatment.*

Table 4 also shows the nature of feedback received. In both treatments, negative (positive) ratings are associated with low (high) amounts sent by Player A or low (high) return rates of Player B. Table 4 further indicates that under both treatments, Player B

Accepted Manuscript

ACCEPTED MANUSCRIPT

receives more negative feedback than positive feedback, with a relatively higher frequency in Private Feedback treatment (68.04%) than in Public Feedback (51.63%). Player A receives relatively less negative feedback than Player B[11]; moreover, the frequency of negative ratings received by Player A does not differ between the Public Feedback (42.08%) and Private Feedback (46.43%) treatments.

Table 5 Determinants of the probability of receiving a feedback

	Player A (1)	Player B (2)
Public Feedback Treatment	Ref.	Ref.
Private Feedback Treatment	-0.384**	-0.308**
	(0.154)	(0.139)
Periods	-0.0006	-0.0120
	(0.008)	(0.009)
Constant	1.255***	1.432***
	(0.146)	(0.141)
Observations	980	980
Log-likelihood	-424.58	-386.25
$\overline{\chi^2}$	22.48 ***	8.75 ***

Notes: Probit model with individual random effect. Robust standard errors in parentheses.

*** *p<0.01; ** p<0.05; * p<0.1.*

4.3 Influence of feedback on the amount sent and return rate

In this sub-section we investigate whether the impact of received ratings differs across treatments. According to our hypothesis, players should be less concerned by private ratings than by public ratings. Indeed, when feedbacks are observed by future partners, some players may behave more cooperatively to improve their reputation. Therefore players who follow a reputation-building strategy should send or return less when ratings are kept private.

To check this hypothesis, we estimated the effects of receiving ratings. More specifically, we measured the impact of ratings received in period *t-1* on the change in an individual's amount sent or return rate between periods *t-1* and *t*. These results are shown in Table 6. The first two columns display estimates, in which the dependent variable is the change in Player A's amount sent between periods *t-1* and *t* ($s_{A,t} - s_{A,t-1}$). Columns (3) to (6) show the results of estimates, in which the dependent

[11] This result may be explained by the nature of the trust game, which provides the opportunity for Player B to sanction Player A either by a negative rating or by returning a small amount.

Accepted Manuscript

ACCEPTED MANUSCRIPT

variable is the change in Player B's return rate between periods *t-1* and *t* ($r_{B,t} - r_{B,t-1}$) for the whole sample (Columns 3 and 4) and for the sub-sample when considering those who have received a positive amount from player A (Columns 5 and 6). The independent variables include dummies for having received positive (negative) ratings during the previous period (*t-1*) and interaction variables *Positive (Negative) rating received in t-1 × Private Feedback treatment*.[12] The *Amount received in t-1* has also been included as a control variable. We also included a trend variable, as well as a variable for the last five periods. To check whether the impact of received feedback is reduced during the last five periods of the game, we introduced interaction terms *Positive (Negative) rating received in t-1 × last five periods.* Finally in estimates (2), (4) and (6), we introduce an individual's amount sent or return rate during period *t-1* (i.e., $s_{A,t-1}$ or $r_{B,t-1}$ *standardized*) in order to control for potential 'regression toward the mean' effects (Kahneman and Tversky, 1973). These variables have all been standardized to obtain a mean of 0 and a standard deviation of 1 (see Lacomba *et al.*, 2014).[13]

Column (1) of Table 6 shows that receiving a negative rating increases the amount sent by Player A in the next period. The impact of a negative feedback is significantly reduced when the feedback is private, as shown by the negative and significant coefficient associated with the interaction variable *Negative rating received in t-1 × Private Feedback treatment*. By contrast, a positive rating tends to reduce Player A's amount sent during the subsequent period. To check whether this effect may partially reflect a regression toward the mean effect, we controlled for Player A's amount sent in *t-1* in column (2). This indeed seems to be the case, as demonstrated by the highly significant coefficient associated with the $s_{A,t-1}$ *standardized* variable. The interaction variable *Positive rating received in t-1 × Private Feedback treatment* is not significant in column (1) but becomes significant at 10% after controlling for regression toward the mean effect in column (2).

Columns (3) to (6) of Table 6 provide the results for Player B; they show that Player B tends to increase his/her return rate when receiving a negative feedback during the

[12] In additional estimates (available upon request), we also included Variables *cumulative positive (negative) ratings* corresponding to the sum of positive (negative) ratings received since the beginning of the game. Our main findings are unchanged.

[13] The $s_{A,t-1}$*standardized* variable is introduced in order to capture a pure "regression toward the mean" effect. It is constructed as follows: $\frac{s_{i,t-1} - \bar{s}_{-i,t-1}}{\sigma_{-i,t-1}}$ with $\bar{s}_{-i,t-1}$ the average amount sent by all players A except i in the previous period and $\sigma_{-i,t-1}$ the standard deviation of the amounts sent by these players. By symmetry, the *standardized* variable $r_{B,t-1}$ is constructed as follows: $\frac{r_{i,t-1} - \bar{r}_{-i,t-1}}{\sigma_{-i,t-1}}$.

ACCEPTED MANUSCRIPT

previous period. In contrast, receiving a positive rating has a negative impact on Player B's behavior. Moreover, Player B's response to a negative feedback is reduced when the feedback is private. The tendency to regress toward the mean also exists, as shown by the negative and highly significant coefficient associated with the $r_{B,t-1}$ *standardized* variable.

Table 6 Determinants of amount sent and return rate difference

	Player A		Player B			
			Whole sample		Sub-sample	
	(1)	(2)	(3)	(4)	(5)	(6)
	$s_{A,t} - s_{A,t-1}$	$s_{A,t} - s_{A,t-1}$	$r_{B,t} - r_{B,t-1}$	$r_{B,t} - r_{B,t-1}$	$r_{B,t} - r_{B,t-1}$	$r_{B,t} - r_{B,t-1}$
No received rating in t-1	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Positive received rating in t-1	-1.260***	-0.817**	-11.61***	-7.534**	-11.41**	-8.122**
	(0.402)	(0.360)	(4.086)	(3.593)	(4.681)	(4.112)
Negative received rating in t-1	1.187**	0.793*	9.440**	2.561	13.87***	9.164**
	(0.536)	(0.475)	(3.941)	(3.909)	(4.282)	(4.077)
Positive received rating in t-1 ×	0.119	0.670*	-3.334	0.0530	0.258	2.640
Private Feedback Treatment	(0.363)	(0.389)	(4.341)	(4.676)	(4.531)	(5.076)
Negative received rating in t-1 ×	-1.177**	-0.972**	-5.376*	-1.682	-8.082**	-6.553*
Private Feedback Treatment	(0.500)	(0.484)	(3.243)	(3.548)	(3.425)	(3.522)
Positive received rating in t-1 ×	0.0378	0.0789	-6.994	-8.019	-7.539	-10.24
Five Last Periods	(0.667)	(0.638)	(6.090)	(5.138)	(8.567)	(7.877)
Negative received rating in t-1 ×	0.103	0.144	-4.199	-2.975	-7.154	-8.408
Five Last Periods	(0.596)	(0.588)	(5.200)	(4.628)	(8.152)	(7.499)
Amount received in t-1	-0.0740**	-0.00112	-0.382	0.154	0.549	0.709*
	(0.0309)	(0.0229)	(0.336)	(0.318)	(0.426)	(0.427)
$s_{A,t-1}$ standardized		-1.070***				
		(0.155)				
$r_{B,t-1}$ standardized				-7.246***		-5.543***
				(1.283)		(1.643)
Five last Periods	-0.377	-0.335	1.420	0.945	2.662	3.885
	(0.624)	(0.606)	(4.568)	(3.735)	(7.783)	(6.707)
Periods	-0.0290	-0.0132	0.0434	0.141	0.199	0.286
	(0.0192)	(0.0219)	(0.149)	(0.156)	(0.178)	(0.188)
Constant	1.063***	0.262	2.008	-0.971	-5.450	-6.112
	(0.365)	(0.316)	(3.280)	(3.120)	(3.908)	(3.723)
Observations	931	931	931	931	667	667
R^2	0.1225	0.2273	0.1399	0.2132	0.1564	0.1978
Wald χ^2	50.78	83.68	57.24	76.94	40.48	42.66
Prob > χ^2	0.000	0.000	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.*

Altogether these findings from Tables 5 and 6 indicate that the reduction in the amount sent and return rate in the *Private Feedback* treatment is due both to a reduction of the number of received ratings and the reduced impact of private ratings. This conclusion is summarized in Result 4.

Accepted Manuscript

ACCEPTED MANUSCRIPT

RESULT 4: *The lower trust and trustworthiness in the Private Feedback treatment compared to the Public Feedback treatment results both from lower received ratings and lower impact of these ratings in the Private Feedback treatment.*

4.4. A variant of feedback system with costly ratings

In this sub-section, we test whether our main results hold when one introduces a cost to rate one's partner. Indeed, it is more realistic to assume that individuals incur a cost to rate their partners. This cost may be interpreted as the opportunity cost of rating, as measured by the amount of time and effort devoted to this task (see Masclet and Penard (2012) for a discussion on this point). Precisely, in a second experiment, we ran the same Private and Public Feedback treatments with the exception that assigning a rating point was costly for the rater, who incurred a direct monetary cost of 1 unit (corresponding to 1/10 of the initial endowment). This second experiment consisted of 12 sessions, as summarized in Table 7.

Table 7 Characteristics of the second experiment

Sessions	Treatment	Number of subjects per session
1	Costly Public Feedback	10
2	Costly Public Feedback	10
3	Costly Public Feedback	10
4	Costly Public Feedback	10
5	Costly Public Feedback	10
6	Costly Public Feedback	10
7	Costly Private Feedback	10
8	Costly Private Feedback	10
9	Costly Private Feedback	10
10	Costly Private Feedback	10
11	Costly Private Feedback	10
12	Costly Private Feedback	8
Total		118

Figure 3 shows the time series of Player A's average amount sent per period for the No Feedback, Costly Public Feedback and Costly Private Feedback treatments. Our findings show that introducing a feedback mechanism increases the amount sent by Player A. Amounts sent are 2.24 units in the No Feedback treatment, and respectively 4.57 units and 3.6 units in the Costly Public and Costly Private Feedback treatments.

Accepted Manuscript

ACCEPTED MANUSCRIPT

The Mann-Whitney tests suggest that the level of amount sent is significantly higher in the Costly Public Feedback treatment than in the No Feedback (z =2.714; p =0.0066) and in the Costly Private Feedback treatment than in the No Feedback treatment (z = 2.000; p = 0.0455). Figure 3 also reveals that the amount sent is higher in the Costly Public Feedback treatment than in the Costly Private Feedback treatment, although the difference is not statistically significant (z = 0.882; p = 0.3776).

Figure 3 Evolution of player A's average amount sent
(With costly rating)

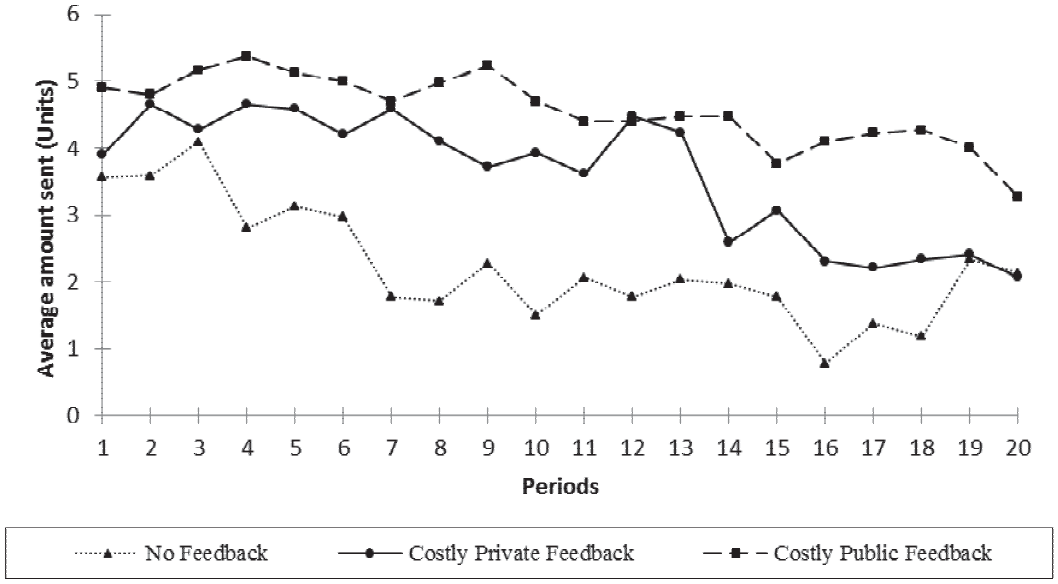


Figure 4 shows Player B's average return rate for the No Feedback, Costly Public Feedback and Costly Private Feedback treatments. We observe a higher return rate in the Costly Public Feedback treatment (24.29%) than in the No Feedback treatment (11.87%), (z = 2.286; p = 0.0233). Figure 4 also indicates that the return rate under the Costly Private Feedback treatment (15.93%) is higher than under the No Feedback treatment, although the difference is not significant (z = 1.286; p = 0.1985). The difference in return rate between the Costly Private Feedback (15.93%) and Costly Public Feedback (24.29%) treatments is significant (z = -1.922; p = 0.0547).

Figure 4 Evolution of player B's average return rate (With costly rating)

ACCEPTED MANUSCRIPT

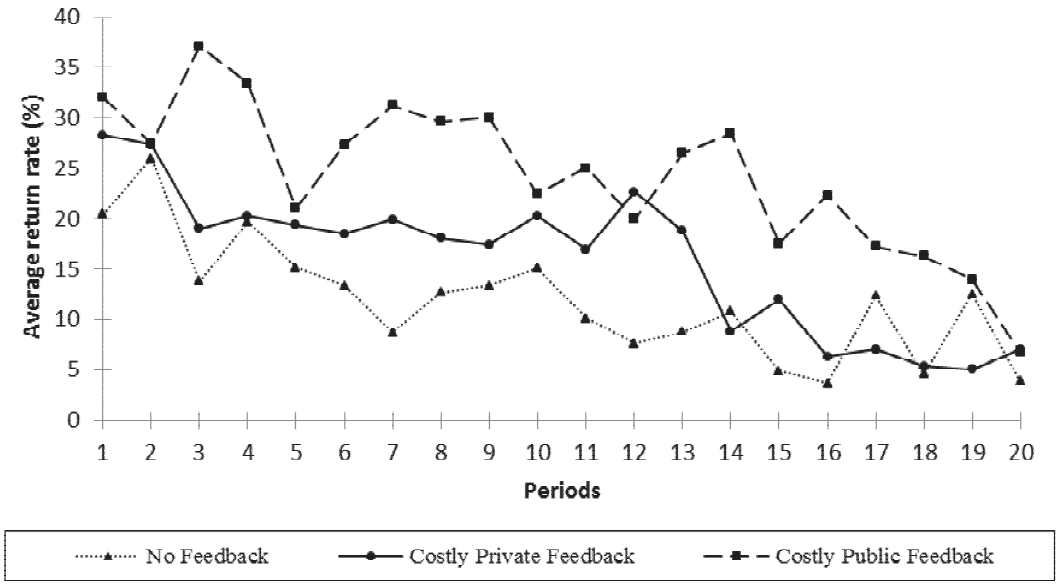


Table 8 displays GLS estimates on the determinants of the amount sent by Player A ($s_{A,t}$) and on the determinants of Player B's return rate ($r_{B,t}$). Consistent with our previous findings without rating cost, column (1) of Table 8 indicates that the amount sent by Player A increases by 2.323 units in the Costly Public Feedback treatment and by 1.355 units in the Costly Private Feedback, as compared to the No Feedback treatment. Column (2) of Table 8 shows that the amount sent by Player A is lower with a Private Feedback mechanism than with Public Feedback, although this difference is not significant.

Column (3) of Table 8 indicates that a costly 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 costly public ratings. The coefficient associated with the Private Feedback treatment is also positive but smaller than the coefficient for Public Feedback treatment and moreover is not significant. Column (4) of Table 8 confirms that the average return rate is significantly lower by 8.365 units in the Costly Private Feedback treatment than in Costly Public Feedback.

These differences in trust and trustworthiness across treatments mainly result from the fact that people assign significantly less ratings in the Costly Private Feedback

ACCEPTED MANUSCRIPT

treatment compared to the Costly Public Feedback treatment.[14] In contrast to the treatments without cost, we found no significant difference in the impact of received ratings between treatments. A possible reason is that significantly less ratings were assigned when a cost was introduced (see Tables A1 and A2 in appendix), which made difficult to clearly investigate the precise impact of received ratings (see table A3 in appendix).

Table 8 Determinants of $s_{A,t}$ and $r_{B,t}$ (With costly rating)

	Player A		Player B	
	(1) All Treat.	(2) Feedback Treat.	(3) All Treat.	(4) Feedback Treat.
No Feedback Treat.	Ref.		Ref.	
Costly Public Feedback Treat.	2.323***	Ref.	12.42***	Ref.
	(0.547)		(3.366)	
Costly Private Feedback Treat.	1.355**	-0.968	4.059	-8.365***
	(0.569)	(0.638)	(2.697)	(3.151)
Periods	-0.107***	-0.106***	-0.947***	-1.050***
	(0.0168)	(0.0234)	(0.0878)	(0.122)
Constant	3.367***	5.677***	21.81***	35.31***
	(0.367)	(0.506)	(2.440)	(2.849)
Observations	1820	1180	1820	1180
R^2	0.1060	0.0457	0.1053	0.0990
Wald χ^2	58.04	21.63	156.24	104.00
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.*

5. Discussion and conclusion

In this paper, we have compared the effectiveness of private and public feedback mechanisms on trust and trustworthiness in the context of a trust game experiment. The aim was to isolate the social (dis)approval effects and reputational effects of feedback mechanisms. Three main findings can be highlighted from our experiments.

[14] Player A receives ratings from Player B in 31% of cases in the Costly Public Feedback system and in 24.66% of the Costly Private Feedback. 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 Costly Public and Private Feedback treatments (z = 2.567; p = 0.0103).

Accepted Manuscript

ACCEPTED MANUSCRIPT

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 such mechanisms, individuals do not hesitate to evaluate their counterpart and that both trust and trustworthiness are significantly improved in the presence of rating systems.

Second, we found that both trust and trustworthiness are significantly lower with a private feedback mechanism compared to a public feedback mechanism, which indicates the important role of rating observability in the effectiveness of feedback mechanisms. Private feedbacks alone do not appear to be sufficient in improving cooperation over the long run, suggesting that reputation plays a critical role in the success of feedback mechanisms.

Third, in both the private and public treatments, ratings are strongly correlated with the amount sent or returned by the partner, which shows that feedback mechanisms also play their role of rewarding cooperative behavior and punishing opportunistic behavior. Individuals assign positive (negative) ratings when amounts received are high (low). However, fewer ratings are assigned with a private feedback mechanism as compared to a public one. We have also observed that receiving a negative rating induces an increase in the amount sent or returned during the subsequent period. This effect is lower in the private feedback treatment.

Our results imply that the lower levels of trust and trustworthiness in the Private Feedback treatment are due to both a reduction in the number of ratings received and a lower impact of received ratings on subsequent decisions. Even if social (dis)approval matters, reputation mechanisms remain critical to induce honest behavior and improve efficiency in markets characterized by anonymity and imperfect information.

To some extent our experimental setting underestimates the role of a reputation mechanism due to the fact that participants do not have the possibility of refusing the partner that has been randomly assigned to them. In other words, they have no outside option.[15] A possible extension of this work may consist in introducing an exit option for both players such that participants may decide to ostracism those who have a bad reputation. Alternatively, participants may choose their partner after observing all profiles (see for instance Chen and Hogg, 2005).

Another extension could be to let players choose which information in their feedback profile they want to disclose to their partners. For instance, we could allow them to

[15] Another limit of our design is that player A's reputation is less relevant for player B since she/he can directly condition her decision based on the observed amount received from player A. Note however that this reflects actual information asymmetry between sellers and buyers in online marketplaces.

Accepted Manuscript

ACCEPTED MANUSCRIPT

reveal only their score (i.e., the cumulative sum of positive and negative points) or the last rating received. This setting is in line with eBay's recent change that allows buyers to make their feedback profile either private or public.

This study has thus paved the way to examine the interplay of (dis)approval and reputation in environments characterized by anonymity and information asymmetry.

Acknowledgments

The authors thank anonymous referees for their helpful remarks. They also thank the participants of the Workshop on Conflicts in Rennes, of the ASFEE in Fort de France, of the ESA in New-York and of the Hotelling Seminar in Paris for helpful comments. Finally, financial support from the *MSHB* project is gratefully acknowledge, as well as Elven Priour for programming the experiment.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Appendix

Table A1 Feedback structure per treatment (with rating cost)

		Received by A from B	*A's amount sent (s_A)*	Received by B from A	*B's return rate (r_B)*
Costly 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*
Costly 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 A2 Probability of receiving a feedback (With and without rating cost)

	(1) Player A	(2) Player B
Public Feedback Treatment	Ref.	Ref.
Private Feedback Treatment	-0.288***	-0.506***
	(0.0950)	(0.0942)
Feedback Cost Treatment	-1.680***	-2.081***
	(0.0976)	(0.0979)
Periods	-0.0206***	-0.0328***
	(0.00547)	(0.00583)
Constant	1.407***	1.773***
	(0.108)	(0.111)
Observations	2,160	2,160
Log-likelihood	-1101.997	-942.69
χ^2	47.80***	28.58***

*Notes: Probit model with individual re. *** p<0.01; ** p<0.05; * p<0.1.*

Accepted Manuscript

ACCEPTED MANUSCRIPT

Table A3 Determinants of amount sent and return rate difference (with rating cost)

	Player A		Player B	
	(1) $s_{A,t} - s_{A,t-1}$	(2) $s_{A,t} - s_{A,t-1}$	(3) $r_{B,t} - r_{B,t-1}$	(4) $r_{B,t} - r_{B,t-1}$
No received rating in t-1	Ref.	Ref.	Ref.	Ref.
Positive received rating in t-1	-1.322***	-1.097**	-10.95**	-5.190
	(0.373)	(0.454)	(3.475)	(3.570)
Negative received rating in t-1	1.203**	0.858*	12.30***	5.476***
	(0.517)	(0.483)	(2.208)	(2.290)
Positive received rating in t-1 ×	0.689	1.184*	-1.544	-0.0702
Private Feedback Treatment	(0.670)	(0.683)	(8.217)	(7.422)
Negative received rating in t-1 ×	-0.629	-0.772	-3.860	-0.354
Private Feedback Treatment	(0.624)	(0.676)	(3.697)	(3.715)
Amount received in t-1	-0.113***	-0.0215	-2.802***	-1.660***
	(0.0302)	(0.0275)	(0.350)	(0.292)
$s_{A,t-1}$ standardized		-0.919***		
		(0.170)		
$r_{B,t-1}$ standardized				-7.741***
				(0.861)
Periods	-0.0393***	-0.0175	-0.248***	-0.0763
	(0.0114)	(0.0133)	(0.0786)	(0.0932)
Constant	0.850***	0.266	12.13***	6.759***
	(0.208)	(0.238)	(1.982)	(1.963)
Observations	1121	1121	1121	1121
R^2	0.0823	0.1633	0.1786	0.2931
Wald χ^2	42.28	43.88	128.33	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.*

Accepted Manuscript

ACCEPTED MANUSCRIPT

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 & Organization 48, 197–216.

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., 2012. Reputation on the Internet, in The Oxford Handbook of the Digital Economy, ed. by M. Peitz and J. Waldfogel, Oxford University Press, chap. 13, 343-354.

Accepted Manuscript

ACCEPTED MANUSCRIPT

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

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 & Organization 55, 31-44.

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.

Dellarocas, C., Dini, F., Spagnolo, G., 2006. Designing Reputation (Feedback) Mechanisms. Handbook of Procurement, Cambridge University Press.

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.

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

Accepted Manuscript

ACCEPTED MANUSCRIPT

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., 2014. Emotions, Sanctions and Cooperation. Southern Economic Journal 80(4), 1002-1027.

Johnson, N., Mislin, A., 2011. Trust games: A meta-analysis. Journal of Economic Psychology 32, 865–889.

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. Games and Economic Behavior 86, 40–57.

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.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Livingston, J., 2005. How Valuable Is a Good Reputation? A Sample Selection Model of Internet Auctions. Review of Economics and Statistics, 87, 453-465.

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

Melnik, M., Alm, J., 2002. Does A Seller's eCommerce Reputation Matter? Evidence from eBay Auctions. The Journal of Industrial Economics 50, 337–350.

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.

Accepted Manuscript