



University of Caen

University of Rennes 1

Is the ebay feedback system really efficient ? an experimental study

David Masclet (CREM – CNRS – CIRANO)
Thierry Pénard (CREM – CNRS- University of Rennes 1)

jan 2008

Series : Industrial Economics

WP 2008-03

Working paper



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Research Center in Economics and Management
UMR CNRS 6211

Is the eBay feedback system really efficient?

An experimental study

David Masclet^{*}
Thierry Pénard[†]

Abstract *The eBay Feedback Forum is claimed to be a crucial component of the success of eBay. Many empirical studies have found that this feedback system exerts a deterrent effect on the opportunistic behavior the Internet's anonymity may incite buyers and sellers to adopt. The feedback system in place on eBay is however far from being perfect and may be especially vulnerable to strategic ratings (or non-ratings) that might reduce the informational content of feedback profiles. This article aims to examine the efficiency of the eBay feedback system, through a set of experiments based on the trust game. Our experimental design consists of four different treatments. The baseline treatment corresponds to a finite repeated simultaneous trust game. The second treatment, called “eBay rating” is identical to the baseline treatment except that we added a second stage in which the players have the opportunity of rating their partner. In this treatment, each participant is given the choice to either evaluate immediately or wait, knowing that only one rating will be accepted. The third treatment, called “Sequential rating” is identical to the “eBay rating” treatment, except that the order in which players evaluate one another is randomly determined by the computer. Finally in the fourth treatment, called “Simultaneous rating”, both players are required to make their rating decisions simultaneously. Our experimental results indicate that the eBay feedback system could be improved by either constraining partners to leave ratings simultaneously or by predetermining the rating sequence.*

JEL classification: C92, C72, L14, L86.

^{*} CNRS, CREM, CIRANO, Marsouin. David.Masclet@univ-rennes1.fr

[†] CREM, Université de Rennes 1, Marsouin. Thierry.Penard@univ-rennes1.fr

We would like to express our gratitude to K. Abbink, N. Nikkiforakis, and conference participants at the JEE (Strasbourg, 2006), ESA (Nottingham, 2006) and AFSE 2006, as well as E. Priour for scheduling and running the experimental sessions. We were able to carry out these experiments thanks to financial support provided by the Brittany Regional Council, as part of the Marsouin program.

1. Introduction

"The majority of people are honest and mean well. [...]. But you can unfortunately, on occasion, run into unscrupulous folks, [...] Our approach is to eliminate them systematically in order to protect the honest ones, and your active participation is vital to this effort. Sign up with eBay and make use of our evaluation procedure to leave comments on other members. Feel free to compliment those members who are deserving and cite grievances when merited."

Pierre Omidyar, *eBay* founder

The online auction web site eBay, founded in 1995 by the computer programmer Pierre Omidyar, is a good example of the success of electronic marketplaces. In 2006, a total of 222 million users were registered on eBay, with \$52 billion-worth of transactions. Millions of collectibles, appliances, computers, furniture, equipment, vehicles, and various other (even rare or very valuable) items are sold daily on eBay.

The success of such electronic marketplaces constitutes a challenge for economists. A number of the features of online markets, such as geographical distance and anonymity between buyers and sellers, make opportunistic behavior much easier than in traditional markets with "face-to-face" transactions. Sellers can be opportunistic by cheating on the quality of the product (for example by exaggerating its quality), or on delivery (e.g. not shipping, shipping items other than those described, shipping counterfeit merchandise or shipping slowly). Buyers can also be dishonest regarding the payment sent to the seller (for example by delaying payment).

However electronic marketplaces can reduce opportunistic behaviors by screening participants and monitoring transactions. But such mechanisms can only be implemented in small online marketplaces (like business-to-business places), where the anonymity of partners is limited. On C2C (customer-to-customer) marketplaces like eBay or Amazon, the huge number of participants and daily transactions makes it difficult to centralize monitoring. An alternative method is to let partners self-monitor their transactions by providing them a decentralized reputation-building mechanism. The eBay Feedback Forum, which is claimed to be a crucial component of the success of eBay, is a good example of such a decentralized mechanism. The idea is that the

traders themselves are often in a better position to monitor and punish (or reward) their partners. In the eBay forum, both buyers and sellers have the opportunity of rating each other. They can send a "positive", "negative", or "neutral" rating and leave a comment.¹ Each *eBay* user is therefore characterized by his or her feedback profile (i.e. the historical record of all the ratings they have received), which is available for consultation by all other users. The buyer and seller thus hold information on the reputation or reliability of their partner at the time of concluding the transaction. The Feedback Forum therefore plays both a punishment and a signaling role, since each trader can punish (reward) her partner by leaving negative (positive) ratings, but it also allows each trader to construct a publicly-observable reputation. The threat of having a bad public reputation may provide the trader with sufficient incentives to be honest.

Many empirical studies have found that this feedback system exerts a deterrent effect on the opportunistic behavior the Internet's anonymity may incite buyers and sellers to adopt. The empirical results show that a seller with good ratings can expect to sell an item more quickly and at a better price.² Houser and Wooders (2005) find that a 10% rise in the number of positive ratings recorded for a seller is associated with a 0.17% in the price that the seller can command, whereas a 10% rise in neutral or negative ratings lowers the price obtained by 0.24%. It is in the interest of both partners for their subsequent transactions to be as honest as possible in order to generate positive ratings, or at least avoid negative ratings, and so to improve or maintain their reputation.

The feedback system in place on *eBay* is however far from being perfect and has proven especially vulnerable to strategic ratings (or non-ratings) that reduce the informational content of

¹ In addition to leaving a overall feedback rating (positive, neutral, or negative) of the seller, buyers can also leave anonymous detailed seller ratings in four areas: item as described, communication, shipping time, and shipping and handling charges. The detailed seller rating system is based on a one- to five-star scale. Five stars is the highest rating, and one star is the lowest rating. Even if these detailed seller ratings do not affect the overall feedback score, they can provide additional information about the seller's performance.

² Resnick *et al.* (2006) and Dellarocas (2006) provide summaries of this work.

feedback profiles (Dellarocas *et al.*, 2006).³ For example, a buyer or seller can submit an unjustified positive rating, to encourage the transaction partner to reciprocate with a positive rating. Analogously, a participant may elect not to enter a "justified" negative rating for fear of receiving an "unjustified" negative rating in return. Dellarocas *et al.* (2006) and Klein *et al.* (2005) provide empirical evidence of these phenomena of strategic reprisals and reciprocity at work on *eBay*, which ultimately serve to increase artificially the number of positive ratings and reduce negative ratings (see also Resnick and Zeckhauser, 2002). Dellarocas *et al.* (2006) suggest that “the most efficient way to curb the retaliation/reciprocation problems is probably one suggested recently, consisting simply in keeping the left feedbacks secret until the period in which feedbacks can be given expires. Then agents cannot react anymore, and the feedbacks (if any) can be made public.”

In this paper we experimentally investigate the extent to which eBay Feedback Forum could be improved in substituting the current rating rules for alternatives that limit the possibility of strategic ratings: for example a rule requiring partners to evaluate simultaneously (i.e. to keep their ratings secret until the feedback period expires). More specifically, our analysis has three aims. First, we investigate the effects of rating on trust and cooperation between traders, by comparing marketplaces with and without feedback systems. Second, we try to understand sellers' and buyers' motives for rating their partners: I may wish to assign a negative (positive) rating to punish (reward) an unfair (fair) transaction with respect to payment, quality, delivery etc. (we call this Direct reciprocity); alternatively, I may want to assign a negative (positive) rating because I myself have received a negative (positive) rating (i.e. Indirect reciprocity); last, I may be willing to assign a positive rating because I expect that my partner will subsequently

³ Many complaints have been made about eBay's system of dealing with fraud. The complaints are generally that eBay sometimes fails to respond when a claim is made. Some complaints also concern eBay feedback. The eBay feedback system is used to combat fraud. However this system has a number of weaknesses, including the fact that small and large transactions carry the same weight in the feedback summary. It is therefore easy for a dishonest user to initially build up a deceptively positive rating by buying or selling a number of low-value items, such as e-books, recipes, etc., then subsequently switching to fraud.

reciprocate by sending me a positive rating (i.e. Strategic reason). While the first motive is directly related to the transaction, this is not the case for the two last motives that might reduce the informational content of feedback profiles. Finally, the third aim of this paper is to investigate whether slight modifications of the current eBay feedback system could improve efficiency. In particular, does the introduction of feedback rules that reduce rating that is not directly related to the transaction improve the informational content of ratings, and hence stimulate cooperation among partners?

To answer these questions, we run an experiment inspired by the trust game devised by Berg *et al.* (1995), in which Player A (the buyer) selects the amount of allocation he wishes to send to Player B (the seller). Player B actually receives three times the amount sent and then must decide how much to return to the first player. This game yields a good approximation of how an *eBay* transaction might be conducted inasmuch as *eBay* practice dictates that one of the commercial partners (the buyer) makes payment to the other (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.⁴ In comparison to the standard trust game, we add a second step, once the investment decisions have been made, during which participants are given the option of evaluating their partners.⁵

Three feedback systems are tested. In the first, each player can opt to submit a rating either immediately or at a subsequent point in time. This option most closely resembles the system currently in use on *eBay*. The second procedure imposes the order in which players evaluate their partners (sequential rating), and the last procedure calls for players to

⁴ The fact that both buyer and seller's decisions are not binary reflects the degree of trust. Indeed for example the seller can be dishonest on the degree of quality of the item sent to the buyer or on the number of days for delivery.

⁵ The first stage of this game is a drastic simplification of "real" transactions on eBay. Note however that in this study we mainly focus on the second stage (i.e. evaluation stage) and in particular the relationship between evaluation and trust that emerges from this first stage.

simultaneously submit ratings (simultaneous ratings). The comparison between these different treatments should allow us to disentangle the different motives for rating partners. In particular, a simultaneous feedback system should limit both strategic ratings and indirect reciprocity, and thereby enhance informational content relative to the current system. Simultaneous ratings prevent players from adopting strategic behaviors, since they cannot assign a positive rating in order to trigger a positive response from the partner (Dellarocas *et al.*, 2006). Simultaneous decisions also reduce indirect reciprocity, since players cannot punish (reward) a partner for having received a negative (positive) rating. Finally, the feedback system with sequential rating should provide an intermediate solution by preventing the first player (only) from indirectly reciprocating and the second player (only) from adopting strategic behavior.

Our analysis builds on previous experimental work. Our approach is related to Bolton *et al.* (2004), who ran an experiment using a two-stage game where buyers decide whether to send money and sellers then decide whether to ship the item. The authors compare a treatment with and without reputation. In the reputation condition, players were informed of each other's past play. Bolton *et al.* (2003) find that trust and trustworthiness were significantly higher under a reputation system.⁶ Our paper is also related to Keser (2003), the only previous paper, to our knowledge, to consider the effect of the eBay feedback system in the context of a trust-game experiment. Keser examined the effects of reputation by comparing three different treatments. In the baseline treatment, subjects play a repeated trust game under a stranger matching protocol. The reputation treatments (short- and long-term reputation) are similar to the baseline treatment, except that there is additional stage in which player A can rate player B by assigning her a costless positive, negative or neutral rating. At the beginning of each period player A will be therefore informed about player B's previous ratings. Keser finds that introducing this feedback system significantly increases cooperation in the trust game, in particular when subjects have full

information (long-run reputation). Our experiment builds on Keser (2003), with the notable exception that we allow both the buyer and the seller to rate their partners and we introduce a cost of evaluation. In addition, our experiment provides an in-depth analysis of the different motives for rating one's partner, and provides suggestions for the amelioration of the current eBay feedback system.

To anticipate our results, we find that trust is significantly improved by the use of a reputation feedback system. However, our results also indicate that trust runs deeper when the feedback system is more constrained than is *eBay's* current system. In particular we find that ratings are largely driven by indirect reciprocity (i.e. assigning a negative (positive) rating for having received a negative (positive) rating) and strategic motives (i.e. giving a positive rating while expecting the trading partner to reciprocate) in the eBay's current system. Consistent with this, positive ratings are generally assigned immediately (the strategic reason) while negative feedback is given later because of the fear of retaliative negative feedback (a form of last minute feedback). *eBay* operations could thus be improved by modifying the rating rules at the end of each transaction. For example, a system in which individuals are not informed their partner's decision before taking their own decision provides better results both in terms of trust and earnings.

The experimental protocol is presented in the next section. The experimental results are discussed in Section 3, and Section 4 concludes.

⁶ Bolton *et al.* (2006) also run experiments investigating how the interaction between market competition and reputation creates trust between sellers and buyers. See also Gazzale and Khopkar (2007) that use the same experimental design, but introduce the possibility for sellers to observe buyer's past feedback provision.

2. Experimental design

2.1. Overview

Our experimental design consists of four different treatments. The baseline treatment corresponds to a finite repeated simultaneous trust game. The game lasts for 20 periods. At the beginning of each period, participants A and B each receive a 10-unit allocation. Player A (the buyer) selects an amount between 0 and 10 units to send to B (the seller), while at the same time B determines the sum to be returned, which is between 0 and the amount received (which latter is the amount sent by A multiplied by three).⁷ Player A's gain is then equal to $10 - \text{amount sent} + \text{amount returned}$, and Player B's gain is $10 + 3 * \text{amount received} - \text{amount returned}$.

The second treatment, called “eBay rating”, is identical to the baseline treatment except that we added a second stage in which the players have the opportunity of rating their partner. Both players (A and B) can decide to rate their partner by assigning either a negative (-1) or positive (+1) point. However, leaving a rating costs 1 unit (i.e. 1/10th of the initial allocation).⁸ The rated player does not incur any direct cost or benefit, although the negative or positive points received are recorded on the player's feedback profile. This profile contains a historical record of all ratings, along with a score that represents the cumulative sum of positive and/or negative points obtained over all of the previous periods. At each new period, the player's profile is transmitted to their subsequent partner, so that each player is aware of their partner's ratings and has an idea of his partner's reputation. In the “eBay rating” treatment, the rating rules correspond quite close to those currently practiced on *eBay*, which leave buyers and sellers with a fair amount of flexibility to rate one another in the 90 days following the transaction: they can either rate immediately or wait until the rating period has almost elapsed before sending their ratings

⁷ This is a *cold* procedure, where A and B play simultaneously. Player A chooses the amount to send B, while B determines the amount to return for all potential amounts received from A. The advantage of this *cold* procedure is that it places the two players in a more symmetrical position than a so-called *hot* procedure (whereby A chooses first, and subsequently B), which could provide Players A and B with justified (i.e. non-strategic) reasons to evaluate their partner positively or negatively.

points. In this treatment, each participant is given the choice to either evaluate immediately or wait, knowing that only one rating will be accepted. This option is reflected in the experimental protocol by splitting the rating stage into two phases: each player has the possibility of proceeding with the rating straight away in Phase 1 or waiting until Phase 2. If the participant waits until Phase 2, he is made aware of his partner's choice (i.e. either an immediate positive/negative rating, or no rating), prior to ultimately deciding whether to evaluate his partner. With this system, players can implement various types of strategies. They can opt to submit a positive rating immediately in order to incite a positive rating in return (strategic motive), provided the partner has decided to wait or, on the other hand, to wait so as to punish any partner who gives a negative rating in Phase 1 (indirect reciprocity).⁹

The third treatment, called "Sequential rating" is identical to the "eBay rating" treatment described above, except that the order in which players evaluate one another (i.e. rating in phase 1 or in phase 2) is exogenously predetermined by a computer in each period. This variant constrains partners' rating freedom along with their possibility of adopting strategic behavior. For the player who is designated to rate first (i.e. in phase 1), the motivations are narrowed to indirect positive reciprocity, while the player in the second position only has negative indirect reciprocity (i.e. reprisal) as a potential strategic incentive.

The last treatment, called "Simultaneous rating", is even more restrictive, since both players are required to make their rating decisions simultaneously. In the *eBay* context, this rule would correspond to keeping buyer and seller ratings secret until the rating period has expired; this system would, in theory, eliminate both strategic ratings and indirect reciprocity (Dellarocas *et al.* (2006)).

⁸ On *eBay*, this cost does not take the form of a direct monetary expense, but rather an opportunity cost related to the amount of time or effort devoted to this task.

⁹ The waiting preference on the part of a player wishing to send a negative rating to an opportunistic partner can also be strategic, *i.e.* in order to avoid reprisals (in the form of a negative rating) from the punished partner.

These experimental treatments enable us to compare the performance and level of trust generated by different rating policies. Does the introduction of feedback rules reducing strategic rating incentives improve the informational content of ratings and hence stimulate cooperation among partners? Before presenting our results, we briefly present the theoretical predictions of our experimental games. Since players deal with different partners each period, it can be proved by backward induction that in the second stage, neither player should rate their partner, inasmuch as rating is expensive (regardless of the feedback system). In the absence of rating, the game reduces to a standard trust game with a trivial subgame perfect equilibrium: since the second player will always benefit by keeping everything, the first player therefore never sends anything and, in the end, each partner's gain equals their initial endowment. This situation is collectively suboptimal because by sending part of his endowment, the first player could have increased the gains of both players; gains would be maximized were the first player to send his entire endowment.

2.2. Procedure and parameters

The experiment consists of 25 sessions, with 10 participants in each session.¹⁰ There were at least six sessions, and thus independent observations, for each treatment. Seven sessions were conducted under the baseline and Simultaneous rating treatments, and six for both the sequential and the eBay rating treatments, giving a total of 252 participants. All of the sessions were held at the Center for Research in Economics and Management (CREM), University Rennes I, Rennes, France. The experiment was computerized using the Ztree program developed at the University of Zurich.¹¹ The subjects were undergraduate students from a variety of majors. Roughly one-third were Economics students in the first two years of their University studies, and all but a

¹⁰ With the exception of two sessions of the baseline treatment and one session of the SEQUENTIAL treatment which contained eight players.

small number of the remaining two-thirds were students in law, management, and medicine. None of the subjects had participated in an economic experiment previously. No individual participated in more than one session. On average, a session lasted 100 minutes, including initial instructions and subject payment.

At the beginning of the experiment, the instructions were distributed and read to the subjects.¹² All subjects were then required to answer a number of questions concerning the rules of the game and how earnings are determined. The experimenter then announced and explained the correct answers. Subjects could indicate whether they had any questions about the process and the experimenter would answer them in private.

Each session had twenty interaction periods. Each period within a session proceeded under identical rules. At the beginning of the experiment, each participant was assigned the role of player A or player B. They kept this role during the entire session. The computer network then matched the subjects into pairs of players, with one player A and one player B. A stranger matching protocol was used in all of the sessions: at the end of each period, the composition of the groups changed so that individuals were rematched with another partner on a random basis. Average player remuneration was 18 euros.

Table 1 presents a summary description of the sessions. The two first columns show the session number and the number of subjects who took part in the session. The third column indicates the treatment.

[Table 1 : about here]

¹¹ See Fischbacher (1999) for a description of the Ztree computer program.

¹² Game instructions are available upon request from the authors.

3. Experimental results

This section is organised as follows. Subsection 3.1 discusses the patterns of players' investments in the trust game and considers the impact of rating on the first-stage decisions. Subsection 3.2 analyses the determinants of rating in each rating system. In particular, we examine the extent to which the current eBay feedback system could be improved.

3.1. Trust and trustworthiness with and without feedback systems

Table 2 displays the means and standard deviations of the investments of players A and B in each treatment.

[Table 2 : about here]

Player A's average investment is the highest in the simultaneous treatment (4.36), followed by the Sequential (4.17), eBay (3.00) and Baseline (2.24) treatments, but there is considerable heterogeneity between groups in all of the treatments. A Mann-Whitney pairwise test comparing investment between treatments, under the assumption that each session is a unit of observation, reveals greater investment in the treatments with reputation (Sequential and Simultaneous) than in the Baseline treatment (two-tailed tests: $z=-2.714$ with $p<0.01$; and $z=-2.747$ with $p<0.01$). However the difference between the Baseline and the eBay treatments is not significant (two-tailed test: $z=-1.14$ with $p>0.1$), although we do observe higher levels of investment in the eBay treatment. These results indicate that introducing a feedback system significantly increases trust and trustworthiness between partners. However, it appears that the type of feedback rule matters, and that the most sophisticated system (like the eBay rating system) is not necessarily the most efficient. Investment is actually significantly higher in the Simultaneous rating system compared

to the eBay rating treatment (two-tailed test: $z = 1.715$ with $p < 0.1$).¹³ Figure 1 illustrates these differences in the investment by player A.

[Figure 1 : about here]

Figure 1 displays average investment by A-players per period for each treatment. The investment level is higher under treatments with rating. Further, average investment in the Simultaneous treatment is higher than in the eBay treatment for all the periods (except period 19). Investment decreases over time. When we decompose the periods into quarters, and compare average investment in periods 1-5 and 16-20, we find that average investment falls significantly in all treatments. This decrease is statistically significant in the baseline (Wilcoxon test: $z = 2.366$, $p < 0.05$). The same test also indicates significant declines in the eBay treatment ($z = 1.99$, $p < 0.05$) and the simultaneous treatment ($z = 2.366$, $p < 0.05$). Finally, the decline is not significant in the sequential treatment ($z = 0.524$, $p > 0.1$).

[Figure 2 : about here]

Figure 2 indicates the frequencies of each investment level. We can see that player A sent 0 units in about 45% of the cases in the baseline treatment. This frequency is significantly lower for the treatments with ratings, in particular in the Sequential (25.5%) and Simultaneous (26%) treatments. In contrast, the frequency of choosing high investments is higher for feedback treatments than for the baseline treatment. For example, the frequency of choosing 10 units is 14.7% in the Simultaneous treatment and only 5% in the baseline treatment.

Turning next to the return of player B, Figure 3 shows player B's investment for each level of player A's investment. For all treatments, the level of player B's investment is strongly and positively correlated with player A's investment.

[Figure 3 : about here]

¹³ The differences between eBay and Sequential, and between Sequential and Simultaneous, are not significant.

Table 2 above shows the average investments by player B and the average relative amounts returned in each treatment. B-players invest significantly more under reputation treatments than in the baseline treatment. They return an average of 4.21 (Simultaneous treatment) and 3.98 units (sequential treatment), which is significantly higher than the 1.45 units in the baseline treatment ($z=-2.875$ with $p<0.01$, and $z=-2.429$ with $p<0.05$, respectively for the simultaneous and sequential treatments). However the average return in the eBay treatment (2.8 units) is not significantly different from that obtained in the baseline treatment ($z=-1.571$ with $p>0.1$). Comparing the different rating systems, we note that the average investment level in the Simultaneous treatment is significantly greater than in the eBay treatment ($z=1.64$ with $p<0.1$). The other differences between rating systems are not statistically significant.

Figure 4 shows player B's average return over time for each treatment. This confirms that Player B's average investment is higher under treatments with a feedback system, and in particular under the Simultaneous treatment compared to the baseline treatment. It also shows that average return falls over time in all treatments. Comparing periods 1-5 and 16-20, this fall is significant in all treatments ($p=0.0180$, $p=0.0464$, $p=0.0747$ and $p=0.0280$ for the baseline, eBay, sequential and simultaneous treatments, respectively).

[Figure 4 : about here]

To sum up, the results of our experiments are all consistent with the hypothesis that introducing a feedback system improves cooperation, in particular when the rating system prevents the adoption of strategic behavior.

3.2 Measuring the efficiency of the different rating systems

To compare the impact of the three feedback systems on the amounts sent by players, we estimate a GLS panel model with random effects. The determinants of the amounts invested and

returned by players in period t are: the *amount received* in period $t-1$ from the previous partner, and a set of variables describing the current partner's profile. The "*positive rating in $t-1$* " ("*negative rating in $t-1$* ") variable indicates whether or not the partner received a positive (negative) rating in the previous period. These variables are interpreted in comparison with the omitted variable "*No rating in previous periods*". The *cumulative positive rating* variable takes the value of 1 if the cumulative sum of ratings since the beginning of the game is positive and 0 otherwise. The *cumulative negative rating* variable is constructed symmetrically.

The impact of feedback systems on player investments is measured by means of a dummy variable for each rating system (*Simultaneous, Sequential, eBay*). We also introduced a trend variable (*period*) as well as an indicator variable for the final period (*period_20*). Table 3 shows the results of the estimation of the GLS model.

[Table 3 : about here]

The amounts sent depend on the amount received during the previous period. Consequently, Player A is more strongly inclined to trust player B (i.e. to send him a higher share of his endowment) were he to have received a greater sum from his previous partner. Similarly, Player B returns more as the amount received previously by his partner rises. The negative coefficient on the "period" variable (trend) also shows that the investment level drops over time, even with the introduction of a feedback system. Furthermore, the amount sent is tied to the partner's feedback profile (i.e. his past ratings). The results show that players take into account the full history of past ratings and do not only focus on the most recent rating. Player A will therefore increase investment if the number of positive ratings received from prior partners exceeds the number of negative ratings. The marginal effects suggest that participants place greater emphasis on positive than on negative ratings in determining how much to send to their partner.

We also find that changing the current eBay rating policy in favor of more restrictive rules (ratings submitted in a predetermined order or simultaneous ratings) leads to greater trust and trustworthiness.

3.3. Motivations for rating one's partner and limitations of the current system

In this section, we consider the motivations for rating one's partner. In the context of a trust game, the motivation of ratings may be complex, depending on both emotional reactions to the amount received from the partner (direct reciprocity), reactions to received ratings (indirect reciprocity) and anticipated reactions to ratings (strategic reasons). Since the Sequential and Simultaneous systems aim to reduce both strategic manipulation and indirect reciprocity, we should see lower rating levels in the Sequential and Simultaneous treatments compared to the current eBay system.

Comparing the individual rating patterns in the three systems yields some interesting and intuitive results. Figure 5 illustrates the relative frequencies of overall ratings (both positive and negative rating) in all the rating treatments.¹⁴ This shows that as the game is repeated, the number of ratings tends to fall over time. Furthermore, consistent with our conjecture, Figure 5 shows that the average rating is higher under the eBay treatment than under the Sequential and Simultaneous treatments.

[Figure 5 : about here]

This result is also supported by Table 4. which shows that the frequency of rating is lower for the simultaneous (20%) and the sequential (23%) treatments compared to the eBay treatment (27%). The frequency of overall rating in the eBay treatment is significantly higher than in the

¹⁴ Similar results were obtained when we distinguished between positive and negative ratings.

Simultaneous treatment ($z=2.017$, $p=0.0437$). The differences between Sequential and eBay, and between Sequential and Simultaneous, are not significant.

[Table 4 : about here]

Table 5 provides further information about the determinants of rating in each treatment. First, it indicates that both buyers and sellers evaluate their partners. However, it appears that A-players generally assign more negative ratings while B-players assign more positive ratings.¹⁵ Second, consistent with direct reciprocity motivations, ratings are strongly correlated with investment levels. Table 5 shows that traders assign negative ratings for low investment levels. In contrast positive ratings are more likely to be assigned to high investment levels. Third, Table 5 also indicates that most of the negative ratings are assigned in phase 2 of the eBay and Sequential treatments while positive ratings are generally given in the first stage. This result indicates that traders assign negative ratings in the second phase to escape retaliation (a *last minute feedback* strategy) while evaluating positively in first phase may incite the partner to reciprocate by evaluating positively.

[Table 5 : about here]

In order to provide more formal evidence of the informational bias and inefficiency contained in the current *eBay* feedback system, we estimate a probit in Table 4 for the probability of negatively rating one's partner within the eBay treatment using a selection-bias correction.¹⁶ We explain rating motivations by the following explanatory variables: the amount sent or returned by the partner and a dummy variable for whether the participant decided to rate during

¹⁵ A potential explanation of this difference is that B-players have two ways of negatively reciprocating: by assigning negative ratings and by returning lower levels of investment.

¹⁶ As the probability of submitting a negative (positive) rating actually depends upon the probability of submitting an rating. We considered two separable decisions using a two-step estimation procedure : first the decision to evaluate someone and second, conditional on the decision to evaluate, the choice of assigning a negative evaluation. We first estimate the evaluation probability using a random-effects Probit model; we then explain the negative evaluation decision, conditional on the decision to evaluate corrected for a potential selection bias via the inverse of the Mills

Phase 1. We also consider interactions for whether the rating posted by the partner is positive (or negative) and whether this information was known at the time his own rating was posted. These latter variables enable us to identify the presence of *strategic* rating motivations based on reciprocity (either positive or negative).

[Table 6 : about here]

Table 6 shows that the probability of negatively (positively) rating one's partner decreases (increases) with the amount received. Moreover, the negative and significant coefficient on "*Rating during Phase 1*" shows that rating immediately reduces the probability of a negative rating. This reveals the presence of strategic rating, with players posting a positive rating as quickly as possible, to incite their partner to rate them positively in return. This also means that players who want to rate their partner negatively tend to wait for the second phase (if possible) to avoid reprisals. Furthermore, the positive and highly significant coefficient on the interaction "*Received a negative rating AND knows this result*" means that players use negative ratings as a means of reprisal against partners who assign them a negative rating.

These results support our conjecture that eBay's current feedback system produces informational bias and can be improved by reducing strategic rating motivations (i.e. by constraining the partners to rate each other simultaneously). All of these results are consistent with Resnick and Zeckhauser (2002) and Dellarocas *et al.* (2004), who found reciprocity and retaliation in eBay ratings. Such behavior could explain the overrepresentation of positive rating in eBay users' profiles.

ratio (the "IMR" variable). The estimation results of the probit selection are not presented here (available upon

3.4. Welfare levels

Table 7 shows the profits of players A and B in each treatment. In all treatments, player B's profits are significantly higher than player A's (Wilcoxon tests, $p < 0.005$). Profit dispersion is also higher for B-players than for A-players. This result is confirmed in Figure 6, which displays average profit over time in each treatment.

[Table 7 and Figure 6: about here]

Player A's profit is on average 9.21 in the baseline treatment compared to 9.44, 9.57 and 9.62 in the eBay, Sequential and Simultaneous treatments, respectively. The difference in total earnings between the baseline treatment and the Simultaneous treatment is significant at the $p < 0.1$ level, according to a Mann-Whitney rank-sum test. However, the two other treatments with rating systems do not generate significantly higher earnings than the Baseline treatment.

Player B's average profit also increases from 15.27 in the baseline treatment to 16.01, 18.3 and 18.65 in the eBay, Sequential and Simultaneous treatments. The differences between the baseline and the Simultaneous treatment as well as the Sequential treatment are significant ($z = -2.492$ with $p = 0.012$, and $z = -2.429$ with $p = 0.0127$, respectively). However, the difference between the baseline treatment and the eBay treatment is not significant ($z = -0.429$, $p = 0.668$). These results indicate that the introduction of feedback systems is welfare-improving, if reciprocation and retaliation are ruled out.

4. Conclusion

This experimental study has provided several insights into the impact of feedback systems on trust and trustworthiness in electronic marketplaces. First and foremost, it has shown the benefit of rating systems like those introduced on *eBay*. The results establish that participants

request).

(whether buyers or sellers) rely upon their partner's profile when making decisions. This effective use of information provided by feedback profiles, as empirically observed on *eBay*, is identified in a controlled experimental setting. The second key finding is that the design of rating policy influences both the level of trust and efficiency of a *eBay*- like marketplace. From this perspective, implementing more flexible feedback mechanisms can erode the trust between partners, by generating excessive negative or positive rating reciprocities. In contrast, a system in which ratings are kept secret serves to limit strategic ratings and proves more effective in enhancing trust.

This study has thus paved the way to the examination of the appropriate design of virtual communities and electronic marketplaces. It also confirms the interest electronic marketplace managers have in performing laboratory experiments prior to implementing or changing the design of their platforms (for example *eBay*'s rules and policies).

References

- Berg J., Dickhaut, J. and MacCabe K. (1995), "Trust, Reciprocity and Social History", *Games and Economic Behavior*, 10, pp. 122-142.
- Bolton G., Loebbecke C. and A. Ockenfels (2006) "How Social Reputation Networks Interact with Competition in Anonymous Online Trading: An Experimental Study", University of Cologne, Working Paper Series in Economics N°32.
- Bolton G. Katok E. and A. Ockenfels (2004) "How effective are electronic reputation mechanisms? An experimental investigation", *Management Science*, 50 (11), pp. 1587-1602.
- Dellarocas C. (2006), "Reputation Mechanisms", in T. Hendershott, (ed.), *Handbook on Economics and Information Systems*. Elsevier.
- Dellarocas C. (2003), "The Digitization of Word-of-Mouth: Promise and Challenges of Online Feedback Mechanisms", *Management Science*, 49 (10), pp. 1404-1427.
- Dellarocas C. Dini, F. and Spagnolo, G. (2006) "Designing Reputation (Feedback) Mechanisms", in *Handbook of Procurement*, Cambridge University Press.
- Dellarocas C. Fan M. and Wood, C. (2004), "Self-Interest, Reciprocity and Participation in Online Reputation Systems", *MIT Sloan*, Working Paper 4500-04.
- Gazzale, R. and Khopkar, T. (2007) "Remain Silent and Ye Shall Suffer: Seller Exploitation of Reticent Buyers in an Experimental Reputation System", Working Paper.
- Houser, D. and Wooders, J. (2005), "Reputation in Auctions: Theory and Evidence from *eBay*" Forthcoming in *Journal of Economics and Management Strategy*.
- Keser, C. (2003), "Experimental Games for the Design of Reputation Management Systems", *IBM Systems Journal* 42(3), pp. 498-506.
- Klein, T. Lambertz, C. Spagnolo, G. et Stahl, K. (2005) "Last Minute Feedback", Stockholm School of Economics.
- Resnick P and Zeckhauser, R. (2002), "Trust Among Strangers in Internet Transactions: Empirical Analysis of *eBay*'s Reputation System", in Michael Baye, ed., *The Economics of the Internet and e-Commerce*. Elsevier Science, pp. 127-157.
- Resnick P and Zeckhauser, R. Swanson, J. and lockwood K. (2006), "The Value of Reputation on *eBay*: A Controlled Experiment", *Experimental Economics*, 9(2), pp. 79-101.

Tables

Table 1: Characteristics of the Experimental Sessions

Session Number	Number of Subjects	Treatment	Session Number	Number of Subjects	Treatment
1	10	Baseline	14	10	Sequential
2	10	Baseline	15	10	Sequential
3	8	Baseline	16	10	Sequential
4	8	Baseline	17	8	Sequential
5	10	Baseline	18	10	Sequential
6	8	Baseline	19	10	Sequential
7	10	Baseline	20	10	Simultaneous
8	10	eBay	21	10	Simultaneous
9	10	eBay	22	10	Simultaneous
10	10	eBay	23	10	Simultaneous
11	10	eBay	24	10	Simultaneous
12	10	eBay	25	10	Simultaneous
13	10	eBay	25	10	Simultaneous
			Total	252 subjects	

Table 2. Player A's investment and Player B's return

	Baseline	eBay	Sequential	Simultaneous
Player A's investment	2.24 (2.91)	3.02 (3.09)	4.17 (3.50)	4.36 (3.70)
Player B's investment	1.45 (3.16)	2.8 (4.67)	3.98 (5.35)	4.21 (6.32)
Relative return for player A in %	11.8%	19.1%	22.4%	22.8%
Observations	1280	1200	1160	1400

Table 3: Determinants of the amounts invested by participants (GLS panel model with random effects)¹⁷

Variable	All treatments	All treatments	Treatments with rating	
	Amount sent by Player A in period t (1)	Amount sent by Player B in period t (2)	Amount sent by Player A in period t (3)	Amount sent by Player B in period t (4)
Amount received in $t-1$	1.912*** (0.231)	1.645*** (0.030)	1.657*** (0.255)	0.173*** (0.037)
Cumulative positive ratings (partner's profile)			0.496*** (0.066)	0.486*** (0.093)
Cumulative negative ratings (partner's profile)			-0.335*** (0.052)	-0.373*** (0.124)
Positive rating in $t-1$ (partner's profile)			0.068 (0.274)	0.215 (0.394)
Negative rating in $t-1$ (partner's profile)			-0.135 (0.185)	-0.14 (0.446)
eBay rating	0.673** (0.366)	1.148*** (0.290)	Ref.	Ref.
Sequential rating	1.740*** (0.370)	2.154*** (0.297)	0.961** (0.388)	0.913*** (0.323)
Simultaneous rating	1.923*** (0.353)	2.360*** (0.285)	1.116*** (0.371)	1.57*** (0.312)
Period	-0.107*** (0.011)	-0.113*** (0.020)	-0.082*** (0.018)	-0.152*** (0.034)
Period_20	0.179 (0.266)	-0.737 (0.491)	-0.103 (0.310)	-0.918 (0.613)
Constant	3.113*** (0.284)	2.323*** (0.312)	3.938*** (0.319)	3.409*** (0.402)
Observations	2394	2394	1786	1786
R ²	0.16	0.08	0.17	0.08

Notes: Standard errors are shown in parentheses; *** statistically significant at the 1% level; ** at the 5% level; and

* at the 10% level.

¹⁷ Our results are robust with respect to other estimation techniques like ordered probit, fixed effect models, and OLS regressions that do not include random effects.

Table 4. Frequencies of rating per treatment

Treatments	Sessions	Frequencies of ratings	Frequencies of positive ratings	Frequencies of negative ratings
eBay treatment	1	0.155	0.03	0.125
	2	0.285	0.045	0.24
	3	0.395	0.185	0.21
	4	0.24	0.19	0.05
	5	0.275	0.12	0.155
	6	0.285	0.145	0.14
	Average	0.27	0.12	0.15
Sequential treat.	7	0.31	0.095	0.215
	8	0.235	0.16	0.075
	9	0.225	0.05	0.175
	10	0.225	0.13	0.095
	11	0.255	0.15	0.105
	12	0.155	0.06	0.095
	Average	0.23	0.10	0.13
Simultaneous treatment	13	0.19	0.07	0.12
	14	0.225	0.065	0.16
	15	0.175	0.055	0.12
	16	0.21	0.07	0.14
	19	0.225	0.1	0.125
	20	0.225	0.13	0.095
	Average	0.20	0.08	0.12

Table 5 . Ratings per treatment depend on average investment and returns.

Treat.	Rating	Rating of player A			Player B's invest.	Rating of player B			Player A's invest.
		All	Seq rating			All	Seq rating		
			Phase 1	Phase 2			Phase 1	Phase 2	
Simultaneous	No rating	77.71%			3.64	78.71%			4,23
	Negative rating	16%			2.8	11.57%			1,71
	Positive rating	6.29%			14.81	9.71%			8,53
Sequential	No rating	75.34%			3.42	77.76%			3,69
	Negative rating	17.59%	20.55%	50%	3.14	7.41%	70.27%	72.46%	2,25
	Positive rating	7.07%	79.45%	50%	12.14	14.83%	29.23%	27.54%	7,60
eBay	No rating	72.83%			2.49	72.67%			2,89
	Negative rating	20.17%	30.51%	58.70%	1.98	10.50%	61.54%	90.28%	1,5
	Positive rating	7%	69.49%	41.30%	11.23	16.83%	38.46%	9.72%	6,33

Table 6: The determinants of ratings (Probit with random effects for the probability of negative rating)

Variable	(1) A's rating of B	(2) B's rating of A
Amount received	-0.241*** (0.043)	-0.430*** (0.067)
Rating during Phase 1	-1.574*** (0.471)	-0.785** (0.328)
Received a positive rating AND knows this result	0.004 (0.428)	-0.358 (0.649)
Received a negative rating AND knows this result	1.785** (0.827)	0.812** (0.403)
Constant	3.810*** (0.789)	2.607** (1.070)
Observations	149	153
Log-likelihood	-32.27	-49.52

Notes: Standard errors are shown in parentheses; *** statistically significant at the 1% level; ** at the 5% level; and * at the 10% level.

Table 7. Players' profits

	Baseline	eBay	Sequential	Simultaneous
Player A's profit	9.21 (2.66)	9.44 (3.66)	9.57 (3.99)	9.62 (5.10)
Player B's profit	15.27 (7.22)	16.01 (7.21)	18.3 (7.91)	18.65 (8.94)
Observations	1280	1200	1160	1400

Figures

Figure 1: Player A's average investment in Each Treatment, by Period

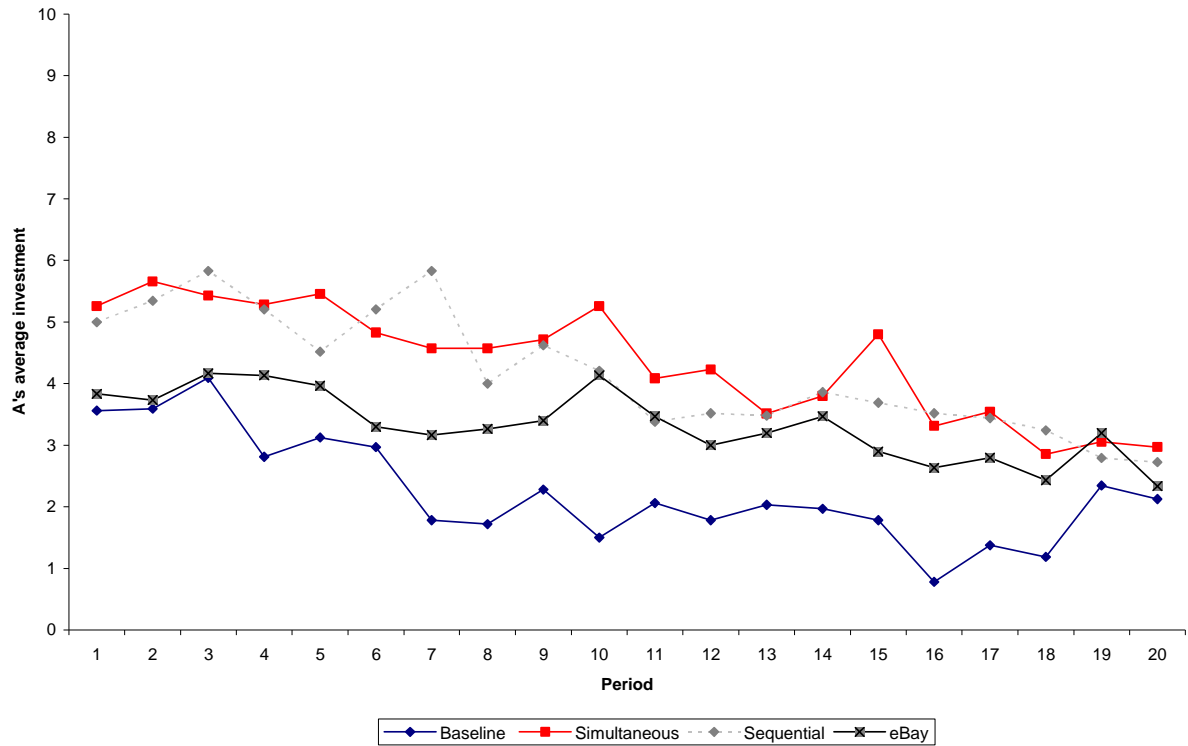


Figure 2: Distribution of investment levels for player A

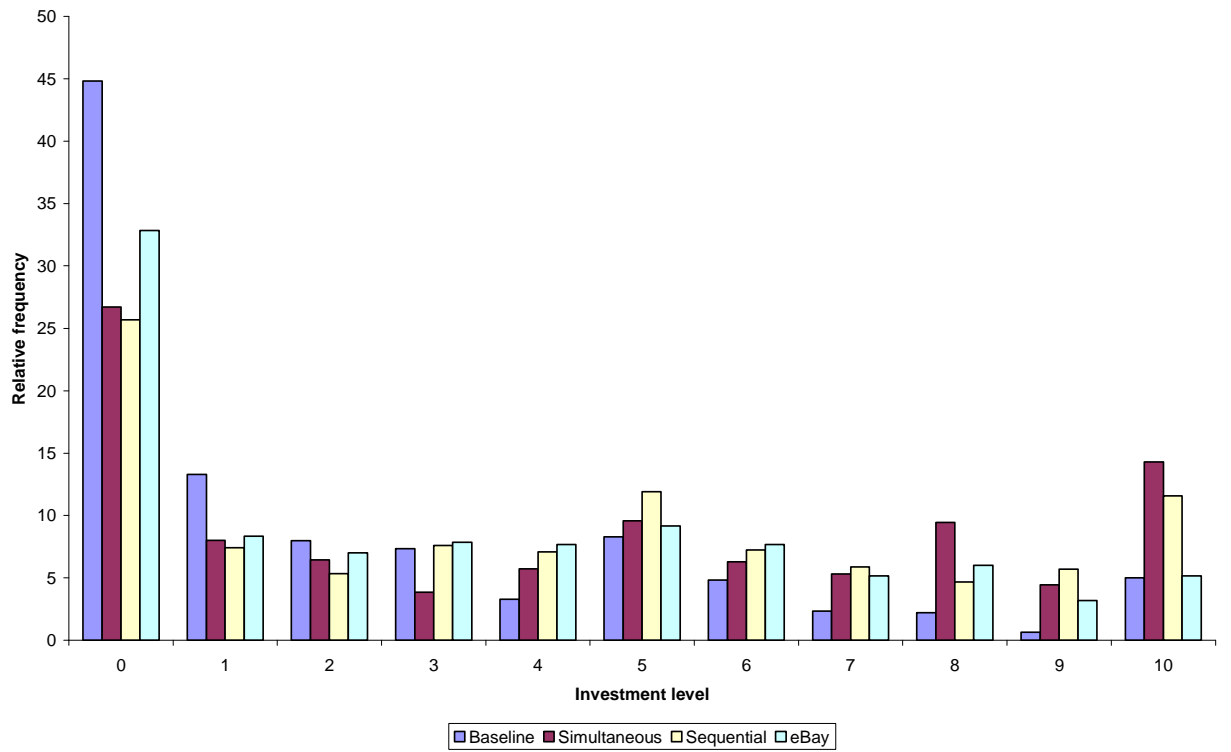


Figure 3. Player B's return for each player A's investment level

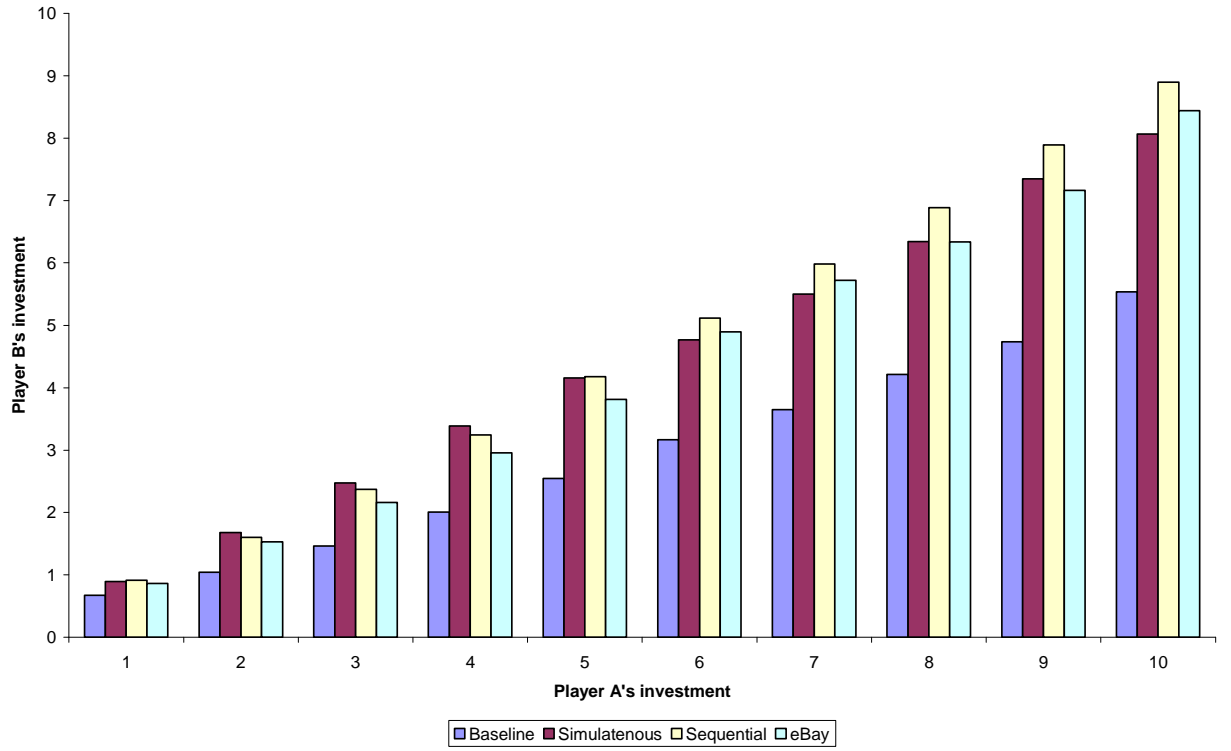


Figure 4: Player B's Average return in each treatment, by period

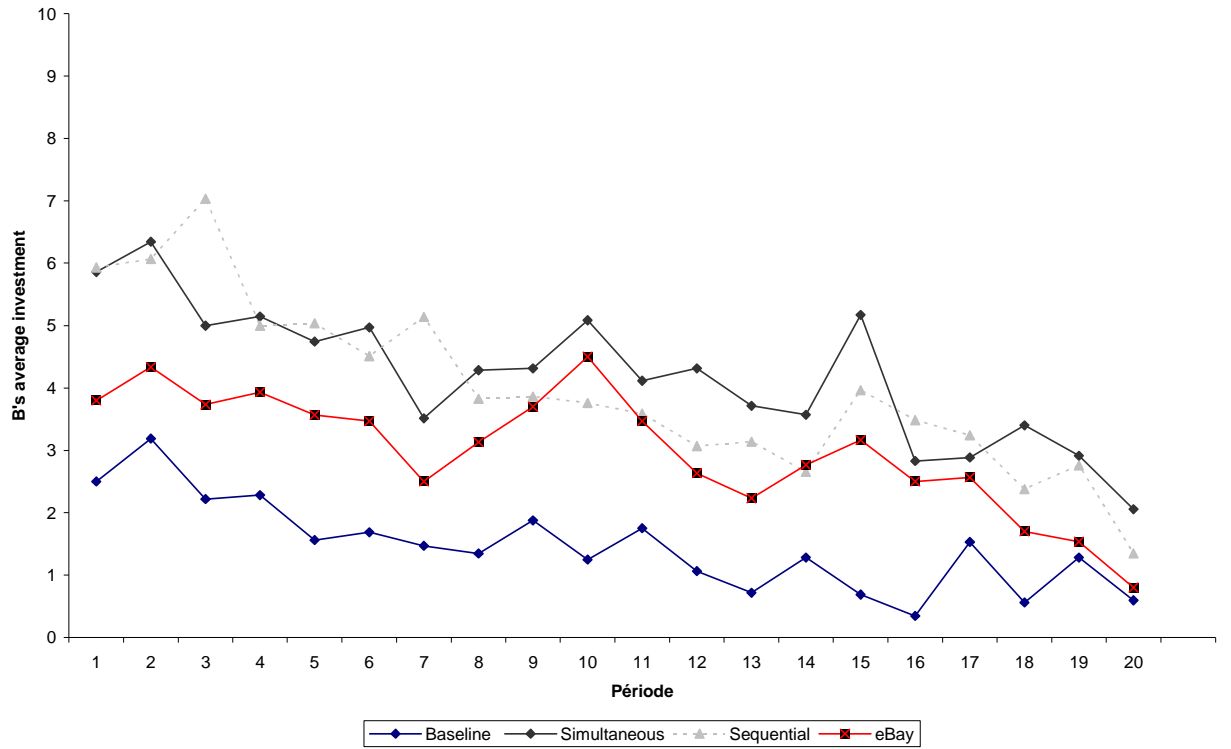


Figure 5. Average rating over time

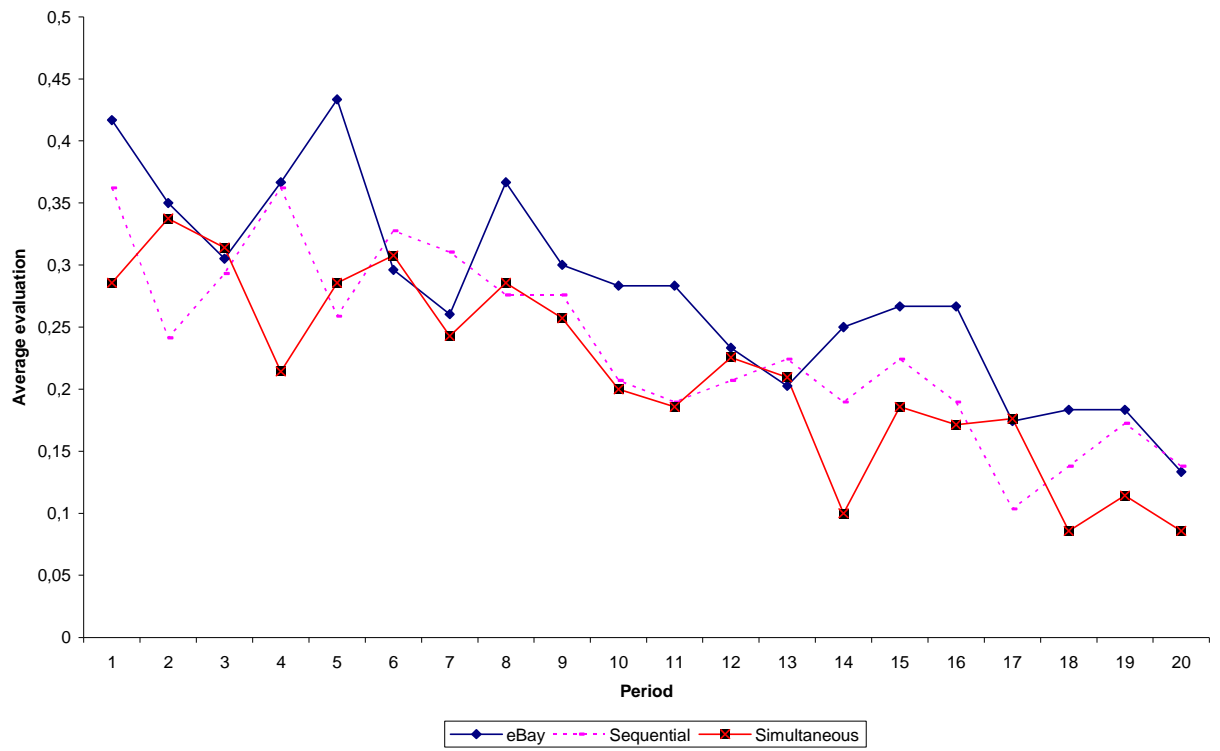


Figure 6. Average payoff over time

