RC22654 (W0211-254) November 25, 2002
Other

IBM Research Report

Managing Brands in e-Business: an Experimental Study in Trust and Reputation Management

Claudia Keser
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Delhi - Haifa - India - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g. , payment of royalties). Copies may be requested from IBM T. J. Watson Research Center , P. O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at http://domino.watson.ibm.com/library/CyberDig.nsf/home .

MANAGING BRANDS IN E-BUSINESS: AN EXPERIMENTAL STUDY IN TRUST AND REPUTATION MANAGEMENT*

Claudia Keser

IBM T. J. Watson Research Center, Yorktown Heights, NY,
and
Centre Interuniversitaire de Recherche en Analyse des Organsisations (CIRANO),
Montreal

November 20, 2002

Abstract: Transactions on online markets require a great deal of trust among anonymous trading partners. To mitigate some of the risks involved in anonymous transactions, several online market sites have implemented reputation management mechanisms that differ in structure and function. In a series of experiments, we examine the impact of two simple reputation management mechanisms on the evolution of trust and trustworthiness in a model e-business environment.

Résumé: Les transactions par Internet exigent une confiance considérable entre les intervenants anonymes. Pour minimiser les risques associés aux opérations anonymes, les responsables de plusieurs marchés en ligne ont mis en place des mécanismes de gestion de réputation différant en structure et fonction. Par une série d'expériences, nous évaluons l'impact de deux mécanismes de gestion de réputation relativement simples sur la confiance et la coopération dans un environnement modélisé de commerce électronique.

Keywords: Trust, Reputation, Experimental Economics, E-commerce
Mots-clés: Confiance, Réputation, Économie Expérimentale, Commerce Électronique

JEL Classification: C72, C91

* I thank CIRANO/LUB in Montreal for their hospitality and in particular Claude Montmarquette and Jacinthe Majeau for their help in running the experiments. Thanks are also due to Prince Stanley who developed the experimental software, which is based on z-Tree by Urs Fischbacher at the Institute for Empirical Research in Economics at the University of Zurich.

1. Introduction

Trust may play a critical role in the development of the Internet as a marketplace. For example, in the emerging world of business-to-business collaboration, successful collaborations require trust in their partners to behave ethically. A company that shares internal data such as sales reports, production schedules, product designs and logistical details with a supply-chain partner must trust the partner with that information. Other examples where trust plays a role are informal online markets where individuals may buy and sell a wide variety of goods and services. In these markets, single, isolated trades often take place between anonymous counterparts. There may be no opportunity for inspection of the item to be traded. Thus, each of the trading parties might be tempted to cheat. As a buyer of Beanie Babies at eBay (http://www.eBay.com), for example, I face some risk that the seller has not accurately described the condition of his or her Beanie Babies, will not pack them properly, or will not deliver them in a timely fashion. To manage this kind of risk several approaches have been proposed (see, for example, Kollock 1999, and Malaga 2001.) For example, third party escrow services could be used. They have the disadvantage, though, that they are time-consuming and costly (service charges). It is sometimes advised to reduce some of the risk related with online trading by frequent communication with the trading partner and by insisting on the revelation of enough information to make the trading partner identifiable. However, there seems to be little hope of actually tracking down a trading partner, given the opportunities to disguise identities due to, for example, free e-mail services. As a more powerful approach, many of the online market sites have developed reputation management systems that allow the trading parties to submit a rating of the counter party's performance in a specific transaction, which will be made available to all visitors of the site. A positive rating of my trading partner is likely to increase my trust in the performance of the counter party.

EBay, for example, uses a reputation management system, called the Feedback Forum that allows participants in a transaction to rate each other with a "+1" for a positive comment, a "-1" for a negative comment or a "0" for a neutral comment.[1] All ratings that an eBay user has received from other eBay users are summed up to build his

[1] In addition to this rating, they may also leave a textual comment.

or her or her Feedback Rating number. This number is given in parenthesis after each seller or bidder's name. A user who has 100 positive comments thus gets a Feedback Rating number of 100. However, another user with 150 positive and 50 negative comments would also receive 100. Any user whose Feedback rating number reaches –4 is suspended from participation. The Feedback Rating number is part of the user's Feedback Profile. The full Feedback Profile can be obtained by clicking on the number. It contains the full list of textual comments and a summary table, which clearly identifies the most recent ratings and comments.[2]

In this paper, we examine the effect of simple reputation management mechanisms on trust and trustworthiness using the method of experimental economics. Our experiments are based on the *trust game* introduced by Berg, Dickhaut, and McCabe (1995). In this game, trust is measured by the amount that one of two players, the *investing player*, unilaterally invests by sending it to the other, the *trusted player*. The trusted player receives three times the amount invested and may then return some amount to the investing player. The amount he returns provides a measure of the trusted player's trustworthiness. Our intuitive hypothesis is that the introduction of a rating system into the trust game, in which the investing player rates the other player's trustworthiness, should increase both trust and trustworthiness. Kollock (1999) argues that reputation management systems lead to lower levels of fraud. He cites a 1997 eBay summer report stating that over 99.99 percent of eBay auctions were completed successfully. Kollock also argues that traders with negative reputations are selected out: not only does the software, as described above, prohibit their further trading but there also is a reluctance by other people to trade with them. Due to the fact that in our experiment, the two players are exogenously matched, the reluctance to trade can show only in a lower investment (trust) level. In our experimental study, besides examining effects on trust and trustworthiness, we investigate strategic aspects of rating and reputation building given the specific reputation management mechanism. One variant of the rating system manages short-run reputation while the other manages long-run reputation and we are

[2] Amazon (http://www.amazon.com) uses a different feedback system to rate sellers: "Any time you make a purchase from a seller at Amazon marketplace, Auctions, or zShops, you're encouraged to rate the seller's performance and leave a short comment. The average ratings accompany a seller's name in every reference and appear as one to five stars, with five stars being the best."

interested in whether the long-run reputation management mechanism is more effective than the short-run reputation management mechanism in increasing trust and trustworthiness. Dingledine, Freedman and Molnar (2001) for example observe a shortcoming of the full history, long-run reputation management mechanism used by eBay:

> *"In the eBay case, a group of people engaged in auctions and behaved well. As a result, their trust ratings went up. Once their trust ratings were sufficiently high to engage in high-value deals, the group suddenly 'turned evil and cashed out.' That is, they used their reputations to start auctions for high-priced items, received payment for those items, and then disappeared, leaving dozens of eBay users holding the bag."*

A related question with respect to the long-run reputation management mechanism to be addressed in our study is whether it financially pays to build up a positive reputation. The results of empirical eBay studies suggest that buyers are willing to pay more for a good coming from a highly rated seller. Kalyanam and McIntyre (2001), for example, find in their study on eBay auctions of Palm Pilot personal digital assistants that reputation has significantly positive returns. Houser and Wooders (2000) also show that sellers in eBay auctions with a high reputation score receive higher bids than those with a lower score.

Experimental and field studies are to be considered as important complements. In the experimental economics laboratory we have the advantage of being able to control the environment to a large extent at a relatively low cost. In our study we can, for example, directly compare the levels of trust and trustworthiness in an environment without any reputation management system to the levels of trust and trustworthiness in the same environment modified by the introduction of a specific rating system. Furthermore, we can compare the functionality of various rating systems.

Our experimental design, thus, involves three treatments, *a baseline and two rating treatments.* The baseline treatment is one in which participants repeatedly play the trust game, remaining in the role of either the investing player or the trusted player but

interacting in each repetition with another unidentified participant. The latter implies that *strangers* interact with each other. La Porta et al. (1997), for example, argue that trust is more essential to ensure cooperation between strangers than between partners who interact frequently and repeatedly. Among partners, reputation building and opportunities for future punishment could support cooperation even with low levels of trust. In our baseline treatment this kind of cooperation would be difficult to build up as reputation building and punishment could work only indirectly through an effect on the entire population. Our reputation treatments are similar to the baseline treatment: they also involve the interaction of strangers, but allow the investing player, at the end of each repetition, to rate the trustworthiness (cooperation) of the trusted player based on the amount returned. A player's trustworthiness may be rated as *positive, neutral* or *negative.* The trusted player is informed about his rating. In the *long-run reputation treatment*, the investing player is informed in the beginning of each repetition, before he makes his investment decision, about the most recent rating and the distribution of all previous ratings of his current trusted party. In the *short-run reputation treatment* he is informed of the most recent rating only.

The following Section 2 presents some definitions of trust. In Section 3 we describe the trust game, previous experimental results and our experimental design. The results of our experiments are presented in Section 4. Section 5 summarizes the results and concludes the article with a discussion of their relevance in e-commerce and marketing.

2. The role of trust, reputation and related concepts

It is difficult to distinguish trust from related concepts, which on the surface resemble trust. Yamagishi and Yamagishi (1994) argue that the most comprehensive definition of trust would be *taken-for-grantedness* of the reality, implying that trust is considered a psychological mechanism for reducing complexity in the environment (Luhmann 1988). However, trust is typically assigned another role: trust provides a solution to the problem caused by *social uncertainty.* Social uncertainty is defined to exist when I am incapable of correctly determining the intentions of other persons who have an incentive to act against my own best interest. We will thus limit our attention to

trust in other beings and organizations. Barber (1983) distinguishes between two types of trust, *trust in another person's competence* and *trust in another person's goodwill*. The former is the expectation of technically competent role performance from those involved with us in social relationships and systems, while the latter is the expectation that partners in interaction will carry their duties in certain situations to place others' interests before their own. Yamagishi and Yamagishi suggest denoting the expectation of competency as *confidence*, and to define trust as the expectation of goodwill and benign intent. They further distinguish between trust and *assurance*, where they define assurance as the expectation of benign behavior for reasons other than goodwill of the other person. In other words, trust is based on the inference of the interaction of another person's traits and intentions, whereas assurance is based on the knowledge of the incentive structure surrounding the relationship. They give a nice example:

> *Suppose I have a special tie with the Mafia, and my trading partner knows this. I am certain that he will not cheat on me; he knows that if he does he will be quickly sent to a mortuary. My expectation of the partner's "honesty" is based on the fact that acting "honestly" is in his own interest, not on the belief that he is a benevolent person. Here, assurance exists but no trust.* (Yamagishi and Yamagishi 1994, p. 132)

Note that in the trust game by Berg, Dickhaut, and McCabe (1995) the amount sent by the investing player yields a measure of trust in the goodwill of the other player. However, when we extend the trust game by the introduction of a rating mechanism, assurance will play some role. The investing player knows that, at least initially, the trusted player might want to build up a good reputation. Thus, we expect in the experiments with a reputation management system to observe higher investment levels than in the baseline experiments without such a mechanism. The difference in the trust levels of the experiments with a reputation management system and the baseline experiment may be considered a measure for assurance.

Yamagishi and Yamagishi (1994) discuss commitment as another concept distinct from trust and assurance. To solve the problem of social uncertainty people form

mutually committed relations. This reduces social uncertainty and thus the need for trust. In a repeated prisoners' dilemma situation, for example, it is possible to induce others to cooperate by the use of a tit-for-tat strategy (reciprocity) (see Axelrod 1984, Selten Mitzkewitz, and Uhlich 1997, Keser 2000). Commitment plays an important role in repeated trust games with partners, such as in Cochard, Van Phu, and Willinger (2000). Thus, the higher investment level in their experiment than in the previous one-shot experiments on the same game.

Reputation may play two different roles in social interactions involving trust. The first role is informational. It makes the recipient of positive reputation information trust more. Trust has been defined above as an expectation that (potential) partners have goodwill in their dealings with us. We do not have perfect information about their intentions, which we have to infer from available information, as for example their reputation. The second role of reputation is a kind of sanctioning. The attribution of a negative reputation may work as a sanctioning mechanism to punish dishonest behavior. This makes the owner of reputation act in a more trustworthy way. Thus, we expect in the experiments with a reputation management system to observe more trustworthiness than in the baseline experiments without such a mechanism.

3. The Experiments

The experiments are based on the trust game, originally called investment game, presented by Berg, Dickhaut, and McCabe (1995), which we discuss in the first subsection. In the second subsection, we present a brief summary of previous experimental results. To examine the impact of simple reputation management mechanisms among strangers in our experiments, we compare the results of two reputation treatments to those of a baseline treatment. The treatment design and organization of the experiments are presented in the third subsection.

3.1 The trust game

In the trust game there are two players. Let us call them player A and player B. Both players have an endowment of 10 Experimental Currency Units (ECUs). There are two decision stages. In the first stage of the game, player A has the opportunity to send

(invest) part or all of his endowment to player B, but he need not send anything. Player B will receive three times this amount. Then, in the second stage of the game, player B may return any amount between zero and the received amount to player A, but he need not return anything. Then the game is over. The profit of player A, in ECU, is 10 minus the amount he sent to player B plus the amount that player B returned to him. The profit of player B, in ECU, is 10 plus the amount received (i.e. three times the amount that player A sent) minus the amount that player B returned to player A.

The game-theoretical solution to this game can be found by backward induction (Selten 1965). In the *subgame perfect equilibrium* solution player A sends nothing to player B, anticipating that the latter, being fully rational, would never return anything. In other words, no transaction takes place between the two players. Thus, each player remains with a profit equal to his initial endowment of 10 ECU. Obviously, this solution is socially inefficient: due to the tripling of the amount invested by the A-player, the two players can earn profits that sum up to 40 ECU if player A invests 10 ECU (his entire endowment), while the sum of their profits is only 20 ECU in the subgame perfect equilibrium solution.

3.2 Previous experimental results

The seminal experiment by Berg, Dickhaut, and McCabe (1995) on a single play of the trust game shows, that participants in the role of the investing player do trust in the other player and invest positive amounts. The trusted players on the aggregate tend to return the amount that the other player invested.[3] These results contradict the game-theoretical solution of the game of zero investment and a zero return (independently of the investment level). The experiment has been replicated by Ortmann, Fitzgerald, and Boeing (2000), Meidinger, Robin, and Ruffieux (1999), Croson and Buchan (1999), Buchan, Croson, and Jonson (2000), Buchan, Croson, and Dawes (2000), Willinger, Keser, Lohman, and Usunier (*forthcoming*). While the qualitative results in these other experiments are the same, the quantitative details have been shown to depend on gender and culture. Also Fehr, Gächter, and Kirchsteiger (1997), Fehr, Kirchsteiger, and Riedl

[3] Another result of this study is that the provision of social history, that is, the provision of information on the amounts invested and returned in previous experimental sessions, significantly increases investment and return in the current session.

(1993, 1998), Fehr, Kirchler, Weichbold, and Gächter (1998) report that participants in experiments, structurally similar to the trust game, often respond reciprocally when they receive a *gift*.[4] For a nice and comprehensive survey of experimental trust games, see Chapter 3 on "Social Preferences in Dictator, Ultimatum, and Trust Games" of Colin Camerer's forthcoming book on Behavioral Game Theory.

In order to interpret the observed behavior as some kind of economic equilibrium behavior different from the subgame perfect equilibrium solution[5], Bolle (1998) defines trust as the anticipation of reciprocating behavior. He reviews several experimental studies that examine this kind of trust, including the one by Berg, Dickhaut, and McCabe (1995), with the focus on the question of whether trust pays. That trust pays would present a minimal requirement for a game-theoretic equilibrium. His answer is *yes*, trust pays: the net value from trust is generally not inferior to the net value from mistrust, although it is often close to zero.

In an experimental trust game where two partners interact repeatedly, Cochard, Van Phu, and Willinger (2000) observe higher levels of trust and reciprocity than in the one-shot game, except for the last round. Their interpretation is that some of the trusted players try to build up a reputation of being cooperative players in early rounds.

3.3 Experimental design

In the experiments we examine three different treatments. In the baseline treatment ten participants, five of them in the role of player A and five of them in the role of player B during the entire experiment, play twenty repetitions (called *periods*) of the trust game. In each of the twenty repetitions the A-B pairs are re-matched such that the same A-B pair will never interact in two consecutive periods. The participants have complete information about the parameters and the rules of the game but they can never identify the other player. The subgame perfect equilibrium to the twenty-fold repetition of the trust game prescribes, for each repetition, that player A sends nothing to player B who would not return anything. If we ignore the potential effects of a small population of

[4] The research by Ernst Fehr and his co-authors reveals the potential of trust and reciprocity to enforce high effort levels in situations of incomplete employment contracts, where the firm offers a fixed wage and the worker has considerable discretion in determining his effort level.

[5] Note that the theories of inequality aversion by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) are, among others, motivated by the behavior observed in trust game experiments.

participants and the re-matching of the same A-B pairs in later periods, this treatment might also be considered as consisting of twenty one-shot trust games.

The reputation treatments are similar to the baseline treatment. However, in each period of the trust game, there is an additional third decision stage, in which player A rates, based on player B's return, the *cooperation* (or, trustworthiness) of the B-player with whom he currently interacts. If the A-player did not send anything to player B, the latter cannot return anything, and thus player A cannot rate his cooperation in that period. The rating can be *positive*, *neutral*, or *negative*. Player B is informed about the rating given to him by player A. Based on the reputation information given to player A, we distinguish between the short-run reputation treatment and the long-run reputation treatment. In the latter, in the beginning of each repetition, player A is informed about (i) the distribution of previous ratings and (ii) the most recent rating of the B-player with whom he is going to interact in that repetition, while in the former player A is informed about the most recent rating only. The subgame perfect equilibrium solution for both of these treatments is again for player A to invest nothing and for player B to return nothing, and rating never takes place.

The experiments were run in November 2001 (baseline and long-run reputation treatments) and in March 2002 (short-run reputation treatment) in the experimental economics laboratory of CIRANO and LUB. This laboratory has privacy conditions sufficient to assure that participants can not even discern whether other subjects are idle or engaged in input: heavy curtains surround and eliminate visual observation, and a heavy carpet eliminates audible clues to activities. We ran 32 experimental sessions, each with ten participants. Thus a total of 320 subjects participated in the experiments. The participants were randomly recruited from a subject pool of students of several universities in Montreal. The language was French. Subject payment was on average $ CN 30.

We organized eight sessions of the baseline treatment and twelve sessions of each of the two reputation treatments. In the beginning of each session, instructions were distributed and read aloud (an English translation of the instructions for the reputation treatment is in the Appendix). Then several questions were presented on each subject's computer screen to test the subject's understanding of the instructions. Only when all

participants had correctly answered to all of the questions could the experiment start. At the end of the each experiment, each subject was paid in cash, depending on his or her profit in the game multiplied by a conversion rate of seven Cents per ECU. A show-up fee of $ CN 5 was added.

4. Results

In the first part of this section we focus on the comparison of trust, trustworthiness and profits in the three treatments. In the second part we examine reputation effects on investment, rating behavior, and the profitability of building and maintaining a positive reputation. The non-parametric data analysis is based on SPSS 10.0 for windows (*Wilcoxon* signed ranks tests, and Mann-Whitney *U*-tests) and Siegel (1987, *Binomial* test). We always consider two-tailed asymptotic significance levels, requiring a 10 percent significance level. The outcome of each session is considered as one independent observation. The tests are thus based on eight independent observations of the baseline treatment and twelve independent observations of each of the two reputation treatments. For the regression analysis we use Limdep 7.0.

4.1 Comparison of the treatments

In the following, we consider the A-players' investments (trust), the B-players' returns (trustworthiness) and both players' profits.

4.1.1 Trust

Table 1 presents the averages, standard deviations and medians of the investments by the A-players over all periods in each session of the three treatments. The median investments in the baseline, the short-run reputation and the long-run reputation treatments are three, five, and seven ECUs, respectively. Comparing the average investments in all three treatments, the Kruskal Wallis test allows us to reject the null hypothesis of no difference ($p = 0.017$). Pairwise comparisons yield significantly higher investment levels in the reputation treatments (5.15 ECUs in the short-run reputation treatment and 6.05 ECUs in the long-run reputation treatment) than in the baseline treatment where the average investment is 3.91 ECUs (U-tests, $p = 0.090$ and $p = 0.004$,

respectively). This implies a 32 percent increase in the short-run reputation treatment and a 55 percent increase in the long-run reputation treatment relative to the baseline.[6] The difference in the investment levels of the two reputation treatments is statistically not significant ($p = 0.248$) although Figure 1, which exhibits for each of the three treatments the development of investments over time, reveals that in each period the average investment in the long-run reputation treatment is higher than in the short-run reputation treatment. Furthermore, we observe a similar pattern in the beginning and toward the end of the game: during the first periods the trust level increases (trust needs, to some extent, to be built up), while the trust level decreases dramatically in the final periods (trust breaks down).[7] Thus we expect the difference of the two treatments to show more clearly in the intermediate phase of the game. Decomposing the twenty periods of interaction into quarters, that is, periods 1-5, 6-10, 11-15, and 16-20, we observe a significant difference in the average investments of the two reputation treatments in the third quarter ($p = 0.061$) while the difference in the other three quarters is not significant.

For a rough examination of the evolution of investment over time, we compare average ECU investments in periods 1-10 to those in periods 11-20. They decrease from 4.27 to 3.56 in the baseline treatment. This decline is statistically significant (Wilcoxon test, $p = 0.05$). In the reputation treatments there is no significant tendency for the average investment to decline, although we observe an overall decline from 5.37 to 4.94 ($p = 0.126$) in the short-run reputation treatment and an overall decline from 6.14 to 5.96 ($p = 0.433$) in the long-run reputation treatment. Thus, only in the baseline treatment do investments by the A-players significantly tend to decline over time.[8]

---

[6] Note that the investment level in our baseline treatment is below the average investment of 5.16 observed by Berg, Dickhaut, and McCabe (1995) in their experiment where participants played only a single round of the trust game.

[7] This is a typical *end game effect*: from the vast literature on finitely repeated public goods experiments, for example, we know that cooperation, actively established by the players in the beginning of their interaction, almost always breaks down toward the end (e.g., Keser and van Winden 2000). See also Selten and Stoecker (1986).

[8] In each of the reputation treatments, the standard deviation significantly increases from periods 1-10 to periods 11-20 (Wilcoxon tests, $p = 0.077$ in the short-run reputation treatment and $p = 0.002$ in the long-run reputation treatment). The increase in the baseline treatment is statistically not significant ($p = 0.123$). The standard deviation in the long-run reputation treatment is significantly higher than in the baseline treatment (U-test, $p = 0.031$). The differences between the short- and long-run reputation treatments and between the short-run reputation and the baseline treatments are statistically not significant ($p = 0.133$ and $p = 0.643$, respectively).

Figure 2a shows the relative frequencies with which investments of zero, one, two, etc, up to ten ECUs were chosen on the aggregate over all periods in each of the three treatments. In the baseline treatment the mode is at zero ECU, while it is at ten ECUs in each of the reputation treatments. When we consider the choices in the first period only, we observe the mode at an investment of two ECUs both in the baseline and the short-run reputation treatment, while there are modes at investments of ten and five ECUs in the long-run reputation treatment. From the very first period on, the average investment is significantly higher in the long-run reputation treatment (5.03 ECUs) than in the baseline treatment (3.75 ECUs) (U-test, $p = 0.056$). The first period investment level in the short-run reputation treatment is 4.25 and significantly different neither from the one in the baseline treatment ($p = 0.446$) nor from the one in the long-run reputation treatment ($p = 0.173$). When we consider the choices in the last period only, there is a clear mode at zero ECU in all treatments (37.5 percent in the baseline, 48.3 percent in the short-run reputation and 51.7 percent in the long-run reputation treatment. In the reputation treatments we thus have the typical end-game effect discussed in Footnote 7 above.

Figure 2b, which presents the cumulative distribution of the investment decisions in each of the three treatments, visualizes our previous result that the "expected" investment level is higher in the two reputation treatments than in the baseline treatments: it reveals that investments in both reputation treatments first-order stochastically dominate investments in the baseline treatment.

4.1.2 Trustworthiness

Table 2 shows the average ECU investment by the A-players, the tripled amount received by the B-players, and the average absolute and relative amount returned by the B-players in each session of both treatments. We observe that the B-players in the reputation treatment return averages of 7.10 ECUs (short-run reputation) or 8.88 ECUs (long-run reputation), each of which is significantly higher than the amount of 3.81 ECUs returned on average in the baseline treatment (U–tests, $p = 0.025$ and $p = 0.001$,

respectively).[9] Given that the B-players in the reputation treatment tend to receive more from the A-players than the B-players receive in the baseline treatment, this result is not surprising. Thus, we consider the returned percentage of the amount received (*relative return*). The B-players' relative return is 46 percent in the short-run reputation treatment and 49 percent in the long-run reputation treatment, while it is only 32 percent in the baseline treatment.[10] The comparison of the relative returns in the three treatments based on the Kruskal Wallis test yields a significant difference ($p = 0.002$). Pairwise comparisons show that the difference between each of the two reputation treatments and the baseline treatments is statistically significant (U-tests, $p = 0.005$ for the short-run reputation treatment and $p = 0.002$ for the long-run reputation treatment). In other words, the relative returns are significantly higher in the reputation treatments than in the baseline treatment. The difference in the two reputation treatments is statistically not significant ($p = 0.248$). However, Figure 3, which shows the development of the relative returns in each of the treatments over time, suggests that end game behavior that appears in both reputation treatments and initial similarities might largely drive this result. Thus, we decompose the twenty periods of interaction into quarters, that is, periods 1-5, 6-10, 11-15, and 16-20. We observe that in the intermediary two quarters the relative return is significantly higher in the long- than in the short-run reputation treatment ($p = 0.073$ and $p = 0.015$ for the second and the third quarter, respectively).[11]

On the aggregate, the B-players in the baseline experiment return about one third (32.46 percent) of the received amount. In other words, they return roughly the amount that was invested by the A-player, neither more nor less. Comparing the B-players' average absolute returns to the average investments by the A-players, we find no significant difference (Wilcoxon test, $p = 0.575$). This is in keeping with the previous experimental results discussed in Section 3.2 above. In contrast to this, in the reputation treatments, the B-players return almost one half (46 percent in the short-run reputation

[9] The difference in the returns of the short- and long-run reputation treatments is statistically not significant ($p = 0.248$).

[10] This implies a 51 percent increase in the long-run reputation treatment and a 41 percent increase in the short-run reputation treatment.

[11] For a rough examination of the development of the relative returns over time, we compare the average relative returns in periods 1-10 to the average relative returns in periods 11-20. In none of the treatments do we observe a significant de- or increase (Wilcoxon tests).

treatment and 49 percent in the long-run reputation treatment) of the received amount. In other words, they return more than was invested by the A-players (138 and 147 percent, respectively). Comparing the B-players' average absolute returns to the average investments by the A-players, we find a significant difference (Wilcoxon test, $p = 0.006$ and $p = 0.003$, respectively).

Figure 4 exhibits the relative return for each investment level (from 1 to 10). The return of roughly a third in the baseline treatment and of roughly one half in the reputation treatments seems more or less independent of the investment level. This is in keeping with the previous results of Berg, Dickhaut, and McCabe (1995) for a single play of the trust game.

4.1.3 Profits

Table 3 exhibits the profits of players A and B in each session of each treatment. In all three treatments the B-players gain significantly higher profits than the A-players (Wilcoxon tests, $p = 0.012$ in the baseline treatment, and $p = 0.002$ in the reputation treatments).

The A-players gain on average 9.90 ECUs per period in the baseline treatment and 11.95 and 12.83 ECUs in the short- and long-run reputation treatments, respectively. We observe a significant difference in the A-players' profits in the three treatments (Kruskal Wallis test, $p = 0.001$). Pair-wise comparisons show that the differences between the baseline treatment and each of the reputation treatments are statistically significant (U-tests, $p = 0.004$ and 0.001, respectively). This implies that the A-players significantly profit from the introduction of the rating system, whether it is building long- or short-run reputations.

The B-players' average profit increases from 17.93 ECUs in the baseline treatment to 18.36 and 19.27 ECUs in the short- and long-run reputation treatments. However, there is no significant difference in the B-players' profits in the three treatments (Kruskal Wallis test, $p = 0.523$). As the differences between the baseline treatment and each of the reputation treatments are statistically not significant (U-tests, $p = 0.699$ and $p = 0.316$, respectively), the B-players do not significantly profit from the introduction of the rating system. Note that this result, together with the previous observation that A-players gain

significantly less than B-players but are better off with a rating system than without, implies more equitable profits of A- and B-players.

Comparing the two reputation treatments, the A-players' profits show no significant difference (U-tests, $p = 0.106$). Neither do the B-players' profits ($p = 0.386$).

For a rough examination of the development of the profits over time we compare the average profits in periods 1-10 to the profits in periods 11-20. In none of the treatments do we observe a statistically significant de- or increase, neither for the A-players nor for the B-players (Wilcoxon tests).

4.2 Rating and reputation effects

In the following we will analyze the effects of the B-player's previous ratings on the A-player's investment, the rating criteria of the A-player, and whether the building up of a positive long-run reputation tends to positively affect a B-player's profit.

4.2.1 Effect of B's previous ratings on A's investment

In the first part of this subsection we examine the impact of a B-player's most recent rating on the A-player's investment in both evaluation treatments. In the second part we analyze the effect that the distribution information of all previous ratings of a B-player (the overall rating) had on the A-player's investment in the long-run reputation treatment. In the third part we use regression analyses to jointly examine both effects.

Most recent rating

To evaluate the impact of the B-player's *most recent rating* on the A-player's investment, consider Table 4. The average investment obviously increases with the most recent rating. In the long-run reputation treatment, considering the values for each independent group, we observe a significant increase from neutral to positive and a significant increase from negative to positive (Wilcoxon tests, $p = 0.008$). However, the increase from negative to neutral is not significant ($p = 0.158$). We conclude, thus, that the most recent rating is significantly important only when it is positive. In the short-run reputation treatment, all increases are statistically significant ($p = 0.023$ from negative to neutral, $p = 0.003$ from neutral to positive, and $p = 0.002$ from negative to positive).

Overall rating

To examine, in the long-run reputation treatment, the impact of the distribution of the previous ratings let us, in an ad hoc way, map the distribution into a unique single-valued measure, called the *rating score*. The mapping is not reversible as each rating score can represent many potential distributions. The rating score is determined, similar to the eBay Feedback Rating number, by

(i) giving a value of +1 to each positive rating, a value of –1 to each negative rating and zero to each neutral rating in the previous periods

and

(ii) summing these numbers.

In periods 2 to 20, we observe rating scores ranging from –12 to 19. The Spearman rank correlation coefficient between the individual rating scores in each period and the respective average investments by the A players is 0.844 and thus significantly positive at the 1 percent level (ignoring that our observations are not strictly independent here). We conclude that the average investment significantly increases with the B-player's rating score.

Comparing the average rating scores per session in periods 1-10 to those in periods 11-20 we observe that of the 12 sessions nine had a positive average rating score in periods 1-10 and increased that score in periods 11-20, while one session with a negative average rating score in periods 1-10 made it to a positive average rating score in periods 11-20. Only two sessions built up a (more) negative reputation over time: they had a negative rating score in periods 1-10 that became more negative in periods 11-20. We conclude that there is a significant increase in the rating score from the first sequence of 10 periods to the second sequence of 10 periods (Binomialtest, 5 percent level).

In the following, we simplify the analysis even further by distinguishing simply between a positive, zero or negative score. Recall, however, that this is not the way in

which the distribution of previous ratings was presented to the participants in our experiment. Table 5 reveals the average investment contingent on the overall rating score being positive, neutral, or negative. Again, we observe that the average investment increases with the overall rating score increasing from negative to neutral and to positive. Considering the values for each independent group, we observe a significant increase from neutral to positive (Wilcoxon test, $p = 0.006$), from negative to positive ($p = 0.002$), and from negative to neutral ($p = 0.099$).

Regression results

The analysis so far suggests that both the overall and the most recent ratings tend to influence player A's investment in the long-run reputation treatment. A regression analysis will provide us with further evidence of the impact of the B-player's ratings on the A-player's investment. Let us use a simple linear regression model to explain the investment In_i of A-player i in period t by a vector of explanatory variables x_{it}, the corresponding parameter vector b and a random variable e_{it}:

$$In_{it} = b\, x_{it} + e_{it}$$

The explanatory variables are defined in Table 6. The regression results for each of the three treatments are presented in Table 7. In all three treatments, we observe a significantly positive constant and a significantly positive influence of *LIn*, the A-player's investment in the preceding period, while the impact of *LRR*, the B-player's relative return in the previous period is insignificant. In the baseline treatment we additionally identify a significantly positive coefficient of *LIn*LRR*, the interaction effect of the player's own investment and the B-player's relative return in the preceding period, and a significantly negative trend (*Period* effect). In the two evaluation treatments, the coefficient of the dummy variable *DnoLIn* is significantly positive, which implies that the A-player tends to make a positive investment after having invested zero in the preceding period. Furthermore, there is a significantly negative last period effect. In the long-run evaluation treatment we identify a significantly positive coefficient of *LIn*LRR*, the crossed effect of the player's own investment and the receiving B-player's relative return

in the preceding period, and a significantly negative trend (*Period* effect); both of these coefficients are insignificant in the short-run evaluation treatment. In the short-run evaluation treatment *RecentPo*, the B-player's positive evaluation in the preceding period has a significantly positive impact on the A-player's investment, while *RecentNe*, the B-player's negative evaluation in the preceding period has a significantly negative impact on the A-player's investment. It appears as if in the short-run evaluation treatment the A-player's are largely focused on whether the most recent evaluation of the B-player they are interacting with was positive or negative but they do not take so much account of their own previous experience regarding the relative returns they received from the B-players (in particular, the variable *LIn*LRR*). In the long-run evaluation treatment, of the most recent evaluation, only *RecentPo* has a significantly positive impact, which is in keeping with the results of the non-parametric analysis presented above. A positive sum of previous evaluations *SumPo* significantly increases while a negative sum of previous evaluations *SumNe* significantly decreases the A-player's investment.

4.2.2 Rating by the A-player

For the following analysis let us define two potential norms of cooperation in the trust game, the *equal split of the gross surplus* or the *equal split of the net surplus* of the investment. Table 8 presents for each investment level (from 1 to 10 ECUs), how much player B should return (i) if he equally splits the amount he received (the *gross surplus* of investment) and (ii) if he equally splits the profit of both players (the *net surplus* of investment). If case (i) applies player B should return one-half of the amount received, while he should return two-third of the amount received if case (ii) applies.

To examine how the A-players rated the B-players, based on the relative return they received, consider Table 9 for each independent session. On average over all sessions, A-players gave a negative/neutral/positive rating for a relative return of 20.91/49.56/61.96 in the short-run reputation treatment and for a relative return of 25.04/50.79/61.76 percent in the long-run reputation treatment. Thus, on the average, A-players gave a negative rating if they received less than one third of the amount that player B received or, in other terms, less than they invested. They gave a positive rating if the B-player returned more than one half of what he received or, in other terms, if player B was more

generous than splitting the gross surplus equally. They gave a neutral rating if the B-player returned about one half of what he received or, in other terms, if player B equally split the gross surplus.

4.2.3 Long-run reputation and profits

Table 10 presents for each session of the long-run reputation treatment the sum of the overall rating scores in the beginning of the final period, and the average profits of the players in that session with an overall positive, neutral or negative rating score. The comparison of the positively and negatively rated B-players shows no important difference. When we aggregate over all sessions the positively rated B-players make a higher profit than the negatively rated B-players: 39 of the 60 B-players have an overall positive rating, and their average per period profit is 19.29 ECUs; while 16 of the 60 B-players have an overall negative rating, and their average per period profit is 18.66 ECUs. However, in seven of the sessions the negatively rated B-players gain more than the positively rated B-payers, while the opposite is true in four of the sessions. This suggests an externality or session effect, implying a positive correlation between the number of positively rated B-players and the A-players' investment level in a session. However, there is no statistically significant correlation between the sum of the overall rating scores of a session and the average profit for the B-players or the average investment by the A-players in that session.

5. Conclusion

5.1 Summary of the results

The introduction of a reputation management system increases both the level of trust (investment) and the level of trustworthiness (returns), and thus the overall efficiency. Trust in the absence of a reputation management system does not yield net gains for the investing party, while it does so in the presence of a reputation management system. While in the absence of a reputation management system we observe a relatively low trust level that even tends to decrease over time and a continuously low level of trustworthiness, we observe continuously high levels of trust and trustworthiness in the

presence of a reputation management system. The introduction of a reputation management mechanism leads to a significant gain for the investing party but not for the trusted party. This implies more equitable profits of the investing and the trusting party through the reputation management system.

Short- and long-run reputation management systems show significant differences in their functionality in the intermediate phase of interaction: the long-run reputation management system leads to higher trust and trustworthiness than the short-run reputation management system.

To receive a positive rating the trusted party needs to be more generous than sharing the gross surplus evenly.

Both short- and long-run reputation tend to have an impact on the amount that will be sent to the trusted party. If the distribution of all previous ratings of the trusted party is available, its most recent rating has a significant impact on the amount invested by the trusting party only if the most recent rating is positive.

The trusted parties care for their long-run reputation if it is at stake.

It is not clear whether it individually pays for a trusted party to develop and maintain a positive reputation.

5.2 Importance of the results and method for e-commerce

We have built a framework that allows us to examine the performance of various other reputation management mechanisms and compare it to the performance of the mechanisms presented in this study that are similar to eBay's feedback forum. Recall that some traders exploit eBay's current reputation mechanism. They built up a reputation by buying or selling cheap items solely to enable them to cheat buyers on high priced ones. Thus, eBay might want to consider modifications of its rating system such that, for example, a rating received for a higher priced item is valued more heavily in the user's feedback rating number than a rating received for a lower priced item. One might also be interested in the function of the positive-neutral-negative rating relative to the function of the one-to-five star mechanism used by Amazon. Note that compared to field experiments (e.g., the one presented by Resnick, Zeckhauser, Swanson, and Lockwood

2002) laboratory experiments allow for a better control; they are also faster, cheaper and less risky.

Our conjecture is that reputation management systems will likely become more important in the future as the Internet develops as a market place. They clearly are a contributing factor to success, as we can see at sites such as eBay. Their existence and performance can make a big difference in buyers' satisfaction or firms' profit and thus in their reliance on online markets.

5.3 Other potential applications

Other important applications of this research on reputation management mechanisms are in marketing. For example, trust and reputation are at the heart of the discussion of buyer-seller relationships. Firms are increasingly concentrating on fewer but more intense relationships with their suppliers for their inputs and their channel members for distribution. This implies the adoption of cooperative systems such as electronic data interchange or joint marketing programs, which make the firms vulnerable to the opportunistic behavior by their partners. In a meta analysis, Geyskens, Steenkamp, and Kumar (1998) provide quantitative evidence that trust plays a crucial role in the building of cooperative marketing channels. They show that whether trust develops depends on how parties feel and behave and on the realized outcomes. The production of trust through reputation management systems also is an important issue for brand management. Using data from the market for red Bordeaux wines, Landon and Smith (1997) show that reputation (past quality) and collective reputation (average group quality) play a major role in consumers' decision making. Their results also indicate that reputation has a strong impact on consumers' willingness to pay. They conclude that " […] consumers place considerable value on mechanisms that disseminate information on the past quality performance of firms." (page 313). Interestingly, "[…] consumers primarily base their purchase decision on persistent, rather than short-run, movements in quality."

References

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

Barber, B., 1983, *The Logic and Limit of Trust*, New Brunswick, NJ: Rutgers University Press.

Berg J., J. Dickhaut, and K. McCabe, 1995, Trust, reciprocity, and social history, *Games and Economic Behavior* 10, 122-142.

Bolle, F. 1998, Does trust pay?, working paper, Europa-Universitaet Viadrina, Frankfurt (Oder).

Bolton, G. E., and A. Ockenfels, 2000, ERC: A theory of equity, reciprocity and competition, *American Economic Review* 90, 166-193.

Buchan, N., R. Croson, and R. M. Dawes, 2000, Who's with me? Direct and indirect trust and reciprocity in China, Japan, Korea, and the United States, working paper.

Buchan, N., R. Croson,, and E. Johnson, 2000, Trust and reciprocity: an international experiment, working paper.

Cochard, F., N. Van Phu, and M. Willinger, 2000, Trust and reciprocity in a repeated investment game, working paper.

Croson, R., and N. Buchan, 1999, Gender and culture: international experimental evidence from a trust game, *American Economic Review* 89, 386-391.

Camerer, C., *forthcoming*, *Behavioral Game Theory: Experiments on Strategic Interaction*, Princeton University Press.

Dingledine, R., M. J. Freedman, and D. Molnar, 2001, Accountability, in: A. Oram (ed.), *Peer-to-Peer: Harnessing the Benefits of a Disruptive Technology*, O'Reilley & Associates.

Fehr, E., S. Gächter, and G. Kirchsteiger, 1997, Reciprocity as a contract enforcement device: experimental evidence, *Econometrica* 65, 833-860.

Fehr, E., G. Kirchsteiger, and A. Riedl, 1993, Does fairness prevent market clearing? An experimental investigation, *Quarterly Journal of Economics* 108, 437-460.

Fehr, E., G. Kirchsteiger, and A. Riedl, 1998, Gift exchange and reciprocity in competitive experimental markets, *European Economic Review* 42, 1-34.

Fehr, E., E. Kirchler, A. Weichbold, and S. Gächter, 1998, When social norms overpower competition: gift exchange in experimental labour markets, *Journal of Labor Economics* 16, 324-351.

Fehr, E., and K. M. Schmidt, 1999, A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics* 114, 817-868.

Fischbacher, U., 1999, z-Tree (Zurich Toolbox for Readymade Economic Experiments): Experimenter's Manual, Working Paper No. 21 of the Institute for Empirical Research in Economics, University of Zurich.

Geyskens, I., J.-B. E. M. Steenkamp, and N. Kumar, 1998, Generalizations about trust in marketing channel relationships using meta-analysis, *International Journal of Research in Marketing* 15, 223-248.

Houser, D., and J. Wooders , 2000, Reputation in internet auctions: theory and evidence from eBay, University of Arizona.

Kalyanam, K., and S. McIntyre, 2001, Returns to reputation in online auction markets, mimeo, presented to IBM Consulting Group.

Keser, C., 2000, Strategically planned behavior in public goods experiments, Working Paper, CIRANO, Scientific Series 2000s-35.

Keser, C., and F. van Winden, 2000, Conditional cooperation and voluntary contributions to public goods, *Scandinavian Journal of Economics* 102, 23-39.

Kollock P., 1999, The production of trust in online markets, in: E. J. Lawler, M. Macy, S. Thyne, H. A. Walker (eds.), *Advances in Group Processes* (Vol. 16), Greenwich, CT: JAI Press.

Landon, S., and C. E. Smith, 1997, The use of quality and reputation indicators by consumers: the case of Bordeaux wine, *Journal of Consumer Policy* 20, 289-323.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny, 1997, Trust in large organizations, *American Economic Review* 87, 333-338.

Luhmann, N., 1988, Familiarity, confidence, and trust, in: D. Gambetta (ed.), *Making and breaking cooperative relations*, Oxford: Blackwell.

Malaga, R. A., 2001, Web-based reputation management systems: problems and suggested solutions, *Electronic Commerce Research* 1, 403-417.

Meidinger, C., S. Robin, and B. Ruffieux, 1999, Confiance, réciprocité et cheap talk, *Revue Economique* 1, 5-44.

Ortmann, A., J. Fitzgerald, and C. Boeing, 2000, Trust, reciprocity, and social history: a re-examination, *Experimental Economics* 3, 81-100.

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood, 2002, The value of reputation on eBay: a controlled experiment, mimeo.

Selten, R., 1965, Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetraegheit, *Zeitschrift fuer die Gesamte Staatswissenschaft* 21, 301-324.

Selten, R., M. Mitzkewitz, and G. Uhlich, 1997, Duopoly strategies programmed by experienced players, *Econometrica* 65, 517-555.

Selten, R., and R. Stoecker, 1986, End behavior in sequences of finite prisoner's dilemma supergames, *Journal of Economic Behavior and Organization* 7, 47-70.

Siegel, S., 1987, *Nichtparametrische statistische Methoden*, Eschborn: Fachbuch-handlung fuer Psychologie.

Willinger, M., C. Keser, C. Lohmann, and J.-C. Usunier, *forthcoming*, A comparison of trust and reciprocity between France and Germany: experimental investigation based on the investment game, *Journal of Economic Psychology*.

Yamagishi, T., and M. Yamagishi, 1994, Trust and commitment in the United States and Japan, *Motivation and Emotion* 18, 129-166.

APPENDIX:

Instructions used for the long-run reputation treatment

You are participating in an experiment on decision-making. In this experiment you can earn money; the amount you earn depends both on your own decisions and those of other participants. Please do not speak to any other participant until the experiment is completed.

The experiment consists of **twenty periods**. There are two player types, called A and B. **You will be the same type player (either an A-type or a B-type) in all of the periods** of the experiment. In the beginning of each period, each A-player will be randomly paired with a B-player for interaction during that period. The **A-B pairs will be re-matched in each period in such a way that no one will ever interact with the same other participant in any two consecutive periods**.

In the beginning of each period, each A-player and each B-player will receive an **endowment of 10 Experimental Currency Units (ECUs)**.

The A-player sends Each A-player will then have the opportunity to send part or all of his/her endowment to the B-player with whom he/she is matched, but needs not send anything. On the computer screen the A-player will be asked to enter an integer number between zero (sending nothing) and ten (sending all of the endowment). To confirm the decision he/she will then have to click on <Ok>. If the A-player decides to send a positive amount to B, the experimenter will **triple** that amount so that the B-player will receive three times the amount that the A-player sent. The A-player will only loose the amount that he/she sent, not triple that amount.

The B-player returns When all A-players have confirmed their decisions, each B-player will be informed of the amount that A sent and the tripled amount that he/she will receive. If a B-player receives a positive amount, he/she will then have the opportunity to send something back to the A-player with whom he/she is matched, but needs not send anything. Each B-player may return any integer amount between zero and the amount received (which is three times the amount sent by the A-player). On the computer screen the B-player will be asked to enter an integer number between zero (returning nothing) and the amount received. To confirm the decision he/she will then have to click on <OK>.

Profits The **profit of each A-player**, in each period, is equal to his/her endowment of 10 ECUs minus the number of ECUs he/she sent to the B-player plus the number of ECUs returned by the B-player. The **profit of each B-player**, in each period, is equal to his/her endowment of 10 ECUs plus the amount received (three times the number of ECUs sent by the A-player) minus the number of ECUs he/she returned to the A-player.

The A-player evaluates the B-player

When all B-players have confirmed their decisions, then each A-player will be informed of the number of ECUs that the B-player with whom he/she is matched returned to him/her. Each A-player that sent a positive amount of ECUs to the B-player in that period will then be asked to evaluate the cooperation of the B-player. Based on the amount returned by the B-payer, the A-player may assign the B-player a **positive, neutral or negative rating**. To do so, the A-player must click on the appropriate rating button on the screen and then confirm the rating by clicking on <Continue>. Each A-player that did not send any ECU to the B-player in that period will not be asked to evaluate the cooperation of the B-player. When all eligible A-players have confirmed their ratings, then each B-player will be informed of his/her evaluation and A- and B-players will be informed of their profits in that period.

After reviewing the results of the period, each participant will have to click on <Continue> to go on to the next period. The next period will start after 60 seconds even if some participants have not clicked on <Continue>. In the subsequent period, the procedure will be the same, each participant will remain the same type player, but each participant will be paired and interact with a different player.

Your total profit in this experiment will equal the sum of your profits from all periods.

During the entire experiment you will find a **table summarizing your own activity history** at the bottom of the screen. This table shows for each previous period your initial endowment, the number of ECUs that the A-player sent to the B-player, the tripled amount that the B-player received, the number of ECUs that the B-player returned to the A-player, your profit and your cumulated profit. The B-player will also find there the rating that was assigned to him/her by the A-player. The table does not allow you to identify the other player with whom you interacted in each period. In later periods of the experiment, you may use the scroll function that will appear at the right side of this table, to review the very early periods.

Above the activity table, each participant will find a **rating table**. The rating table **on each A-player's screen** will show the ratings previously assigned to the B-player with whom the A-player is currently paired. The rating table **on each B-player's screen** will show the ratings previously assigned to that B-player. The rating tables will show the number of positive, neutral, and negative evaluations assigned to B-players, and the number of periods in which no rating was assigned (when no ECU was sent to the B-player in a period). It will also reveal the most recent rating that was assigned to the B-player. Recall that an A-player who did not send any ECU to the B-player could not evaluate the B-player's cooperation. The B-player's most recent rating is thus not necessarily from the previous period.

Now, you will be asked to answer some questions to demonstrate that you understand these instructions. As soon as everybody has correctly answered all of the questions, the experiment can start. **Thank you for your participation**.

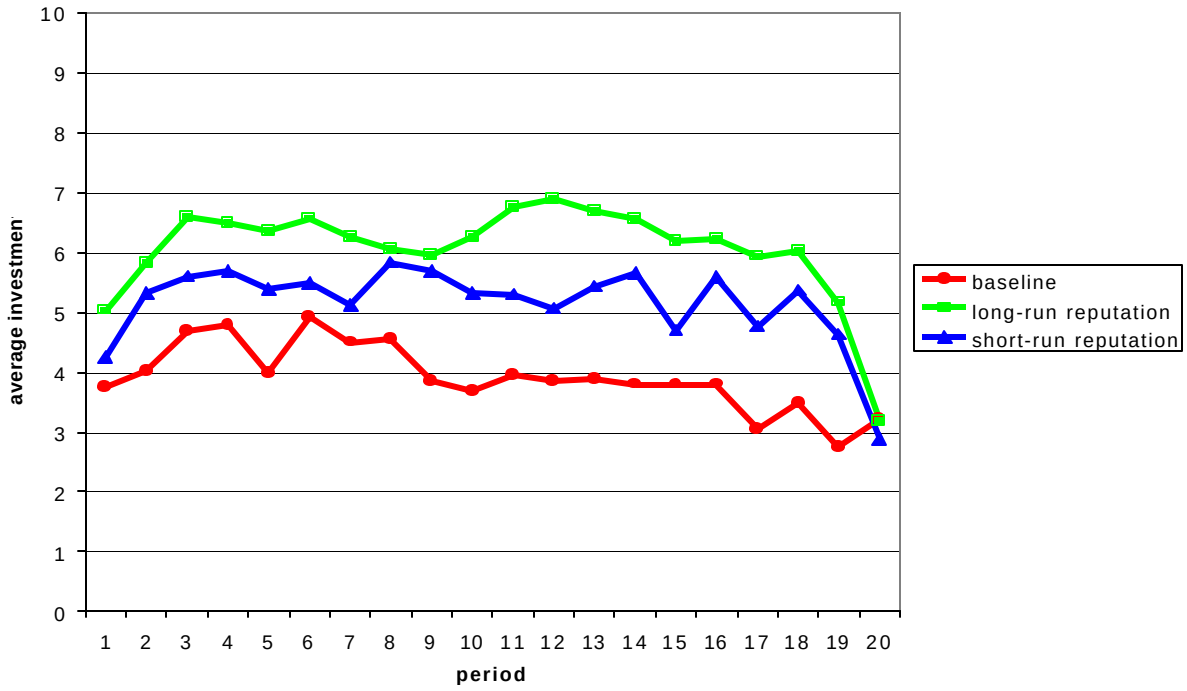

Figure 1: Average investments over time in the three treatments

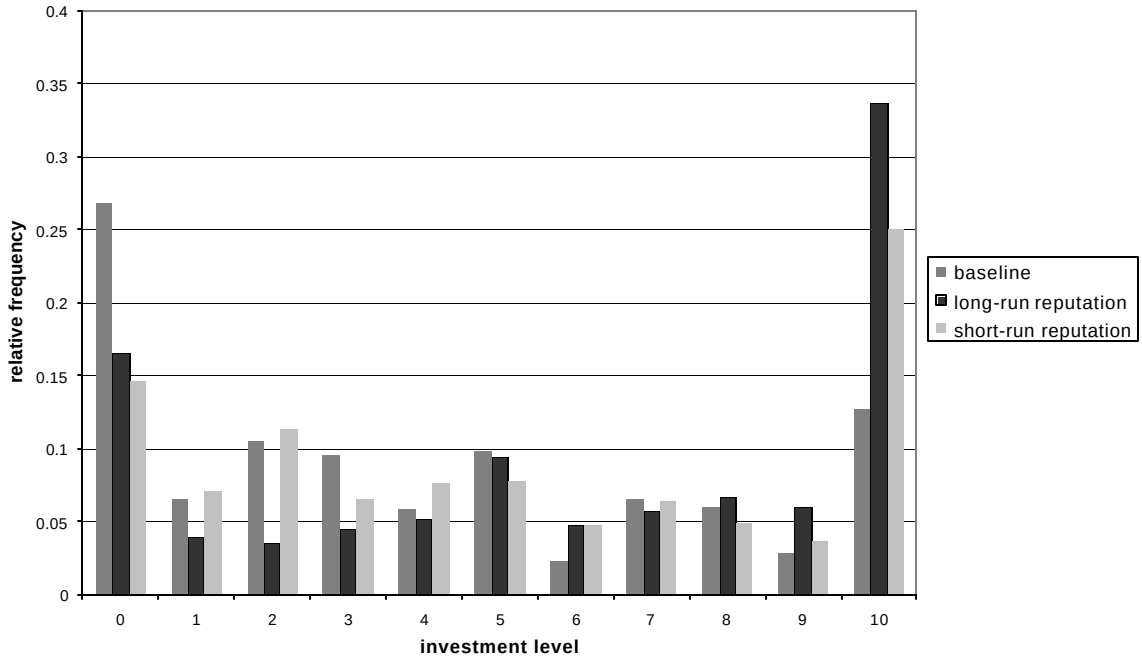

Figure 2a: Distribution of investment decisions over all periods in the three treatments

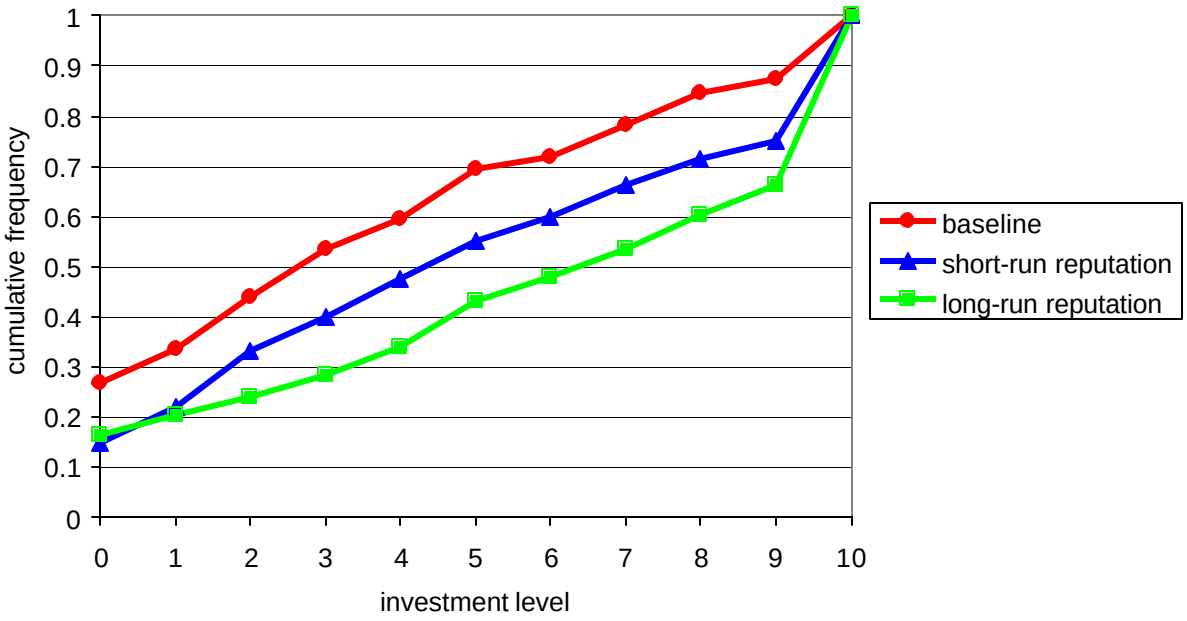

Figure 2b: Cumulative distribution of investment decisions over all periods in the three treatments

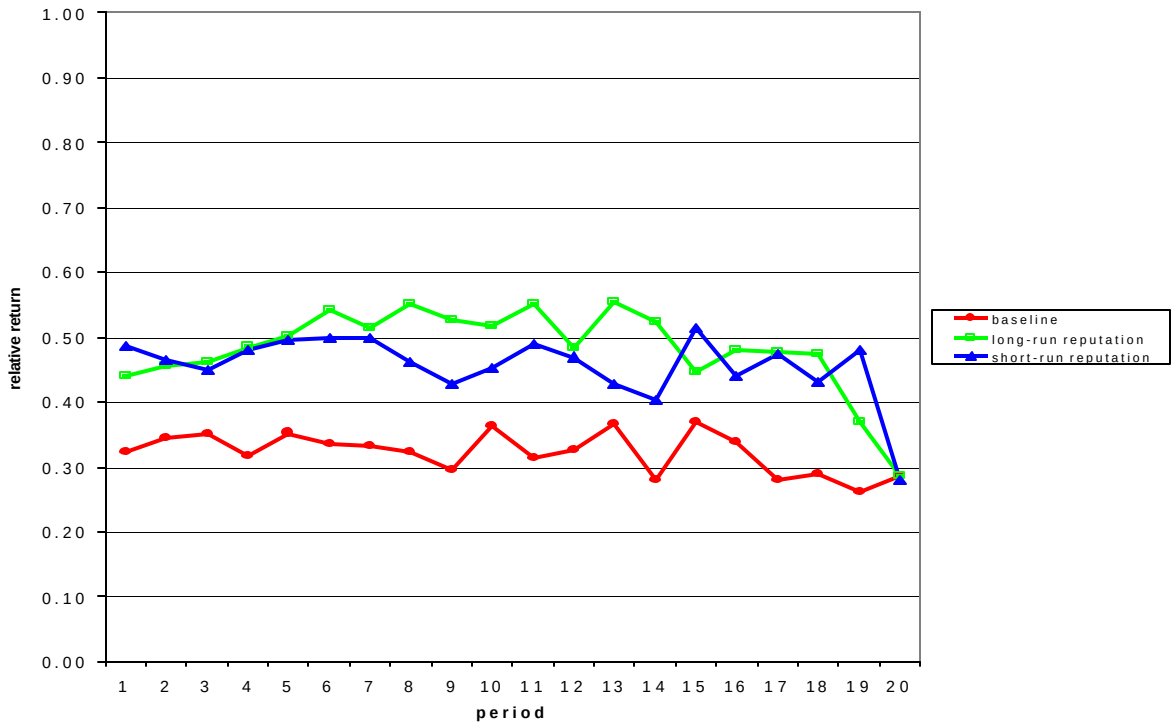

Figure 3: Average relative returns (based on received amount) over time in the three treatments

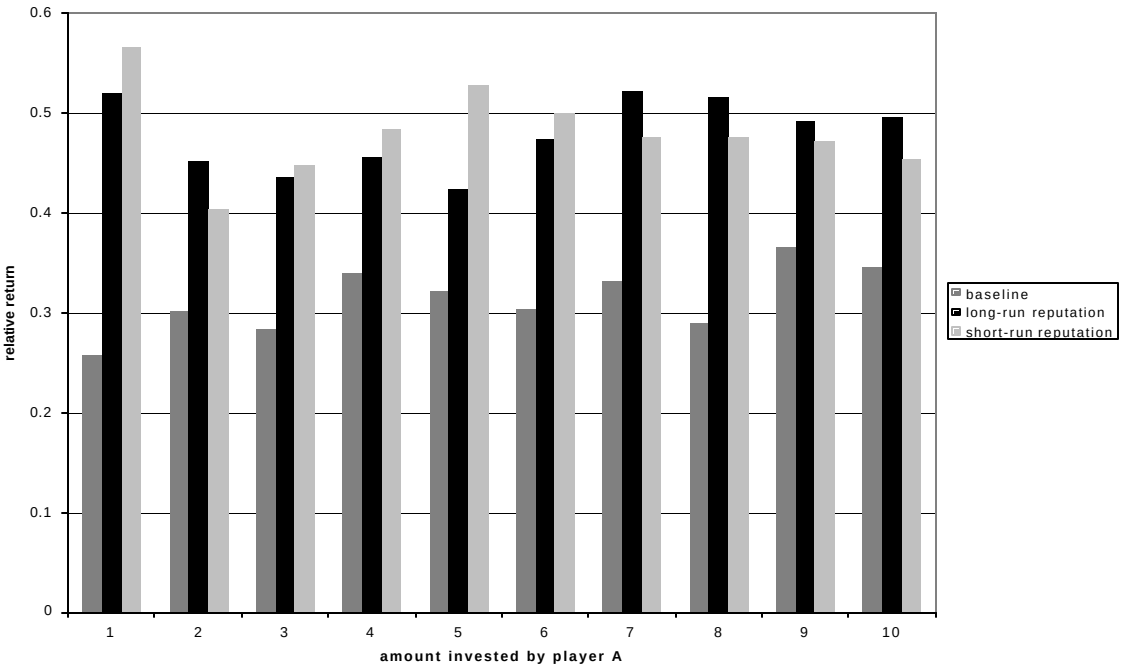

Figure 4: Average relative returns (based on received amount) for investment levels 1 to 10 in the three treatments

Table 1:
Player A's investment (average, standard deviation, median)

Treatment	*Session*	*Average*	*STD*	*Median*
Baseline	1	5.91	3.69	6
	2	4.72	3.46	5
	3	5.26	3.63	5
	4	2.45	3.51	0
	5	1.19	1.73	0
	6	3.74	2.53	3
	7	3.25	3.47	2
	8	4.79	3.28	5
	All	3.91	3.53	3
Short-run Reputation	1	6.11	3.66	7
	2	3.42	3.22	2
	3	3.89	4.03	2
	4	6.17	3.73	7.5
	5	3.45	2.88	2
	6	4.83	3.40	4
	7	2.98	2.70	3
	8	3.88	3.40	3
	9	6.28	2.74	6.5
	10	6.73	3.79	9
	11	7.82	3.58	10
	12	6.27	3.31	7
	All	5.15	3.71	5
Long-run Reputation	1	7.92	2.78	9
	2	6.22	3.89	7
	3	4.79	3.71	5
	4	6.00	3.74	6
	5	6.69	3.90	8
	6	4.83	3.87	4
	7	5.71	4.49	7.5
	8	6.48	3.56	8
	9	8.06	2.28	8
	10	3.92	3.51	4.5
	11	5.87	3.96	5
	12	6.11	3.49	6
	All	6.05	3.80	7

Table 2:
Player B's (relative) return

Treatment	*Session*	*A invested*	*B received*	*B returned*	*% returned*
Baseline	1	5.91	17.73	5.14	28.99
	2	4.72	14.16	5.36	37.85
	3	5.26	15.78	7.05	44.68
	4	2.45	7.35	1.26	17.14
	5	1.19	3.57	0.84	23.53
	6	3.74	11.22	2.54	22.64
	7	3.25	9.75	3.18	32.62
	8	4.79	14.37	5.12	35.63
	All	3.91	11.74	3.81	32.46
Short-run reputation	1	6.11	18.33	8.61	46.97
	2	3.42	10.26	4.50	43.86
	3	3.89	11.67	2.74	23.48
	4	6.17	18.51	9.38	50.68
	5	3.45	10.35	4.11	39.71
	6	4.83	14.49	7.92	54.66
	7	2.98	8.94	3.88	43.40
	8	3.88	11.64	5.19	44.59
	9	6.28	18.64	10.16	53.93
	10	6.73	20.19	10.71	53.05
	11	7.82	23.46	10.36	44.16
	12	6.27	18.81	7.64	40.62
	All	5.15	15.46	7.10	45.93
Long-run reputation	1	7.92	23.76	12.00	50.51
	2	6.22	18.66	9.81	52.57
	3	4.79	14.37	7.07	49.20
	4	6.00	18.00	9.58	53.22
	5	6.69	20.07	8.55	42.60
	6	4.83	14.49	7.62	52.59
	7	5.71	17.13	8.24	48.10
	8	6.48	19.44	10.32	53.09
	9	8.06	24.18	12.05	49.83
	10	3.92	11.76	3.38	28.74
	11	5.87	17.61	7.78	44.18
	12	6.11	18.33	10.12	55.21
	All	6.05	18.15	8.88	49.00

Table 3:
Average profits per period

Treatment	*Session*	*Profit player A*	*Profit player B*
Baseline	1	9.23	22.59
	2	10.64	18.80
	3	11.79	18.73
	4	8.81	16.09
	5	9.65	12.73
	6	8.80	18.68
	7	9.93	16.57
	8	10.33	19.25
	All	9.90	17.93
Short-run reputation	1	12.50	19.72
	2	11.08	15.76
	3	8.85	18.93
	4	13.21	19.13
	5	10.66	16.24
	6	13.09	16.57
	7	10.90	15.06
	8	11.31	16.45
	9	13.88	18.68
	10	13.98	19.48
	11	12.54	23.10
	12	11.37	21.17
	All	11.95	18.36
Long-run reputation	1	14.08	21.76
	2	13.59	18.85
	3	12.28	17.30
	4	13.58	18.42
	5	11.86	21.52
	6	12.79	16.87
	7	12.53	18.89
	8	13.84	19.12
	9	13.99	22.13
	10	9.46	18.38
	11	11.91	19.83
	12	14.01	18.21
	All	12.83	19.27

Table 4:
Average investment by the A-players depending on their corresponding B-player's most recent rating

Treatment	*Most recent rating of B*	#	*Average investment by A*
Short-run reputation	*Negative*	432	3.62
	Neutral	200	5.08
	Positive	503	6.61
	Not yet rated	65	4.28
Long-run reputation	*Negative*	386	4.82
	Neutral	211	5.89
	Positive	537	7.12
	Not yet rated	66	5.11

Table 5:
Average investment by the A-players depending on their corresponding B-player's overall rating

Overall rating of B	#	*Average investment by A*
Negative	293	4.23
*Zero **	167	5.62
Positive	293	6.87

* Including "Not yet rated"

Table 6
Explanatory variables of econometric models

Variable	Definition
LIn	A-player's investment in the preceding period
DnoLIn	1 if A-player made zero investment in preceding period, 0 otherwise
LRR	Relative return experienced by A-player in preceding period
*LIn*LRR*	Interaction effect between *LIn* and *LRR*
Last	1 in last period of the game, 0 otherwise
Period	Period number
RecentPo	1 if B-player's most recent evaluation was positive, 0 otherwise
RecentNe	1 if B-player's most recent evaluation was negative, 0 otherwise
SumPo	1 if sum of B-player's previous evaluations is positive, 0 otherwise
SumNe	1 if sum of B-player's previous evaluations is negative, 0 otherwise

Table 7
OLS estimates for the A-player's investment in periods 2 to 20

Explanatory variable	Baseline		Short-run evaluation		Long-run evaluation	
Const	1.7523*	1.8435*	1.8481*	1.5671*	2.7274*	2.3660*
DnoLIn	0.3083	--	1.6065*	1.6680*	1.0953**	1.3673*
Lin	0.5230*	0.4941*	0.6664*	0.7067*	0.4055*	0.4420*
LRR	0.9583	--	-0.2199	--	-0.5969	--
*LIn*LRR*	0.4009*	0.4154*	0.0919	--	0.3392*	0.2631*
Last	0.5838	--	-1.6947*	-1.8524*	-1.7617*	-1.7595*
Period	-0.0497**	-0.0371**	-0.0184	--	-0.0559*	-0.0566*
RecentPo	--	--	0.8197*	0.8273*	0.8836*	0.9739*
RecentNe	--	--	-1.5154*	-1.5186*	-0.1646	--
SumPo	--	--	--	--	0.8646*	0.8697*
SumNe	--	--	--	--	-1.1860*	-1.2381*
Adj. R^2	0.42610	0.42661	0.52844	0.52829	0.37475	0.37539

* significant at 1 percent level, ** significant at 5 percent level

Table 8:
Amounts in ECU to be returned by player B for
equal split of the gross surplus or equal split of the net surplus

Investment [ECU]	*Amount received [ECU]*	*Equal split of gross surplus*	*Equal split of net surplus*
1	3	1.5	2
2	6	3	4
3	9	4.5	6
4	12	6	8
5	15	7.5	10
6	18	9	12
7	21	10.5	14
8	24	12	16
9	27	13.5	18
10	30	15	20

Table 9:
Average relative return (in percent) for which the A-players gave a negative, neutral or positive rating

Treatment	*Session*	*Negative rating*	*Neutral rating*	*Positive rating*
Short-run reputation	1	27.97	55.12	65.81
	2	7.71	41.43	59.38
	3	7.19	33.85	41.09
	4	28.21	45.02	66.13
	5	17.04	47.94	53.89
	6	25.92	64.74	71.68
	7	13.66	43.89	69.35
	8	16.71	42.50	59.79
	9	21.68	61.78	68.74
	10	25.05	59.51	67.69
	11	33.91	45.47	52.77
	12	21.81	37.66	55.02
	All	20.91	49.56	61.96
Long-run reputation	1	25.51	46.32	59.18
	2	38.78	47.16	58.53
	3	12.32	51.03	58.10
	4	15.24	62.78	63.63
	5	26.19	43.58	50.36
	6	33.87	55.95	60.76
	7	16.52	53.16	69.10
	8	28.04	54.14	6775
	9	31.63	53.98	59.33
	10	9.35	44.02	75.76
	11	16.69	41.67	57.51
	12	40.29	63.15	73.01
	All	25.04	50.79	61.76

Table 10:
Sum of overall ratings in the beginning of the final period, and the average per period profits [in ECUs] of the players in that session with an overall positive, neutral or negative rating

		Overall positive rating		*Overall neutral rating*		*Overall negative rating*	
Session	*Sum of ratings*	*# B-players*	*Profit*	*# B-players*	*Profit*	*# B-players*	*Profit*
1	35	4	21.96	1	20.96	0	-
2*	32	4	18.35	0	-	1	20.85
3*	31	4	16.73	0	-	1	19.60
4	30	3	19.33	0	-	2	17.05
5	21	4	22.71	0	-	1	16.75
6*	18	4	16.43	0	-	1	18.65
7*	12	4	17.94	0	-	1	22.70
8*	6	3	18.67	1	19.85	1	19.75
9*	16	2	21.00	2	22.83	1	23.00
10	-12	2	21.15	0	-	3	16.53
11	25	4	20.54	0	-	1	17.00
12*	-13	1	15.60	1	18.85	3	18.87
All		39	19.29	5	21.06	16	18.66

* The negatively rated B-players gain more than the positively rated B-players