PRELIMINARY AND INCOMPLETE

Is Reputation Good or Bad? An Experiment

Brit Grosskopf
Texas A&M University

Rajiv Sarin
Texas A&M University

August 19, 2005

Abstract

We design and conduct an experiment to determine whether reputation is bad. Our design nests that of earlier papers which test a model of good reputation. Our results suggest that the prediction that reputation is good is more robust than the bad reputation prediction. While we observe some evidence in accordance with the bad reputation prediction, we did not observe the complete breakdown of the market.

1 Introduction

Economic agents are concerned about their reputations. These reputations are often made up of the past choices of an agent which are observed by other market participants. For example, if an entrepreneur has re-paid the loans she has taken in the past then future lenders, observing this, will be more willing to offer loans to the entrepreneur. An entrepreneur, recognizing the positive effects of repaying earlier loans may become willing to re-pay them. Reputation here is good and it benefits the entrepreneur even though she would have preferred to renege on the loans had her choices not been observed. The lenders also gain by the entrepreneur caring for her reputation.

Some recent work, however, suggests that reputation may not always be beneficial. The harmful effects of reputation are nicely illustrated with the following example taken from Ely and Valimaki (2003). Consider a mechanic who meets a sequence of automobile owners who observe the previous actions of the mechanic. A friendly mechanic always prefers[1] the correct or appropriate repair whether it be that a tune up is required or whether the engine needs replacement. The unfriendly mechanic always prefers an engine replacement. Notice that when the friendly mechanic appropriately changes an engine, future automobile owners are apt to suspect that she may be an unfriendly mechanic. It may, hence, serve the (future) interests of the friendly mechanic to perform a tune up even when an engine replacement is required. If the motorists understand that both the friendly and the unfriendly mechanics are prone to perform an inappropriate repair then the motorists may stop going to the mechanic altogether. Hence, a mechanic who is, in fact, friendly suffers (as do the motorists).

More abstractly, the theory that reputation is bad considers short run players who are uncertain about the long run player's preferences.[2] If the short run players think there is a possibility that the long run player may have preferences that are not aligned with theirs then the short run players may refuse to interact with a long run player whose preferences are actually aligned with theirs. This result emerges when there is a possibility that the actions chosen by an aligned or friendly long run player can be interpreted by future short run players as coming from an unfriendly one. In such a case the theory shows that the friendly long run player may choose inappropriate actions to signal her preferences to future short run players. These harm the current short run player. It is precisely because even the friendly long run player may choose the inappropriate action that the short run players refuse to interact with her.

Whereas the ultimate prediction of the bad reputation theory is that markets collapse it has two important building blocks. Firstly, motorists are assumed to be more prone to believe the mechanic is bad after they observe that she has recommended an engine replacement. Therefore, observing that the mechanic has performed an engine replacement should reduce the likelihood that motorists bring their automobiles to the mechanic. Secondly, friendly mechanics are assumed to recognize the perverse effect on future

[1] That is, gets a higher payoff from.

[2] This theory has been developed in a parallel manner to the theory of good reputation by Ely and Valimaki (2003). The results of Morris (2001) are similar.

motorists of seeing her do an engine job and may therefore perform a tune up even when an engine job is required. The prediction that the market collapses additionally requires that motorists recognize that both the friendly and unfriendly mechanics may choose the inappropriate action and hence stop bringing their automobiles to the mechanics entirely.

Many (potential) markets are susceptible to the implications of bad reputation result. Any market in which the seller provides diagnosis before providing the good or service are prone to the negative implications of the result. This includes markets in which sellers are mechanics, lawyers, management consultants and doctors. Establishing that these effects are present in real world markets is, however, not completely straightforward. How does one know that a market doesn't exist, or has failed, because of the bad reputation effect or for some other reason like inadequate or decreased demand? It, hence, seems desirable to see if this effect emerges in the controlled environment of an economic experiment.

We know of no test of the bad reputation result. In this paper we design and report on an experiment to test whether reputation is bad. Our design of the bad reputation experiment nests the design of earlier experiments that tested the literature which suggested that reputation was good.[3][4] These papers found that the unfriendly long run players do choose in a manner so that the short run players want to interact with them. This allows the long run player to appropriate surplus that would not have been available had her preferences been known or had she been unable to build a reputation. The focus of these earlier studies was on the sequential equilibrium predictions of the good reputation model. In contrast, our focus is more on the harmful (or beneficial) effects of reputation.[5]

We find evidence that short run players do condition their behavior on the reputation (i.e., past actions) of the long run player. This hurts the friendly long run player when her (appropriate) actions can be misconstrued as coming from an unfriendly long run player. This effect is clearest when

[3] These papers include Camerer and Weigelt (1988), Neral and Ochs (1992) and Brandts and Figueras (2003).

[4] The theory was pioneered by Kreps and Wilson (1982) and Milgrom and Roberts (1982) and further developed by Fudenberg and Levine (1989, 1992).

[5] A recent paper by Ely, Fudenberg and Levine (2005) theoretically investigates the conditions under which reputation is bad.

the long run player chooses such an action the first time. However, this effect is not monotonic. Short run players do not respond by withdrawing from the market after additional observations of such action. We also find evidence that friendly long run players choose, or learn to choose, the inappropriate action to signal to future short run players that they are friendly, but in so doing hurt the current short run player (and get a lower payoff for themselves in the current period).

Whereas the theory of bad reputation predicts that these reputational concerns should lead to a market collapse, we do not observe this. There seem to be several reasons for this result. Firstly, the short run players initially enter the market too often. The excessive entry leads the long run player to signal (by choosing the inappropropriate action) only infrequently. In fact, even the unfriendly long run players often choose the action that benefits the short run player at a cost to themselves. This does not seem to occur because of the long run player wishing to signal her type, but rather because of some other-regarding behavior. In turn, this keeps large numbers of short run players in the market.

Even though markets do not collapse, we do observe that the market size falls as the length of the interaction (and reputation) increases. This suggests that if long enough reputations can develop then the market might collapse. As the bad reputation game was repeated we did not observe "learning" to be taking place. That is, the start of each repetition of the bad reputation game saw market size rise. Our experiment suggests that longer interactions, rather than a larger number of repetitions of the interactions of given size, may cause markets to collapse.

Lastly, our results suggest that reputation is not bad. Long run players always earn more than what the theory of bad reputation would predict they earn. Other regarding preferences and some level of bounded rationality among the market participants might be the reason for this finding. This is in accordance with observations from other experiments where people often do better than theory would predict.[6] When our bad reputation design is suitably specialized to the good reputation design, we do observe that reputation is good. From our experiment it, thus, appears that the prediction

[6]For example, cooperation in finitely repeated Prisoners' Dilemma games or moving across in Centipedes games which is not predicted by the theory.

that reputation is bad is not as robust as the prediction that reputation is good.

Our paper is structured as follows. In the next section we report our experimental design. Section 3 presents the hypotheses we consider. Section 4 reports our experimental results. Section 5 concludes. An Appendix includes summary statistics of our data, estimation results and a sample set of instructions.

2 Experimental Design

To help explain our design we adapt the story used by Camerer and Weigelt (1988). This will most clearly reveal the manner in which their design is nested in ours.

The long run player, called the entrepreneur (E), meets a sequence of six short run players called banks (B). Banks can observe the projects ($e1$ or $e2$) chosen by entrepreneur in the past but do not know if they were the appropriate projects.[7] Each project is the appropriate one with 50% probability. Which project is appropriate depends on the state of the world (which may be either N or P) and only the entrepreneur observes the realized state of the world.

The entrepreneur's preferences are unknown to the banks.[8] A good (type X) entrepreneur always prefers the appropriate project. The bad (type Y) entrepreneur always prefers project $e2$ independently of the state of the world. Banks decide whether to give the entrepreneur a loan or not (i.e. choose $b1$ or $b2$). If the bank refuses to give a loan (i.e. choose $b1$) then the interaction with the entrepreneur ends. If a loan is given, the entrepreneur has to decide on a project.

The introduction of states of the world, which determine the appropriateness of a project, differentiate our design from that of Camerer and Weigelt (1988), henceforth CW, Neral and Ochs (1992) and Brandts and Figueras (2003). If the only possible state is P (which occurs with probability 1)

[7] The appropriate project is the one that maximizes the banks stage game payoff conditional on the state of the world.

[8] We assume a player's payoff function completely describes her preferences.

then our design collapses to that of CW. The uncertainty regarding the state and, consequently, whether the entrepreneur chooses the appropriate project is instrumental in the emergence of the bad reputation result.

The extensive form representation of the stage game interaction between the bank and entrepreneur is reproduced below as Figure 1. Observe that E participants of type X have preferences aligned with the B players. They prefer $e2$ in state N and $e1$ in state P. E participants of type Y always receive a higher payoff from choosing $e2$, independently of the state of the world. If the probability of state of the world being P is 1 then the game tree reduces to the good reputation game studied by CW.[9]

We designed four treatments, in which B players repeatedly interacted with E players.[10] Three of these treatments were aimed to test the bad reputations predictions and one set out to test the good reputation predictions. For each treatment we have four independent cohorts of 6 B players and 6 E players who participate in 10 repetitions (sequences) of a 6 period game. Players were randomly assigned to their roles (B or E) and these stayed the same throughout the experiment. Each of the six B players met each of the six E players once in a randomly predetermined order that changed from sequence to sequence. States of the world were determined independently for each E player and each round. The states of the world were also randomly generated and kept fixed for each of the three bad reputation treatments. Whereas the E players knew their type (i.e. preferences) and the state of the world (in the bad reputation treatments), the B players only knew the probability with which they would meet an E player with particular preferences and the probability of each state of the world. Before choosing an action in any stage game the B players were informed about the previous choices of the E player they were meeting in the current round (if there were any). We will begin by describing the details of our three bad reputation treatments followed by the good reputation treatment.

The Benchmark Bad Reputation Treatment: There were 6 B

[9] Note that we changed the payoff numbers in comparison to these authors. We did this to avoid the possibility of negative payoffs and have "simpler" numbers for the players to work with. Our numbers also allowed us to keep the exchange rate from numbers to dollars the same for the subjects.

[10] We used a between subject design, i.e. each subject participated only once in our experiment.

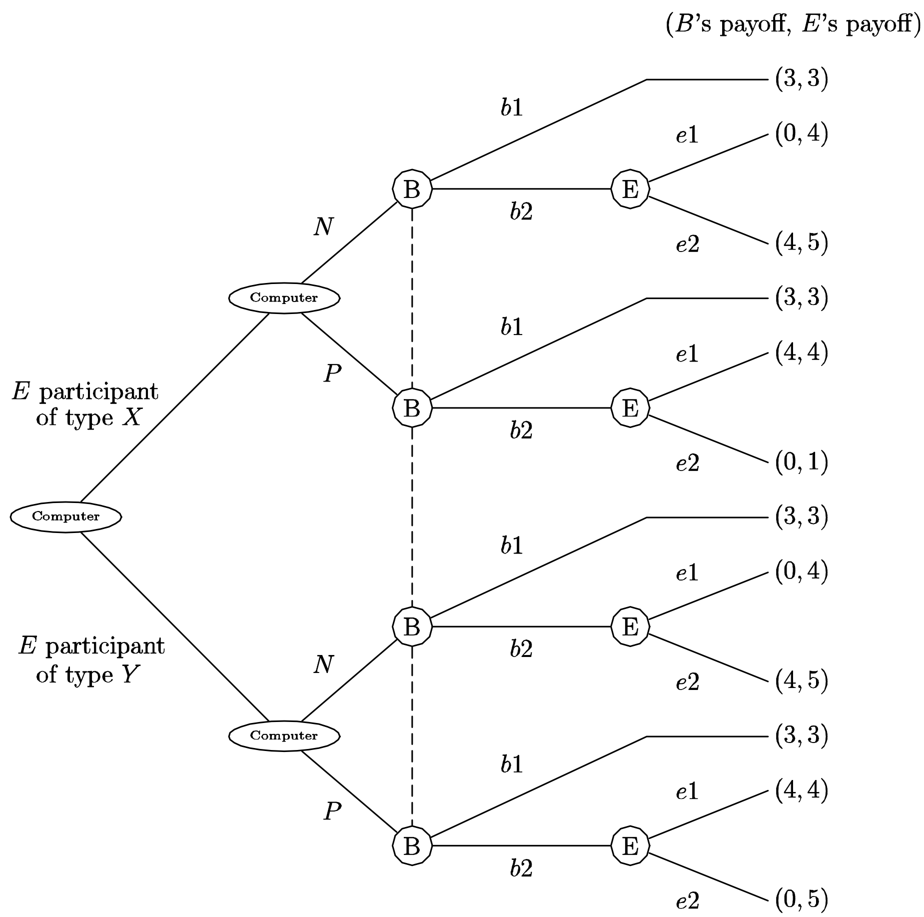

Figure 1: The bad reputation stage game

players and 3 E players of each type. Hence, a B player met an E player of each type in any stage game with 50% probability. The E players of type Y were computerized players who were programmed to always choose $e2$ when given a loan ($b2$).[11] These E players were therefore committed (nonstrategic) types. E players of type X were human (strategic) players. The probability of both states of the world was equally likely, i.e. $\Pr\{N\} = \Pr\{P\} = 1/2$.

The Bad Reputation – All Human Treatment: This treatment is an all human version of the benchmark bad reputation treatment, where again the states of the world were equally likely, i.e. $\Pr\{N\} = \Pr\{P\} = 1/2$. In contrast to the benchmark treatment, both E types were played by human participants. That is, the E players of type Y were also strategic. We considered this treatment because having all players be real (strategic) allows a better test of the "real world" relevance of the theory.[12]

The Bad Reputation – 66.7% Treatment: This treatment is a version of the benchmark treatment in which there are 4 unfriendly E players and 2 friendly E players. Hence, a B player meets an unfriendly E player in four of the six stage games. The unfriendly E players are computerized and therefore are committed (nonstrategic) types. Again, both states of the world were equally likely, i.e. $\Pr\{N\} = \Pr\{P\} = 1/2$. The increase in the proportion of unfriendly players was motivated by a desire to see if the forces leading to the bad reputation result are more stark in this setting. We suspected this to be the case, as banks would be more unwilling to give loans when the proportion of (computerized) unfriendly players is higher and this would lead the (human) friendly players to feel a greater urgency to signal by choosing the inappropriate action. Hence, we expected the market collapse prediction to be more easily observable in this treatment.

The Good Reputation Treatment: Lastly, we ran a good reputation treatment in which the probability of the state of the world being P was 1. Since the strategic long run player in the good reputation model is the unfriendly one, we computerized the committed friendly long run player. The E players of type X were programmed to always choose $e1$ upon a choice of

[11] As computers didn't receive payoffs the corresponding entries were omitted in the presentation of the game.

[12] The bad reputation result of market collapse is also predicted in this setting even though the arguments involved are somewhat more intricate. We refer the readers to Ely and Valimaki (2003) for the detailed argument.

$b2$ by the B players. Observe that the unfriendly players were computerized in the bad reputation treatment and the friendly players were computerized in the good reputation treatment. There were the same number of long run players of each type in both treatments, i.e. 3 friendly and 3 unfriendly. We ran the good reputation treatment to verify that we could recover the results of the earlier experiments that suggest reputation is beneficial for the long run player. Table I summarizes our experimental treatments.

TABLE I:
SUMMARY OF THE EXPERIMENTAL DESIGN

	Bad Reputation Benchmark	Bad Reputation All Human	Bad Reputation 66.7%	Good Reputation
E player – type X (friendly)	Real	Real	Real	Computerized
E player – type Y (unfriendly)	Computerized	Real	Computerized	Real
$\Pr\{X\}$	1/2	1/2	1/3	1/2
$\Pr\{N\}$	1/2	1/2	1/2	0
# cohorts	4	4	4	4

All our experimental sessions were conducted at the Economic Research Laboratory at Texas A&M University, from June 2004 until July 2005. Our participants were undergraduate students of all but economics major. All sessions were computerized, using a computer interface programmed with zTree (Fischbacher, 1999). Instructions were read aloud and questions were answered in private.[13] After reading the instructions and having questions answered, all participants had to answer a set of questions that were meant to check on whether the instructions had been understood. All answers were checked and corrected by the experimenters and remaining questions were answered. Throughout the sessions participants were not allowed to communicate with one another and dividers separated the individual computer terminals. An experimental session typically lasted about two hours. Participants were paid their accumulated earnings at the end of the session. Payoffs in points were converted to dollars at the rate of 1 point = $0.12, plus a $5 show-up fee.

[13] A sample set of instructions can be found in Appendix C.

Our instructions did not mention any frame or context. We provide examples ("stories") in the text of the paper to help illustrate and motivate the game played by the participants and to make our discussion more easily comparable with the earlier good reputation literature.

3 Hypotheses

3.1 The Benchmark Bad Reputation Treatment

Consider, first, the short run B players. In the benchmark bad reputation treatment banks should be indifferent between giving the entrepreneur a loan or not if they believe that the type X (human) E players always choose the appropriate project.[14] To see this observe that B players get 3 from choosing $b1$ and they get an expected payoff of 2 from giving a loan and meeting a computerized (unfriendly) E player and an expected payoff of 4 from giving a loan to a human (friendly) E player who always chooses the appropriate project (see Figure 1). As meeting both types of E players is equally likely the B players get an expected payoff of 3 from not giving a loan ($b1$) and 3 from choosing $b2$.[15]

Given our assumption of risk neutrality it seems reasonable to expect that the B players would choose each of their choices 50% of the time. Arguments could be given for loan giving to be higher or lower than 50%. However, it is useful to have a benchmark initial loan giving to see how it changes over time and 50% seems to be the most reasonable expectation in this regard. Hence, our first hypothesis is:

H1: The proportion of loan giving in the initial period is 50%.

Our next two hypotheses also concern the behavior of the B players. If the B players meet an E player who has only chosen $e2$ in her previous interactions then the B players should believe that it is more likely that the E

[14] We assume, as do the other studies in this area, that players are risk neutral and do not discount future earnings.

[15] Notice also that if the B players suspect that friendly E players sometimes choose the inappropriate action (to signal to future B players) then the B players would be better off choosing $b1$. We do not expect that B players would initially harbor such concerns about the friendly E players.

player is of type Y (computerized). This should lead them, strictly speaking, to not give a loan or in any case to give a loan with lower probability.[16] Hence, we should observe that the probability with which B players give a loan should decline as they meet E players who have larger numbers of $e2$ choices in the past and no previous $e1$ choice. Accordingly, our second hypothesis is:

H2: $\Pr\{b2\}$ *is a decreasing function of the number of* $e2$ *the* E *player has chosen in the past in the case she has never previously chosen an* $e1$*.*

If the B player meets an E player who has previously chosen an $e1$ then the B player should become convinced that the E player has preferences aligned to theirs (i.e. is a human E player of type X) and hence offer a loan with increased probability. Accordingly, our third hypothesis is:

H3: $\Pr\{b2|e1\} > \Pr\{b2|\text{no } e1\}$.

Consider, next, the long run E players. The only E players we are concerned about in the benchmark treatment are the friendly (human) E players. The result that markets breaks down requires them to understand that B players are less likely to offer them a loan if they have only chosen $e2$ in the past. Given our choice of payoffs, E players should initially choose $e1$. If the state of the world is P then the choice of $e1$ is the appropriate choice and, actually, costless. However, E players understanding the positive signalling value of choosing $e1$ in the current period should choose $e1$ even when the state of the world is N and this is the first time they get to make a choice (note that signalling once is enough since all past choices of the E players are part of their reputation). That is, the E player should sometimes choose the inappropriate project just to signal her type to future B players. Hence, our fourth hypothesis is:

H4: $\Pr\{e1|\textit{when first loan received in state } N\} > 0$*.*

The next hypothesis in the benchmark bad reputation treatment is that the market breaks down. That is, we should observe an increasing amount

[16] We can update the prior of an E player being good after she chooses $e2$ in the first round and so forth. If an E player chooses $e2$ throughout, the priors in each round are $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{3^2}$, $\frac{1}{3^3}$, $\frac{1}{3^4}$, $\frac{1}{3^5}$. Therefore, theory would predict a B player to not give a loan after a single observation of $e2$.

of $b1$ over time. Notice that there are two senses in which this may be true. It could be that loan giving delines over time within a sequence or it could be that loan giving declines across the sequences. We discuss both possibilities when we present the results. At this stage we state, somewhat ambiguously, our fifth hypothesis as:

H5: $\Pr(b1)$ *increases over time.*

To determine whether reputation is bad we contrast the earnings of the players in the experimental treatments with reputation that we conducted to what they would have theoretically earned if there was no possibility of building a reputation.[17]

If the long run players' previous choices were not known by the short run players when they made their choices then, in the bad reputation game, the long run player would choose the action from which she obtains her maximum payoff of the stage game, when given the choice. B players would earn an expected payoff 3 when they give loans and 3 when they do not. As B players are indifferent it seems reasonable to expect them to give a loan 50% of the time as we argued earlier. If this was the case the (human) long run players would earn $(.5)(3) + (.5)(.5(5) + .5(4)) = 3.75$. We say that reputation is *bad* if the long run players earn less than 3.75 on average in the stage game (per round). Our seventh hypothesis therefore is:

H6: Reputation is bad if the human long run players earn less than 3.75 points per round on average.

3.2 Additional Hypotheses in the other Treatments

The All Human Treatment: As our objective was to see the real world relevance of the theory, we will focus on how behavior in the all human treatment differs from our benchmark treatment. All the hypotheses we consider for the benchmark case appear pertinent for this treatment also.

The 66.7% Treatment: Compared to our benchmark treatment, loan giving in the first period doesn't seem as attractive in this treatment.

[17] In future work, we plan to run an experiment to test specifically for the effect of reputation in the good and bad reputation games (i.e. considering treatments in which the past choices of the long run player are not observed by the short run payers before they choose).

Specifically, since a choice of $b1$ would yield a payoff of 3, whereas a choice of $b2$ would give an expected payoff of $1/3 \cdot 4 + 2/3 \cdot 2 = 8/3 < 3$, the B players should not give a loan at all. In comparison with our benchmark treatment we, therefore, expect lower loan giving in the first period in this treatment. Accordingly:

H1': The proportion of loan giving in the initial period is < 50%.

The Good Reputation Treatment: The final hypothesis concerns the good reputation treatment and how the results there compare with the bad reputation treatment. The game played by subjects in the good reputation treatment is strategically very different from what they confronted in the bad reputation treatment. Correspondingly, the predictions of the theory are equally different. The predicted path of loan giving and the predicted behavior of the human long run players are, hence, not straightforward to compare between the two treatments. To compare the two treatments, we focus on the comparison between player earnings. Specifically, we contrast the earnings of the players in the experimental treatments with reputation to what they would have theoretically earned if there was no possibility of building a reputation.

If the long run player's previous choices were not known by the short run players when they make their choices then the long run player would choose the maximum payoff of the stage game when given the choice. In this good reputation game the B players would therefore earn an expected payoff of 2 when they give loans and obtain 3 when they do not. Hence, without reputations we expect both the long and the short run player to earn 3. We say that reputation is *good* if the (human) long run players earn a payoff greater than 3 on average in the stage game (per period).[18] Our seventh hypothesis is therefore:

H7: Reputation is good if the human long run players earn more than 3 points per round on average.

[18] This should also be the case for the short run players. However, the reputation literature so far has mainly focused on the effects of reputation on the long run players.

4 Experimental Results

In this section we present summary statistics of the data and nonparametric statistical tests of our hypotheses. We also present results of random effect probit estimations that take into account the heterogeneity observed among the players. The details of the probit analyses are presented in Appendix B. As before, we will begin with our Benchmark Treatment.

4.1 The Benchmark Treatment

Loans were given in the first round 60% of the time if the data from all cohorts and all sequences are aggregated. The Binomial test of the hypothesis that loan giving was 50% (*H1*) in the first period was rejected on this aggregate data (see Table II for details). Loan giving behavior, however, differed quite substantially among the cohorts. The Binomial test of the hypothesis that loan giving was 50% could not be rejected for 3 cohorts.[19]

Table II:
First Round Loan Giving in the Benchmark Treatment

Cohorts		z-values
c1	51.67% (31/60)	-0.129
c2	75% (45/60)	-3.744**
c3	53.33% (32/60)	-0.387
c4	60% (36/60)	-1.420*
all	60% (144/240)	-3.034**

Note: z–values of the Binomial Test are given in the third column,
**(*) indicates significance on the 5% (10%) level, two–tailed.

Our second hypothesis asks us to see how loan giving responds to the observation of $e2$. After one observed $e2$ (and no previous $e1$) loan giving declines to 50% (see Table III for details). This decline is significant (Test

[19] If we only look at initial loan giving in sequences 1–3, no cohorts' initial loan giving was significantly different from 50% at the 5% level using a Binomial test.

of equality of proportions, $z = 2.48$, $p < 0.05$).[20] Furthermore, this decline in observed in every cohort. After two observations of $e2$ and no previous $e1$ loan giving further declines to 45%. This, however, is not a significant decline (Test of equality of proportions, $z = 1.05$) and the behavior differs between the cohorts. In two cohorts, loan giving actually increases (to an even higher level than in the first period). After three or more observations of $e2$ and no previous $e1$ aggregate loan giving rises to 58% (significantly different to 45%, Test of equality of proportions, $z = -2.49$, $p < 0.05$).[21] Two cohorts continue their decline and the two other cohorts decline but still remain high due to their increased loan giving observed after two $e2$ and no previous $e1$.[22]

Additional tests of the second hypothesis are conducted through the aforementioned probit analysis. After one $e2$ is observed and no previous $e1$ loan giving significantly declines by 9%. With a second $e2$ observed (and no previos $e1$) there is a decline of 8% in loan giving (which implies that there is no further decline compared to one $e2$). In fact, a likelihood ratio test confirms that the coefficients in front of one observed $e2$ and two observed $e2$ (with no previous $e1$) are not significantly different from oneanother. There is no significant additional effect from observing more than two $e2$ and no previos $e1$. This suggests that loan giving responds in the manner suggested by the theory (and that we cannot reject *H2*). Note, however, that the responsiveness of loan giving to observing additional $e2$ (and no previous $e1$) declines.[23]

[20] The specific test statistic is $Z = (p_1 - p_2)/S_{p_c}$, where p_i is the proportion of loan giving in subsample i. S_{p_c} is an estimate of the standard error of $p_1 - p_2$, $S_{p_c} = \{p_c(1 - p_c)[(1/N_1) + (1/N_2)]\}^{0.5}$. p_c is an estimate of population proportion under the null hypothesis of equal proportions, $p_c = (p_1 N_1 + p_2 N_2)/(N_1 + N_2)$, where N_i is the total number in subsample i.

[21] We aggregated over three or more $e2$ (with no previous $e1$) because we only had a limited number of observations for these cases.

[22] We actually ran cohorts 1 and 2 at the same time and 3 and 4 at the same time on another day. Therefore, it cannot be that instructions or any experimenter effect was driving the differences since 1 and 3 and 2 and 4 seem rather similar.

[23] In fact, the probit analysis also picks up an effect which suggests that if the B player has chosen to not give a loan in a certain period then he is more likely to give a loan in the succeeding period. We shall comment on this seeming unwillingness of the B players to withdraw from the market further in the sequel.

Table III:
Choices of $b2$ After $e2$ in the Benchmark Treatment

Cohorts	$1 \times e2$	$2 \times e2$	$>2 \times e2$
c1	40.95% (43/105)	25% (18/72)	12.50% (1/8)
c2	60% (48/80)	76.79% (43/56)	69.01% (49/71)
c3	50% (44/88)	23.53% (16/68)	21.05% (4/19)
c4	50.54% (47/93)	67.92% (36/53)	63.64% (35/55)
all	49.73% (182/366)	45.38% (113/249)	58.17% (89/153)

Our third result reveals that loan giving increases to 83% after the observation of at least one $e1$ (see Table IV for details). This is a significant increase compared to the proportion of initial loan giving (Test of equality of proportions, $z = -5.92$, $p < 0.05$). Again there is considerable heterogeneity among the cohorts, with one cohort giving loans 100% of the time after observing $e1$ while another one does so only 57% of the time. The probit analysis reveals that after observing at least one $e1$ loan giving rises by 18%. Hence, *H3* cannot be rejected.[24]

Table IV:
Choices of $b2$ After observing at least one $e1$
in the Benchmark Treatment

Cohorts	
c1	82.61% (57/69)
c2	86.49% (64/74)
c3	100% (81/81)
c4	57.38% (35/61)
all	83.16% (237/285)

To summarize the behavior of the B players, we can reject H1 on the aggregate data. There is too much loan giving. While all cohorts give more than 50% loans in the first period, only one gives significantly more. We

[24] Another interesting observation picked up by the probit analyis is that if a B player had chosen $b2$ in the previous period and obtained a positive payoff then the probability to give a loan increases by 12%.

cannot reject *H2* or *H3*. The B players respond in the direction suggested by the theory to both the observation of $e2$ (with no previous $e1$) and to the observation of $e1$.

Our next result concerns the behavior of the entrepreneurs. Recall that for markets to collapse the second critical ingredient is for a friendly player to choose the inappropriate action sometimes. Our data reveals that this happens. In fact, we observe that 13% of the time the long run players choose the inappropriate action. This is significantly positive (Test of equality of proportions, $z = 3.43$, $p < 0.05$).[25] As can be seen in Table V there is again considerable heterogeneity among the cohorts and the B players. In fact the data in this regard comes entirely from the choices of two players (out of 12), one in each of the two cohorts where signalling happens. The signalling player in cohort 1 took 3 sequences to learn, i.e. never inappropriately chose $e1$, i.e. chose $e1$ in state P and $e2$ in state N in the first three sequences. For the remaining 7 sequences she always chose $e1$ the first time she was given the opportunity. And, in accordance to the theory she only chose $e1$ once within a sequence. The "signalling" player in cohort 3 chose $e1$ the first time he was given the opportunity (regardless of the state of the world). He also did so only once within a sequence.

TABLE V:
CHOICES OF $e1$ IN STATE N WHEN FIRST TIME CHOICE IN THE BENCHMARK TREATMENT

Cohorts	
c1	18.18% (4/22)
c2	0% (0/18)
c3	31.81% (7/22)
c4	0% (0/21)
all	13.25% (11/83)

There are two ways in which we can test whether loan giving decreases over time: within a sequence and across sequences. Across sequences, there

[25] For this and all following tests when the comparison is 0% we use the same N as in the proportion observed, i.e. here we test 11/83 vs. 0/83.

is no indication that loan giving decreases.[26] Overall B players offer a loan 60% of the time (see Table VI for details). As we will later see this is almost the same proportion of loan giving as occurs in the good reputation treatment. However, the probit analysis reveals that loan giving declines by 2% per round within a sequence. Hence, there is some evidence that markets might break down as the size of the reputations increase (even though this behavior is not learnt across sequences).

Table VI:
Overall Choices of $b2$ in the Benchmark Treatment

Sequences	c1	c2	c3	c4	all
1–3	49.07%	67.59%	59.26%	50.93%	56.71%
	(53/108)	(73/108)	(64/108)	(55/108)	(245/432)
4–7	51.39%	78.47%	56.94%	59.72%	61.63%
	(74/144)	(113/144)	(82/144)	(86/144)	(355/576)
8–10	48.15%	72.22%	53.70%	65.74%	59.95%
	(52/108)	(78/108)	(58/108)	(71/108)	(259/432)
all	49.72%	73.33%	56.67%	58.89%	59.65%
	(179/360)	(264/360)	(204/360)	(212/360)	(859/1440)

Average payoff over the experiment (excluding show up fees) were $21.64 for the B players and $28.46 for the E players (or type X). See Table VII for details. This suggests that the B players earned roughly 18 points per sequence or 3 points per round and the E players earned 24 points per sequence or 4 points per round. The B players earn roughly their outside option even though they give a loan over 60% of the time. The higher loan giving among the B players allows for the earnings of the E players to be above what they would have theoretically been in the absence of reputation. In fact, we only observe 8.33% of the E players earning less than 3.75, which is significantly less than 50% (Binomial test, $p = 0.003$, one–tailed). Hence, for the majority of players this suggests that reputation is not bad and we can reject *H6*.[27]

[26] We calculated Spearman correlation coefficients between observed loan giving and sequences for each cohort and found no significant relationship.

[27] As for the B players, we observe 45.83% (11/24) of them earning less than 3. This is not significantly smaller than 50% (Binomial test, $p = 0.419$, one–tailed).

Table VII:
Average Payoffs in the Benchmark Treatment

	B players		*E* players (type *X*)		*E* players (type *Y*)[28]	
Cohorts	Points	$	Points	$	Points	$
c1	181.83	21.82	232.33	27.88	232.33	27.12
c2	177.33	21.28	244	29.28	244	31.84
c3	184	22.08	241.67	29	241.67	27.44
c4	178	21.36	230.67	27.68	230.67	30.56
all	180.29	21.64	237.17	28.46	237.17	29.24

4.2 Other Bad Reputation Treatments

4.2.1 The All Human Treatment

Initial period loan giving was much higher in this treatment. A loan was given in the first round 80% of the time in the aggregate. This is significantly different from 50% (see Table VIII for details) and from the initial loan giving in the Benchmark Treatment.[29] The probit analysis reveals that, in comparison to the Benchmark Treatment, there is initially about 16% more loan giving in the All Human Treatment. *H1* is easily rejected in this treatment. For 3 of the 4 cohorts the amount of loan giving was significantly different 50% and all give more than 50% loans.

[28] Note that *E* players of type *Y* in the Benchmark Treatmark were computerized. Therefore, these payoffs are only hypothetical, but we report them here for comparison reasons.

[29] Aggregating over all sequences, loan giving in the All Human Treatment is (marginally) significantly higher than in the Benchmark treatment (Robust rank order test, $\grave{U} = -2.160$, $p < 0.1$, two–tailed). However, if we only look at sequences 1–3, loan giving is significantly higher in the All Human Treatment (Robust rank order test, $\grave{U} = -3.364$, $p < 0.05$, two–tailed)

Table VIII:
First Round Loan Giving in the All Human Treatment

Cohorts		z–values
c1	96.67% (58/60)	-7.100**
c2	60% (36/60)	-1.420*
c3	96.67% (58/60)	-7.100**
c4	65.00% (39/60)	-2.195**
all	79.58% (191/240)	-9.102**

Note: z–values of the Binomial Test are given in the third column, **(*) indicates significance on the 5% (10%) level, two–tailed.

After observing one $e2$ (and no previous $e1$) loan giving increases slightly to 83% and declines in two of the cohorts. Compared to the initial loan giving there is no significant difference (Test of equality of porportions, $z = -0.926$). After observing two $e2$'s and no previous $e1$'s loan giving declines to 74% with a decline being observed in three of the four cohorts (see Table IX for details). This decline is significant (Test of equality of proportions, $z = 2.265$, $p < 0.05$). After more than two $e2$'s and no previous $e1$'s loan giving declines to 67% on the aggregate. This decline in loan giving is observed in each cohort. However, compared to the previous decline, i.e. after two $e2$, this was not significantly different (Test of equality of proportions, $z = 1.532$). It may be noted that the decline was monotonic in the two cohorts which began with a very high proportion of loan giving (97% in each) and was more erratic in the two groups that started with loan giving below 70% in the first round. In comparison with the benchmark treatment, loan giving remained higher in the all human treatment even after B players observed $e2$'s.

The probit analysis reveals that neither the observation of one $e2$ and no previous $e1$ nor the observation of two $e2$ and no previous $e1$ change the probability to give a loan. In conclusion, *H2* is rejected.

Table IX:
Choices of $b2$ After $e2$ in the All Human Treatment

Cohorts	$1 \times e2$	$2 \times e2$	$>2 \times e2$
c1	83.33% (40/48)	58.06% (36/62)	50% (29/58)
c2	70.49% (43/61)	76.09% (35/46)	61.54% (32/52)
c3	93.48% (43/46)	88.64% (39/44)	70.73% (68/82)
c4	90% (36/40)	78.79% (26/33)	62.5% (20/32)
all	83.08% (162/195)	73.51% (136/185)	66.52% (149/224)

After observing an $e1$ loan giving increased to 83% (see Table X for details). This is not significantly different from the 80% initial loan giving (Test of equality of proportions, $z = -1.067$). In comparison with the benchmark bad reputation treatment there was a significant difference in the manner in which B players responded to observing an $e1$. The probit analysis reveals that while there is still an increase in the probability to give a loan after at least one $e1$ has been observed, this increase is significantly smaller in the All Human Treatment than it was in the Benchmark Treatment by about 10%.[30] Hence, *H3* is not rejected.[31]

Table X:
Choices of $b2$ after observing at least one $e1$ in the All Human Treatment

Cohorts	
c1	89.23% (116/130)
c2	69.37% (77/111)
c3	84.92% (107/126)
c4	84.97% (147/173)
all	82.78% (447/540)

[30] It seems that B players understand that a choice of $e1$ in the All Human treatment can also come from a strategic unfriendly player. Therefore, they seem to respond less strongly and increase their loan giving by less than the B players in the Benchmark Treatment.

[31] (Probit analyses): The increase in the probability to give a loan after a successful $b2$ in the previous period is the same as in the Benchmark Treatment.

In comparison to the benchmark treatment the E players, however, were marginally less likely to signal and we only observe roughly 8% of inappropriate choices by the E players (see Table XI for details).[32] Given the high loan giving it appears that the E players did not feel that they needed to signal their preferences. There were two instances when the E players chose the inappropriate action even though they had previous signalled their type. This may be understood here as choosing $e1$ is not a perfect indicator of the type of the player.

Table XI:
Choices of $e1$ in State N when first time Choice in the All Human Treatment

Cohorts	
c1	10.53% (2/19)
c2	5% (1/20)
c3	0% (0/19)
c4	5% (1/20)
all	7.69% (4/78)

Regarding *H5*, as in the benchmark treatment, there is a 2% decrease per round in the probability to give a loan, while there is no decline across sequences.

Table XII:
Overall Choices of $b2$ in the All Human Treatment

Sequences	c1	c2	c3	c4	all
1–3	80.56%	76.85%	89.81%	81.48%	82.18%
	(87/108)	(83/108)	(97/108)	(88/108)	(355/432)
4–7	74.30%	61.18%	86.11%	78.47%	75.17%
	(107/144)	(89/144)	(124/144)	(113/144)	(433/576)
8–10	80.56%	69.44%	88.89%	81.48%	80.09%
	(87/108)	(75/108)	(96/108)	(88/108)	(346/432)
all	78.06%	68.61%	88.06%	80.28%	78.75%
	(281/360)	(247/360)	(317/360)	(289/360)	(1134/1440)

[32] The Test of equality of proportions yields $z = -1.77$, $p < 0.1$ if we compare these choices with the Benchmark treatment, and $z = 2.03$, $p < 0.05$ if we compare it to 0/78.

Given the high amount of loan giving in this treatment it can be expected that the long run players earned more than they would have had there been no possibility to build a reputation. Earnings of the friendly long run players was \$30.18 which amounts to them earning 4.17 points per round (see Table XIII for details). Unfriendly long run players (of type Y) earned even more, \$31.42 which amounts to them earning 4.33 per round. Reputation was not bad in this treatment.[33]

TABLE XIII:
AVERAGE PAYOFFS IN THE ALL HUMAN TREATMENT

	B players		E players (type X)		E players (type Y)	
Cohorts	Points	\$	Points	\$	Points	\$
c1	198.83	23.86	251	30.12	260.33	31.24
c2	190.5	22.86	245.67	29.48	250.67	30.08
c3	197.5	23.70	260.67	31.28	276.33	33.16
c4	218.17	26.18	248.67	29.84	259.67	31.16
all	201.25	24.15	251.5	30.18	261.75	31.41

Further insight is gained by observing that even the short run B players earned more than they would have in the absence of reputation, earning roughly 3.33 points per round. Clearly, the long run players did better than in a world without reputation by sharing some of their potential gains with the short run players. In particular, long run players of type Y often choose $e1$ when the state is P (46% of the time), thereby sacrificing some short term payoff (see Table XIV for details). Furthermore, such a choice is better explained as other-regarding behavior than simply a device to build a reputation. If reputational concerns were driving the E players of type Y to choose $e1$ in state P then we would expect this to happen only once in a sequence. However, 57% of the $e1$ choices in state P are repeated choices. Furthermore, we never noticed such signalling by the Y type players when the state of the world was N, i.e. 100% of choices of E players of type Y in state N are choices of $e2$.

[33] 0% (0/12) E players of type X earned less than 3.75 and 95.83% (23/24) B players earned more than 3 (Binomial test, $p < 0.001$, one–tailed).

Table XIV:
Choices of $e1$ in State P by E Players of Type Y in the All Human Treatment

Cohorts	Overall choices of $e1$	Repeated choices of $e1$
c1	42.03% (29/69)	65.52% (19/29)
c2	26.67% (16/60)	43.75% (7/16)
c3	32.05% (25/78)	60% (15/25)
c4	82.61% (57/69)	54.39% (31/57)
all	46.01% (127/276)	56.69% (72/127)

4.2.2 The 66.7% Treatment

Loan giving in the initial round is 35% when aggregated over all cohorts and this is significantly different from 50% (see Table XV for details).[34] All four cohorts gave fewer than 50% loans with three of them doing so significantly. Furthermore, the probit analysis reveals that compared to the Benchmark Treatment there is initially roughly 16% less loan giving in the 66.7% Treatment. We therefore cannot reject *H1'*.

Table XV:
First Round Loan giving in the 66.7% Treatment

Cohorts		z–values (50%)	z–values (33.3%)
c1	35% (21/60)	-2.195**	0.411
c2	45% (27/60)	-0.646	2.054**
c3	33.33% (20/60)	-2.453**	0.137
c4	26.67% (16/60)	-3.486**	-1.232
all	35% (84/240)	-4.583**	0.616

Note: z–values of the Binomial Test are given in the third column, ** indicates significance on the 5% level, two–tailed.

At an aggregate level observing one $e2$'s and no $e1$'s did not lead to a significant decline in loan giving even though two cohorts exhibit lower loan

[34] Interestingly, it is not significantly different from 33.3% (see third column of Table XIV) which suggests the tendency of B players to use some kind of probability matching.

giving (see Table XVI for details). After observing two $e2$'s and no $e1$'s loan giving remained at 34% in the aggregate with a decline occurring in two cohorts.[35] With more than two $e2$'s and no $e1$'s being observed loan giving was 37% in the aggregate with a decline again exhibited in two cohorts.[36] Loan giving declined monotonically in two cohorts.

Table XVI:
Choices of $b2$ after $e2$ in the 66.7% Treatment

Cohorts	$1 \times e2$	$2 \times e2$	$> 2 \times e2$
c1	21.92% (32/146)	12.5% (5/40)	0% (0/0)
c2	44.21% (42/95)	50% (26/52)	48.65% (18/37)
c3	40.96% (34/83)	45.24% (19/42)	40% (12/30)
c4	35.56% (32/90)	23.68% (9/38)	0% (0/14)
all	33.82% (140/414)	34.30% (59/172)	37.04% (30/81)

After observing an $e1$ loan giving almost doubles, i.e. increases to 64% (see Table XVII for details). If we compare this to the initial loan giving this is a significant increase (Test of equality of proportions, $z = -5.944$, $p < 0.05$).

Table XVII:
Choices of $b2$ after observing at least one $e1$ in the 66.7% Treatment

Cohorts	
c1	72.41% (21/29)
c2	63.64% (35/55)
c3	60% (27/45)
c4	64.58% (31/48)
all	64.41% (114/177)

[35] This is also picked up by the Probit since neither coefficient is significantly different from zero.

[36] The respective z–values of the Test of equality of proportions are $z = 0.307$ when initial loan giving is compared to loan giving after observing one $e2$, $z = -0.113$ if we compare loan giving after one $e2$ with loan giving after two $e2$ and $z = -0.425$ if we compare loan giving after two $e2$ with loan giving after more than two $e2$.

Our Probit analysis reveals that there is no difference in the response after observing at least one $e1$ and no previous $e2$, i.e. the probability to give a loan increases like in the Benchmark Treatment by about 17%. Additionally, if the B player had chosen not to give a loan in the previous round, he is much more likely to do the same again, i.e. the probability to give a loan after having chosen $b1$ in the previous round decreases by 10%. The increase in the probability to give a loan after a successful $b2$ in the previous period is the same as in the Benchmark Treatment.

In comparison to the benchmark treatment, E players were significantly more likely to choose the inappropriate action here (Test of equality of proportions, $z = -1.98$, $p < 0.05$). This accords with what we expected. See Table XVIII for details.

Table XVIII:
Choices of $e1$ in State N when first Choice in the 66.7% Treatment[37]

Cohorts	
c1	5.26% (1/19)
c2	15.79% (3/19)
c3	15.79% (3/19)
c4	22.22% (4/18)
all	14.67% (11/75)

However, even in this treatment the market does not break down (see Table XIX for details). In fact, our Probit analysis reveals that unlike in the benchmark treatment or in the All Human treatment, the coefficient of the round variable is not significantly different from zero. Furthermore, loan giving across sequences did not decline.

[37] There were 10 more choices of $e1$ in state N when it was not the first time to make a choice.

Table XIX:
Overall Choices of $b2$ in the 66.7% Treatment

Sequences	c1	c2	c3	c4	all
1–3	32.41%	57.41%	49.07%	36.11%	43.75%
	(35/108)	(62/108)	(53/108)	(39/108)	(189/432)
4–7	29.86%	50%	42.36%	31.25%	38.37%
	(43/144)	(72/144)	(61/144)	(45/144)	(221/576)
8–10	36.11%	42.59%	31.48%	39.81%	37.5%
	(39/108)	(46/108)	(34/108)	(43/108)	(162/432)
all	32.5%	50%	41.11%	35.28%	39.72%
	(117/360)	(80/360)	(148/360)	(127/360)	(572/1440)

To see if reputation is bad we notice that without any possibility to build a reputation the long run players should behave myopically. In response to such behavior the short run players should choose their outside option. Hence, without reputation each player should earn 3 points per round. The earnings for both the B and the E players was lower in this treatment than in any other. The short run players did slightly worse than their outside option earning 2.83 points per rounds (see Table XX for details). In fact, 75% (18/24) earn less than 3, which is significantly more than 50% (Binomial test, $p = 0.011$, one–tailed). The long run players, however, managed to do considerably better than there outside option. They earn roughly 3.67 points per round with none of them earning less than their outside option. The long run players therefore did not experience the bad reputation effect.

This result appears to be driven here by the "bounded rationality" of the B players. They give too many loans. And, unlike in the all human treatment, extra loan giving cannot be fully reciprocated since E players of type Y are computerized and therefore committed to choose $e2$ independently of the state of the world.

Table XX:
Average Payoffs in the 66.7% Human Treatment[38]

	B players		E players (type X)		E players (type Y)	
Cohorts	Points	$	Points	$	Points	$
c1	174.17	20.90	211.5	25.38	214	25.68
c2	170	20.40	230.5	27.66	234.5	28.14
c3	172	20.64	220	26.40	225	27
c4	168.5	20.22	217.5	26.10	214	25.68
all	171.17	20.54	219.88	26.39	221.88	26.63

4.3 The Good Reputation Treatment

We begin by discussing the earnings of the players in this treatment. Recall that theory would predict that in the absence of the ability to have a reputation the short run player should choose her outside option and hence both players should earn 3 points per round. Our results reveal that both players do better. The long run player earning 3.67 points per round and the short term player earning 3.33 points per round (see Table XXI for details). Individually, we observe 100% (12/12) of E players of type Y and 87.5% (21/24) of B players earning more than 3. This is significantly greater than 50% (Binomial test, $p < 0.001$, one–tailed). This suggests that reputation is good.

Table XXI:
Average Payoffs in the Good Reputation Treatment[39]

	B players		E players (type X)		E players (type Y)	
Cohorts	Points	$	Points	$	Points	$
c1	220.17	26.42	237	28.44	243	29.20
c2	181.67	21.80	200.33	24.04	206.33	24.76
c3	202.5	24.30	219	26.28	222.67	26.72
c4	190.67	22.88	220.67	26.48	222.33	26.68
all	198.75	23.85	219.25	26.31	223.67	26.84

[38] Note that E players of type Y in the 66.6% Treatment were computerized just like in the Benchmark Treatment. Therefore, these payoffs are only hypothetical, but we report them here for comparison reasons.

[39] Note that E players of type X in the Good Reputation were computerized. Therefore, these payoffs are only hypothetical, but we report them here for comparison reasons.

Comparable to the All Human Treatment, we observe strong evidence for other regarding behavior. Almost 31% of loans are paid back in the very last round of a sequence (see Table XX for details). This is roughly consistent with earlier good reputation studies (see, in particular, Brandts and Figueras (2003), who report a similar magnitude for this variable).

Table XXI:
Choices of $e1$ in Last Round
in the Good Reputation Treatment

Cohorts	
c1	57.14% (12/21)
c2	50% (1/2)
c3	0% (0/11)
c4	13.33% (2/15)
all	30.61% (15/49)

The Probit analysis for the Good Reputation treatment reveals that rounds play a specific role and one that is distinct from the bad reputation treatments.[40] In particular, the decline in loan giving is not monotone over the rounds. As compared to the first round, loan giving in rounds 2 and 3 does not decline (ceteris paribus). There is a decline in rounds 4, 5 and 6, with the decline being the same in rounds 4 and 5 of about 14% (a LR test could not reject the hypothesis that both coefficients were equal to one another). The decline in round 6, 30%, is significantly higher than in rounds 4 and 5 (LR test). If no $e2$ has been observed, B players are 15% more likely to give a loan than in the initial round and if they observe an $e2$ then the probability to give a loan declines by about 17%. As in all bad reputation treatments, B players are much more likely to give a loan if they have good experience with it, i.e. got a positive payoff from doing so.

[40] As sequential equilibrium would predict. In fact, the SE prediction for our case are for B players to always give a loan in the first 4 rounds and mix with 50% probability in round 5 and 6.

5 Preliminary Conclusions

Short run players typically respond to information in the manner suggested in the theory of bad reputation. Some long run players also signal their types. These are the two fundamental building blocks of the prediction that markets should collapse. However, we observe that this does not lead to a collapse of the market.

Three factors give rise to this result. Firstly, that the short run players give loans too often and don't respond strongly enough to information. Such behavior could be explained by some kind of bounded rationality like being excessively optimistic or not being introspective enough. It could also reflect some kind of other regarding behavior, coupled with the expectation that people of the other side of the market also have such preferences. Our experiment cannot distinguish between these rival explanations.

The long run players did not seek to signal to future short run players their preferences strongly enough and seemed to exhibit other-regarding preferences. Their other-regarding behavior kept the short-run players in the market and the long run players earned more than they would have if they had chosen strategically (and in line with the theory of bad reputation).

A final explanation of why we did not obtain market collapse if that our experiment did not allow for the friendly long run player to build a long enough history. We do observe market participation to decline within any interaction. Hence, one could postulate that markets might actually collapse if this interaction is long enough. The short run players, however, overcame this tendency for markets to collapse by starting each new interaction with higher market participation. They were, hence, able to overcome the predictions of the model of bad reputation and all were able to benefit from the market.

Our study actually suggests the opposite conlusion to that of reputation being bad. Reputation is good and not bad. Market participants are able to benefit from reputation when it will allow it and they do not succumb to the harmful effects of reputation. That is, the finding that reputation is good is quite robust.

Appendix A: Data

Table:XXII Some More Data

	Initial loan giving in Good Reputation			
c1	96.67% (58/60)			
c2	45% (27/60)			
c3	51.67% (31/60)			
c4	58.33% (35/60)			
all	62.92% (151/240)			
	Choices of $b2$ after any $e2$			
	Benchmark	All Human	66.7%	Good Reputation
c1	33.51% (62/185)	62.5% (105/168)	19.89% (8/29)	52.94% (9/17)
c2	67.63% (140/207)	69.18% (110/159)	46.74% (20/55)	17.24% (10/58)
c3	36.57% (64/175)	87.21% (150/172)	41.94% (18/45)	5.88% (1/17)
c4	58.71% (118/201)	78.10% (82/105)	28.87% (17/48)	19.70% (13/66)
all	50% (384/768)	74.01% (447/604)	8.56% (63/177)	20.89% (33/158)
	Choices of $b2$ after $e2$ (Good Reputation)			
	$1 \times e2$	$2 \times e2$	$> 2 \times e2$	
c1	50% (7/14)	0% (0/2)	0% (0/0)	
c2	17.07% (7/41)	36.36% (4/11)	0% (0/5)	
c3	6.67% (1/15)	0% (0/2)	0% (0/0)	
c4	28.57% (16/56)	0% (0/2)	0% (0/0)	
all	24.6% (31/126)	22.22% (6/27)	0% (0/6)	

Appendix B: Details of Probit estimations

B1. Explanatory Variables for Bad Reputation Probits

The dependent variable **choice** was encoded as 0 for $b1$ and 1 for $b2$. The following are the explanatory variables that we used in our random effects Probit estimations for all bad reputation treatments.

round 1,...,6 (independent of sequence)

e1 Equals 1 if at least one $e1$ was observed but no previous $e2$, otherwise 0.

one_e2 Equals 1 if at one (and only one) $e2$ was observed and no previous $e1$.

two_e2 Equals 1 if at two (and only two) $e2$ was observed and no previous $e1$.

prev_success Equals 1 if B in the previous round of the same sequence had chosen $b2$ and received a payoff greater than 0, otherwise 0.

prev_outside Equals 1 if B player in a previous round of a sequence has chosen $b1$.

treatment Equals 1 for observations from the respective other treatment, i.e. when Benchmark is compared with the All Human Treatment, then the All Human observation get a 1 and Benchmark observations get a 0 (respectively for the comparison with the 66.7% Treatment).

t_variable Any of these variables are interaction variables with the treatment variable, i.e. previously mentioned variables are multipled by the variable treatment. They are 0 for all observations of the Benchmark Treatment and take the value of the variable for all observations of the comparative treatment.

We started off by also including **above_two_e2** (equal to 1 if more than two $e2$'s had been observed an no previous $e1$) as an explanatory variable. However, a likelihood ratio test confirmed that the restriction of above_two_e2 being equal to 0 could not be rejected. We therefore did not include this variable in our random effects Probit estimates that we report here. Including **prev_success** and **prev_outside** in our model significantly improved the explanatory power of our model.[41]

B2. Bad Reputation - Estimation Results

We first run separate random effects Probit regression for each treatment (Results are given in Table XXIII). In order to compare these results with the Benchmark treatment we then pooled the data from the Benchmark treatment with the data of the treatment we wanted to compare it to. Results are given in Table XXIV.

Table XXIII:
Results of Random Effects Probit estimations
Separate for each Treatment

	Benchmark	All Human	66.7%
Constant	0.4199**	0.7851**	-0.5391**
	(0.1010)	(0.1406)	(0.1059)
Round	-0.1049**	-0.1867**	-0.0252
	(0.0279)	(0.0343)	(0.0301)
e1	0.8488**	0.5406**	0.8958**
	(0.1438)	(0.1159)	(0.1462)
one_e2	-0.3732**	0.1166	-0.0838
	(0.1018)	(0.1545)	(0.1054)
two_e2	-0.3138**	-0.1747	0.0244
	(0.1187)	(0.1381)	(0.1406)
prev_success	0.5559**	0.3263**	0.3984**
	(0.1154)	(0.1438)	(0.1135)
prev_outside	0.3444**	0.3967**	-0.0984
	(0.1191)	(0.1645)	(0.1018)
-Log likelihood	*801.5418*	*569.6330*	*779.3993*

Note: We report bootstrapped standard errors in parentheses (1000 replications), ** (*) indicates significance on the 5% (10%) level

[41] We did not include any hypotheses about these effects as theory was silent on them.

Table XXIV:
Results of Random Effects Probit estimations
For Pooled Data

	Benchmark vs. All Human	Benchmark vs. 66%
Constant	0.3568**	0.3236**
	(0.0912)	(0.1195)
Round	-0.0827**	-0.0749**
	(0.0263)	(0.0260)
e1	0.7951**	0.8002**
	(0.1307)	(0.1303)
one_e2	-0.3079**	-0.2912**
	(0.1008)	(0.0963)
two_e2	-0.3388**	-0.3771**
	(0.1167)	(0.1120)
prev_success	0.5011**	0.5043**
	(0.1042)	(0.1078)
prev_outside	-0.1476	0.1138
	(0.1102)	(0.1073)
Treatment	0.7428**	-0.6277**
	(0.1309)	(0.1189)
t_round	-0.0638	0.0665*
	(0.0401)	(0.0365)
t_e1	-0.3056*	0.0045
	(0.1684)	(0.1897)
t_one_e2	0.4051**	0.2133*
	(0.1722)	(0.1325)
t_two_e2	0.2871*	0.3485**
	(0.1738)	(0.1683)
t_prev_success	-0.1948	-0.0386
	(0.1624)	(0.1511)
t_prev_outside	-0.2403	-0.3999**
	(0.1805)	(0.1395)
-Log likelihood	*1489.5052*	*1677.2007*

Note: We report bootstrapped standard errors in parentheses (1000 replications), ** (*) indicates significance on the 5% (10%) level

B3. Explanatory Variables for the Good Reputation Probit

As in for all bad reputation estimations, the dependent variable **choice** was encoded as 0 for $b1$ and 1 for $b2$. The following are the explanatory variables that we used in our random effects Probit estimations for the Good Reputation treatment.

R2 Equals 1 if round=2.

R3 Equals 1 if round=3.

R4 Equals 1 if round=4.

R5 Equals 1 if round=5.

R6 Equals 1 if round=6.

e1 Equals 1 if at least one $e1$ was observed and no previous $e2$.

e2 Equals 1 if at least one $e2$ was observed.

prev_success Equals 1 if B in the previous round of the same sequence had chosen $b2$ and received a payoff greater than 0, otherwise 0.

We started off by also including **prev_outside** (equal to 1 if B had chosen $b1$ in the previous round of the same sequence) as an explanatory variable. However, a likelihood ratio test confirmed that the restriction of prev_outside being equal to 0 could not be rejected. We therefore did not include this variable in our random effects Probit estimates that we report here. Furthermore, we conducted several likelihood ratio tests in order to determine whether it is the different number of $e1$ and no previous $e2$ or the specific rounds that best model our data. We did the same for $e2$. It turned out that segregated rounds and aggregtaed $e1$'s and $e2$'s best describe our data.

B4. Good Reputation – Estimation Results

Table XXIII:
Results of Random Effects Probit estimations
for the Good Reputation Treatment

	Good Reputation
Constant	0.3556**
	(0.1141)
R2	-0.2827
	(0.1692)
R3	-0.3712
	(0.1941)
R4	-0.5554**
	(0.2085)
R5	-0.5459**
	(0.2043)
R6	-1.2916**
	(0.2096)
e1	0.6876**
	(0.1455)
e2	-0.6906**
	(0.2031)
prev_success	0.3625**
	(0.1092)
-Log likelihood	*600.3349*

Note: We report bootstrapped standard errors in parentheses (1000 replications), ** (*) indicates significance on the 5% (10%) level

6 Appendix D: Instructions

Welcome to the Economic Research Laboratory. This is an experiment in decision making. Texas A&M University has provided funds for this research. During the course of the experiment, you will be called upon to make a series of decisions. If you follow the instructions carefully and make good decisions, you can earn a considerable amount of money which will be paid to you in cash at the end of the experiment.

It is important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your cooperation.

We will first jointly go over the instructions. All figures have been reproduced on the screens in front of the room. After we have read the instructions, you will have time to ask clarifying questions. Please do not touch the computer or its mouse until you are instructed to do so. Thank you.

Participants

There are two groups of 12 people who participate in today's experiment. Each of you has been randomly assigned to a group and to a role which could be either a so-called **B** or **E** participant. You will remain in the same group and same role throughout the entire duration of the experiment. After we have completed reading the instructions you will be informed about your role on your computer screen.

Timing and Pairing

Participants only interact with members of their own group throughout the experiment. In each group there will be 6 participants assigned to a Brole and 6 participants assigned to an Erole. Furthermore, Eparticipants have been randomly assigned to be either of type **X**or of type **Y**. Eparticipants will be of the same type for the entire duration of the experiment.

The experiment is divided into 10 "**sequences**" of six "**rounds**" each. In each round of a sequence, each B participant will be paired with an Eparticipant of their group. The E participant may be of type X or Y. The B participants will not be informed who they have been paired with and, in particular, they will not know whether they are paired with an E participant of type X or Y. The order in which B participants meet E participants within a sequence has been randomly determined in advance in such a manner that each B participant meets each E participant exactly once during one sequence. This order changes from sequence to sequence.

Within a Sequence

As mentioned earlier, a sequence will consist of six rounds. Each round will proceed as follows: Each B participant is either paired with an E participant of type Xor of type Y. Each B participant begins the round by choosing one of two alternatives. These alternatives are labeled $b1$ and $b2$. The choices of the B participants are transmitted to the E participants. Figure 2 illustrates this decision. As you can see, if B's choice is $b1$ there is no decision to be made by the E participant. If the choice of B is $b2$, then E participants can decide between choosing $e1$ or $e2$.

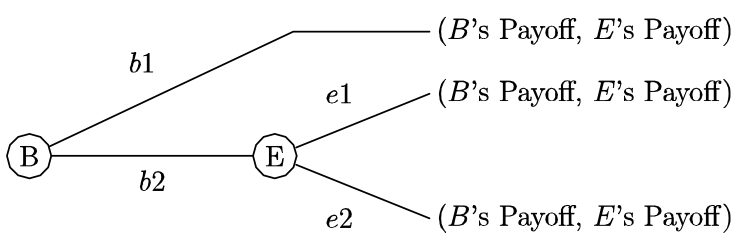

Figure 2: Sequencing of choices

Choices are made by selecting the corresponding alternative on the computer screen and clicking the "OK" button using the mouse (see Figures 3 and 4).

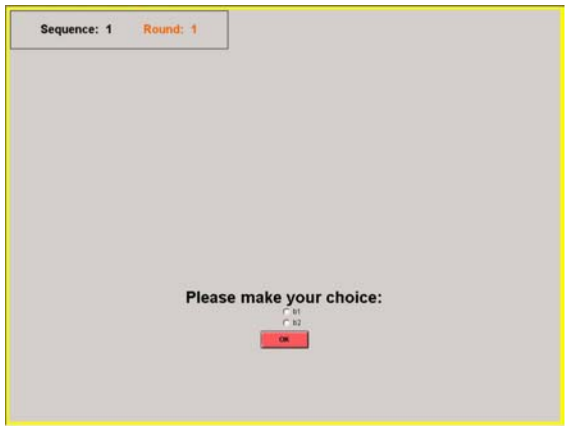

Figure 3: Decision screen of B participants. (The top left-hand side of each computer screen will keep track of the sequence number and the round number within each sequence.)

Figure 4: Decision screen of E participants (after B has chosen $b2$). We will explain the state of the world in a second.

Information

At the beginning of each round (before any choices are made) a computer program will randomly determine whether the state of the world is **N** or **P** with both states being equally likely, i.e. on average 50% of the time the state of the world is N and 50% of the time the state is P. This state of the world is determined for each Eseparately and can therefore be different for different E participants. Only Eparticipants will be informed about the state of the world (see Figure 3). B participants will not be informed.

After each round, all participants will be informed about their own payoff. Nobody will be informed about the identity or the payoff of the participant they are paired with.

However, at the beginning of each round (starting with round 2 in each sequence) and before B participants make their decisions, each B participant will be informed about **all previous rounds' choices** of the E participant they are paired with in the current round (see Figure 4). They will not be informed about whether the E participant is of type Xor of type Y. After six rounds have been completed, a new sequence of six rounds will start. No information will be given about choices and payoffs from previous sequences.

Figure 4: Information to B participants (beginning with round 2 of each sequence)

Determination of your Earnings

In each round, your earnings depend on your choice, the choice of the participant with whom you are paired and the state of the world. Figure 5 describes how your earnings are related to these choices. All earnings are expressed in points. Points will be converted to dollars at the following rate, (1 point ≡$0.12). At the end of the experiment, your total points from all rounds in all sequences will be converted into dollars at the rate specified above, and will privately be paid to you in cash.

Let's slowly go over Figure 5. The figure is to be read from the left to the right. The first node indicates a random draw of the computer program that decides whether a B participant will be paired with a E participant of type X or Y in a given round.

At the beginning of each round there will be another random draw that determines whether the state of the world is N or P for each E participant separately. Each state of the world is equally likely, i.e. the probability that the state of the world is N is 50%. E participants will be informed about the state of the world, but B participants will not.

After the state of the world is determined, B participants are asked to make their choice. B participants will not know whether they are paired with an E participant of type X or Y. This is indicated by the dashed lines linking the choice nodes of the B participants. However, B participants will know the past choices of the E participants they are currently paired with. The payoffs that result from the different choice combinations are written at the right hand side of Figure 5 in the following form: (B's payoff, E's payoff).

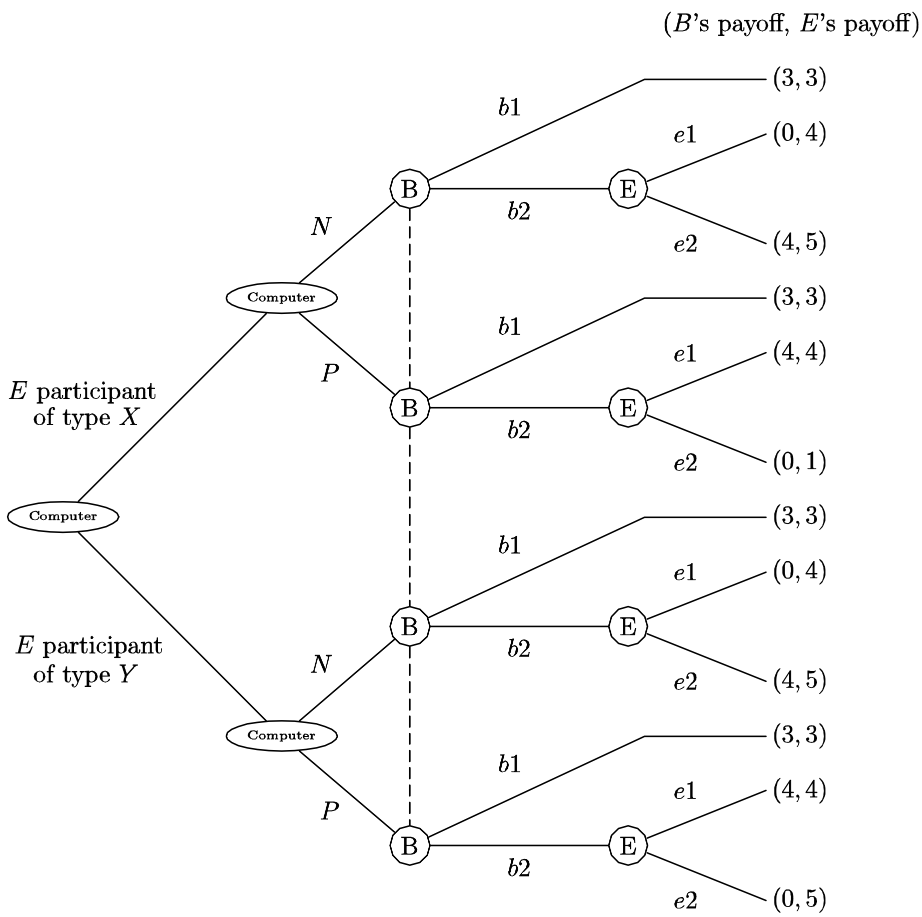

Figure 5: Sequencing of decisions and payoffs within a round

B always chooses first, selecting either $b1$ or $b2$. If B chooses $b1$, then E has no choice, and B and E each earn 3 points. If B chooses $b2$, then an E participant has a choice between $e1$ and $e2$. The payoff that the participants earn will depend on E's choice, the type of the E participant and the state of the world. For example, if E is of type X and the state of the world is N, then when B chooses $b2$ and E chooses $e1$, B will receive 0 points and E will receive 4 points. If instead E chooses $e2$ in the same state, then B will receive 4 points and E will receive 5 points. If the state of the world is P, then when B chooses $b2$ and an X-type E chooses $e1$, B will receive 4 points and E will receive 4 points. If instead E chooses $e2$ in the same state, then B will receive 0 points and E will receive 1 point.

If the E participant is of type Y, and the state of the world is N, then when B chooses $b2$, and E chooses $e1$, B will receive 0 points and E will receive 4 points. If instead E chooses $e2$ in the same state, then B will receive 4 points and E will receive 5 points. However, if the state of the world is P, then when B chooses $b2$ and the Y-type E participant chooses $e1$, B will receive 4 points and E will receive 4 points. If instead E chooses $e2$ then B will receive 0 points and E will receive 5 points.

Summary

1. The experiment consists of 10 sequences.
2. There are two groups of twelve (12) participants each. Each group has six (6) B participants and six (6) E participants. Of the six (6) E participants, three (3) will be of type X and three (3) will be of type Y.
3. Within a sequence:
   (a) There are six (6) rounds.
   (b) Within a round:
      i. Each B participant is paired with a E participant.
      ii. State of the world is determined (separately for each E) and only the E participant is informed about it.
      iii. B participants choose after observing **all** previous choices that the E participant they are currently paired with has made so far within the sequence.
      iv. E participants choose.
      v. Payoffs are determined.
   (c) New pairings of B and E participants.
   (d) Go back to step (b) until each B participant has met six different E participants, and each E participant has met six different B participants.
4. New sequence begins (until 10 sequences are completed).

Please remember not to discuss your role, choices or your results with any other participant between rounds or at any other time during the experiment.

Are there any questions? Please raise your hand and an experimenter will come to you. Do not ask a question out loud. Thank you.

Before we start the experiment, we would like you to answer a few question that are meant to review the rules of today's experiment. Please raise your hand once you are done and an experimenter will attend to you.

1. In which roles can participants in this experiment be?__________
2. Who chooses first? __________
3. How many rounds are in each sequence?__________
4. How often is the state of the world determined in a sequence?__________
5. Suppose you are a B participant and it is your turn to make a choice in round 6. Do you know all the choices that the E participant you are currently paired with has made in rounds 1 through 5? __________ Do you know whether it is an X-type E participant or a Y-type E participant? __________
6. What are the possible payoffs (in points) a B participant can receive if she has chosen $b2$ and the E participant she is matched with has chosen $e1$?__________Which one of these will be realized if the state of the world is P?__________
7. What are the payoffs to the B and E participants if the B participant chooses $b1$?__________

REFERENCES

1. Brandts, J. and N. Figueras (2003): "An exploration of reputation formation in experimental games," *Journal of Economic Behavior and Organization*, 50, 89 – 115.

2. Camerer, C. and K. Weigelt (1988): "Experimental tests of a sequential equilibrium reputation model," *Econometrica*, 56, 1–36.

3. Ely, J. D. Fudenberg and D. Levine (2005): "When is reputation bad?" Harvard Institute of Economic Research Discussion Paper No. 2035 .

4. Ely, J. and J. Valimaki (2003): "Bad reputation," *Quarterly Journal of Economics*, 118, 785 – 814.

5. Fudenberg, D. and D. Levine (1989): "Reputation and equilibrium selection in games with a patient player," *Econometrica*, 57, 4, 759 – 778.

6. Fudenberg, D. and D. Levine (1992): "Maintaining a reputation when strategies are imperfectly observed," *Review of Economic Studies*, 59, 3, 561 – 579.

7. Kreps, D. and R. Wilson (1982): "Reputation and imperfect information," *Journal of Economic Theory*, 27, 2, 253 – 279.

8. Milgrom, P. and J. Roberts (1982): "Predation, reputation and entry deterrence," *Journal of Economic Theory*, 27, 280 – 312.

9. Morris, S. (2001): "Political correctness," *Journal of Political Economy*, 109, 21, 231 – 265.

10. Neral, J. and J. Ochs (1992): "The sequential equilibrium theory of reputation building: A further test," *Econometrica*, 60, 1151-169.

11. Siegel, S. and N.J. Castellan Jr. (1988): *Nonparametric Statistics for the Behavioral Sciences*, 2nd edition, McGrawHill.