Is Reputation Good or Bad? An Experiment*

Brit Grosskopf Rajiv Sarin

Department of Economics
Texas A&M University

June 21, 2007

Abstract

We experimentally test whether the possibility of building a reputation impacts behavior in the manner suggested by theory. Our design varies whether reputation building is possible or not by making the past choices of the long run player observable to the short run players or not. Our unified reputation framework theoretically allows for either the good or the bad reputation prediction to emerge. We observe that reputation is not necessarily bad, but neither is it as good as previously thought.

1 Introduction

Questions of reputation pervade the analyses of markets. Should a monopolist develop a reputation to fight entry in her market? Should an entrepreneur develop a reputation of re–paying loans? How much should such

*This paper has benefited from comments from audiences at the Brown Bag Seminar at Texas A&M University and seminars at various universities and conferences. We thank them all. We also thank Paan Jindapon and Elizabeth Watson for indispensible help in running the experimental sessions. Financial support from NSF SES#0649218 is gratefully acknowledged.

reputations be worth? Many important policy questions also boil down to questions of reputation. How important is it to punish criminals harshly to maintain a reputation for toughness? Should a country ever negotiate with one associated with terrorism? Questions of reputation also pervade thinking about developing societies in which reputation is often seen as an alternative to commitment in the absence of a suitable institutional framework.

In the now classic approach to study reputation, pioneered by Kreps and Wilson (1982) and Milgrom and Roberts (1982), a sequence of short run players interact with a long run player whose preferences or type is unknown to them. The short run players observe the previous choices of the long run player. The availability of this information allows the long run player to build a reputation about her type. This literature suggests that the long run player's ability to build a reputation may allow her to obtain a higher payoff than what would have been obtained by her had her previous actions not been observable (or her type been known).[1] Therefore, reputation is thought to be *good*.

A long run player, recognizing this beneficial effect of reputation, would seek to provide this information. This implication of the theory is evident on the internet where vendors provide a record of their past choices and, in fact, pay to be able to provide this information. This evidence corroborates one of the salient findings of the early literature on reputation. But, is providing such information necessarily beneficial?

Recent work by Ely and Valimaki (2003) suggests that reputations need not be beneficial for the long run player.[2] That is, if the previous actions of the long run player are observed by successive short run players then the long run player may do worse than if such information was not available to the short run players. Therefore, reputation is considered *bad*. In fact, the theory predicts that the short run players may not interact with the long run players and, hence, markets may not exist because of the bad reputation effect. Consequently, this effect may be difficult to observe in the field.

[1] For precise bounds on how much a reputation can benefit the long run player see Fudenberg and Levine (1989, 1992).

[2] A similar point is made by Morris (2001) in a framework in which there are two "long" lived players.

The objective of this study is to experimentally test whether the possibility of having reputations impacts behavior in the manner suggested by theory. We test this in two ways. First, we vary whether reputation building is possible or not by making the past choices of the long run player observable to the short run players or not. Second, we design a reputation framework which allows for both the good and the bad reputation result to emerge.[3] Our design allows us to explicitly test whether the long run player, of the relevant type, gets a higher or lower payoff than what she would have obtained in the absence of the possibility of having a reputation. Hence, we are able to test whether the prediction that reputation is good is more or less robust than the prediction that reputation is bad.

We obtain two surprising results under the assumption that monetary payoffs represent preferences. First, long run players earn more than predicted in the good framework when their past choices are not observed. Second, long run players earn more than predicted in the bad framework when their past choices are observed. In both cases this happens because the short run players, who may be thought of as Banks, give loans when theory predicts otherwise. This allows the long run player, who may be thought to be an Entrepreneur, to earn more than theory predicts.

It is often observed in experiments that monetary payoffs do not represent preferences.[4] We, therefore, attempt to determine the preferences of the long run player (in the good framework) by designing a suitable Dictator–like game. We also use the same game to elicit the beliefs of the short run players regarding the long run player's actions. We use these "revealed preferences" to further analyze the experimental results.[5]

We find that, in the good framework when past choices of the long run player are not observed, revealed preferences can explain why the short run players give loans. This is because the short run players correctly believe the long run player will often choose according to the short run players' interests

[3] Our reputation framework nests that used in earlier studies of the good reputation result (Camerer and Weigelt, 1988; Neral and Ochs, 1992 and Brandts and Figueras, 2003).

[4] For example, money amounts cannot represent the preferences of the Responder in Ultimatum games (who rejects small offers) or that of Trustees in the Trust game (who return money to Trustors). See, Güth, Schmittberger and Schwarze (1982) and Berg, Dickhaut and McCabe (1995).

[5] Camerer and Weigelt (1988) refer to these as "homemade priors."

(i.e., choose appropriately) and repay the loans. Using the long run players' revealed preferences and the expectation thereof by the short run players we are, however, unable to explain why short run players give loans in the bad framework when past choices of the long run player are observed. To be able to explain this behavior we need to acknowledge that the short run players themselves seem to be guided by preferences for appropriate choices and are not only concerned with monetary payoffs. Our results, hence, suggest that the prediction that reputation is good is more robust than the prediction that it is bad. However, reputation is not as good as predicted because the long run players are able to perform much better than theory predicts when their past choices are not observed.

Our paper provides the first study of the impact of information about the long run players' previous choices in the classic model of reputation. The first paper to carefully study this model of reputation was Camerer and Weigelt (1988). We adopt from them a version of the trust (or investment) game in which the short run players are Banks and the long run players are Entrepreneurs. Their paper, however, focuses on how sequential equilibrium predicts behavior in the experiment. Their design varies the probability with which the long run player may be of either type but does not alter the information available to the short run players. Neral and Ochs (1992) and Brandts and Figueras (2003) further study the Camerer and Weigelt reputation framework. Together these studies provide some support of sequential equilibrium though they do not address the impact of information upon which short run players form reputations.

A study that varies the availability of information regarding the long run player's choices is Bolton, Katok and Ockenfels (2004). They allow buyers (short run players) to observe or not the previous choices of the sellers (long run players). They find that reputation helps the long run players realize higher payoffs. The basic game they study is also a variant of the trust game. Their setting is different from ours and Camerer and Weigelt's in that the buyers have no uncertainty about the sellers' preferences. Uncertainty regarding these preferences is a critical part of the classic reputation models. Another related study is Anderhub, Engelmann and Güth (2002). The basic game they study is the trust game in which the trustor has incomplete information about the trustee's preferences. However, both players are modelled as long run who are playing a repeated trust game (under incomplete information).

None of the previous work has studied the impact of information about the long run players' previous choices in the bad reputation setting. Our paper provides the first such test of this theory. Our design of the bad reputation theory nests the earlier work on the good reputation model and clearly reveals why the starkly different results are obtained in closely related settings.[6]

The paper is organized as follows. In the next section we present our experimental design. Section 3 derives the theoretical predictions. Section 4 reports the experimental findings. Section 5 discusses some of the salient and surprising features of these results and Section 6 concludes. Appendix A presents details of the Probit analyses we performed. Appendix B and C present sample sets of instructions. Appendix D contains additional round–by–round data.

2 Experimental Design

To help explain the design of the experiment we adapt the story pioneered by Camerer and Weigelt (1988) and adopted by Neral and Ochs (1992) and Brandts and Figueras (2003). This will most clearly reveal the manner in which their design of the good reputation framework is nested in our reputation framework.

A 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 the entrepreneur in the past but do not know if they were the appropriate projects.[7] Each project is appropriate 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 state of the world.

The entrepreneur's preferences are unknown to the banks.[8] A friendly

[6] A recent study that investigates the conditions under which reputation may be bad is Ely, Fudenberg and Levine (2007).

[7] The appropriate project is the one that maximizes the bank's payoff conditional on the state of the world.

[8] We assume that payoffs completely describe a player's preferences. The validity of this assumption is discussed in Section 5.

(type X) entrepreneur always prefers the appropriate project. The unfriendly (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 $b2$ or $b1$). If the bank refuses to give a loan (i.e., chooses $b1$) then the interaction with the entrepreneur ends. If a loan is given, the entrepreneur has to decide on a project.

The extensive form of the "stage game" interaction between the entrepreneur and a bank is illustrated below in Figure 1. Observe that E players of type X have preferences aligned with the B players. They prefer $e2$ in state N and $e1$ in state P. E players of type Y always receive a higher payoff from choosing $e2$, independently of the state of the world.

The introduction of states of the world, which determine the appropriateness of a project, extends the good reputation framework of Camerer and Weigelt (1988), henceforth CW, to a more general reputation framework. This framework allows for both the good and the bad reputation result to emerge. If the only possible state is P (which occurs with probability 1) then our design collapses to that of CW.[9] The uncertainty regarding the state, and, consequently, whether the entrepreneur chooses the appropriate project, is instrumental in the emergence of the bad reputation result.

We employ a 2×2 design, in which we vary the framework, either good or bad, and either allow for reputation building, by giving information about the past choices of the long run player, or not (see Table I for a summary). We use a between–subject design, i.e., each subject participates only once in our experiment. For each treatment we have four independent cohorts of 6 B players and 6 E players who participate in 10 repetitions (or, sequences) of a 6 period (or, round) game.

Players are randomly assigned to their roles (B or E) and these stay the same throughout the session. Each of the six B players meet each of the six E players once per sequence in a randomly predetermined order that changes from sequence to sequence. States of the world are determined independently for each E player and each round. The states of the world are randomly

[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 subjects in different roles.

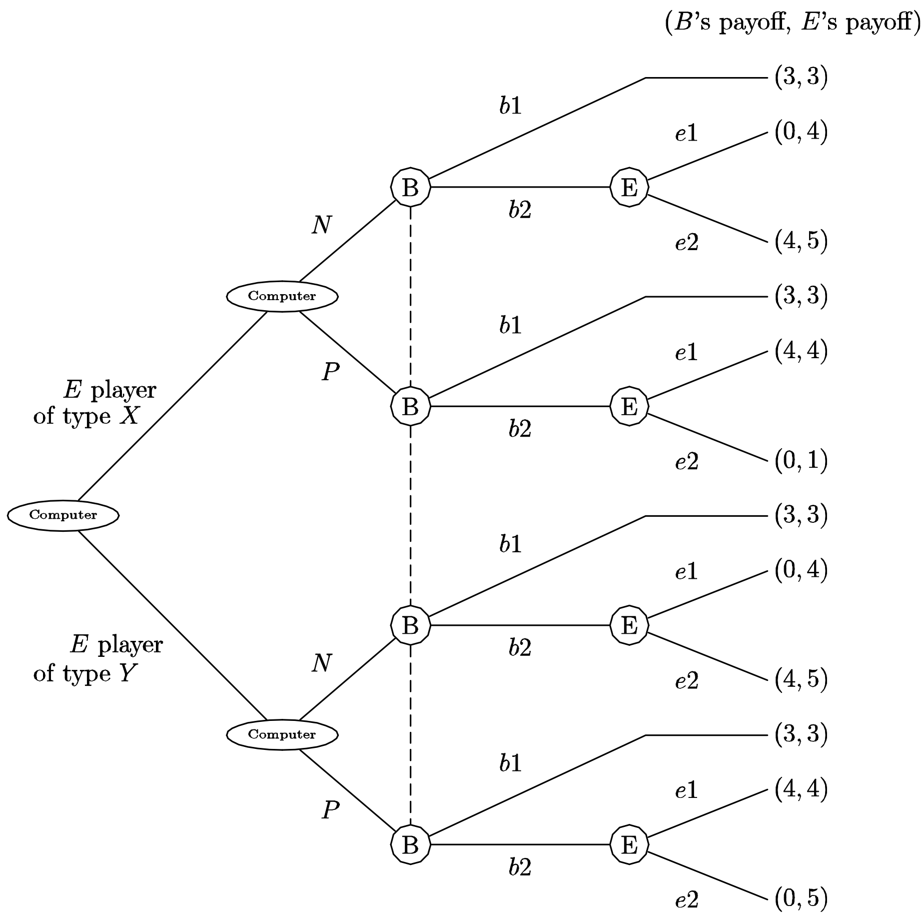


Figure 1: The reputation stage game

generated and kept fixed for all bad reputation treatments. Whereas the E players know their type (i.e., preferences) and the state of the world (in the bad reputation treatments), the B players only know the number of E players of each type in their group and the probability of each state of the world.[10]

In the treatments where reputation building was possible, B players were informed about the previous choices of the E player they were meeting in the current round before they made a choice (if there were any). In the treatments without reputation building, B players were not informed about the previous choices of the E players.

TABLE I:
THE 2×2 EXPERIMENTAL DESIGN

	Framework	
	Good	Bad
With Reputation	4 cohorts	4 cohorts
Without Reputation	4 cohorts	4 cohorts

Each cohort of each framework consisted of 6 B players and 3 E players of each type, i.e., 3 friendly and 3 unfriendly ones.[11]

The Good Framework: The probability of the state of the world P was 1 in the Good Framework. Theory in the Good Framework focuses attention on the choices of the (strategic) long run player who is unfriendly. The other type of the long run player is assumed to be committed to always choose the appropriate action. In order to implement this assumption we computerized the committed friendly long run player. The computerized E players of type X were programmed to always choose $e1$ upon a choice of $b2$ by the B players.[12]

[10] All players were informed about this.

[11] See Grosskopf and Sarin (2006) for implementations with other proportions of E players.

[12] As computers did not receive payoffs the corresponding entries were omitted in the presentation of the game. All participants were informed about the presence of computerized players.

The Bad Framework: Each state of the world was equally likely, i.e., $\Pr\{N\} = \Pr\{P\} = 1/2$. Since the (strategic) long run player on whom interest is focused in the Bad Framework is the friendly one, we computerized the committed unfriendly long run player. The E players of type Y were computerized players who were programmed to always choose $e2$ when given a loan ($b2$). Observe that the unfriendly players were computerized in the Bad Framework and the friendly players were computerized in the Good Framework. Table II summarizes the differences between the good and the bad frameworks.

TABLE II:
COMPARISON OF THE FRAMEWORKS

	Good	*Bad*
E player – type X (friendly)	Computerized	Real
E player – type Y (unfriendly)	Real	Computerized
$\Pr\{X\}$	1/2	1/2
$\Pr\{N\}$	0	1/2

All our experimental sessions were conducted at the Economic Research Laboratory at Texas A&M University, from June 2004 until November 2005. Our participants were undergraduate students of all, but economics, majors. 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 test whether the instructions had been understood. All answers were checked and corrected by the experimenters and remaining questions were answered. Throughout the sessions the subjects 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 given a score sheet to record their choices and earnings for the entire session. Participants were paid their accumulated earnings at the end of the session.

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

Payoffs in points were converted to dollars at the rate of 1 point = \$0.12. Average earnings were \$25.37. An additional \$5 show–up fee was paid.

Our instructions did not mention any frame or context. We provide "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.

We supplement the analyses of the reputation games with data from one–shot decisions in which 34 students participated in a pen and paper experiment in October 2005. Half of them were randomly chosen to be deciders and half of them were randomly chosen to be receivers.[14] After initially meeting in the same room and going over the instructions, deciders and receivers were separated into two rooms. Deciders had to make a sequence of 6 choices. The results of one of the six decisions are discussed later in the paper. Instructions can be found in Appendix C. We paid for 2 of the six choices (determined by two roles of a die) and 1 point was equal to \$1. We paid a \$7.50 show–up fee for this much shorter experiment in which the average earnings were \$15.26 including the show–up fee.

3 Theoretical Predictions

3.1 Good Framework with Reputation

The human (type Y) long run player in the last (sixth) round has no reputational concerns. Hence, if given a loan she chooses $e2$. Consider the strategy of the short run B player in this last round. If he observes that the long run E he is meeting has previously chosen $e2$ then he knows the E player is of type Y and has interests that are not aligned with his. He therefore should choose $b1$ and get a payoff of 3.[15] If the E player he meets has no prior history (which happens if she was never previously given a loan) then the B player should choose $b1$ since 3 is greater than 2, the expected payoff from loan giving.[16] If the E player he meets has never chosen $e2$ but has

[14] Upon entering the classroom, the subjects picked tokens with either an odd or an even number. Besides serving as IDs for the session, these tokens determined who would be a decider and who would be a receiver.

[15] As is common in the analyses of such games, we assume risk neutrality, additive utility and no discounting of future payoffs.

[16] We use loan giving and the choice of $b2$ interchangeably throughout the paper.

sometimes chosen $e1$ then the updated beliefs of the short run player have to take into account that a long run player of type Y might have strategically chosen the appropriate project (i.e., $e1$) in previous rounds. Let's assume that in round T the B player thinks he meets a computerized long run player (type X) with probability p_T. Then the gain from choosing $b2$ exceeds the sure gain from choosing $b1$ if $4p_T > 3$. This means that as long as E's reputation (p_T is a measure of what type of player E is thought to be) is greater than the threshold of $\frac{3}{4}$, B will choose $b2$.

In round $T-1$, given that E is given a choice (i.e., B has chosen $b2$), a human E player could choose $e2$ and get 5 and another 3 in T (since B will for sure choose $b1$ in the last round if E chooses $e2$ in $T-1$). Or a human E can play a mixed strategy, choosing $e1$ with probability s_{T-1} and $e2$ with probability $1-s_{T-1}$. E will want to choose s_{T-1} so that when B observes E's choice of the appropriate project ($e1$) in period $T-1$ and updates his beliefs about E, his updated posterior probability p_T is above the threshold of $\frac{3}{4}$ (i.e., B will give a loan in T). Then E's total expected earnings from periods $T-1$ and T are $(4+5)s_{T-1}+(5+3)(1-s_{T-1}) = s_{T-1}+8$. Therefore, intuitively, E would like to choose s_{T-1} as large as possible, provided s_{T-1} makes the posterior probability p_T above B's threshold of $\frac{3}{4}$. More precisely, if the probability p_T is exactly at this threshold, B will be indifferent between lending and not lending. B will choose a mixed–strategy probability which makes E indifferent between choosing the appropriate and the inappropriate project, and E will choose a mixed–strategy probability which makes the posterior probability p_T equal to its threshold. If B uses Bayes's rule to update probabilities, the posterior probability p_T is given by $p_T = p_{T-1}/(p_{T-1}+s_{T-1}(1-p_T))$. For this posterior p_T to exceed the lending threshold of $\frac{3}{4}$ requires $s_{T-1} < \frac{1}{3}\frac{p_{T-1}}{1-p_{T-1}}$. A rational E will choose s_{T-1} to make this hold with equality.

In round $T-1$, B will choose $b2$ if and only if his expected payoffs from lending are greater than his expected payoff from not lending. We get $4(p_{T-1}+s_{T-1}(1-p_{T-1})) > 3$. Since E will choose $s_{T-1} = \frac{1}{3}\frac{p_{T-1}}{1-p_{T-1}}$, we can substitute this and find B's lending threshold in period $T-1$ as $p_T > \left(\frac{3}{4}\right)^2$. In period $T-k$, E's mixed strategy should satisfy $s_{T-k} = \left[\left(\frac{4}{3}\right)^k - 1\right]\frac{p_{T-k}}{1-p_{T-k}}$ and B's lending threshold is $p_{T-k} > \left(\frac{3}{4}\right)^{k+1}$. This threshold is smaller in earlier rounds because even human long run players (type Y) are likely to choose the appropriate project in early periods, therefore B's require less assurance that E is of type X to convince them to lend. Since each sequence begins with the commonly known prior of 0.5 (3 E players of type X and 3 E players of type Y), in every sequential equilibrium B lends and the human E players choose the appropriate project, as long as the prior of 0.5 is above the threshold of $\left(\frac{3}{4}\right)^{k+1}$(with $k = 5$ in the first round). But as k gets smaller (towards the end of the sequence), the threshold goes up. In the period before the threshold is supposed to be higher than 0.5, the E player has to do something to enhance her reputation – that is, to increase B's posterior probability that E is a X type – or else B would refuse to lend in the remaining periods. That is why the E player starts mixing in the last period when the B player is still giving a loan for sure. With the choice of our parameters this happens in period 4 and we get a probability for choosing the inappropriate project of 0.222. (That is a probability of 0.222 makes the belief of the B player that the E player is of type X equal to 0.5, at which point the B player is indifferent between giving a loan or not).

As mentioned before, once mixed–strategy play begins, E's choice of s_T in equilibrium makes B indifferent between strategies and vice versa. B chooses to lend with probability m, where m makes E indifferent between mixing and choosing the inappropriate project for sure. The expected payoff of E from mixing in period $T-1$ is $s_{T-1}\left[4 + 5m + 3(1-m)\right] + (1-s_{T-1})(5+3)$. Setting this equal to 8 (5 in period $T-1$ and 3 in period T), which is the expected payoff from choosing the inappropriate project, gets us a value of m that is equal to 0.5. This mixing probability is optimal in every period where the posterior probability p_{T-k} equals B's threshold. Table III summarizes the round–by–round theoretical predictions.

We assume the off–equilibrium beliefs of the B player are such that if he observes that the E player has previously chosen e_2 then he becomes convinced that the E player is of type Y. Hence, our off–equilibrium predictions in this case are that $\Pr\{b_1|e_2 \text{ in } t < 4\} = 1$.[17] Table IV summarizes the expected payoff predictions for B players and E players of type Y.

[17] Note that for round 4 through 6 this prediction is on the equilibrium path, since E players are predicted to mix and B players are predicted to not give a loan if they observe $e2$. However, for rounds 1 through 3 E players are predicted to always choose $e1$.

Table III:
Summary of the Theoretical Predictions
(Good Framework with Reputation)

	R1	R2	R3	R4	R5	R6
Lending Probabilities $Pr\{b_2\}$	1	1	1	1	0.5	0.5
Probabilities of Inappropriate Project $Pr\{e_2 \mid Y-\text{type}\}$	0	0	0	0.222	0.571	1

Table IV:
Expected Payoff Predictions
(Good Framework with Reputation)

	R1	R2	R3	R4	R5	R6	Average
B player	4	4	4	3.56	3	3	3.59
E player of type Y	4	4	4	4.22	3.61	3.33	3.86

3.2 Good Framework without Reputation

Without reputation, the long run E player (of type Y) should always choose $e2$ and, hence, the B player should always choose $b1$. There is no market interaction and the long run player is predicted to receive a payoff of 3. Note that theory predicts reputation to be good as the (human) long run player is predicted to earn more in the treatment in which reputation building is possible in comparison to when it is not possible.

3.3 Bad Framework with Reputation

The human (type X) long run player in the last (sixth) round has no reputational concerns. Hence, if given a loan she chooses $e2$ if the state of the world is N and $e1$ if the state of the world is P. Consider the strategy of the B player in the sixth round. If he observes that the E he is meeting has previously chosen $e1$ then he knows the E player is of type X and has interests aligned with his. He should choose $b2$ and get a payoff of 4. If the E player he meets has never chosen $e1$ but has sometimes chosen $e2$ then the updated beliefs of the short run player are such that he believes the E player is of type X with strictly lower than 50% probability and hence the B player gets a higher payoff from choosing $b1$. If the E player he meets has no prior history (which happens if she was never previously given a loan) then the B player is indifferent between $b1$ and $b2$.

Given the strategy of the B player in the last round, it is easy to calculate the strategy of the E player in the second last round. If she has never previously been given a loan then she should choose $e1$ if given a loan, independently of the state of the world. To see this note that she obviously chooses $e1$ if the state of the world is P. If the state is N then she gets a payoff of 4 in the current period from choosing $e1$ and an expected payoff of 4.5 in the final period whereas she would obtain 5 in the current period and 3 in the last if she chooses $e2$.

If she has previously been given a loan then her optimal response depends on her previous choice(s). If she has previously chosen $e1$ then she should choose the appropriate action ($e2$ if state N and $e1$ if state P). If she has never previously chosen $e1$ but only chosen $e2$ then she shouldn't have been given a loan (we are already off the equilibrium path). We assume she chooses $e1$, independently of the state, in order to ensure that she gets a loan in the final period. The optimal response to the E player's strategy in the 5th period is for the B player to choose $b2$ if the E player he meets has previously chosen $e1$ and to choose $b1$ if she has not previously chosen $e1$.

The equilibrium unravels. In period 1 the long run player chooses $e1$ regardless of the state and the B player chooses $b1$. This means that there is no market interaction except possibly in the last period. Table V summarizes the round–by–round theoretical predictions.

TABLE V:
SUMMARY OF THE THEORETICAL PREDICTIONS
(Bad Framework with Reputation)

	R1	R2	R3	R4	R5	R6
Lending Probabilities $Pr\{b_2\}$	0	0	0	0	0	$x \in [0,\ 1]$
Probabilities of Inappropriate Project $Pr\{e_1\|$ state $N \wedge$ X–type$\}$	1	1	1	1	1	0

Off the equilibrium path, we suppose for human E players that $\Pr\{e1|b2 \wedge\ t < 6\ \wedge$ no previous $e1\} = 1$, and for B players that $\Pr\{b2|e1$ in any previous round$\} = 1$. That is, if a B player observes that the E player has previously chosen $e1$ then he believes the long run player is of the human type (and hence chooses $b2$ in response). Table VI summarizes the expected payoff predictions for B players and E players of type X.

TABLE VI:
EXPECTED PAYOFF PREDICTIONS
(Bad Framework with Reputation)

	R1	R2	R3	R4	R5	R6	Average
B player	3	3	3	3	3	3	3
E player of type X	3	3	3	3	3	[3, 4.5]	[3, 3.25]

3.4 Bad Framework without Reputation

If reputation building is not possible then the B player is predicted to be indifferent between giving a loan or not and the human E player chooses the appropriate action in each period in which she is given a loan. Expected payoffs for the (human) long run player are therefore predicted to lie in the interval [3, 4.5]. As B players are indifferent it seems reasonable to suppose that they give a loan 50% of the time. If this was the case the E player of type X would earn $0.5 \cdot 3 + 0.5 \cdot (0.5 \cdot 5 + 0.5 \cdot 4) = 3.75$ per period. This yields the prediction that the (human) long run player earns less in the treatment in which reputation building is possible in comparison to when it is not possible.

4 Experimental Results

We begin by describing the payoffs players earned in the different treatments to ascertain the effects of reputation in each framework. To further understand these results we focus on the behavior of the short run players in subsection 4.2 and long run players in subsection 4.3. In these latter two sections we begin with the point predictions of the theory in the first round and then proceed to see the manner in which the players respond to the information that (possibly) becomes available.

4.1 Payoff Comparisons

How well does reputation fare in the end? Is it good or is it bad or both? To answer this question we compare the payoffs of our participants in the reputation treatments with the respective ones in the no reputation treatments. We also compare obtained payoffs with those predicted by theory. Table VII contrasts average per round earnings in the four different treatments (theoretical predictions are given in parentheses). While we report aggregated data over cohorts in our tables throughout the paper, any nonparametric tests are done using session–level data.[18] In our parametric analysis, reported later in the paper, we will control for individual effects.

Surprisingly, per round earnings in the Good Framework are not significantly different when reputation building is possible from when it is not (Robust Rank Order tests, B players, $\grave{U} = 1.2649$, $n.s.$, E players $\grave{U} = 0$, $n.s.$). Comparing average earnings with the ones that are theoretically predicted we find that E players do not do better or worse than predicted in the Good Framework with reputation (Binomial test, one–tailed $p = 0.194$). However, subjects in the Good Framework without reputation do much better than what was theoretically predicted. In fact, we observe every single E player earning more than the theoretically predicted 3. For the short run

[18] See Siegel and Castellan (1988) for the nonparametric tests used in this paper.

players we find that B players in the Good Framework with reputation do worse than predicted (Binomial test, one–tailed $p = 0.003$), while their peers without reputation do just as well as predicted (Binomial test, one–tailed $p = 0.154$).

TABLE VII:
PAYOFF COMPARISONS ACROSS TREATMENTS
(Predictions in parentheses)

	Good Framework	
	With Reputation	Without Reputation
B's payoff	3.31 (3.59)	3.10 (3)
E's payoff	3.73 (3.86)	3.77 (3)
	Bad Framework	
	With Reputation	Without Reputation
B's payoff	3.00 (3)	2.99 (3)
E's payoff	3.95 ([3,3.25])	3.70 ([3,4.5])

Even more surprisingly, our experimental E players in the Bad Framework managed to obtain a significantly higher payoff when reputation building was possible than when it was not (Robust Rank Order test, $\grave{U} = -2.5022$, $p = 0.05$). There is no significant difference between the earnings of the B players in the Bad Framework with and without reputation building possibilities (Robust Rank Order test, $\grave{U} = -0.5394$, *n.s.*).

Comparing average earnings with those theoretically predicted we find that E players in the Bad Framework with reputation do significantly better than predicted by theory. In fact, all players are earning more than 3.25, the upper bound of the prediction. All earnings of E players without reputation fall into the theoretically predicted interval. For the short run players we find that they neither earn more nor less than theoretically predicted in either the Bad Framework with reputation (Binomial test, one–tailed $p = 0.581$) or the Bad Framework without reputation (Binomial test, one–tailed $p = 0.271$).

Summarizing, the experimental subjects manage to do better than theory would predict. In particular, the long run players in the Bad Framework with reputation do not just do as good as without reputation, they do even better!

Similarly, subjects in the Good Framework without reputation manage to make as much as their peers in the case when reputation building is possible.

To get a better understanding of behavior in the experiment we now turn to look at the behavior of the short and the long run players.

4.2 Behavior of Short Run Players

Theory predicts starkly differing behavior across treatments in the first round. As we shall see the observed behavior of the short run players is very different from the theoretical predictions. This leads to play in the subsequent periods that is very different from the predicted behavior. Rather than check how actual behavior contrasts with the theoretical predictions in the later rounds we analyze how the short run players respond to the information that becomes available.

4.2.1 First Round Behavior

We will start off by analyzing initial loan giving behavior, i.e. first round behavior of the short run B players. For a first look, we shall aggregate over all sequences and all cohorts and look at the observed frequency of $b2$ choices.

When we compare within frameworks, Table VIII shows that there is no difference between initial loan giving in the Good Framework when reputation building is possible compared to when it is not (Robust Rank Order test, $\grave{U} = -0.1240$, $n.s.$). This contradicts the theoretical predictions. Initial loan giving in the Bad Framework is marginally higher when there is a possibility to build a reputation than when there is not (Robust Rank Order test, $\grave{U} = 1.5861$, $p = 0.1$). This is in opposite direction to the theoretical predictions. Clearly, observed loan giving behavior is way off the sequential equilibrium predictions.

When there is a possibility of reputation building there is initially too little loan giving in the Good Framework and too much in the Bad Framework in comparison to the theoretical predictions. Surprisingly, there is no significant difference between the initial loan giving in the Good and Bad Frameworks with reputation (Robust Rank Order test, $\grave{U} = -0.3843$, $n.s.$).

When there is no possibility of reputation building, we find that loan giving is marginally higher in the Good Framework than in the Bad Framework, (Robust Rank Order test, $\grave{U} = -1.5861$, $p = 0.1$). Note that B players are predicted to be indifferent between giving a loan or not in the Bad Framework without reputation. An observational approximation of this statement can be seen in a prediction of 50%. Conducting Binomial tests on session level data we find that in 2 out of the 4 sessions initial loan giving is not significantly different from 50%.

TABLE VIII:
INITIAL LOAN GIVING BY TREATMENT
($b2$ choices in $t = 1$ aggregated over all sequences, theoretical predictions are given in parentheses)

	Framework	
	Good	Bad
With Reputation	63% $\frac{151}{240}$ (100%)	60% $\frac{144}{240}$ (0%)
Without Reputation	60% $\frac{143}{240}$ (0%)	48% $\frac{115}{240}$ ([0,100%])

If we compare initial loan giving in the first three sequences with initial loan giving in the last three sequences in all treatments (see Table IX) we find that there seems to be a slight increase in the observed frequency of initial loan giving. However, Sign tests of session level data show that none of these increases are significant.[19] This suggests that there is *no learning going on over the sequences* and initial loan giving is too low throughout in the Good Framework with reputation and too high in the Bad Framework with reputation when compared to the sequential equilibrium predictions. Furthermore, it is too high in the Good Framework without reputation.

[19]Furthermore, there is no significant correlation between initial loan giving and the sequence number in any of the four treatments.

Table IX:
Initial Loan Giving Over Time

		Framework Good	Bad
With Reputation	First 3 Sequences	56% $\frac{40}{72}$	53% $\frac{38}{72}$
	Last 3 Sequences	68% $\frac{49}{72}$	61% $\frac{44}{72}$
Without Reputation	First 3 Sequences	49% $\frac{35}{72}$	44% $\frac{32}{72}$
	Last 3 Sequences	61% $\frac{44}{72}$	47% $\frac{34}{72}$

4.2.2 Behavior across Rounds (Within a Sequence)

Next we analyze how short run B players react to the reputation of the long run E player (i.e., information about past choices). We begin with the Good Framework. Table X summarizes observed frequencies of loan giving conditional on the E players' history. Note that, while the B players in the treatment without reputation did not know the history of the E players, we include conditional frequencies for comparison in order to see whether the ones we observe with reputation are just random or truly driven by the history.

Table X:
Observed Conditional Loan Giving
(Good Framework)

	With Reputation	*Without Reputation*
After any $e2$	21% $\frac{33}{158}$	35% $\frac{91}{257}$
After only $e1$	72% $\frac{621}{861}$	39% $\frac{305}{787}$

We see that B players continue to give loans to E players who have previously chosen inappropriately in 21% of the cases. However, a spotless history (i.e., only choices of $e1$) gets rewarded with a loan 72% of the time. There is no such differentiation in the case when no information about past choices is provided.

The data provided in Table X aggregates over all rounds whereas the predictions for the Good Framework differ considerably according to which round is being played. Statistical tests on session level data do not seem to be the appropriate ones here. Observed frequencies vary not only across rounds but more importantly across individuals. Therefore, we conduct parametric random effects Probit estimations. For a detailed description of the variables see Appendix A. Results are provided in Table XI.

TABLE XI:
RESULTS OF RANDOM EFFECTS PROBIT ESTIMATIONS
(Good Framework with Reputation)

	Coefficients	Marginal Effects
Constant	0.3556** (0.1141)	0.5880
R2	-0.2827 (0.1692)	
R3	-0.3712 (0.1941)	
R4	-0.5554** (0.2085)	-0.1378
R5	-0.5459** (0.2043)	-0.1354
R6	-1.2916** (0.2096)	-0.3063
e1	0.6876** (0.1455)	0.1515
e2	-0.6906** (0.2031)	-0.1710
prev_success	0.3625** (0.1092)	0.0842
-Log likelihood	*600.3349*	

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

We included dummies for the history observed. The first dummy, "e1", identifies E players that have only chosen $e1$ previously (i.e., never $e2$). The second one, "e2", picks out E players who have chosen $e2$ at least once in the past. Besides including these history dummies, we included specific round dummies in our random effects Probit estimations. We additionally included two variables that theory is silent about. The first variable, "previous success" (B player gave a loan in the previous round and got a positive payoff), was included to capture whether short run B players truly see the interaction with different E players as distinct experiences. The second variable, "previous outside" (B player did not give a loan in the previous period), aimed at accounting for the desire of participants to do something instead of just staying out. We started off by including both variables, however, the inclusion of "previous outside" did not improve the power of our model (Likelihood Ratio test). We therefore did not include it in the estimation that we are presenting here.

As can be seen from Table XI (and in particular from the marginal effects) loan giving in the second and third round does not significantly decrease. The decreases in rounds 4 and 5 are significant and about 14%. There is no significant difference between round 4 and 5 (Wald test). This means that loan giving decreases in round 4 but it does not decrease further in round 5. However, there is clearly a strong end effect with the probability of giving a loan decreasing to 31% by round 6. If only $e1$ was observed then loan giving increases by 15% and as soon as one $e2$ is observed loan giving decreases by 17%. These results suggest that players are responding to the information in the intuitive manner. However, we observe that loan giving increases by 8% by having successfully given a loan in the previous period.[20]

Let's turn to the Bad Framework. Table XII shows summary statistics (aggregated over all participants, rounds and cohorts) of the responsiveness of B players to the history of the E players. B players in the Bad Framework with reputation give a loan in 50% of the cases when they are paired with an E player that has chosen one (and only one) $e2$ in the past. This represents a 10% drop in comparison to initial loan giving. This percentage drops slightly after a B player observes two (and only two) $e2$ choices of the E player. The probability to give a loan goes up to 83% when the B players meet an E player who has only chosen $e1$ previously. There is not much variation in the probability of loan giving in the Bad Framework without reputation.

[20] As discussed above, theory does not predict this effect. Reinforcement models (e.g., Roth and Erev, 1995), however, would tend to predict such an effect.

Table XII:
Observed Conditional Loan Giving
(Bad Framework)

	With Reputation	*Without Reputation*
After one $e2$	50% $\frac{182}{366}$	46% $\frac{169}{369}$
After two $e2$	45% $\frac{113}{249}$	45% $\frac{95}{212}$
After $e1$	83% $\frac{236}{284}$	41% $\frac{90}{219}$

As with the Good Framework, we conduct parametric random effects Probit estimations in order to account for individual heterogeneity. Table XIII indicates the manner in which B players respond to the history of the E players. We see a significant 18% increase in loan giving after at least one $e1$ has been observed. Loan giving decreases significantly in response to an observation of one $e2$ (and no previous $e1$) by 9% and in response to an observation of two $e2$ (and no previous $e1$) by 8%. However, while these two decreases are significantly positive, they are not significantly different from one another (Wald test). That is, B players do not grow increasingly vary by observing more and more $e2$'s as we might have intuitively expected. The results of our random effects Probit estimations also reveal that there is a slight decrease in loan giving within a sequence which is significant. The probability of loan giving goes down by roughly 2.5% per round. Again, the B players seem to respond to information in the intuitive manner.

Table XIII:
Results of Random Effects Probit Estimations
(Bad Framework with Reputation)

	Coefficients	Marginal Effects
Constant	0.4199**	0.6035
	(0.1010)	
Round	-0.1049**	-0.0253
	(0.0279)	
e1	0.8488**	0.1770
	(0.1438)	
one_e2	-0.3732**	-0.0918
	(0.1018)	
two_e2	-0.3138**	-0.0769
	(0.1187)	
prev_success	0.5559**	0.1228
	(0.1154)	
prev_outside	0.3444**	0.0788
	(0.1191)	
-Log likelihood	*801.5418*	

Note: Bootstrapped standard errors in parentheses (1000 replications),
** indicates significance on the 5% level

As in our random effects Probit analysis for the Good Framework, we included two variables that theory is silent about, i.e., "previous success" and "previous outside." We observe that if a B player successfully gave a loan in the previous round and received a payoff of 4, then he is 12% more likely to give a loan than somebody who gave a loan and received a payoff of 0. We also find some evidence that our participants "want to do" something, i.e. if they have stayed out (not given a loan in the previous round) they are 8% more likely to enter (i.e., give a loan) than if they had not stayed out. By conducting Likelihood Ratio tests we found that the inclusion of these two additional variables significantly improved our Probit estimations.

4.3 Behavior of the Long Run Players

The good reputation result is sustained with the unfriendly long run player choosing appropriate action in the earlier rounds. Similarly, the bad reputation result is driven by the friendly long run player choosing the inappropriate action in earlier periods. Do we observe such behavior in the laboratory? Table XIV summarizes our findings in this regard.

TABLE XIV:
FREQUENCY OF LONG–RUN PLAYERS' CHOICES
THAT ARE NOT IN HER SHORT–TERM INTEREST
(Predictions in parentheses)

	Good Framework	
	With Reputation	Without Reputation
$e1$ if $t < 4$	85% $\frac{190}{224}$ (100%)	60% $\frac{127}{213}$ (0%)

	Bad Framework	
	With Reputation	Without Reputation
$e1$ in state N if $t < 6$ and no previous $e1$	13% $\frac{11}{83}$ (100%)	0% $\frac{1}{77}$ (0%)

Table XIV shows that while we find some support for strategic choices of the unfriendly E player in the Good Framework (85%) and the friendly E player in the Bad Framework (13%), the observed frequencies are very different from the predicted 100%. However, it is also clear that the good reputation predictions, if anything, are supported more strongly than the bad reputation predictions. In fact, all observations of choices of inappropriate projects in the Bad Framework with reputation come entirely from 2 E players (out of 12). Both were in different sessions, such that we have 2 sessions where there was one E player each who chose "strategically." One of the two always chooses the inappropriate project when first given the chance and the other E player learns after three sequences to do so.

5 Discussion

The two most surprising results are the higher than expected earnings of the long run players in (i) the Good Framework without reputation and (ii) the

Bad Framework with reputation (see Table VII). We turn now to discuss these results and other findings of our experiment.

5.1 Good Framework

Without reputation, initial loan giving was 60% (see Table VIII). This did not change significantly as the players gained experience with the game (see Table IX). That is, there was little learning. This is surprising as theory predicts loan giving of 0%. How could such a high level of loan giving be sustained?

When reputation building is not possible this result can be explained by the short run players believing that the long run players would often choose $e1$ (i.e., choose the appropriate project). Such beliefs can be sustained if the short run players believed that the long run players had some other regarding preferences (and so monetary payoffs didn't fully capture the players' preferences). To test this conjecture we conducted one–shot decision tasks in a Dictator game like setting which mimic the choices faced by the long run player if given a loan.[21]

Specifically, we asked the subjects whether they would prefer \$5 for themselves while the other receives \$0, or \$4 for themselves while the other receives \$4. This decision mimics the choice set of a bad E player in the Good Framework when she is given a loan. In our one–shot experiments we observe that 82% (14/17) chose the \$4/\$4 split. It is interesting to note that the real receivers (the ones that were not making any choices) expected the deciders to be less other–regarding than they actually ended up being. Only 59% (10/17) thought that the decider would choose the \$4/\$4 split.

This suggests an explanation for the behavior we observe in the spirit of "homemade priors" used by Camerer and Weigelt (1988).[22] That is, roughly 60% of the short run players who give loans do so because they believe the long run players will choose the appropriate project and hence make giving a loan seem more attractive than not. Alternatively, it may be argued that if the short run players believe that 60% of the unfriendly types will choose the

[21] Instructions can be found in Appendix C.

[22] This explanation does not rely on any repeated game effects occuring in the experiment.

appropriate action then they must behave as if they are facing a population of 80% friendly types (as 50% are computerized and friendly and 60% of the remaining will behave friendly). This would lead them to prefer giving a loan (since, giving a loan has an expected payoff of $0.8\,(4) + 0.2\,(0) = 3.2 > 3$ which is the payoff from not giving a loan). As expected by the short run players, long run players respond to loan giving by choosing $e1$ 60% of the time in the first five periods (see Table XV in Appendix D).[23] This allows the short run players to earn a little more than predicted and the long run players to earn much more than predicted.[24]

If we note that, without the information on the previous choices of the long run player, the agents are involved in one shot interactions and that this interaction is like the much studied trust game (Berg, Dickhaut and McCabe, 1995) then our results on too much loan giving can also can be understood as the long run being trustworthy (and the short run player anticipating this). Our findings agree with the well established results in the trust games literature that the return to trust is approximately zero (for the trustor). Additionally, we also observe that trust on the part of the short run players allows the long run player to earn significantly more than theoretically predicted.

In the treatment with reputation, initial loan giving was not very different from the treatment without reputation (see Table VIII) and there was no learning with repetitions of the game. After the initial loan giving, we see that the behavior of the short run players responds to information in the treatments with reputation in the intuitive manner: loan giving decreases in response to observing the long run player has chosen the inappropriate project (i.e., $e2$) and increases otherwise. Loan giving decreases in later rounds with the decrease being sharp in the last round. Long run players choose the appropriate project at a higher rate (85%) in the first four periods and then choose the inappropriate project at increasing rates in the last two periods (see Table XVI in Appendix D). Behavior is, hence, close to the theoretical predictions. The manner of play ensured that the long run players

[23] There is a strong end effect in round 6 when the appropriate project is only chosen 20% of the time.

[24] Other regarding behavior by the long run players is also revealed by the frequency with which the appropriate project is chosen in the very last round of a sequence. This happens 31% (15/49) of the time in the treatment with reputation and 20% (9/46) of the time in the treatment without reputation.

earned roughly the predicted amount while the short run players earned a little less than predicted (but significantly so). We believe the slightly lower earnings of the short run players reflect their lower than predicted loan giving in the earlier periods coupled with the lower than predicted amount of appropriate choices in these periods.

5.2 Bad Framework

In the treatment with reputation, short run players' initial loan giving is 60% (see Table VIII) whereas theory predicts no loan giving. This does not change significantly as the subjects gain experience with the game (see Table IX). There is no learning.

Can any explanation in the spirit of homemade priors be used to understand this result? Recall that short run players are indifferent between giving a loan and not when all long run players choose the appropriate action. We actually observe that some long run players (13%, see Table XIV) do choose the inappropriate action. As short run players typically think that the long run players would choose the inappropriate action more frequently than they actually do, this would result in homemade priors of the short run players which imply over 13% of the long run players to choose inappropriately. Such homemade priors should lead the short run players to not give a loan.

Note that our explanation of too much loan giving in the previous sub–section relied on the long run players often choosing appropriately and of short run players anticipating this and "rationally" responding to it. Trusting is rational for the short run player if the long run player is trustworthy. Being trustworthy corresponds to choosing appropriately. The explanation for why there is so much loan giving in the Bad Framework with reputation cannot solely rely on the revealed preferences of the long run players. That is, we cannot explain trust solely on the basis of the trustworthiness of the other player. Indeed, to explain the level of loan giving we need to acknowledge that the short run players also have a tendency to choose appropriately, with the appropriate choice for the short run player being to give a loan. Being trusting does not require the trustworthiness of the other player and seems to provide some utility in and of itself.[25]

[25] The findings of Ashraf, Bohnet and Piankov (2006) agree with this.

For the short run player, it seems reasonable to argue that, the incentive to choose appropriately is strongest when he knows he is meeting a human long run player. Our data suggest that 60% of the short run players choose appropriately when they believe that with 50% probability they are meeting a human long run player. This lending proportion goes up as the short run players become more confident that they are dealing with a human and decreases otherwise. This tendency to choose appropriately is augmented by the additional and distinct effect we observe that the short run players do not like to repeatedly choose their outside option (see Table XIII).

After the initial round of loan giving, we see that the behavior of the short run players responds to information in the treatment with reputation in the intuitive manner: loan giving decreases in response to observing the long run player has chosen $e2$ and increases significantly otherwise.

As theoretically predicted, long run players in the treatment without reputation never choose inappropriately. The short run players give loans roughly 50% of the time. The earnings are in accordance with those predicted.

6 Conclusions

The long run players do better than theory predicts in two situations. Firstly, in the Good Framework when past actions of the long run player are not observed. Secondly, in the Bad Framework when past choices of the long run player are observed. In both instances, theory predicts the long run players cannot earn more than what they earn when the short run players choose their outside option. In both instances they do better because of the willingness of the short run players to offer loans (and not withdraw from the interaction).

In the Good Framework, we find that long run players choose the action that maximizes the short run players' payoff, even when that is at some cost to themselves. That is, they choose the appropriate action. Hence, we can argue that short run players give loans because they correctly anticipate that monetary payoffs will not accurately reflect the long run players' preferences. Interestingly, the other regarding preferences of the long run players help them earn a higher payoff while the short run players do no better than predicted![26] The finding that the payoffs of the short run players is not higher than predicted also reflects the well–known and robust finding in the trust literature that there is roughly a zero return to trust.

In the Bad Framework, when past actions of the long run player are observed, the anticipation of other regarding preferences of the long run player cannot explain the data. Rather, we find that the short run players themselves may be choosing the appropriate action. Giving a loan is the appropriate action for the short run player. This corresponds to the finding in trust games, that the trust people exhibit cannot be explained by the trustworthiness of others. Indeed, being trusting and trustworthy both correspond to choosing the appropriate actions in our framework.

Our findings, hence, suggest that theoretical predictions describe behavior better when they are in agreement with other regarding preferences and the anticipation of such preferences. In our reputation experiment, predicted behavior is closer to observed behavior when theory prescribes appropriate choices by the long run player as it does in the Good Framework with reputation and the Bad Framework without reputation. Theoretical predictions do less well when theory prescribes inappropriate choices as it does in the Good Framework without reputation and the Bad Framework with reputation. While trust and trustworthiness do not interfere with the theoretical predictions in the Good Framework, they mitigate the market destroying prediction in the Bad Framework.

[26] That people do better (i.e., earn more) than theory predicts has been found in a number of experiments (e.g., finitely repeated Prisoners Dilemma (Selten and Stoecker, 1986), Centipede game (McKelvey and Palfrey, 1992) and social dilemma games (Ostrom, 2003)).

Appendix A: Details of Probit Estimations

A1. Explanatory Variables for Good Reputation Probit

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 Framework with reputation.

R2 Equals 1 if round=2, otherwise 0.

R3 Equals 1 if round=3, otherwise 0.

R4 Equals 1 if round=4, otherwise 0.

R5 Equals 1 if round=5, otherwise 0.

R6 Equals 1 if round=6, otherwise 0.

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

e2 Equals 1 if at least one $e2$ was observed, otherwise 0.

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 also considered 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 the random effects Probit estimates that we report. 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 aggregated $e1$'s and $e2$'s best describe our data.

A2. Explanatory Variables for Bad Reputation Probit

As in our good 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 Bad Framework with reputation.

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 one (and only one) $e2$ was observed and no previous $e1$, otherwise 0.

two_e2 Equals 1 if two (and only two) $e2$ was observed and no previous $e1$, otherwise 0.

prev_success Equals 1 if the B player 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 the B player in a previous round of the same sequence has chosen $b1$, otherwise 0.

We started off by also including **above_two_e2** (equal to 1 if more than two $e2$'s had been observed and 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 the random effects Probit estimations that we report. Including **prev_success** and **prev_outside** in our model significantly improved the explanatory power of our model.

Appendix B: 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 B role and 6 participants assigned to an E role. Furthermore, E participants have been randomly assigned to be either of type **X** or of type **Y**. E participants 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 E participant 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 X or 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 $e2$.

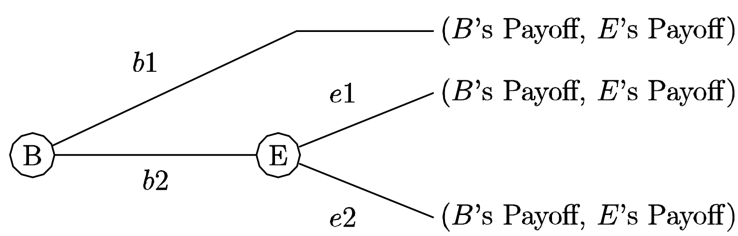


Figure 2: Sequencing of choices

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).

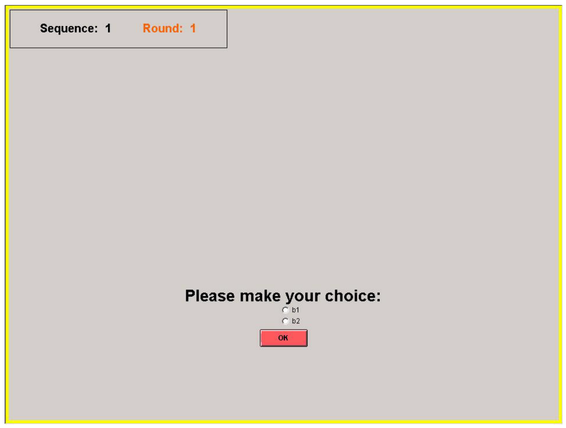


Figure 3: Decision screen of B participants. (The top left-hand side of each computer screen will keep track of the sequence number and the round number within each sequence.)

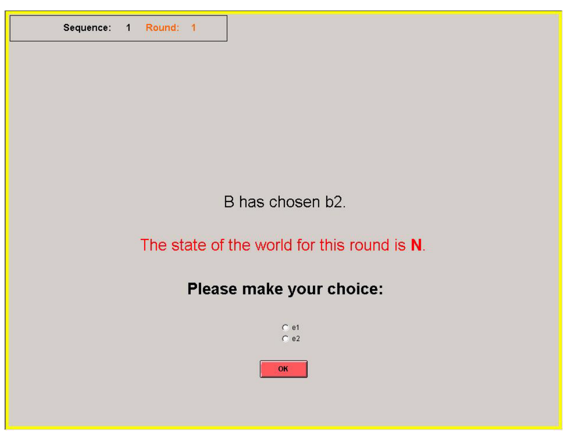


Figure 4: Decision screen of E participants (after B has chosen $b2$). We will explain the state of the world in a second.

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 E separately and can therefore be different for different E participants. Only E participants 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 X or 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.

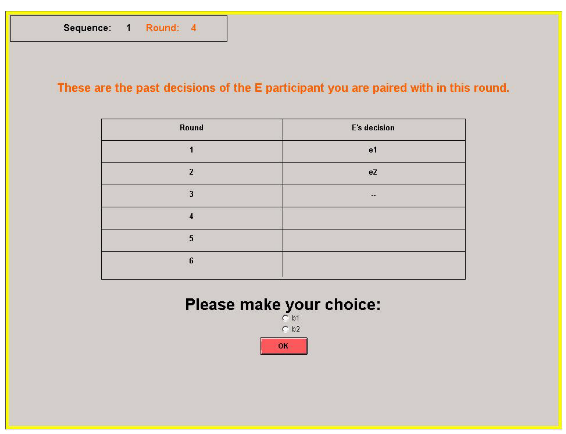


Figure 4: Information to B participants (beginning with round 2 of each sequence)

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 $\equiv$\$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).

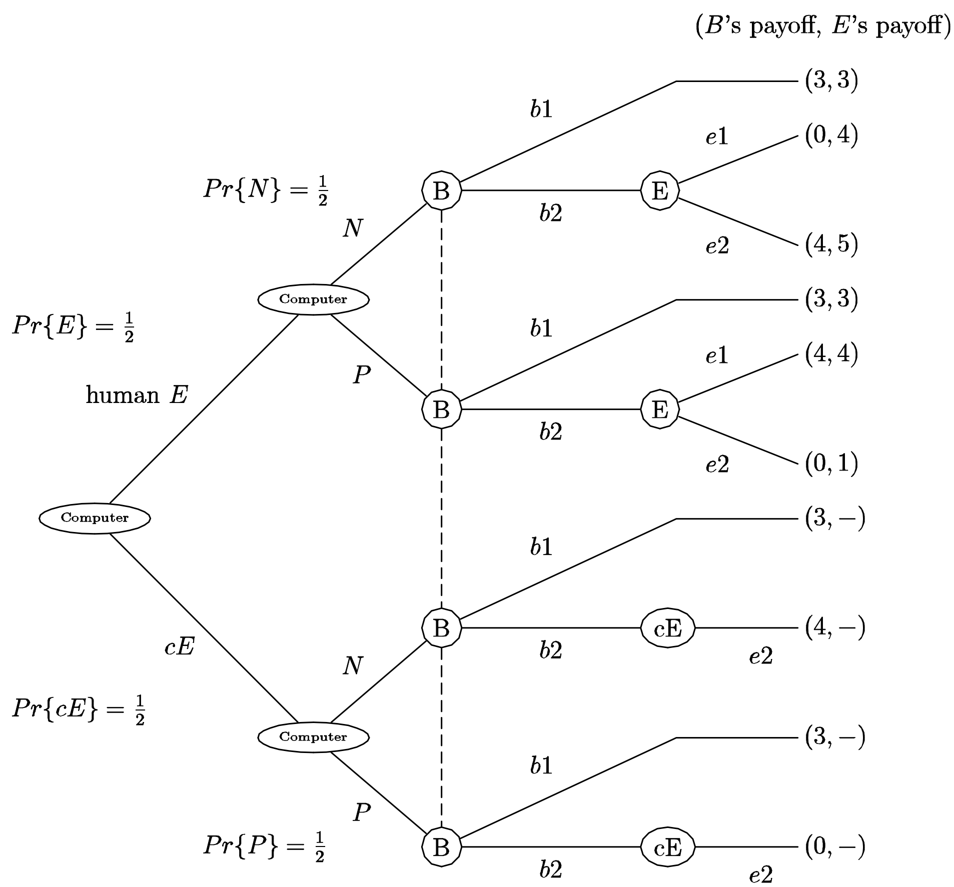


Figure 5: Sequencing of decisions and payoffs within a round

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.

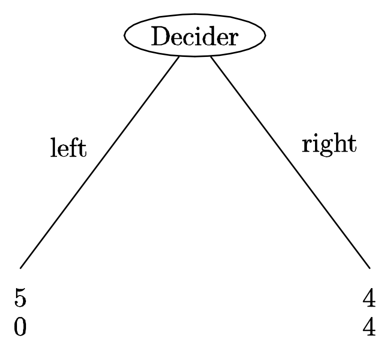
Before we start the experiment, we would like you to answer a few question that are meant to review the rules of today's experiment. Please raise your hand once you are done and an experimenter will attend to you.

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 $b1$?__________

Appendix C: One–Shot Decision

Decision 1

Please choose either left or right in the following decision. If you choose left, your payoff will be $5 and that of the receiver you are paired with $0. If you choose right, then you will receive $4 and the receiver you are paired with will receive $4.



Decision 1

Appendix D: Round by Round Data

TABLE XV:
OBSERVED FREQUENCIES OF LOAN GIVING
(Choices of $b2$)

	Good Framework					
	R1	R2	R3	R4	R5	R6
With Reputation	$\frac{151}{240}$ (63%)	$\frac{159}{240}$ (66%)	$\frac{163}{240}$ (68%)	$\frac{154}{240}$ (64%)	$\frac{153}{240}$ (64%)	$\frac{106}{240}$ (44%)
Without Reputation	$\frac{143}{240}$ (60%)	$\frac{155}{240}$ (65%)	$\frac{153}{240}$ (64%)	$\frac{139}{240}$ (58%)	$\frac{134}{240}$ (56%)	$\frac{89}{240}$ (37%)
	Bad Framework					
	R1	R2	R3	R4	R5	R6
With Reputation	$\frac{144}{240}$ (60%)	$\frac{134}{240}$ (56%)	$\frac{152}{240}$ (63%)	$\frac{142}{240}$ (59%)	$\frac{144}{240}$ (60%)	$\frac{143}{240}$ (60%)
Without Reputation	$\frac{115}{240}$ (48%)	$\frac{116}{240}$ (48%)	$\frac{111}{240}$ (46%)	$\frac{108}{240}$ (45%)	$\frac{94}{240}$ (39%)	$\frac{99}{240}$ (41%)

TABLE XVI:
OBSERVED FREQUENCIES OF CHOICES OF $e1$

	Good Framework					
	R1	R2	R3	R4	R5	R6
With Reputation	$\frac{58}{73}$ (79%)	$\frac{68}{77}$ (88%)	$\frac{64}{76}$ (84%)	$\frac{59}{72}$ (82%)	$\frac{42}{68}$ (62%)	$\frac{15}{49}$ (31%)
Without Reputation	$\frac{39}{67}$ (58%)	$\frac{46}{74}$ (62%)	$\frac{42}{72}$ (58%)	$\frac{43}{65}$ (66%)	$\frac{36}{59}$ (61%)	$\frac{9}{46}$ (20%)

References

1. Anderhub, V., D. Engelmann and W. Güth (2002): "An experimental study of the repeated trust game with incomplete information," *Journal of Economic Behavior & Organization*, 48, 197 – 216.

2. Ashraf, N., I. Bohnet and N. Piankov (2006): "Decomposing trust and trustworthiness," *Experimental Economics*, 9, 193 – 208.

3. Berg, J., J. Dickhaut and K. McCabe (1995): "Trust, reciprocity, and social History," *Games and Economic Behavior*, 10, 122 – 142.

4. Bolton, G. E., E. Katok and A. Ockenfels (2004): "How effective are electronic reputation mechanisms? An experimental investigation," *Management Science,* 50, 1587 – 1602.

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

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

7. Ely, J., D. Fudenberg and D. Levine (2007): "When is reputation bad?," *Games and Economic Behavior*, forthcoming.

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

9. Fischbacher, U. (1999): "z–Tree – Zurich toolbox for readymade economic experiments: Experimenter's manual," mimeo.

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

11. Fudenberg, D. and D. Levine (1992): "Maintaining a reputation when strategies are imperfectly observed," *Review of Economic Studies*, 59, 561 – 580.

12. Grosskopf, B. and R. Sarin (2006): "In search for bad reputation – A sequence of experiments," mimeo.

13. Güth, W., R. Schmittberger and B. Schwarze (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3, 367 – 388.

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

15. McKelvey R. D. and T. R. Palfrey (1992): "An experimental study of the centipede game," *Econometrica*, 60, 803 – 36.

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

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

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

19. Ostrom E. (2003): "Toward a behavioral theory linking trust, reciprocity and reputation," in *Trust and Reciprocity*, E. Ostrom and J. Walker, eds. Russell Sage Foundation.

20. Roth, A. and I. Erev (1995): "Learning in extensive–form games: Experimental data and simple dynamic models in the intermediate term," *Games and Economic Behavior*, 8, 164 – 212.

21. Selten, R. and R. Stoecker (1986): "End behaviour in sequences of finite prisoner's dilemma supergames: A learning theory approach," *Journal of Economic Behavior & Organization*, 7, 47 – 70.

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