Social Norms, Information and Trust among Strangers: Theory and Evidence*

John Duffy[a] Huan Xie[b] and Yong-Ju Lee[c]

December 2009

Abstract

How do norms of trust and reciprocity arise? We investigate this question by examining behavior in an experiment where subjects play a series of indefinitely repeated trust games. Players are randomly and anonymously matched each period. The parameters of the game are chosen so as to support trust and reciprocity as a sequential equilibrium when no reputational information is available. The main questions addressed are whether a social norm of trust and reciprocity emerges under the most extreme information restriction (community-wide enforcement) or whether trust and reciprocity require additional, individual-specific information about a player's past history of play and how long that history must be. In the absence of such reputational information, we find that a social norm of trust and reciprocity is difficult to sustain. The provision of reputational information on past individual decisions significantly increases trust and reciprocity, with longer histories yielding the best outcomes. Importantly, we find that making reputational information available at a small cost may also lead to a significant improvement in trust and reciprocity, despite the fact that most subjects do not choose to purchase this information.

JEL Codes: C72, C91, C92

Keywords: Social Norms, Trust Game, Random Matching, Trust and Reciprocity, Information, Reputational Mechanisms.

---

* We thank Andreas Blume, Giovanna Devetag, Guillaume Frechette, Soiliou Namoro, Lise Vesterlund, Roberto Weber, and seminar participants at Carnegie-Mellon University, Iowa State University, Shanghai Jiao Tong University, the University of Copenhagen and the 2009 CEA annual meeting for helpful comments and suggestions on earlier drafts.

[a] Department of Economics, University of Pittsburgh, jduffy@pitt.edu

[b] Department of Economics, Concordia University, huanxie@alcor.concordia.ca

[c] Samsung Research Institute of Finance, yj0612.lee@samsung.com

1. Introduction

Trust is a key element in sustaining specialization and trade. In many economic transactions, trust emerges among essentially anonymous agents who have little recourse to direct or immediate punishment. For instance, in electronic commerce, it is easy to create new identities, and buyers and sellers often engage in what are, essentially, one-shot transactions. Nevertheless, it is claimed that there are more than 100 million users on eBay! In the credit card market, individual card-holders frequently display little loyalty to any particular bank or card issuer, freely switching balances between credit cards. Similarly, few tourists repeatedly return to the same vacation area to consume again in the same hotel or restaurant.

Given the anonymous and infrequent nature of economic transactions in these markets, an important question is how such markets can work efficiently. In particular, what is the incentive for sellers in electronic markets to deliver the goods purchased, or of the quality promised, knowing that they are unlikely to meet the same buyer again? What is the incentive for borrowers to repay credit card debts if they can switch to another lender next time? What is the incentive for hotels and restaurants in vacation areas to provide good service, knowing that the same consumers are unlikely to ever return?

One possibility is that such incentive problems can be solved by a legal process. However, in many instances, the cost of litigation would far exceed the benefit from the transaction; in such instances legal considerations can simply be ruled out. On the other hand, we do observe that in all of these markets there exist *reputation systems* that collect and disseminate information about market participants. For instance, in most electronic markets there is an online feedback system that allows buyers to rate their prior transaction experiences with sellers and this information is publicly (and typically freely) available. In the credit card market, third party credit bureaus collect information about the customers of all banks and credit card companies and provide the information to other financial institutes, typically for a small fee. Travel guides and websites (e.g. Tripadvisor) provide feedback from tourists about hotels and restaurants in vacation areas.

In this paper we examine two mechanisms by which trust and the reciprocation of trust might be sustained in a population of anonymous strangers. We first examine the hypothesis that trust might be attached to a society as a whole; the fear of the destruction of that trust might suffice to enforce trustworthy behavior by *all* members of the society as shown by Kandori (1992). On the other hand, such a mechanism might be too fragile and so we also examine the possibility that trustworthiness resides at the *individual* rather than at the *societal* level. In particular, we ask whether the provision of information on individual reputations for trustworthiness engenders greater trust than in the case where such information is absent. We further explore whether the free provision of reputational information is responsible for our findings or whether the *availability* of acquiring such information at a small cost suffices to sustain greater trust and reciprocity.

To explore these issues, we conduct an experiment that uses a version of the two-player sequential "trust" (or "investment") game (Berg et al., 1995). In our version of this game, the first mover or "investor" decides whether to invest his endowment with the second mover, the "trustee," resulting in an uncertain payoff. Alternatively, the investor can simply keep his endowment. If the investor invests (or "trusts"), the endowment is multiplied by a fixed factor that is greater than 1 and it falls to the trustee to decide whether to keep the whole amount or return some fraction of it to the investor (or "reciprocate"), keeping the rest for himself. Subjects are asked to play this game for several indefinite sequences (supergames), each consisting of a number of periods. In each period, they are randomly and anonymously matched with one another. Within this framework, we examine several different treatments. In our baseline treatment (and in fact, in all of our treatments), the trust game is parameterized in such a way that, given the number of participants and random anonymous matching, a social norm where all investors invest (trust) and all trustees return part of the investment (reciprocate) constitutes a sequential equilibrium without *any* information provided to investors regarding the identity of their current trustee or that trustee's past history of play. In a second treatment, everything is the same as in the baseline treatment except that, prior to making a decision, the investor can observe the trustee's action choice in the prior period (Keep or Return). In a third treatment, everything is the same as in the second treatment except that, prior to making a decision, the investor can observe a longer history of the trustee's prior choices as well as the frequency the trustee chose to return in the current supergame. Finally, in a fourth treatment, everything is the same as in the third treatment, except that the investor must first choose whether to pay a small cost to view the trustee's history of actions for the current supergame. If the investor does not pay this information cost then, from the investors' perspective, the game is similar to our first baseline treatment where the investor has no knowledge of the prior actions of the trustee with whom he is matched. If the investor does pay for this information, then, from the investor's perspective, the game is similar to that of our third treatment. Importantly, in our fourth treatment, the trustee does not know whether the investor has purchased information about the trustee's past behavior.

In the first treatment, where no individual information is available, we are able to test the theoretical possibility that a social norm of trust and reciprocity can be sustained by anonymous, randomly matched agents out of the fear that deviating from such a norm would precipitate a contagious wave of distrust and retaliatory non-reciprocation. We find that there is very little trust and reciprocity in this baseline treatment. Our second treatment asks whether "minimal" reputational information at the individual level can improve matters, specifically whether additional information on the prior-period behavior of trustees (second-movers) causes these players to reciprocate (Return) more often and if so, whether this change in trustees' behavior engenders greater trust on the part of investors who move first. We find that, when minimal information on the trustee's prior-period choice is provided *following* the absence of such a reputational mechanism (treatment 1 to treatment 2), it leads to a large and significant increase in both trust and reciprocity. However, reversing the order, when minimal information about trustees is initially provided and then removed (treatment 2 to treatment 1) we find no significant difference in the levels of trust and reciprocity between these two treatments. When the amount of information about trustees is increased (in our third treatment) to include the frequency with which the trustee has played return in *all* prior periods of the current supergame ("full" information), we find that such order effects disappear: the provision of the longer history of information about trustees leads to significant increases in trust and reciprocity relative to the absence of such information, regardless of treatment order. Finally, in our fourth treatment, where investors must decide whether to purchase full information on the prior decisions of their matched trustee in the current supergame (provided freely in our third treatment), we find that on average, only one-fourth of investors choose to purchase this information so that the other three-fourths are in the dark about the prior behavior of their current trustee. Nevertheless, trust and reciprocity is significantly higher in this costly information treatment as compared with the baseline, no-information treatment.

We conclude that the emergence of trust and reciprocity resides with the *availability* of individual reputational information as provided, for example, by a credit bureau and not through society-wide enforcement of a social norm of good behavior. We further conclude that *longer* histories are more beneficial than shorter histories in the promulgation of reputational concerns.

2. Related Literature

We are not the first to explore the mechanisms supporting trust and reciprocity among anonymous strangers. Our research draws upon several prior theoretical and experimental studies.

2.1 Cooperation in the Infinitely Repeated Prisoner's Dilemma Game under Random Matching

With anonymous random matching, Kandori (1992) shows that cooperation may be possible if all players adhere to a "contagious strategy" in which individuals who have not experienced a defection choose "Cooperation," and individuals who have either experienced a defection by their opponent or have defected themselves in the past choose "Defection." Specifically, he shows that for an infinite horizon and for any fixed population size, we can define payoffs for the Prisoner's Dilemma game that sustain cooperation in a sequential equilibrium.

As pointed out by Kandori (1992), there are two substantial problems associated with a "contagious equilibrium." First, when the population is large, the argument applies only to games with extreme payoff structures. Second, a single defection causes a permanent end to cooperation and this fragility may make the equilibrium inappropriate as a model for trade.

Ellison (1994) extends Kandori's work and remedies these problems by introducing a public randomization device which adjusts the severity of the punishment. Compared to Kandori's (1992) results, the equilibrium in Ellison (1994) does not require excessive patience on the part of players and applies to more general payoff structures. Furthermore, given public randomizations, the equilibrium strategy supports nearly efficient outcomes even when players make mistakes with a small probability.

Duffy and Ochs (2009) conduct an experimental test of Kandori's (1992) contagious equilibrium using groups of subjects who play an indefinitely repeated two-person Prisoner's Dilemma under different matching protocols and different amounts of information transmission. Their results show that, under fixed pairings there appears to develop a social norm of cooperation as subjects gain experience, while under random matching, experience tends to drive groups toward a far more competitive norm, even when some information is provided about the prior choices of opponents. Thus they conclude that random matching works to prevent the development of a cooperative norm in the laboratory. Camera and Casari (2009) address the same issue of cooperation under random matching, but focus on the role of private or public monitoring of the anonymous (or non-anonymous) players' choices and find that such monitoring can lead to a significant increase in the frequency of cooperation relative to the case of no monitoring.

In contrast to these papers, in this study we examine the indefinitely repeated "trust" game instead of the Prisoner's Dilemma game. Unlike the Prisoner's Dilemma game, the trust (or "investment") game (Berg et al., 1995) we study in this paper has 1) sequential moves and 2) no strictly dominant strategies. In particular, the first mover has an incentive to choose "trust" (rather than no trust) if he believes the second mover will reciprocate, while the second mover has an incentive to cheat (not reciprocate) if the first mover trusts him, but is indifferent between cheating or reciprocating otherwise. This game is more closely related to many real-world *one-sided incentive problems* found, for example, in credit markets or in transactions between buyers and sellers in cyberspace (e-Commerce), where two players move sequentially and only the second mover always wants to deviate from reciprocation in the one-shot game.[1] The one-sided incentive problem of the trust game may be a more promising environment for the achievement of a social norm of cooperation (trust and reciprocity) under anonymous random matching than the Prisoner's Dilemma game with its two-sided incentive problem. Furthermore we note that most real-world reputation systems are designed to monitor the behavior of "second movers". For these reasons, we think it is important to study the trust game under anonymous random matching and with various levels of information on second movers.

2.2 Repeated Trust Games

---

1 Kandori (1992) has a formal definition of a "one-sided incentive problem" (Definition 4 on page 73). The concept requires that, only one of two parties has an incentive to deviate from the cooperative outcome, and there is a Nash equilibrium such that the payoff from the equilibrium is less than the payoff from the cooperative outcome for the party who has the incentive problem.

Lee and Xie (2009) theoretically extend Kandori's (1992) argument to the development of trust and reciprocity among anonymous, randomly matched players in the infinitely repeated trust game and provide sufficient conditions that support a social norm of trust and reciprocity as a sequential equilibrium in the absence of reputational information. The trust game experiment we report on in this paper satisfies the Lee and Xie conditions in all treatments, so that in the absence of any information about one's randomly determined opponents, a social norm of trust and reciprocity may be sustained by the threat to move to a contagious wave of distrust and confiscation. However, we also explore the notion that some information about opponents' prior behavior may help to sustain social norms of trust and reciprocity, as such information makes it easier for players to discern player types thus enabling reputational considerations.

There are several experimental papers on repeated trust games that relate to this study. Bolton et al. (2004) report on an experiment that evaluates the effectiveness of electronic reputation mechanisms. A trust game with binary choices (buyer-seller game) is played repeatedly for 30 periods in each session. They compare the results from three treatments: a *stranger market*, where individual buyers and sellers meet no more than once and the buyer has no information about the seller's transaction history; a *feedback market*, which has the same matching rule as the stranger market and provides the seller's histories of shipping decisions to the buyer; finally, a *partners market*, where the same buyer-seller pairs interact repeatedly in every period. Not surprisingly, transaction efficiency, trust and trustworthiness (reciprocity) are smallest in the stranger market, greater in the feedback market, and greatest in the partners market.

Brown and Zehnder (2007) conduct an experiment in which they use trust games to study the effect of 1) a public credit registry and 2) relationship banking in a competitive market. The main treatments are whether credit reporting is available or not, and whether the public ID of players is random or fixed; only the latter allows relationship banking. They found that when the relationship banking is not feasible (random ID treatment), the credit market essentially collapses in the absence of credit reporting. However, when bilateral relationships are feasible, as when player IDs are fixed and known, the effect of credit reporting is negligible. Therefore, both credit reporting and relationship banking can significantly improve the performance of credit markets.

Charness, et al. (2009) examine the effect of different kinds of information about trustees. Subjects take turns playing both roles—first mover (investor) or second mover (trustee)—in a finitely repeated version of the trust game we propose in this paper. First movers either receive information on the history of return behavior by their matched trustee or on the history of invest (trust) decisions by their matched trustee when that trustee was in the investor (first mover) role. They find that both types of histories can significantly increase trust relative to the absence of such information.

Finally, Engle-Warnick and Slonim (2006) examine whether and how the exogenously determined length of past relationships affects trust and trustworthiness in new relationships. Participants in their experiment play several supergames. Each supergame consists of a sequence of periods of play of the trust game by the same two players (fixed matches). The lengths of these supergames were drawn prior to the first session. The treatments focus on whether initial sequences of short- or long- supergames impacts on the extent of trust and trustworthiness found in subsequent supergames. They find that initial short-supergame relationships have a negative impact on both trust and trustworthiness in the relationships that immediately follow, while longer-lasting relationships have the opposite effect. As subjects gain experience, the effect declines for trustworthiness (reciprocity) but not for trust.

These papers differ from this paper in several significant ways. Since Bolton et al. (2004), Brown and Zehnder (2007) and Charness et al. (2009) investigate a finitely repeated game and in Engle-Warnick and Slonim (2006) the same players interact in fixed matches in each supergame, none of these studies can rationalize trust and trustworthiness as an equilibrium phenomenon among anonymous, randomly matched players who have no information about the history of play of their partners as is the case in our study.[2] Thus they do not address one of the main questions we pose here: whether the mechanism that supports trust and reciprocity comes about through community-wide enforcement (fear of a contagious wave of distrust and confiscation) or from the provision of information on individual behavior (that affects the behavior of both the observed and those deciding whether to trust). Furthermore, the Bolton et al. (2004), Brown and Zehnder (2007) and Charness et al. (2009) study only the case where information is freely provided, and examine how variations in the content of information matter for sustaining cooperative outcomes as we do as well. However, we go a step further and (in one treatment) consider how behavior is affected if information on individual behavior is costly and trustees don't know whether information about them has been purchased or not. This asymmetric information treatment enables us to consider whether the availability of (costly) information (rather than its content) may suffice to sustain cooperative behavior.

Finally, this paper is also related to the literature exploring the historic development of economic institutions in fostering trade among strangers such as the analysis on medieval trade by Greif (1989, 1993) and Milgrom et al. (1990). These papers model a large number of traders who are randomly paired with each other in each period. Each pair is presumed to play a game similar to the trust game, where one party has an incentive to cheat the other by supplying goods of inferior quality or reneging on promises to make future payments. In this literature, institutions are seen as a way of avoiding the inefficiency of noncooperative equilibria. Greif (1989) and Milgrom et al. (1990) argue that the exchange of information on the identity of cheaters or the development of a mechanism which strengthen the power of enforcement can help sustain cooperation.

2 Many experimental studies find that trust and reciprocity prevail under the conditions of complete anonymity and one-shot interaction. As these behaviors are inconsistent with all participants being payoff maximizers, they are often explained by psychological factors such as fairness, altruism, and inequality aversion etc. See, e.g., Berg et al. (1995), Bolton and Ockenfels (2000), Fehr and Schmidt (1999); Camerer (2003) provides a survey.

3. The Model

We briefly describe the model and its predictions for our experimental design. We adopt the notation of Lee and Xie (2009). The set of players $N = \{1, 2, \ldots, 2n\}$ is partitioned into two sets of equal size, the set of investors $N_I = \{1, 2, \ldots, n\}$ and the set of trustees $N_T = \{n+1, n+2, \ldots, 2n\}$. In each period, each investor is matched with a trustee according to the uniform random matching rule, and they play the binary trust game as a stage game. This procedure is infinitely repeated, and each player's total payoff is the expected sum of his stage game payoffs discounted by $\delta \in (0,1)$.

The trust game we study is depicted in Figure 1.[3] At the beginning of the game, the investor is endowed with an amount $a \in (0,1)$. If the investor decides not to invest, the game ends. The investor's payoff is a (the value of his outside option) and the trustee's payoff is 0. If the investor chooses to invest his endowment, this choice yields an immediate gross return of 1, but the division of this gross return is up to the trustee, who moves second. If the investor has invested, the trustee decides whether to keep all of the gross return for a payoff of 1 for himself and 0 for the investor or to return a fraction $0 < b < 1$ to the investor, earning a payoff of $1 - b$ for himself. Throughout we shall assume that $0 < a < b < 1$.

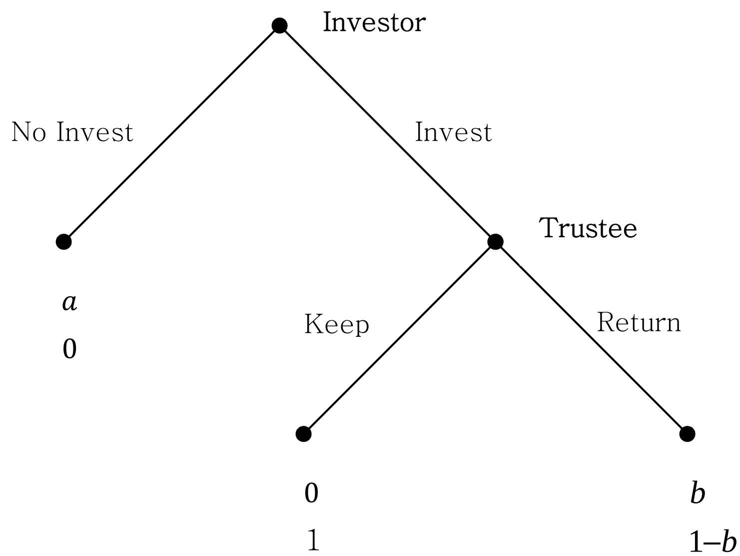

Figure 1: The Trust Game

If the game is played once, the unique subgame perfect equilibrium is for the investor not to invest as the trustee will always choose to play Keep. But since $a < 1$, this equilibrium is not efficient. The efficient outcome, where the investor invests and the trustee returns, *can* be achieved under the conditions of the "contagious equilibrium" of the infinitely repeated game, even if players are anonymously and

[3] In the trust game we study, both players have binary choice sets, a simplification necessary for the theoretical analysis that follows.

randomly re-matched after each period. We now turn to characterizing this contagious equilibrium.

3.1 Contagious equilibrium

Define the action No Invest as a "defection" by an investor and the action Keep as a defection by a trustee. Define *d-type* players as those whose history includes a defection either by themselves or by any of their randomly assigned partners. Otherwise, players are defined as *c-type* (cooperative) players.

Definition: *The "contagious strategy" is defined as follows: An investor chooses Invest if she is a c-type and No Invest if she is a d-type. A trustee chooses Return if he is a c-type and Keep if he is a d-type.*

The idea of the contagious strategy is that trust applies to the community as a whole and cannot be applied to individuals because of random anonymous matchings. Therefore, a single defection by a member means the end of trust in the whole community and a player who experiences dishonest behavior starts defecting against all of his opponents (Kandori, 1992). It is shown below that we can define payoffs for the trust game which allow trust and reciprocity to be a sequential equilibrium for any finite population.

To show that the contagious strategy constitutes a sequential equilibrium, it is sufficient to show that one-shot deviations are unprofitable after any history. In particular, Lee and Xie (2009) provide these conditions in the following lemma which puts constraints on investors' and trustees' incentives not to deviate from the contagious strategy both on-the-equilibrium-path and off-the-equilibrium-path.

Before stating the lemma, we first introduce the terms $f(\delta)$ and $g(\delta)$ which are functions of the period discount factor δ --for details of the construction of these terms see Appendix A. Conceptually, $f(\delta)$ represents the discounted sum of expected future payoffs – the gain-- to a trustee from *not* initiating a contagious wave of defection when all the other players in the community are c-types, and $g(\delta)$ represents the gain to a d-type trustee from deviating from defection (i.e., resuming to play Return) given that there is just one d-type investor and one d-type trustee (himself) in the current period. Thus, $f(\delta)$ and $g(\delta)$ are the discounted, expected payoffs to a trustee from avoiding triggering or slowing down the contagious strategy in the current period in different states of the world (i.e., when there are different numbers of d-type investors and d-type trustees in the community).

Lemma: *The contagious strategy constitutes a sequential equilibrium if*

$$a \geq \frac{n-1}{n} b \qquad (1)$$

and

$$g(\delta) \leq b \leq f(\delta) \qquad (2)$$

Condition (1) controls the investor's incentive to deviate from the contagious strategy off-the-equilibrium-path. Due to the one-sided nature of the incentive problem, Invest is the best response to Return, so the investor has no incentive to deviate on-the-equilibrium-path. Condition (1) requires that a d-type investor defect forever (never go back to Invest), even if she believes there is only one d-type trustee, which is the most favorable situation for investment. The left hand side of inequality (1), a, is the investor's opportunity cost from choosing Invest, and the right hand side of inequality (1) is the expected payoff from Return given that there is only one d-type trustee.

The implication of condition (1) is that the existence of the contagious equilibrium requires a high outside option. For the development of a cooperative social norm, the concept of the contagious equilibrium requires a harsh punishment scheme. Not only are those who deviate from the desired behavior punished, but a player who fails to punish is in turn punished himself (Kandori, 1992). So, an investor must defect forever once she is cheated upon. To control the d-type investor's incentive to start investing again off-the-equilibrium-path, the outside option a must be high enough.

Condition (2) controls the trustee's incentive to deviate from the contagious strategy both on-the-equilibrium-path and off-the-equilibrium-path. Notice that b represents both the trustee's extra payoff from one-period defection and his loss from one-period return at the same time. The first part of condition (2), $f(\delta) \geq b$, requires that the trustee's one-period gain from defection, b, must be less than or equal to the gains from *not initiating* a defection in the current period, $f(\delta)$. Thus, a trustee will not start a defection in the current period. The second part of condition (2), $g(\delta) \leq b$, implies that the period loss from attempting to slow down a contagious wave of defection, b, must be greater than the gains from slowing down the contagion when there are already other d-type players in the community. So this restriction controls the trustee's incentive not to deviate off-the-equilibrium-path. Finally, to show that there always exists a b between $g(\delta)$ and $f(\delta)$, Lee and Xie (2009) show that $g(\delta)$ is less than $f(\delta)$ for any δ greater than 0 given any finite population size. Intuitively, when the trustee in consideration is the only d-type player in the community, the trustee's payoff from slowing down the contagious procedure (i.e., $f(\delta)$) is largest, since the contagious procedure stops completely for the current period if the d-type trustee chooses not to defect. This payoff, however, becomes smaller when there are other d-type players in the community.

The lemma above is used in the proof of the following theorem, which states that we can find values for a and b in the trust game that satisfy the sufficient conditions of the lemma.

Theorem (Lee and Xie 2009): *Consider the model described above where $2n \geq 4$ players are randomly paired each period to play the infinitely repeated trust game. Then for any δ and n, there exist a and b such that (i) $0 < a < b < 1$; and (ii) the contagious strategy constitutes a sequential equilibrium in which (Invest, Return) is the outcome in every period along the equilibrium path.*

While other repeated game equilibria may exist under these conditions, the contagious equilibrium where (Invest, Return) is the outcome in every period is the most efficient of these equilibria, and therefore the focus of our analysis.

3.2 Equilibria when information about trustees is available

In this paper, we consider as an alternative to anonymous, community-wide enforcement, environments where information on an individual trustee's past history of play can be observed by an investor prior to the investor making a decision to invest or not. We focus on the case of one-sided information flow (investors only view information on trustees and not vice versa) as this seems most appropriate for the trust game with its one-sided incentive problem, and because this information set-up also follows that of many real-world examples, e.g., credit markets. Specifically, we consider two different trustee histories that may be available to the investor: 1) "minimal information", where the investor observes only the action chosen by the trustee in the prior period (Keep, Return, or no choice) and 2) "full information", where the trustee's past history of decisions in all prior random matches with investors is revealed to the investor with whom the trustee is currently matched. We further consider an environment where the full information is available to investors only at some cost, $c > 0$. The following propositions apply to these environments with costless or costly information on the trustee's history of play.

Proposition 1: *When information on the past behavior of trustees is free and full, the contagious strategy is not an equilibrium strategy.*

Proof: Consider the case where a d-type investor meets a c-type trustee in the current period. Under full information, the d-type investor can identify the trustee as a c-type player. According to the contagious strategy, the trustee should choose Return-given-Invest, and the investor, being a d-type should choose No Invest. However, given the trustee's strategy, the investor has an incentive to choose Invest since she can not only gain $b-a$ in the current period but she can also slow down the contagious process by not changing the current c-type trustee into a d-type trustee.

If the contagious strategy is no longer an equilibrium strategy, a natural question that arises is what *is* an equilibrium strategy when information is available on trustees? We propose the following:

Proposition 2: *When information on the history of a trustee's play is free and full and $\delta \geq b$, there exists an equilibrium in which the trustee continues to play the contagious strategy but investors play a strategy that is conditional on the information revealed about the trustee. Specifically, an investor chooses Invest* *if the trustee's history of play reveals the trustee to be a c-type and the investor chooses No Invest otherwise.*

Corollary 1: *When minimal information is provided freely, the strategy described in Proposition 2 is an equilibrium strategy only for a knife-edge condition* $\delta = b$ *.*

Proof: See Appendix A.

Proposition 2 and Corollary 1 together indicate that if investors condition their investment decision on information about a trustee's prior behavior, an equilibrium involving complete trust and reciprocity will be easier to sustain in the case of full information than in the case of minimal information. Intuitively, the discount factor cannot be too high in the equilibrium under minimal information, since a d-type trustee will have an incentive to attempt to remove his bad reputation by engaging in one-shot good behavior in the current period so as to appear to be a c-type and attract investment in future periods. This problem does not arise in the case of full information because in that case it is impossible for a d-type trustee to change his type as perceived by investors.

Our final proposition applies to environments where the investor may choose to purchase full information about a trustee's past history of play at a per period cost of $c > 0$. The information purchase decision is private information; the trustee does not know whether or not his matched investor has chosen to purchase information. For this environment, we propose the following asymmetric equilibrium: only a fraction of investors choose to purchase information (or equivalently, investors choose to purchase information with some probability); a fraction of trustees always choose Return and the remaining trustees always choose Keep. For some intuition as to why there is a mixture of behavior in the equilibrium of this environment, suppose that all trustees always chose Return. Then investors would not need to purchase information, since the value of information is to distinguish trustees with a good reputation from those with a bad reputation. However if none of the investors purchased information yet they still invested with a positive probability, then trustees would have strong incentives to defect. Therefore, investors should play a mixed strategy with regard to the information purchase decision, provided the cost is small enough.

Proposition 3: *When information on the history of trustees' play is full and not too costly and* $\delta \geq b$*, there exists an equilibrium characterized by a vector of probabilities,* (q, γ, p)*, where investors purchase information with probability* $q < 1$*, choose Invest if this information reveals the trustee to have always chosen Return, and choose No Invest otherwise. Investors who do not choose to purchase information choose Invest with probability* γ*. Fraction* $p < 1$ *of trustees always choose Return and fraction* $1 - p$ *always choose Keep. The most efficient such equilibrium obtains where* $\gamma = 1$*.*

Corollary 2: *When full information about trustees is available for purchase there also exists an inefficient, pure strategy equilibrium where investors never purchase information or choose Invest and no trustee chooses Return.*
Proof: See Appendix A.

Proposition 3 says that when full information is available and not too costly (the cost conditions are given in the proof of Proposition 3), there exists an equilibrium in which only some investors purchase information about trustees and, consequently some trustees play Keep. Hence, an implication of making full information costly is that trust and reciprocity may be lower than when full information is costless. While there are many equilibria with positive levels of trust and reciprocity when information is costly (these are indexed by γ), we focus our analysis (as we have done previously) on the most efficient of these equilibria, which obtains when investors choosing not to purchase information always choose Invest ($\gamma = 1$).

Of course, as stated in Corollary 2, the inefficient equilibrium where all investors choose not to purchase information and No Invest and all trustees choose Keep always remains an equilibrium possibility. Thus, there is an empirical question as to whether information will be purchased in the costly information environment. We examine the latter question as well as all of our other theoretical predictions by designing and analyzing results from a laboratory experiment. We now turn to this exercise.

4. Experimental Design

Our main treatment variable concerns the information available to investors in advance of their investment decision. We investigate four different "information treatments." In the first, "no information" treatment (henceforth referred to as "No"), investors only know their own history of play and payoff in each period. Nevertheless, in this environment, trust and reciprocity (the play of Invest and Return) can be supported under random anonymous matching via the contagious strategy. In the second, "minimal information" treatment (referred to as "Min"), investors are informed of the prior-period decision of their current paired trustee, i.e., whether that trustee chose Keep or Return in the prior period of the current supergame, in the event the trustee had the opportunity to make a choice in the prior period; if the trustee did not have an opportunity to make a decision in the prior period, the information reported to the investor is No Choice. In the third, "information" treatment (referred to as ``Info''), investors are told the frequencies with which their currently matched trustee chose Keep or Return out of the total number of opportunities the trustee had to make either choice over *all prior periods* of the current supergame – called the Keep or Return ratios. The latter information is all that is necessary to label a trustee as either a c- or d-type, consistent with Propositions 1-3. In addition, investors were also shown the trustee's actual, period-by-period history of play (Return, Keep or No Choice) for up to 10 prior periods of the current supergame.[4] Finally, in the fourth, "costly information" treatment (denoted as Cost), investors are not automatically provided with information on their paired trustee's previous choices as in the Info treatment; instead, the investors can choose to purchase the same information provided in the Info treatment at a small cost.

The parameterization of the stage game used in all experimental sessions is given in Figure 2. In this figure, the terminal nodes of the tree give the number of points each type of subject earned under the three possible outcomes for each stage game played. This parameterization of the game was chosen to be consistent with our theoretical assumption that $a = 35 < b = 45 < 100$ and also satisfies the conditions (1) and (2) of the Lemma in the prior section given the choice of $n = 3$ pairs of players and the induced period discount factor $\delta = .80$ used in all of our experimental sessions.[5] While other parameterizations are possible, we chose a parameterization that is *not* at the boundary of the conditions (1)-(2), but instead

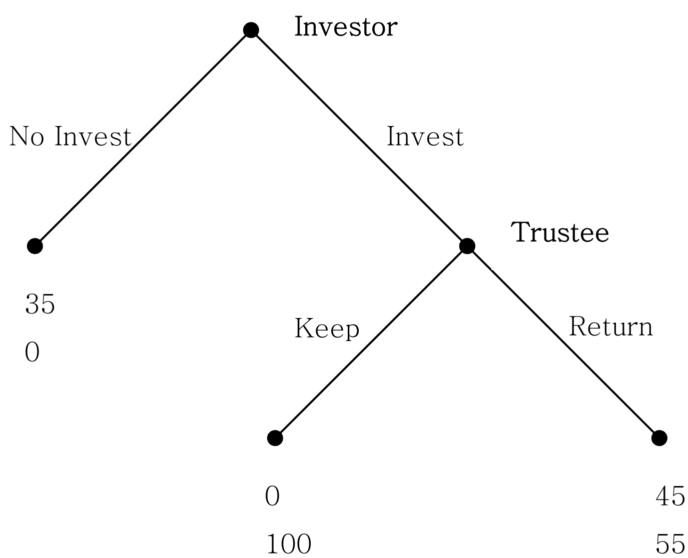

Figure 2: Stage Game Parameterization Used in the Experiment.

lies well within the region supporting trust and reciprocity among randomly matched players.[6] The cost of

[4] While we limited the period-by-period history of actions about a trustee to a maximum of 10 prior periods, the reported frequencies with which a trustee played Keep or Return were ***for all periods of the current supergame*** and this fact was made clear to subjects. Note further that the expected duration of a supergame, given our choice of $\delta = .80$, is just 5 periods.

[5] A computer program used to verify condition (2) is available upon request.

[6] In many experimental implementations of trust games, the trustee is given a positive endowment so as to avoid the possibility that the investor feels compelled (out of some sense of fairness) to invest. While this may be an issue in one-shot games, it seems less relevant in our repeated random-matching trust game, where (as we discuss below) all

purchasing information in the Cost treatment was set at 2 points, and satisfies restrictions given in the Proof of Proposition 3. The experiment was programmed and conducted using the z-Tree software (Fischbacher, 2007).

All of our experimental sessions involve groups of size $2n = 6$. We chose to work with groups of 6 subjects for several reasons. First, and most importantly, condition (1) for the existence of the contagious equilibrium in the trust game (where $a < b$) is more difficult to satisfy when n is large. On the other hand, we did not want the expected frequency of repeat matchings to be as high as in the minimal group size of 4. Finally, we wanted to give the contagious equilibrium a chance to work; it is well known that the contagious equilibrium involving complete trust and reciprocity can collapse due to noise or trembles, and such noise is likely to increase with the size of the group.[7]

An indefinitely repeated *supergame* was implemented as follows. At the start of each supergame, subjects were randomly assigned a role as either the investor or trustee and they remained in that role for all rounds of the supergame.[8] This design gave subjects experience with playing both roles across many supergames. In each period of the supergame, the 3 investors and 3 trustees were randomly and anonymously matched with one another for a single play of the stage game with all matchings being equally likely.[9] At the end of play of the stage game, the results of the game were reported to each pair of subjects and then a 10-sided die was rolled. If the die came up 8 or 9, the supergame was declared over; otherwise the game continued on with another period. Subjects were randomly rematched before playing the next period, though they remained in the same role in all periods of that supergame. We told them that we would play a number of "sequences" (i.e., indefinitely repeated supergames) but did not specify how many. For transparency and credibility purposes, we had the subjects take turns rolling the 10-sided die themselves and calling out the result. Our design thus implements random and anonymous matching, a discount factor, $\delta = .80$, and the stationarity associated with an infinite horizon.

---

players are equally likely to be assigned the role of investor or trustee at the start of each new supergame, and are paid for all periods of all supergames played. Therefore, each subject in our design is effectively given the same "endowment" in expected terms. Related to the real life examples that motivate our paper, e.g., borrowing in credit markets, it also seems more reasonable to assume that only the first mover (bank) has an outside option (endowment); if a transaction does not occur, then the bank keeps its money while the borrower earns 0. Finally, according to the theory we are testing, giving the trustee a positive endowment does not matter for any of our equilibrium predictions. For all of these reasons we did not provide an endowment to the trustee.

[7] Camera and Casari (2009) offer a similar justification for their choice of a group size of 4. Duffy and Ochs (2009) look at groups of size 6 as well as larger groups of size 14 and find cooperation rates under random anonymous matchings are twice as high on average in groups of size 6 as compared with groups of size 14.

[8] In the instructions (Appendix B) we use neutral word "First Mover" for investor, "Second Mover" for trustee, and "sequence" for indefinitely repeated supergame. We also use "A" "B" "C" "D" to denote the investor and trustee's choices. See Appendix for instructions.

[9] This is the same matching protocol used by Duffy and Ochs (2009). Camera and Casari (2009) use a matching protocol wherein no two subjects are matched to play more than one supergame. In all treatments of our design, the assignment of roles (Investor, Trustee) was randomly determined at the start of each new supergame thereby distinguishing one supergame from the next.

In all of the informational mechanisms discussed above, information on the trustee' behavior in previous supergames does not carry over when a new supergame begins. In the treatments where information is available, it is always available from the start of the *second* period of each supergame.

We used a within-subjects design in all sessions. Subjects begin to play under one information condition and were switched to the second condition (and then to the third in some sessions). In practice, there are at least 30 periods under each information condition in all sessions – see Table 1 below. When the total number of periods under one information condition exceeded approximately 30 periods, we made that supergame the last supergame played under that treatment condition. Subjects were only informed of the change in an information condition when the switch took place, i.e., they did not know that a change was coming or our rule for implementing a change.

We have in total 14 experimental sessions which we have divided up into two main sets. The first set of eight sessions examines whether providing investors with no information or minimal information on a trustee's prior behavior affects trust and reciprocity. We conducted four "No_Min" sessions (sessions that began with "no information" and later switch to "minimal information", and four "Min_No" sessions following the opposite treatment order. The other set of six sessions investigates the effect of a longer history of information regarding trustees' prior behavior on trust and reciprocity and whether the possibility to purchase that "full" history of information at a small cost affects the frequency of trust and reciprocity. We conducted three "No_Info_Cost" sessions and three "Info_No_Cost" sessions. (Recall No means no information, Info means information and Cost means costly information). We reversed the order of the first two treatments to examine whether the treatment order matters. The Cost treatment is always the *last* treatment in this second set of sessions, as we wanted subjects to have experience with the full information "Info" treatment before they faced a decision as to whether to purchase that same amount of information at a small cost. The Instructions used in the "Min_No" and "Info_No_Cost" sessions are provided in Appendix B (instructions for the other treatment orderings are similar).

The motivations for this experimental design follow from our theoretical model. First, under our parameterization of the model the contagious strategy supports a social norm of full trust and reciprocity among randomly matched anonymous players when no information on trustees is available. However, we cannot exclude other equilibria, e.g., the social norm of no trust and no reciprocity is another one. Thus it remains an empirical question as to whether community-wide enforcement suffices to support a social norm of trust and reciprocity and whether different informational mechanisms can help select different social norms.

Second, since the collection, storage and dissemination of information is always costly for a society, a question of practical interest is how much reputational information is enough in order to significantly enhance the frequencies of trust and reciprocity. Thus we are not only interested in examining whether there are differences when information is available or not, but also whether any such differences depend on the quantity of information provided. That is one motivation for why we consider both the Min and Info treatments. A second motivation comes from Proposition 2 and its Corollary which predict that full information on trustees can sustain an equilibrium with full trust and reciprocity under more general conditions than the case of minimal information on trustees. Notice further, that the information reported to subjects in the Min treatment nests that of the No treatment while the Info treatment nests that of the Min treatment.

Finally, the Cost treatment recognizes that information on trustees' past history would be costly to gather and that the costs of gathering such information would likely be paid by the information consumers, i.e., the investors. The Cost treatment thus addresses the role of costly reputational information on trust and reciprocity—a more empirically relevant setting. Importantly, trustees are *not informed* as to whether their paired investor purchased information about them or not and this asymmetry of information is public knowledge. Thus, on the one hand, if some fraction of investors choose to purchase information about trustees (and act according to the content of that information) their decisions can potentially provide a positive externality to the whole community due to the anonymity of matching and information purchase decisions. On the other hand, if trustees believe that some fraction of investors will not purchase information, they may behave similarly as in the No information treatment. Our Proposition 3 predicts this kind of mixed equilibrium.

All subjects were recruited from the undergraduate populations of the University of Pittsburgh and Carnegie Mellon University. No subject had any prior experience participating in our experiment. Subjects were given $5 for showing up on time and completing the experiment and they were also paid their earnings from all periods of all supergames played. Subjects accumulated points given their stage game choices (points are shown in Figure 2, the cost of information is set 2 points). Total points from all periods of all supergames were converted into dollars at a fixed and known rate of 1 point = ½ cent.

5. Results

Table 1 provides basic characteristics of all sessions, specifically the number of supergames for each treatment, 1, 2 or 3 of the session, the total number of periods played in each of those treatments, as well as the average payoff earned by subjects for the session and per period. As Table 1 reveals, the two or three treatments of each session involved roughly similar numbers of periods. Subjects earned on average, $17.07 ($13.36 for the first and $22.02 for the second set of sessions) in addition to their $5 show-up fee. The first set of 8 sessions all finished within 1.5 hours and the second set of 6 sessions all finished within 2 hours. In the following subsections, we first report the results from the first and second set of sessions respectively, and then we analyze how investors made use of the various amounts of information about trustees across both sets of sessions.

Session	No. of Supergames Treat1/Treat2/Treat3	Total No. of Periods Treat1/Treat2/Treat3	Avg. Payoff	Avg. Payoff per Period
No_Min1	7 / 9	38 / 46	$12.12	$0.14
No_Min2	5 / 8	34 / 37	$14.39	$0.20
No_Min3	8 / 2	35 / 37	$14.80	$0.21
No_Min4	9 / 8	39 / 35	$13.63	$0.18
Min_No1	9 / 8	44 / 37	$17.11	$0.21
Min_No2	5 / 11	38 / 41	$12.55	$0.16
Min_No3	9 / 7	34 / 38	$12.75	$0.18
Min_No4	13 / 5	41 / 38	$9.51	$0.12
No_Info_Cost1	6 / 7 / 9	42 / 37/ 40	$25.65	$0.22
No_Info_Cost2	11 / 6 / 7	35 / 35 / 36	$22.98	$0.22
No_Info_Cost3	7 / 8 / 8	34 / 32 / 31	$20.75	$0.21
Info_No_Cost1	5 / 9 / 7	44 / 43 / 32	$25.90	$0.22
Info_No_Cost2	8 / 4 / 8	35 / 43 / 35	$17.49	$0.15
Info_No_Cost3	7 / 9 / 7	34 / 33 / 34	$19.36	$0.18
Average	N/A	N/A	$17.07	$0.19

Table 1: All Experimental Sessions

5.1 No_Min and Min_No Sessions

We first analyze whether a social norm of trust and reciprocity is possible when there is no information available on trustees and compare that with the case where "minimal" information on trustees' prior-period action –Keep or Return-- is freely provided. We begin with a within-subject analysis and then move to a between-subject analysis.

Tables 2.1, 2.2 and 2.3 report respectively on the aggregate frequencies of 1) Invest, 2) Return conditional on investment, and 3) the combined frequencies of Invest-and-Return for the four No_Min and four Min_No sessions.[10] To aid in reading these tables, where the treatment orders differ, the clear (unshaded) areas indicate average frequencies for the No treatment and the grey (shaded) areas indicate average frequencies for the Min treatment. These tables provide support for our first experimental finding:

Finding 1: *A social norm of full trust and reciprocity (100% Invest-and-Return) is not sustained in the absence of information on trustees' behavior, contrary to the theoretical possibility described by the contagious strategy.*

[10] We will sometimes equivalently refer to Invest as "trust," Return-given-Invest as "reciprocate" and Invest-and-Return as "trust and reciprocity."

	1st treatment	2nd treatment
No_Min1	0.333	0.362
No_Min2	0.618	0.793
No_Min3	0.476	0.964
No_Min4	0.521	0.676
Avg. of No_Min	0.487	0.699
Min_No1	0.674	0.865
Min_No2	0.561	0.325
Min_No3	0.333	0.746
Min_No4	0.301	0.096
Avg. of Min_No	0.467	0.508

Table 2.1: Frequency of Invest

	1st treatment	2nd treatment
No_Min1	0.526	0.760
No_Min2	0.587	0.841
No_Min3	0.780	0.972
No_Min4	0.525	0.831
Avg. of No_Min	0.605	0.851
Min_No1	0.820	0.854
Min_No2	0.797	0.725
Min_No3	0.588	0.706
Min_No4	0.649	0.364
Avg. of Min_No	0.714	0.662

Table 2.2: Frequency of Return-Given-Invest

	1st treatment	2nd treatment
No_Min1	0.175	0.275
No_Min2	0.363	0.667
No_Min3	0.371	0.937
No_Min4	0.274	0.562
Avg. of No_Min	0.296	0.610
Min_No1	0.553	0.739
Min_No2	0.447	0.236
Min_No3	0.196	0.526
Min_No4	0.195	0.035
Avg. of Min_No	0.348	0.384

Table 2.3: Frequency of Invest-and-Return

While the payoffs for the game were chosen so that the contagious strategy supports an equilibrium of full trust and reciprocity, other equilibrium possibilities cannot be ruled out, for example zero trust and reciprocity by all players remains an equilibrium strategy. What we find is that the frequency of Invest-and-Return (trust and reciprocity, Table 2.3) averages around one-third across all sessions of the No treatment. While this is far from 100%, it is also significantly different from zero (two-tailed t test, $p<.01$). Notice also that the observed frequencies of Invest (trust), Return-given-Invest (reciprocity) are also less than 100% but significantly different from zero in the No treatment (two-tailed t test, $p<.01$).

Our next finding makes use of our within-subjects design to compare behavior with and without minimal information.

Finding 2: *The free provision of minimal information to investors on the prior-period action of their matched trustee leads to a significant increase in trust and reciprocity but only if the provision of this information follows the treatment in which investors receive no information about trustees.*

For the No_Min sessions, the provision of minimal information about the trustee's prior-period play in the second half of the session leads to significantly larger frequencies of Invest, Return, and Invest-and-Return relative to the No treatment (Wilcoxon signed ranks test, $p=0.0625$ for all three tests). However, none of the frequencies of Invest, Return-given-Invest, and Invest-and-Return is significantly different from the levels observed in the No treatment when the minimal information is provided in the first half of the session –the Min_No sessions (Wilcoxon signed ranks test, $p>0.4$ for all three tests).

To explore this finding further, we also report the results of a *between-subjects* session-level analysis using all Min_No and No_Min sessions. In particular, we compare 1) the first treatment of these 8 sessions (i.e., the No treatment of the 4 No_Min sessions and the Min treatment of the 4 Min_No sessions); 2) the No treatment of all 8 sessions; and finally, 3) the Min treatment of all 8 sessions.

Since the subjects were not informed of the second treatment until the change in treatment mid-way through a session, we may regard the *first* treatment of each session as an independent observation that was not influenced by other information treatments. We find that regardless of whether the first treatment is No or Min information, the session-level frequencies of Invest and Invest-and-Return are not significantly different from one another. However, the frequency of Return-given-Invest is significantly higher when minimal information is available in the first treatment (Robust Rank Order test, $p<0.05$). The latter finding implies that the provision of information has a more significant effect on the behavior of trustees than on the behavior of the recipients of minimal information-- the investors!

For the between-subject analyis using only the No treatment session-level data, we find that there are no significant differences in the frequencies of Invest, Return-given-Invest, and Invest-and-Return when the No treatment is in the first half of the No_Min sessions or in the second half of the Min_No sessions.

However, for the between-subjects analysis using only the Min treatment session-level data, all frequencies of Invest, Return-given-Invest, and Invest-and-Return are significantly higher when the minimal information is provided in the second half of the No_Min sessions than in the first half of the Min_No sessions (Robust Rank Order test, $p<0.05$ for all three tests). These findings strongly suggest that when minimal information is provided to subjects who have suffered from the absence of reputational information as in our No_Min sessions, they learn to use the minimal information more effectively than subjects who begin interacting with minimal information and then lose access to that information as in our Min_No sessions.

5.2 No_Info_Cost and Info_No_Cost Sessions

We next report results from the second set of sessions which involved three different treatments. Tables 3.1, 3.2 and 3.3 report the frequencies of Invest, Return-given-Invest, and Invest-and-Return respectively from the second set of sessions. Again, the clear (unshaded) areas show frequencies from the

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	0.516	0.892	0.983
No_Info_Cost2	0.714	0.857	0.824
No_Info_Cost3	0.667	0.813	0.871
Avg. of No_Info_Cost	0.632	0.854	0.893
Info_No_Cost1	0.848	0.659	0.927
Info_No_Cost2	0.829	0.132	0.352
Info_No_Cost3	0.775	0.515	0.471
Avg. of Info_No_Cost	0.817	0.435	0.583

Table 3.1: Frequency of Invest (all periods)

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	0.800	0.970	0.992
No_Info_Cost2	0.680	0.867	0.775
No_Info_Cost3	0.603	0.923	0.889
Avg. of No_Info_Cost	0.694	0.920	0.885
Info_No_Cost1	0.982	0.741	0.989
Info_No_Cost2	0.931	0.471	0.838
Info_No_Cost3	0.911	0.471	0.625
Avg. of Info_No_Cost	0.941	0.561	0.817

Table 3.2: Frequency of Return-Given-Invest (all periods)

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	0.413	0.865	0.975
No_Info_Cost2	0.486	0.743	0.639
No_Info_Cost3	0.402	0.750	0.774
Avg. of No_Info_Cost	0.434	0.786	0.796
Info_No_Cost1	0.833	0.488	0.917
Info_No_Cost2	0.771	0.062	0.295
Info_No_Cost3	0.706	0.242	0.294
Avg. of Info_No_Cost	0.770	0.264	0.502

Table 3.3: Frequency of Invest-and-Return (all periods)

No treatment, the grey areas show frequencies from the Info treatment, and now the light grey areas show frequencies from the Cost treatment (always the 3rd treatment). Recall that in the Info treatment the investor sees the matched trustee's aggregate Ratio of Return in all periods of the current supergame as well as the trustees' actions chosen in up to the 10 most recent periods of the current supergame, and in the Cost treatment such information is available only at a cost of 2 points paid by the investor in advance. Similar to the last subsection, we start with a within-subject analysis and then move to a between-subject analysis.

Notice first that, consistent with Finding 1 for the No_Min and Min_No sessions, the frequencies of trust and reciprocity in the No treatment are significantly different from zero (two-sided t test, $p<0.01$), but the social norm of full trust and reciprocity is not supported in that the aggregate frequency of Invest-and-Return remains low, averaging again around one-third over all sessions.

Comparing the impact of free full information –the Info treatment with the No treatment, we have the following:

Finding 3: *The free provision of a longer history of information to investors on the past behavior of their matched trustee leads to a significant increase in trust and reciprocity relative to the No treatment regardless of whether the provision of information occurs before or after the No treatment.*

Support for Finding 3 is found using a within-subjects analysis of session-level differences in Invest, Return-given-Invest and Invest-and-Return between the No and Info treatments. (one-tailed Wilcoxon signed ranks test, $p<0.05$ for all three tests).

Consider next the Cost treatment where investors are provided with the possibility to purchase information at a cost of 2 points. Consistent with Finding 3, the frequencies of trust and reciprocity in the Cost treatment remain significantly greater than when no information is available (one-tailed Wilcoxon signed ranks test, p<0.05 for all three tests). Furthermore, we also have the following (using a one-tailed Wilcoxon signed ranks test at the 10% significance level):

Finding 4: *The frequencies of Invest (trust) and Invest-and-Return (trust and reciprocity) are not significantly different in the Info treatment and Cost treatment. The frequency of Return-given-Invest (reciprocity) is significantly higher in the Info treatment than in the Cost treatment.*

Finding 4 is consistent with the prediction of our Proposition 3: trustees recognize that when information is costly not all investors will purchase it and consequently they do not play Return as often as in the Info treatment. This finding suggests that trustees' behavior may be more sensitive to the change in information treatments than investors' behavior, an observation that is consistent with the earlier findings from the No_Min and Min_No sessions.

In our design, there is no information available in the first period of a supergame ("sequence") in any treatment, so an alternative way to examine the treatment effect is to compare the frequencies of trust and reciprocity excluding the first periods. We can confirm that Findings 3-4 continue to hold when the first periods are excluded from the data analysis -- the frequencies of Invest, Return-given-Invest, and Invest-and-Return excluding the first periods are given in Appendix C.

Although the frequencies of trust and reciprocity are significantly increased when information is freely provided or available for purchase, one may wonder whether efficiency is enhanced by information provision, especially in the case where information is costly. Table 4 presents the average payoff per period over all subjects under each treatment (taking into account information purchase costs, if any).

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	34	46	49
No_Info_Cost2	41	45	44
No_Info_Cost3	39	44	46
Avg. of No_Info_Cost	38	45	46
Info_No_Cost1	45	39	48
Info_No_Cost2	44	22	29
Info_No_Cost3	43	34	33
Avg. of Info_No_Cost	44	32	37

Table 4: Average Payoff per Period (in points)

Finding 5: *Players' average payoffs are significantly increased when information is freely provided or available for purchase compared with the case where information is absent.*

Support for Finding 5 comes from analysis of the session-level data in Table 4. Using a one-tailed Wilcoxon signed ranks test, we find that p<0.05 for both tests.

We next turn to a between-subjects analysis for our second set of 6 sessions. As before, we treat the first treatment in these 6 sessions as independent observations. Consistent with Finding 3, we find that providing a larger amount of information significantly increases the frequencies of Invest, Return-given-Invest, and Invest-and-Return (one-sided Robust Rank-Order test, p<0.05). This evidence again shows that the amount of information provided about trustees matters for enhancing trust and reciprocity.

We look for order effects by examining the third, Cost treatment of all 6 sessions. We find that none of the frequencies of Invest, Return-given-Invest and Invest-and-Return in the Cost treatment are significantly different across the 6 sessions. This finding implies that there is no significant order effect, that is, whether the subjects experience the No treatment or Info treatment first and then switch to the other one does not significantly affect their behavior in the third Cost treatment. Finally, we also check whether subjects' behavior in the No treatment and Info treatment is different when the same treatment is the first or the second treatment in the session. We find that there is no significant difference for most cases except that the frequency of Invest in the No treatment is marginally significantly higher when the No treatment is the first treatment than when it is the second treatment (Robust Rank-Order test, p<0.1).

Finding 6: *In the second set of experiments, unlike the first set, we do not find strong evidence of treatment order effects.*

Finally, we also examined whether the frequencies of Invest, Return-given-Invest and Invest-and-Return are greater under free, full information as compared with the free, minimal information used in our first set of experiments. In the case where free information (minimal or full) was provided in the *first* treatment of a session, we find that the frequencies of Invest, Return-given-Invest, and Invest-and-Return are all significantly higher in the case of full information (one-sided Robust Rank-Order test, p<0.05 for all three tests). However, when free full or free minimal information are the second treatment of a session, following a first treatment of no information, the same three frequencies are not significantly different from one another (one-sided Robust Rank-Order test, p>0.1 for all three tests). Summarizing, we have:

Finding 7: *The frequencies of Invest, Invest-given-Return and Invest-and-Return under free and full information are greater than or equal to those observed under minimal information.*

Summarizing the results of the paper to this point, we have several important findings. First, absent reputational information on the prior behavior of trustees, a social norm of full trust and reciprocity is not sustained despite the theoretical possibility under the contagious equilibrium. Second, providing minimal information on trustee's prior behavior has an inconsistent effect on Invest, Return-given-Invest and Invest-and-Return relative to case of no information. Third, providing a longer history on trustees' prior behavior has a larger and more consistent effect on trust and reciprocity compared to providing no information or minimal information on trustees' prior-period choices. The effect of the longer history of information is robust regardless of whether that information is provided before or after the No treatment, according to both a within-subjects and between-subjects analysis of session-level data. Finally, making information available at a small cost yields outcomes that are similar to those observed when information is provided automatically and without cost.

5.3 Use of Information

This subsection focuses on how investors' behavior varied with the information they had about trustees' past decisions. We first analyze the effect of no versus minimal information from our first set of experiments. We then examine the effect of minimal versus the larger amount of information provided in our second set of experiments. Finally we compare the case of free versus costly information.

5.3.1 Minimal Information versus No Information

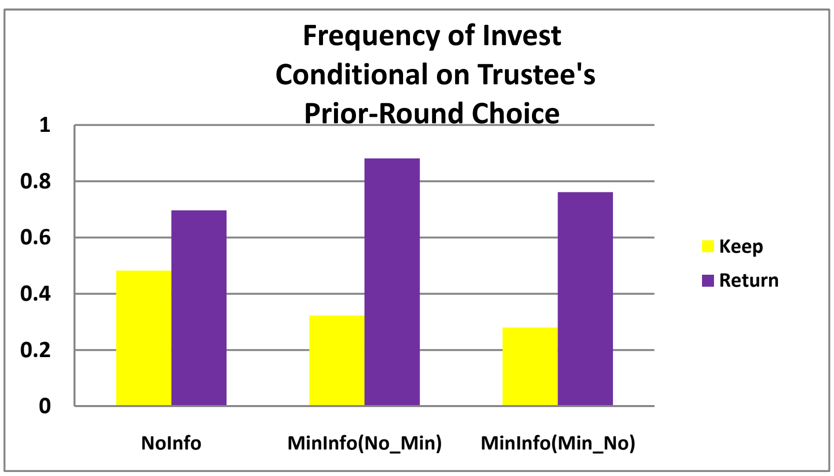

Figure 3: Frequency of Invest Conditional on Trustee's Prior-Period Choice

Figure 3 presents the frequency of Invest conditional on the prior-period choice of the investor's current matched trustee (i.e., Keep or Return) in the No and Min treatments of the No_Min and the Min_No sessions, respectively. In the No information treatment, there is no significant difference in these conditional frequencies of Invest. However in the Min information treatment, the frequency of Invest is significantly higher when the prior period behavior of the trustee is Return than when it is Keep (regardless of the treatment order - Wilcoxon signed ranks test on session-level data p>0.1 in the No treatment and p<.10 in the both Min treatments).

Finding 8: *The frequency of Invest is significantly higher when the matched trustee's prior-period choice is "Return" than when it is "Keep" in the Min treatment but not in the No treatment.*

Finding 8 has to be treated with some caution as the prior behavior of trustees is not an experimental manipulation and may well represent an endogenous response to the prior trust exhibited by investors. Indeed, in the No information treatment we observe higher frequencies of Invest when the current matched trustee returned in the prior period as compared with when that trusted chose Keep, though this difference is not statistically significant. Relatedly, Figure 3 also suggests that investors' behavior is not 100% dependent on the minimal information they do receive about trustees' prior period behavior. Indeed, we observe that investors still invest with a positive probability even if the minimal information shows that their current partner defected (kept) in the prior period, and symmetrically, investors do not fully trust when their partner is revealed to have cooperated (returned) in the prior period. Such behavior is fully consistent with Corollary 1, which states that the conditional investment strategy described in Proposition 2 is *not* an equilibrium strategy in the minimal information environment examined here, where $\delta > b$.

5.3.2 Minimal Information versus Information

In the case of free information on longer histories of trustees' behavior, Proposition 2 states that *there* exists an equilibrium in which the trustee continues to play the contagious strategy but investors play a strategy that is conditional on the information revealed about the trustee. In particular, full trust and reciprocity is an equilibrium possibility in this information environment. In Figure 4 we compare the conditional frequency of Invest decisions by investors in the Min treatment from the first set of experiments with that in the Info treatment from the second set of experiments. For the Min treatment, the conditional frequency is based on whether the investor's currently matched trustee was revealed to have kept or returned in the prior period. For the Info treatment, the conditional frequency is based on whether the investor's currently matched trustee was revealed to have always kept (All Keep) or always returned (All Return) in all prior periods of the current supergame. Figure 4 provides support for the following:

Finding 9: *The frequency of Invest, conditional on a history of "All Return" (Info treatment) is significantly greater than the frequency of Invest conditional on a prior period history of "Return" (Min treatment). There is no significant difference in the frequency of Invest conditional on a history of "All Keep" (Info treatment) or a prior period history of "Keep" (Min treatment).*

Finding 9 is based on the results of a robust rank order test using session-level data, $p \leq .05$. This finding is consistent with the equilibrium described in Proposition 2 (see also Corollary 1), where investors play a strategy that is conditional on the prior history of play of the trustee, but (generally) only in environments where investors have access to a long history of information about the trustee's prior behavior.

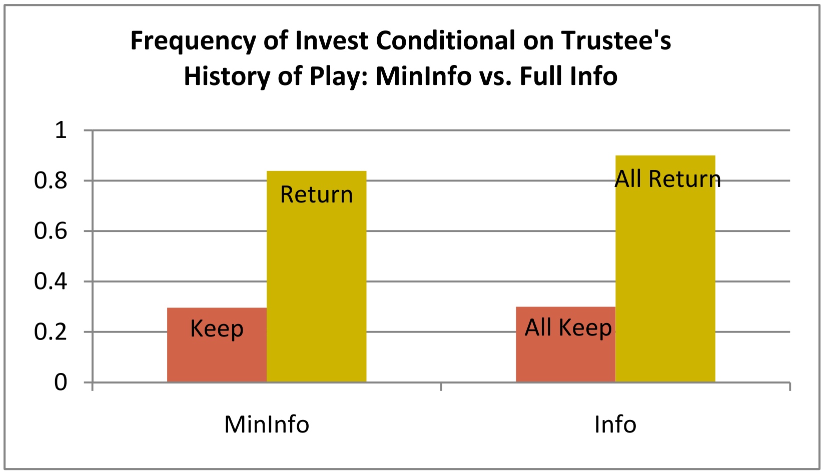

Figure 4: Frequency of Invest Conditional on Trustee's History

5.3.3 (No) Information versus Costly Information

Figure 5 shows the frequency of Invest conditional on the matched trustee's Ratio of Return in the current supergame for the (free) Info and Cost treatment-when information was purchased (top panel) and for the No and Cost treatment-when information was not purchased (bottom panel). Recall that this ratio, (as well as the trustee's history of play in the 10 most recent periods), was revealed to investors in the Info treatment and was available for purchase in the Cost treatment. Here we compare the frequency of Invest in these two treatments when the trustee's Ratio of Return is allocated into 5 non-overlapping bins for the Ratio of Return. Figure 5 provides support for the following finding:

Finding 10: *When investors have information on trustees, frequencies of Invest are generally increasing in the Ratio of Return by the matched trustee.*

We further observe from the top panel of Figure 5 that investors purchasing information do not choose Invest with a high frequency unless the trustee's Ratio of Return is also very high (0.8-1), which is likely owing to the cost these investors had to pay for information about their trustee.

How does behavior in the Cost treatment compare with the equilibrium prediction of Proposition 3?

Table 5 provides some answers.[11] First, the most efficient equilibrium— the benchmark we use in our analysis of the data— has $\gamma = 1$, i.e., investors who do not purchase information choose Invest 100% of the time. As Table 5 reveals, this prediction comes close to holding true in 4 of our 6 sessions – a t-test on the session level data indicates the frequencies are marginally significantly different from 1 (p=.09).

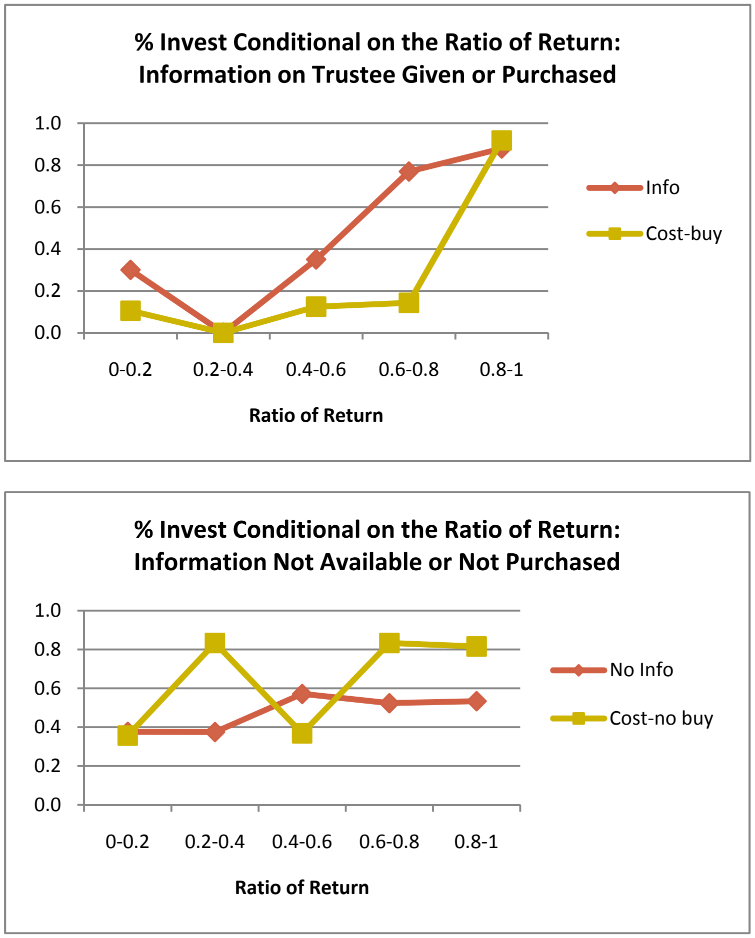

Figure 5: Frequencies of Invest Conditional on Trustee's Aggregate Ratio of Return

Conditional on $\gamma = 1$, the efficient equilibrium prediction calls for investors to purchase information on trustees 56.3% of the time. As Table 5 reveals, the actual frequency of information purchase is significantly lower, averaging 24.49% (t-test p<.01). Combined with the finding that the frequency of trust and reciprocity is not significantly different between the Info and Cost treatment, we conclude that the *mere availability* of information creates a large positive externality for investors. While the overall frequency of Invest in the Cost treatment is 65.55% when information is purchased, as Table 5 reveals this frequency averages 72.48% when information is *not* purchased. Finally, we note that despite the lower-

[11] All analysis here excludes the first periods, where information is not available to purchase.

than-predicted frequency of information purchase by investors, trustees are nevertheless playing Return with high frequency in the Cost treatment. In fact we cannot reject the null hypothesis that the frequency of return given investment is not different from the efficient equilibrium prediction of 94.3% (t-test p>.10).

	% of Invest conditional on No Information Purchase		% of Information Purchase		% of Return given Invest	
Session	Data	Equil. pred.	Data	Equil. pred.	Data	Equil. pred.
No_Info_Cost1	1.000	1	0.226	0.563	0.992	0.943
No_Info_Cost2	0.907	1	0.368	0.563	0.775	0.943
No_Info_Cost3	0.882	1	0.261	0.563	0.889	0.943
Info_No_Cost1	0.905	1	0.013	0.563	0.989	0.943
Info_No_Cost2	0.164	1	0.247	0.563	0.838	0.943
Info_No_Cost3	0.407	1	0.333	0.563	0.625	0.943
Averages	0.725	1	0.245	0.563	0.877	0.943

Table 5: Comparison with Equilibrium Prediction for Cost Treatment

In Table 6 we take a closer look at which subjects bought information and how often they bought information. Table 6 shows the frequency of information purchase decisions by each of the six subjects when that subject was assigned the role of investor in the Cost treatment. In Table 6, frequencies over 0.5 are shown in bold face.

Notice that 9/36=1/4 of subjects purchased information more than 50% of the time (9 entries are in boldface). Approximately 1/2 of subjects (17/36) never chose to buy information. One subject was never assigned as an investor in the Cost treatment. The remaining 1/4 of subjects bought information at least once, but less than 50% of the time.

Session	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
No_Info_Cost1	0/14	0/9	0/7	**21/21**	0/18	0/24
No_Info_Cost2	4/27	**2/2**	5/24	2/6	**19/26**	0/2
No_Info_Cost3	0/8	2/20	**11/11**	**3/5**	**2/2**	0/23
Info_No_Cost1	0/23	N/A	0/4	0/23	1/9	0/16
Info_No_Cost2	0/19	0/14	0/19	4/12	**1/1**	**15/16**
Info_No_Cost3	1/7	5/20	1/3	0/7	**20/20**	0/24

Table 6: The Frequency of Information Purchase for Each Subject

We summarize these results as follows:

Finding 11: *In the Cost treatment, there is heterogeneity in the frequency with which subjects purchase information, with some investors never purchasing information, while others frequently purchase information. On average, the frequency with which investors purchase information is less than half of the predicted level in the efficient equilibrium. Nevertheless, the frequency with which uninformed investors play Invest and the frequency with which trustees play Return-given-Invest are both high, and close to the efficient equilibrium predictions.*

5.4 Individual Behavior

In this section we explore the types of strategies that were adopted by individual investors in making investment decisions in both sets of experiments. Our method is to use a random-effects Probit regression model on pooled data from all sessions of a treatment, where the standard errors have been adjusted to allow for clustering of observations by session number.[12] In all specifications of this model, we always include several variables that capture fixed effects (sequence, period, order). Here, "sequence" refers to the supergame number and "period'' to the period number in a sequence. "Order" is a dummy variable that equals 1 if the session begins with minimal or full information and equals 0 otherwise. Following Camera and Casari (2009), we also introduce a set of regressors (dummy variables) that are used to trace out investor's strategies. In particular, we include a "grim trigger" dummy that equals 1 in all periods of a supergame following the first period of that supergame in which the investor experienced a defection by a trustee and equals 0 otherwise. We also include two "tit-for-tat" dummies that equal 1 in the first and second period respectively following a defection and equal 0 otherwise. Third, we include dummies that represent the information shown to the investor regarding to the current paired trustee's past behavior in the current sequence (Last return, All return, and All no return). "Last return" is a dummy variable that is equal to 1 if information reveals that the current paired trustee returned last period. "All return" has a value of 1 if information reveals that the trustee returned every time the investor invested, and "All no return" has a value of 1 if the information reveals that the trustee never returned given investment. Finally, we create some interactive dummy variables that are the product of these previously defined dummy variables, e.g., Grim and Last return, Buy and All return, Buy and All no return, Buy and Last return. These interactive (multiplicative) dummy variables have a value of 1 only if both components equal 1, and are equal to 0 otherwise. Finally, "Buy information" is a dummy variable that is used only in the Cost treatment, having a value of 1 if the investor chose to purchase information and 0 otherwise.

The Probit Regression results in Table 7 are for the first set of experimental sessions (No_Min) and (Min_No) and reveal several important findings. First, a case can be made that subjects in the No treatment were playing in accord with the predictions of the contagious strategy, as evidenced by the

[12] For this estimation we used the gllamm package in Stata version 11.

significance of the negative coefficient associated with the grim trigger dummy and insignificance of all other left hand side variables (other than the constant term). By contrast, in the Min treatment, subjects were playing according to a mixture of motives, as indicated by the positive and significant coefficient on the last return dummy variable and the negative and significant coefficients on the grim and grim-and-last-return dummy variables. The negative coefficient on the Grim-and-Last return interactive dummy variable implies societal enforcement concerns trump individual reputational information in the Min treatment.

Table 7: Random-Effect Probit Regression on Individual Choice to Invest
(For No_Min and Min_No Sessions)

Dependent variable: 1 = Invest, 0 = No Invest	No Information (Baseline Treatment)	Minimal Information
Constant	.994 (.217)***	-.548 (.541)
Period	-.020 (.012)*	-.009 (.018)
Sequence	.021 (.041)	.038 (.030)
Order	-.366 (.347)	-.289 (.318)
Grim trigger	-1.188 (.335)***	-.765 (.391)**
Tit-for-tat with lag 1	-.050 (.225)	-.305 (.246)
Tit-for-tat with lag 2	-.130 (0.233)	-.229 (.191)
Last return		2.254 (.207)***
Grim and Last return		-1.102 (.498)**
Observations	720	747

Table 8: Random-Effect Probit Regression on Individual Choice to Invest
(For No_Info_Cost and Info_No_Cost Sessions)

Dependent variable: 1 = Invest, 0 = No Invest	No Information (Baseline Treatment)	Full Information	Costly Information
Constant	.761 (.477)	.588 (.853)	4.418 (2.389)*
Period	-.002 (.054)	-.011 (.007)	-.089 (.045)**
Sequence	.030 (.038)	.036 (.069)	-.085 (.105)
Order	-.465 (.370)	-.307 (.515)	-1.662 (.402)***
Grim trigger	-1.227 (.382)***	-1.549 (.210)***	-.969 (.409)**
Tit-for-tat with lag 1	-.973 (.111)***	-1.088 (.622)*	-.519 (.201)***
Tit-for-tat with lag 2	-.419 (.289)	-.515 (.488)	.121 (.403)
All return		.468 (.229)**	
All no return		-1.350 (.449)***	
Last return		.935 (.409)**	
Grim and Last return		1.127 (.337)***	
Buy Information			-.468 (.583)
Buy and All return			1.555 (.659)**
Buy and All no return			-1.062 (.964)
Buy and Last return			.980 (.513)*
Observations	552	528	486

Similar findings emerge from the probit regressions conducted using the second set of experimental sessions, the No_Info_Cost and Info_No_Cost data as revealed in Table 8. In particular, we see that investors in the No treatment continued playing a grim-type strategy, with tit-for-tat also showing up significantly. Investors in the Info and Cost treatments were more likely to choose Invest if the trustees' history showed they had always returned or returned in the last period and were less likely to choose Invest otherwise. By contrast with the Min treatment, in the Info treatment, we observe that the coefficient on the grim-and-last-return interaction dummy is positive; when more information is available, it appears that investors rely less on the global, grim punishment strategy in favor of a more local, individual information-contingent strategy.[13] Notice that buying information by itself does not make an investor more likely to invest, which is consistent with our finding of no difference in the frequencies of investment between those who purchase information and those who do not. We summarize these results as:

Finding 12: *When investors lack information on trustees, they behave in accordance with the contagious strategy, investing less frequently when their history has included a defection (play of Keep by the trustee) than when it has not. When investors have information on trustees, they condition their behavior on this information. With longer histories of information, the importance of the contagious-grim strategy in sustaining trust is diminished in favor of a conditionally trustworthy strategy where investors choose Invest if the trustee's history indicates the trustee is likely to return.*

6. Conclusion

We have studied the development of a social norm of trust and reciprocity among strangers in an indefinitely repeated trust game with random matching. By contrast with the Prisoner's Dilemma game, the trust game captures the one-sided incentive problem that arises in many everyday economic transactions. Indeed, the main focus of most reputation systems is on the behavior of trustees (second movers). The baseline treatment involves no information on trustees' (second mover) behavior, and a treatment variable is the amount of information provided on a trustee's prior history of play. A secondary treatment concerns whether information provision is free or costly. We have provided a simple theoretical model examining these various informational assumptions and have examined the predictions that follow from our model in a controlled laboratory experiment.

Although the parameters of the game were chosen to support a social norm of full trust and reciprocity as an equilibrium in the baseline treatment with no information about trustees, we find that full

[13] The variable "Last return" is highly correlated with "All return", and the longer historical record on trustees favors the local (individual-specific) over the global (grim) response. We use the grim-and-last-return dummy in the full information case to facilitate comparison with the minimal information case.

trust and reciprocity is difficult to sustain when no reputational information is available. Providing reputational information about trustees increases the frequencies of trust and reciprocity, and this effect becomes significantly stronger and more consistent as the amount of information on the prior behavior of trustees is increased. Furthermore, providing the possibility to purchase information at a small cost also significantly increases the frequencies of trust and reciprocity compared with the case where such information is not available. The latter result suggests that it may be the *availability* of information that matters even more than the actual content of that information. After all, only 24% of investors in our Cost treatment purchased information, and most who did not purchase information chose Invest anyway.

Our findings help us to comprehend the emergence and prevalence of reputation systems in economic interactions involving "strangers" and one-sided incentive problems such as the online feedback system found on eBay or credit reports provided by credit bureaus. The significant contribution of this paper is that we have identified the importance of individual, reputational information over community-wide enforcement in increasing and sustaining trust and reciprocity and improving efficiency. This understanding is of obvious importance to the design and operation of economic institutions.

For future research, we are interested in designing and comparing different reputational mechanisms that improve the efficiency of the markets with random and anonymous players. For instance, one comparison is between an online, peer-to-peer feedback system (where information need not be truthful) and a third party credit bureau (where information may be presumed to be more truthful). Online feedback systems involve decentralized voluntary contribution of information about transactions with free dissemination to other users. On the contrary, the traditional credit bureau collects information in a centralized and mandatory method, as in our Min and Info treatments and charges users for access to this information. Although we have shown in this paper that providing reputational information significantly increases efficiency, we still have not solved the more practical issue of how to decentralize or finance the reputational system. We leave such an analysis to future research.

References

1. Berg, J., Dickhaut, J. and McCabe, K. (1995). "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10 (1), 122-42.
2. Bolton, G., Katok, E. and Ockenfels, A. (2004). "How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation," *Management Science*, 50 (11), 1587-1602.
3. Brown, M., Zehnder, C. (2007). "Credit Reporting, Relationship Banking, and Loan Repayment," *Journal of Money, Credit and Banking*, 39 (8), 1883-1918.
4. Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction,* Russell Sage Foundation and Princeton University Press.
5. Camera, G. and Casari, M. (2009). "Cooperation among Strangers under the Shadow of the Future," *American Economic Review,* 99 (3), 979-1005.
6. Charness, G., Du, N. and Yang, C-L. (2009). "Trust and Trustworthiness Reputation in an Investment Game," working paper, UC Santa Barbara.
7. Duffy, J. and Ochs, J. (2009). "Cooperative Behavior and the Frequency of Social Interaction," *Games and Economic Behavior,* 66 (2), 785–812.
8. Ellison, G. (1994). "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies,* 61 (3), 567-588.
9. Engle-Warnick, J. and Slonim, R. (2006). "Learning to Trust in Indefinitely Repeated Games," *Games and Economic Behavior*, 54 (1), 95-114.
10. Fehr, E. and Schmidt, K. (1999). "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.
11. Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Ready-made Economic Experiments," *Experimental Economics*, 10(2), 171-178.
12. Greif, A. (1989). "Reputation and Coalitions in Medieval Trade," *Journal of Economic History*, 49, 857-882.
13. Greif, A. (1993). "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review*, 83 (3), 525-548.
14. Kandori, M. (1992). "Social Norms and Community Enforcement," *Review of Economic Studies*, 59, 63-80.
15. Lee, Y.-J. and Xie, H. (2009). "Social Norms and Trust among Strangers," Working Paper, Concordia University.
16. Milgrom, P., North, D. and Weingast, B. (1990). "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics*, 2 (1), 1-23.
17. Okuno-Fujiwara, M. and Postlewaite, A. (1995). "Social Norms and Random Matching Games," *Games and Economic Behavior*, 9, 79-109.

Appendix A: Definition of $f(\delta)$ and $g(\delta)$ and Proof of Propositions 2 and 3

Definition of $f(\delta)$ and $g(\delta)$: To provide a formal definition of $f(\delta)$ and $g(\delta)$, further notation is necessary. Let X_t be the total number of d-type investors and let Y_t be the total number of d-type trustees at the beginning of period t. The state of the world in period t, Z_t, contains information about the number of d-type investors and d-type trustees in the current period and is defined as a one-to-one and onto function from (X_t, Y_t) to the set of natural numbers $\{1, 2, \ldots, n(n+2)\}$:

$$Z_t = (n+1)X_t + Y_t \quad for \quad X_t + Y_t > 0.$$

Let A be an $n(n+2) \times n(n+2)$ transition matrix when all players follow the contagious strategy. It has elements

$$a_{ij} = \Pr\{Z_{t+1} = j \mid Z_t = i \text{ and all players follow the contagious strategy}\}.$$

For example, $a_{12} = \Pr\{Z_{t+1} = 2 \mid Z_t = 1\} = \Pr\{(X_{t+1}, Y_{t+1}) = (0,2) \mid (X_t, Y_t) = (0,1)\}$ denotes the probability that there are two d-type trustees and no d-type investors in next period given that there is one d-type trustee and no d-type investors in the current period. Similarly, let B be an $n(n+2) \times n(n+2)$ transition matrix when the d-type trustee in consideration deviates from the contagious strategy while all other players continue to follow the contagious strategy, with elements

$$b_{ij} = \Pr\{Z_{t+1} = j \mid Z_t = i \text{, one d-type trustee deviates from the contagious strategy and all other players follow the contagious strategy}\}.$$

Thus, matrix $B - A$ characterizes how the diffusion of d-type players is delayed if one d-type trustee unilaterally deviates from the contagious strategy. Define ρ as an $n(n+1) \times 1$ column vector with ith element equal to the conditional probability that the d-type trustee meets a c-type investor when the state is i in period t. Finally, let e_i be a $1 \times n(n+2)$ row vector with the ith element equal to 1 and all other elements equal to 0.

Define $f(\delta) \equiv \delta e_1 (B - A)(I - \delta A)^{-1} \rho$ and $g(\delta) \equiv \delta e_{n+2} (B - A)(I - \delta A)^{-1} \rho$.

$f(\delta)$ is the increase in the sum of the expected probability (and also the expected payoff) of meeting a c-type investor for the d-type trustee in all the future periods when this d-type trustee chooses to deviate from defection (i.e. continues to return) given that the d-type trustee is the *only* d-type player in the community. Similarly, $g(\delta)$ is the increase in the sum of the expected payoff (and the probability of meeting a c-type investor) in future periods for the d-type trustee from slowing down the contagious process given that currently there is just one d-type trustee and one d-type investor.[14]

[14] The formal derivation of $f(\delta)$ and $g(\delta)$, as well as the formula for each element of matrix A and B can be

Proof of Proposition 2: Consider the following strategy: the trustee chooses Keep if he is a d-type and chooses Return if he is a c-type; the investor chooses Invest in the first period and in subsequent periods if the information reveals the current trustee to be a c-type and chooses No Invest if the information reveals the current trustee to be a d-type.

We show next that a one-shot deviation is not profitable for the investor or the trustee both on the equilibrium path and off the equilibrium path.

For the trustee on the equilibrium path, the payoff from following the strategy above is $\frac{1-b}{1-\delta}$, while the payoff from a deviation is 1, with no payoff in all future periods. So the trustee has no incentive to deviate if $\delta \geq b$.

For the trustee off the equilibrium path, the payoff from following the strategy is 1, the payoff while the payoff from a deviation is $1-b$, so the trustee has no incentive to deviate.

For the investor on the equilibrium path (when the information reveals the current trustee to be a c-type), the payoff from following the strategy is $b + E\pi_1$, and the payoff from deviation is $a + E\pi_2$, where $E\pi_1$ and $E\pi_2$ are the future payoffs given that the investor chooses Invest and No Invest in the current period respectively. If the investor chooses No Invest instead of Invest this period, the current matched trustee will become a d-type instead of remaining a c-type player, so the investor's future payoff given No Invest in the current period, $E\pi_2$, is smaller than the future payoff given Invest in the current period, $E\pi_1$. Therefore the investor has no incentive to deviate on the equilibrium path.

For the investor off the equilibrium path (when the information reveals the current trustee to be a d-type), the payoff from following the strategy is $a + E\pi_3$, and the payoff from deviation is $E\pi_4$, where $E\pi_3$ and $E\pi_4$ are the investor's future payoffs given No Invest and Invest in the current period respectively. Since the current trustee is a d-type, the investor's choice in the current period will not affect the number of d-type trustees in the future. Thus, $E\pi_3$ equals $E\pi_4$ and the investor has no incentive to deviate.

Proof of Corollary 1: We will show that the following strategy under minimal information is an equilibrium strategy, but only for the knife-edge condition where $\delta = b$. Specifically, the trustee chooses Keep if he is a d-type and chooses Return if he is a c-type; the investor chooses Invest in the first period or if the minimal information reveals the current trustee returned last period (is a c-type) and chooses No Invest otherwise.

It is easy to verify that the investor will follow the proposed strategy (similar to the proof of Proposition 1). In particular, consider the incentive constraint for a d-type trustee given investment by the investor in the current period. The d-type trustee's payoff from choosing Keep in the current period is 1. However, by choosing Return in the current period and waiting to play Keep in the next period, his payoff is $1 - b + \delta$ since the investor in the next period will choose Invest given the trustee's record of Return in

found in Lee and Xie (2009).

the current period. Therefore a d-type trustee will not deviate from playing Keep only if $b \geq \delta$.

Next, consider the incentive constraint for a c-type trustee given investment by the investor in the current period. His payoff from choosing Return is $\frac{1-b}{1-\delta}$ and his payoff from choosing Keep is 1. So he will not deviate from playing Return only if $\delta \geq b$.

Therefore under minimal information, the proposed strategy is an equilibrium strategy only if $\delta = b$.

Proof of Proposition 3:

Consider the following strategies: In the first period, the investor always chooses Invest. In subsequent periods, the investor purchases information with probability q. If he does not purchase information, he chooses Invest with probability γ, where $0 \leq \gamma \leq 1$, otherwise if the information purchased reveals the trustee has always played Return, the investor plays Invest and otherwise he plays No Invest. Trustees adopt an asymmetric strategy, with a fraction p of trustees choosing Return given investment in every period, and fraction $1-p$ of trustees choosing Keep given investment in every period.

The trustee's payoff from choosing Return every period is $(1-b) + \frac{\delta}{1-\delta}[q + (1-q)\gamma](1-b)$.

The trustee's payoff from choosing Keep every period is $1 + \frac{\delta}{1-\delta}(1-q)\gamma$.

These two payoffs must be equal in the first period, so

$$(1-b) + \frac{\delta}{1-\delta}[q + (1-q)\gamma](1-b) = 1 + \frac{\delta}{1-\delta}(1-q)\gamma,$$

which implies $q = \frac{b-\delta b+\gamma\delta b}{\delta-\delta b+\gamma\delta b}$.

$$\frac{\partial q}{\partial \gamma} = \frac{(\delta - b)\delta b}{(\delta - \delta b + \gamma\delta b)^2}$$

Notice that q is increasing in γ if $\delta > b$. Thus, if the investor chooses Invest with a higher probability when no information is available, in equilibrium the probability for the investor to purchase information must be higher. The intuition for this finding is that if the investor increases the probability of invest under no information, she must at the same time decrease the probability of no information, otherwise all trustees will switch to always choosing Keep in every period.

In order to show that this is an equilibrium strategy, we also need to show that, given investment, the trustee will not switch from playing Return to playing Keep or from playing Keep to playing Return in every period, i.e., that p is constant. This is easy to verify since once the trustee chooses Keep in one period, then the best he can do (under full information) is to continue to choose Keep in every period. Therefore, a trustee who chooses Keep in the first period has no incentive to switch to the other strategy in later periods. For a trustee who chooses Return in the first period, in every period he is facing the same constraint he faced in the first period, so in every period he will be indifferent between always choosing Return and always choosing Keep given investment.

Next we show that investors have no incentive to deviate from the proposed strategy. Consider first the investors who do not purchase information. These investors will follow a mixed strategy only if their payoff from Invest is equal to their payoff from No Invest, i.e., if $a = pb$. Since investors also randomize as to whether or not they will purchase information, the following equation must hold,

$$a = pb = pb + (1-p)a - c,$$

where the last term of the equation is investor's payoff from purchasing information and c is the cost of purchasing information.

Therefore, in equilibrium the fraction of good trustees is $p = \frac{a}{b}$, and the cost of information is $c = \frac{a}{b}(b-a)$.

In particular, we are interested in two special cases: Investors who do not purchase information either choose Invest with probability zero (i.e., choose No Invest) or they choose Invest with probability one.

Case 1: If the investors choose No Invest with probability one when no information is purchased ($\gamma = 0$), it must hold that $a = pb + (1-p)a - c \geq pb$. Given the cost c chosen in the experiment, $a > pb$ and therefore the investors will choose No Invest in the first period. So in all the subsequent periods, investors should choose not to purchase information and No Invest with probability one. Therefore, the equilibrium reduces to the most inefficient equilibrium.

Case 2: If the investors choose Invest with probability one when no information is purchased ($\gamma = 1$), the investors purchase information with probability $q = \frac{b}{\delta}$ and it must hold for that $pb = pb + (1-p)a - c \geq a$. So the fraction of trustees who always choose Return given investment is $p = \frac{a-c}{a}$, where $c \leq \frac{a}{b}(b-a)$. From the parameters chosen in the experiment, $q = 9/16$ and $p = 33/35$.

Next we turn to the efficiency comparison.

(1) When $0 < \gamma < 1$, according to the equilibrium outcome, the investor's payoff in each period is a. The trustee's average payoff in the first period is $p(1-b) + (1-p) = 1-a$ given that $p = a/b$, and the trustee's average payoff in each subsequent period (denoted by AU) is $AU = p[q + (1-q)\gamma](1-b) + (1-p)[(1-q)\gamma]$, where $[q + (1-q)\gamma](1-b)$ is the expected payoff for a trustee who always chooses Return and $(1-q)\gamma$ is the expected payoff for a trustee who always chooses Keep. In the equilibrium, $q = \frac{b - \delta b + \gamma\delta b}{\delta - \delta b + \gamma\delta b}$ is increasing in γ. It is easy to verify that $(1-q)\gamma$ is

also increasing in γ. Therefore, the most efficient outcome obtains when $\gamma \to 1$ and $AU \to 0.52$ given $p = \frac{a}{b} = 7/9$.

(2) When $\gamma = 1$, according to the equilibrium outcome in Case 2, the investor's payoff in each period is $pb > a$ given $c = 2$ as we chose in the experiment. The trustee's average payoff in the first period is $p(1-b) + (1-p)$, and the trustee's average payoff in each subsequent period $AU = p[q + (1-q)\gamma](1-b) + (1-p)[(1-q)\gamma] = 0.54 > 0.52$ given $p = 33/35$. Therefore the most efficient equilibrium outcome obtains when $\gamma = 1$.

Proof of Corollary 2: As noted in the text, when information is costly, there also exists an inefficient equilibrium where no investors purchases information or chooses Invest and no trustee chooses Return. It is easy to show that deviations from this equilibrium either on or off the equilibrium path are not profitable for either the trustee or the investor.

Not for Publication

Appendix B: Instructions Used in the Experiment

In this appendix we provide the instructions given to subjects in the Min_No sessions. In italics, we show the additional instruction that was provided to subjects in the Info_No_Cost sessions.

Instructions

Overview

This is an experiment in decision-making. The department of economics has provided funds for this research. During the course of this experiment, you will be called upon to make a series of decisions. If you follow these 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. We ask that you not talk with one another for the duration of the experiment.

Specifics

The experiment is divided into a series of sequences. A sequence will consist of an indefinite number of rounds. At the beginning of each sequence, you will be randomly assigned as a First Mover or a Second Mover. Your role will appear on your computer screen and will not change during the sequence. At the beginning of each round you will be randomly paired with another person who is assigned to the other role from your own. That is, if you are a First Mover (Second Mover), in each period you will be randomly paired with a Second Mover (First Mover) with all possible pairings being equally likely.

In each round, you and your paired player will play the game described in the following graph. First, the First Mover chooses between A and B. If the First Mover chooses A, the round is over. The First Mover receives 35 points and the Second Mover receives 0 points. If the First Mover chooses B, then the Second Mover must make a choice between C and D. If the Second Mover chooses C, then the First Mover receives 0 points and the Second Mover receives 100 points. If the Second Mover chooses D, then the First Mover receives 45 points and the Second Mover receives 55 points. Following the first round of a sequence, the First Mover will be told the decision that his/her paired Second Mover has made in the last round, if that player was able to choose between C and D. The First Mover never knows the identity of his/her paired Second Mover.

(*Following the first round of a sequence, at the beginning of each round the First Mover will be told the past choices that his/her paired Second Mover has made in each of the most recent rounds of the current sequence (up to 10 rounds), provided that this Second Mover had a choice to make (the First Mover chose B). The First Mover will be also told the total number of times that his/her paired Second Mover has chosen C or D in the entire sequence, again conditional on having the opportunity to make a choice (First Mover chose B). The First Mover never knows the identity of his/her paired Second Mover*.)

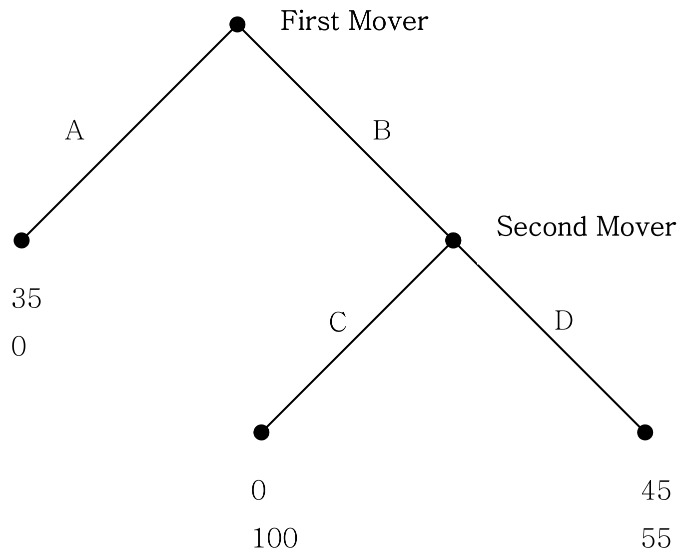

Figure A1: Decisions and Earnings (in Points)

To complete your choice in each round, click on the decision button and then the OK button. The Second Movers need to wait for the First Movers to make a choice between A and B before making their own choice. Then the Second Mover will be told that the round is over (if the First Mover chooses A), or will be asked to make a choice between C and D (if the First Mover chooses B).

The computer program will record your choice and the choice made by the player paired with you in this round. After all players have made their choices, the results of the round will appear on your screen. You will be reminded of your own choice and will be shown the choice of your match, as well as the payoff you have earned for the round. Record the results of the round on your RECORD SHEET under the appropriate headings.

Immediately after you have received information on your choice and the choice of the player with whom you are randomly paired for the round, a ten-sided die with numbers from 0 to 9 on the sides will be thrown by one of you (you will take turns throwing the die) to determine whether the sequence will continue or not. The experimenter will announce loudly the result of each die roll. If a number from 0 to 7 appears, the sequence will continue into next round. If an 8 or 9 appears, the sequence ends. Therefore, after each round there is 80% chance that you will play another round and 20% chance that the sequence will end.

Suppose that a number less than 8 has appeared. Then you will play the same game as in the previous round, but with an individual selected at random whose role is different from yours. You record the outcome and your earnings for the round. Then another throw is made with the same die to decide whether the sequence continues for another round.

If an 8 or 9 appears, the sequence ends. The experimenter will announce whether or not a new sequence will be played. If a new sequence is to be played then you will be randomly reassigned as a First Mover or a Second Mover. The new sequence will then be played as described above.

If the experiment does not end within 2 hours, you will be invited to continue the experiment in the next several days.

Earnings

Each point you earn is worth 0.5 cent ($0.005). Therefore, the more points you earn the more money you earn. You will be paid your earnings from all the rounds of all sequences in cash, and in private, at the end of today's experiment as well as $5 show-up fee.

Final Comments

First, do not discuss your decisions or your results with anyone at any time during the experiment.

Second, your ID# is private. Do not reveal it to anyone.

Third, since there is 80% chance that at the end of a round the sequence will continue, you can expect, on average, to play 5 rounds in a given sequence. However, since the stopping decision is made randomly, some sequences may be much longer than 5 rounds and others may be much shorter.

Fourth, your role as a First Mover or a Second Mover will be randomly assigned when a new sequence begins. Your role will not change for the duration of that sequence.

Finally, remember that after each round of a sequence you will be matched randomly with a player whose role is different from yours. Therefore, the probability of you being matched with the same individual in two consecutive rounds of a game is 1/3 since there are 3 First Movers and 3 Second Movers in the room.

Questions?

Now is the time for questions. Does anyone have any questions?

Quiz

If there are no more questions, please finish the quiz. Your answers to this quiz will not affect your earnings. The purpose of the quiz is to help you understand the instruction better. After everyone has completed the quiz the answers will be reviewed.

Continuation Instruction (Min_No sessions)

From now on until the end of today's experiment, everything is the same as in the original instruction except that there is no information available anymore to the First Mover about the decision that his/her paired Second Mover has made in the last round.

Continuation Instruction (Info_No_Cost sessions)

From now on everything is the same as in the original instructions except that there is no information available anymore to the First Mover about the past choices that his/her paired Second Mover has made in the most recent rounds of the current sequence, or about the frequency with which his/her paired Second Mover has chosen C or D in all rounds of the current sequence.

Continuation Instruction (Info_No_Cost sessions)

From now on until the end of today's session, at the beginning of each round of a sequence except the first round, the First Mover may choose to buy information about the past choices that his/her paired Second Mover has made in the current sequence, at a price of 2 points. The information provided is the same that was provided in the instructions for the first part of today's session: the First Mover is informed of the past choices that his/her paired Second Mover has made in the most recent rounds of the current sequence and about the frequency with which his/her paired Second Mover has chosen C or D in all rounds of the current sequence. The First Mover is free to decide whether or not to buy this information every round. If the First Mover chooses to buy the information, the 2 point cost will be deducted from his/her payoff from the round; otherwise there will be no deduction (and no information on the Second Mover will be provided to the First Mover). As in all previous rounds, the Second Mover will only be informed of his own payoff, and will not be able to observe whether or not the First Mover has chosen to purchase information about the Second Mover.

Not For Publication

Appendix C: Aggregate Frequencies in No_Info_Cost and Info_No_Cost Sessions (first period excluded)

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	0.500	0.867	0.978
No_Info_Cost2	0.639	0.839	0.782
No_Info_Cost3	0.642	0.833	0.841
Info_No_Cost1	0.863	0.627	0.907
Info_No_Cost2	0.790	0.094	0.296
Info_No_Cost3	0.778	0.417	0.432

Table B1: Frequency of Invest (first periods excluded)

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	0.815	0.974	0.989
No_Info_Cost2	0.696	0.877	0.794
No_Info_Cost3	0.654	0.933	0.879
Info_No_Cost1	0.980	0.734	1.000
Info_No_Cost2	0.906	0.455	0.750
Info_No_Cost3	0.889	0.433	0.600

Table B2: Frequency of Return-Given-Invest (first periods excluded)

	1st treatment	2nd treatment	3rd treatment
No_Info_Cost1	0.407	0.844	0.968
No_Info_Cost2	0.444	0.736	0.621
No_Info_Cost3	0.420	0.778	0.739
Info_No_Cost1	0.846	0.461	0.907
Info_No_Cost2	0.716	0.043	0.222
Info_No_Cost3	0.691	0.181	0.259

Table B3: Frequency of Invest-and-Return (first periods excluded)