Social Norms and Trust among Strangers[*]

Huan Xie[†] Yong-Ju Lee[‡]

July 2012

Abstract

We study the development of a social norm of trust and reciprocity among a group of strangers via the "contagious strategy" as defined in Kandori (1992). Over an infinite horizon, the players anonymously and randomly meet each other and play a binary trust game. In order to provide the investors with proper incentives to follow the contagious strategy, there is a sufficient condition that requires that there exist an outside option for the investors. Moreover, the investors' payoff from the outside option must converge to the payoff from trust and reciprocity as the group size goes to infinity. We show that this sufficient condition is also a necessary condition to sustain *any* sequential equilibrium in which the trustees adopt the contagious strategy. Our results imply that a contagious equilibrium only supports trust if trust contributes almost nothing to the investors' payoffs.

JEL Codes: C72, C73, D83

Keywords: Trust and Reciprocity, Infinitely Repeated Game, Random Matching, Contagious Strategy, Sequential Equilibrium.

[†]Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montréal, Québec, Canada H3G 1M8, email: huanxie@alcor.concordia.ca.

[‡]Samsung Research Institute of Finance, Seoul, Korea, Email: yj0612.lee@samsung.com.

1 Introduction

One of the most important issues in economics is how to sustain cooperation when players have a short-term incentive to deviate from cooperative behavior. Economists have long recognized that reputation is an effective means of enforcing cooperation when there exists an institution to track and disseminate information on players' past behavior, or when the group is small so people are intimately familiar with one another's history. These *personal enforcement* mechanisms make quick and substantial retaliations possible. The Folk Theorem in the repeated game literature (Fudenberg and Maskin, 1986) provides a formal model of personal enforcement, showing that any mutually beneficial outcome can be sustained as a subgame perfect equilibrium if the same set of players play the same stage game ad infinitum (Kandori, 1992).

When economic interactions happen among essentially anonymous players, as noted by Kandori (1992), *community enforcement* is required to provide incentives for agents not to deviate from cooperative behavior. In this case, dishonest behavior against one partner causes sanctions by other agents in the community. Kandori (1992) examines an infinite-horizon model where agents are anonymously and randomly paired to each other and play a Prisoner's Dilemma game. Even when each agent knows nothing more than his personal experience, Kandori shows that a community can still sustain cooperation via a "contagious strategy," in which players who defected or experienced defection in the past choose non-cooperation forever. In the equilibrium, a deviator could be indirectly punished if the deviation were to trigger a contagious reaction that destroyed the social norm of cooperation. If the consequences of the eventual destruction of the cooperative norm were sufficiently severe and credible, then the threat of the contagious reactions might sustain a social norm of cooperation. Ellison (1994) extends Kandori's work by introducing a public randomization device and remedies two issues with Kandori's results. First, the equilibrium in Ellison (1994) does not require excessive patience of players and applies to more general payoff structures. Secondly, the equilibrium is stable and efficient with noise, and a single deviation or mistake will not lead to a complete destruction of cooperation due to the property of contagious reaction.[1] [2]

---

[1]See Ghosh and Ray (1996), Greif (1993), Okuno-Fujiwara and Postlewaite (1995) and Tirole (1996) for several related models.

[2]There are also several experimental studies that directly test the contagious equilibrium. Duffy and Ochs (2008) test Kandori's (1992) contagious equilibrium with an indefinitely repeated Prisoner's Dilemma under different matching protocols and different amounts of information transmission. Their results suggest that random matching works to prevent the development of a cooperative norm in the laboratory, even when some information is provided about the prior choices of opponents. Camera and Casari (2008) test the contagious equilibrium with indefinitely repeated Prisoner's Dilemma under random matching, but focus on the role of private or public monitoring of the players' choices. They find that such monitoring can lead to a significant increase in the frequency of cooperation relative to the case of no monitoring. Duffy, Xie and Lee (2011) experimentally test the contagious equilibrium using a trust game and find that trust and reciprocity must rely on the availability of individual reputational information. Societal reputation as suggested by the contagious equilibrium is not enough to sustain a high level of trust and reciprocity even with a group size of six individuals.

The purpose of this paper is to examine whether the concept of contagious equilibrium (Kandori, 1992) can be extended to classes of games other than the Prisoner's Dilemma game. In particular, we focus on a binary trust game modified from Berg et al. (1995). In our version of the trust game, the investor (first mover) first decides whether to invest her endowment with the trustee (second mover) or choose the outside option.[3] If the investor invests, the endowment is multiplied by a fixed factor greater than one and it falls on the trustee to decide whether to keep the whole amount or return some fraction of it to the investor. When the game is played one-shot, the unique equilibrium outcome is for the trustee not to reciprocate and the investor not to trust. Unlike the Prisoner's Dilemma game, there is no dominant strategy for the investor in the trust game. The trust game represents a one-sided rather than a two-sided incentive problem. That is, knowing the partner will cooperate, only the trustee has an incentive to deviate from the cooperative outcome. Therefore, the game captures the feature of some real-life examples such as transactions between buyers and sellers on the Internet or loan repayment in credit markets, which usually have a sequential-move game structure and only the second mover has an incentive problem. Following Kandori (1992), in our model a finite population of anonymous agents are randomly paired with each other in each period and play the binary trust game. Players can neither recognize nor communicate the identity of any of their past opponents, and they do not observe the outcomes of games in which they are not involved or any aggregate information about the entire community.

We find a sufficient condition that provides the investors with proper incentives to follow the contagious strategy. However, this sufficient condition requires that there should exist an outside option for the investors. Moreover, only if the outside option almost equals the maximum possible payoff from trust itself, will the investors choose not to trust a randomly paired trustee after having experienced non-reciprocative trustees in the past. This sufficient condition controls the investors' incentives to follow the contagious strategy off the equilibrium path and so supports a credible threat to keep the trustees from behaving dishonestly. However, the condition becomes more constraining when the community size becomes larger. In particular, the payoff from the outside option must approach the payoff from trust and reciprocity when the community size grows to infinity. This makes the sufficient condition essentially useless since the sustaining of the social norm of trust and reciprocity between anonymous agents is more important in a large economy than in a small community and in that case trust is only possible if its net benefit to investors is almost zero.

Importantly, we further examine a more general version of contagious equilibrium. We assume that the trustees continue to adopt the contagious strategy while the investors are allowed to choose any strategy based on their consistent beliefs. Then we ask how *necessary* the sufficient condition on the outside option is to support such a strategy profile in a sequential equilibrium. In particular, if the investors' outside option is smaller, will it still be possible to support the

---

[3]We denote investor as she and trustee as he for clarity.

2

social norm of trust and reciprocity in a sequential equilibrium? We show that the investors' consistent belief system always assigns probability one to the event that there is *at most one* non-reciprocating trustee, no matter how many defections the investors observe in their private history. Therefore, with a smaller outside option, the investors always choose to trust, and consequently, there is no deterrence to defection of trustees, and the proposed equilibria collapse. So in order to sustain the social norm of trust and reciprocity in any sequential equilibrium involving the contagious strategy, it is necessary to have an outside option for the investors that converges to the payoff from trust and reciprocity in a large community. This necessity of the condition implies that the contagious equilibrium only supports trust if trust contributes almost nothing to the investors' payoffs.

This paper is the first, to the best of our knowledge, that tries to extend the contagious equilibrium to games other than the Prisoner's Dilemma game. Sociologically, the trust game represents social exchange or risk taking while the Prisoner's Dilemma illustrates social conflict. As noted by Ellison (1994), the results of the previous papers rely on the game structure of the Prisoner's Dilemma: it has dominant strategies for both players and simultaneous moves. The negative result presented in this paper seems to be more associated with the fact that the investor in the trust game does not have a dominant strategy rather than the fact that the trust game has sequential moves, as discussed later in the paper.

The rest of the paper is organized as follows. Section 2 describes the infinitely repeated random matching model with a binary trust game. Section 3 defines the contagious strategy and describes the sufficient condition for the investors. Section 4 proves the necessity of the sufficient condition. Section 5 concludes.

2 The Model

We first describe the structure of the repeated random matching game. 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 randomly paired with one trustee and they play the following binary trust game modified from Berg et al. (1995).

At the beginning of the game, the investor is endowed with an amount $a \in (0, 1)$. If the investor decides not to invest (NI), the game ends. The investor's payoff is a (the value of her outside option) and the trustee's payoff is 0. If the investor chooses to invest (I) her 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 and decides whether to keep (K) all of the gross return for a payoff of 1 for himself and 0 for the investor or to return (R) a fraction $0 < b < 1$ to the investor, earning a payoff of $1 - b$ for himself. Throughout we shall assume that $0 < a < b < 1$, so that it is efficient for the investor to invest, the investor prefers the cooperative outcome (I, R) to her

Investor

NI   I

Trustee

a
0   K   R
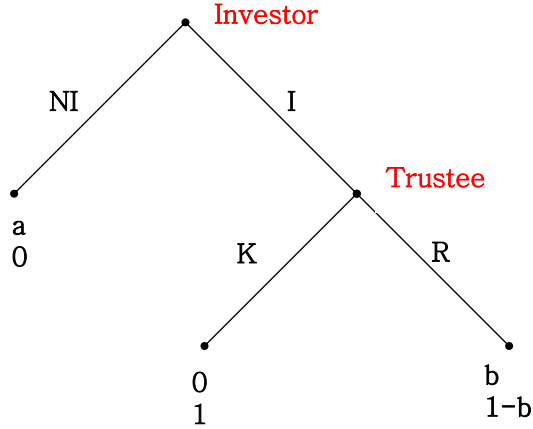
0   b
1   1-b

Figure 1: The Trust Game

outside option, and the trustee has an incentive to keep all the return once the investor invests. The trust game we study is depicted in Figure 1.

This binary trust game well represents the situation in a one-sided incentive problem (Kandori, 1992). Unlike the Prisoner's Dilemma game, there is no dominant strategy for the investor in the trust game since she prefers cooperation to her outside option. The trustee as a second mover, however, has an incentive to deviate from cooperation. The game can also apply to a trading situation in which the buyer needs to make a payment first and the seller chooses whether or not to ship the good after receiving the payment. If the binary trust game is only played once, the unique subgame perfect equilibrium is for the investor not to invest and for the trustee to keep all the return.

In this model we assume that pairings between investors and trustees are independent and uniform in each period. Let $m_t(i)$ denote investor i's match at time t. By uniform random matching, $\Pr[m_t(i) = j] = \frac{1}{n}$ for all investors $i \in N_I$ and trustees $j \in N_T$ and for all t. Each period every pair of investor and trustee play the binary trust game described above as a stage game. This procedure is repeated infinitely and each player's total payoff is the expected sum of his stage payoffs discounted by $\delta \in (0, 1)$.

In the entire paper, we assume that each player only observes the history of action profiles in the stage games that he has experienced. When paired with another player, a player has no idea about the identity of his match or the previous experience of his match or that of any other player. Therefore, a player cannot base his action on his personal experience with the current match. Neither can his choice be based on any information on the plays between his match and other players in the community or on any aggregate information about the entire community.

4

3 A Sufficient Condition for Contagious Equilibrium

In this section, we first define the concept of "contagious equilibrium" in the infinitely repeated trust game with random matching, following the seminal work of Kandori (1992). Define *NI* (No Invest) as the defection of an investor, and K (Keep) as the defection of a trustee. Define d-type investors or trustees as those whose private history includes defection by themselves or their partners. Otherwise, the players are defined as c-type.

Definition 1 *The "contagious strategy" is defined as follows: An investor chooses* I *(Invest) if she is c-type and* NI *(No Invest) if she is d-type. A trustee chooses* R *(Return) if he is c-type and* K *(Keep) if he is d-type.*

Under the contagious strategy, a player who experienced either dishonest or untrusting behavior starts defecting against all of his or her opponents. Since the players are anonymous to each other, trust is applied to the community as a whole rather than to each individual player, and a single defection by a player causes the end of trust in the whole community. In order to show that the contagious strategy constitutes a sequential equilibrium, we need to show that neither the investors nor the trustees have an incentive to deviate from the contagious strategy, when the players hold consistent beliefs as defined in Kreps and Wilson (1982).[4]

In particular, consider the sufficient condition for the investors to follow the contagious strategy in a sequential equilibrium. Since the binary trust game represents a one-sided incentive problem in which the investors prefer the cooperative outcome (I, R) to the outside option, it is easy to verify that the investors have no incentive to deviate from the equilibrium path. In order for the investors to follow the contagious strategy off the equilibrium path, we also need to show that a d-type investor will never choose to trust a trustee again, given any consistent belief on the number of d-type trustees in the community. It turns out that the following condition ensures that a d-type investor will not deviate to go back to choosing Invest,

$$a \geq \left(1 - \frac{1}{n}\right) b. \tag{1}$$

Here $(1 - \frac{1}{n})b$ is the d-type investor's expected payoff for the current period from choosing to invest when she believes there is only one d-type trustee.[5] This is the strongest condition since it

---

[4]In a previous working paper version, we prove the following existence result. For any δ and $n \geq 2$, there exist a and b such that (i) $0 < a < b < 1$; and (ii) the contagious strategy constitutes a sequential equilibrium in which (I, R) is the outcome in every period along the equilibrium path under uniformly random matching. We also characterize a set of sufficient conditions under which the investors and the trustees have no incentives to deviate from the contagious strategy given any possible consistent belief. Since those conditions require a to approach b as n approaches infinity, the focus of this paper is to show that one of these very restrictive sufficient conditions to sustain the contagious equilibrium is also a very restrictive necessary condition. Thus, the existence result and the proof for the sufficient condition are not provided here but are available upon request. Interested readers can also find the previous version of the paper on the authors' websites.

[5]Notice that the investor's payoff in any future period can be assumed to always be the outside option, a, no matter whether she chooses to invest or not for the current period, since we only need to check that the players

is sufficient even if the investor believes there is more than one d-type trustee in the community.

Condition (1) shows that the existence of the contagious equilibrium critically depends on a high enough outside option. Since the concept of the contagious equilibrium is based on community enforcement, dishonest behavior against one partner must cause sanctions by other members in the society. The development of a cooperative social norm requires a harsh punishment scheme: investors must stop trusting everyone after one defection, i.e., they must be sufficiently pessimistic so that they do not want to trust given their beliefs. The role of the high outside option a is to provide the investors with enough incentives to carry out the punishment and to serve as a credible threat for the trustees not to initiate defection.

This condition, however, becomes more restrictive when the population size n becomes larger. It requires that the outside option, a, converge to the investors' payoff from the outcome of trust and reciprocity, b. In that case, the gain from the outcome of trust and reciprocity is almost the same as the gain from taking the outside option for the investors. Then even if the contagious equilibrium can achieve the social norm of trust and reciprocity in a large anonymous community, the benefit to the investors is negligible. That is, a contagious equilibrium supports trust if trust contributes almost nothing to the investors' payoff.

4   Necessity of the Sufficient Condition

The question we pose in this section is how *necessary* the condition $a \geq (1 - \frac{1}{n})b$ is to sustain any (potentially more complicated) sequential equilibrium that is built on the basic idea of a contagious equilibrium. Notice first that this condition is indeed a necessary condition in order for the contagious strategy in Definition 1 above to constitute a sequential equilibrium. The contagious strategy requires any investor, who has ever experienced defection in the past, not to invest again in all future periods. By the remark in Kandori (1992, page 69), if a player sees a defection in the first period, he must believe that everybody else is cooperating, because the definition of consistent belief requires that defections by different players are statistically independent. By the same token, when a player defects when he has seen no defection, he must believe that he is the first person to defect in the society. Similarly, in the trust game, when an investor experiences defection in the first period, she should believe there is only one d-type trustee in the community. In order for her not to invest in any future period, the one-shot deviation principle requires $a \geq (1 - \frac{1}{n})b$.

When $a < (1 - \frac{1}{n})b$, even though an ordinary contagious strategy profile, as defined in Definition 1, cannot constitute a sequential equilibrium, it is still interesting to ask whether there are other more complicated equilibria, built on the basic idea of a contagious equilibrium, which are sequential.[6] Specifically, consider any equilibrium where, along the equilibrium path,

---

will not deviate from the contagious strategy in the current period, by the one-shot deviation principle.

[6]When $a < (1 - \frac{1}{n})b$, there *may* be equilibria involving *mixing* along the equilibrium path, which includes

the investors trust and the trustees return. Suppose that the outside option, a, is right below $(1 - \frac{1}{n})b$, so the sufficient condition is not satisfied. In addition, suppose that the trustees continue to adopt the contagious strategy as defined before. Investors will be willing to invest if they believe there is only one d-type trustee, but will choose not to invest if they believe there is more than one d-type trustee. Since the trustees follow a contagious strategy, if an investor believes there is more than one d-type trustee and starts to defect, defection will spread to other c-type trustees, and eventually, to other investors who previously believed there was no more than one d-type trustee. Consequently, as in the ordinary contagious strategy, the fear of spreading the defection to the entire community may serve as a credible threat for the trustees not to initiate defection in the first place. The question we ask specifically in this section is whether this kind of strategy profile can constitute a sequential equilibrium when $a < (1 - \frac{1}{n})b$. We will show that it *cannot* be a sequential equilibrium, since an investor's consistent beliefs must attach a probability of one to *at most one trustee* defecting in any given period, no matter how many defections the investor experienced before. Therefore, the investors always choose to trust and there is no deterrent to defection by trustees. Thus, the condition $a \geq (1 - \frac{1}{n})b$ is indeed a necessary condition to sustain any sequential equilibrium built on the idea of contagious equilibrium.

Theorem 1 *When $a < (1 - \frac{1}{n})b$, any strategy profile in which trustees follow the contagious strategy cannot be supported as a sequential equilibrium.*

Theorem 1 can be proven using the sequence of lemmas below. Consider any equilibrium where, along the equilibrium path, the investors invest with their paired trustee and the trustees choose to return. If an investor ever sees a defection by one or more trustees, then she may cease to trust her future trustees. Let T be the first period where some investors cease to trust, given that they have seen one or more defections in the past. The key thing to show is that consistent beliefs on the part of the investor must put probability one on *at most one* trustee defecting in period T. It will then turn out that investors will continue to trust their trustees in period T (contrary to the above assumption) when $a < (1 - \frac{1}{n})b$.

Let D_t be the number of different trustees who have ever defected, on or before period t. As in Kreps and Wilson (1982), we model consistent beliefs by considering "trembles." We begin by showing that, in any period $t \leq T$, if there is only one tremble up to and including period t, the number of different trustees, D_t, who defect at least once, up to and including period t, cannot exceed one. In other words, if there is more than one defecting trustee in any period $t \leq T$, it must be the case that there have occurred more than one tremble up to and including period t. Notice that Lemma 1 will be implicitly used in the proof of Lemmas 2-5 below.

some trust. However, in such equilibrium, the payoff from Invest must equal to the payoff from the outside option for the investors. So trust has almost no value to the investors, just as it has almost no value to the investors in an ordinary contagious equilibrium when $a \geq (1 - \frac{1}{n})b$.

Lemma 1 *For any period $t \leq T$, if there is only one tremble up to and including period t, then $D_t = 1$.*

Proof. Consider first a tremble by a *trustee*. That trustee may or may not defect again in future periods. However, by assumption, for all periods strictly before period T, investors will not cease to trust their trustees, no matter how many defecting trustees they meet. Thus, a trustee's defection cannot spread, through an investor, to another trustee. Of course, a trustee's defection before period T may induce investors to defect *in* period T, and so, spread to other trustees in period $T + 1$. But an initial tremble by one trustee will not spread to other trustees until strictly *after* period T.

Next consider an initial tremble by an *investor*. That investor's initial defection may lead a trustee to begin defecting. However, in all future periods strictly before period T, the investor herself, who trembled in the previous period, will go back to trusting unless she is hit by additional trembles, by the definition of T. In addition, by the same logic as in the last paragraph, in all future periods strictly before period T, the defected-against trustee will not spread defection, through other investors, to another trustee. Thus, again, the result of a single tremble is that at most one trustee will begin defecting. ∎

Now, consider any history h_{T-1} faced by an investor as of the beginning of period T (so h_{T-1} does not include what happens in period T itself). If h_{T-1} involves no defections, then the investor believes, with probability at least $1-\epsilon$, that $D_T = 0$, where ϵ is the probability of at least one tremble in the first T periods.

Next we show that, if h_{T-1} involves at least one defection, the investor believes, with probability one, that there is only one trustee defecting in period T. The proof works by focusing on the first period in h_{T-1} in which a given investor observes a defection by a trustee. Thus, fix an investor and a history h_{T-1} for that investor, and suppose that τ is the period of the first defection by a trustee in h_{T-1}.

Lemma 2 *Let D be the total number of defections by trustees in h_{T-1}. Let ϵ_1 be the probability of at least two trembles as of period T, given exactly one tremble as of period τ. Then*

$$A = P(h_{T-1}|D_\tau = 1) \geq \left(\frac{1}{n}\right)^D \left(\frac{n-1}{n}\right)^{T-D-1} + O(\epsilon_1). \tag{2}$$

Proof. First,

$$\begin{aligned} P(h_{T-1}|D_\tau = 1) =& P(h_{T-1}|h_{\tau-1},\, D_\tau = 1)P(h_{\tau-1}|D_\tau = 1) \\ \geq& P(h_{T-1}|h_{\tau-1},\, D_\tau = 1)P(h_{\tau-1}|D_1 = D_\tau = 1), \end{aligned} \tag{3}$$

where the first equality is a basic property of conditional probabilities, and the inequality holds because $h_{\tau-1}$ is a sequence of purely cooperative returns, and so, given $D_\tau = 1$, is least likely

8

if it is also the case that $D_1 = 1$ (i.e., the untrustworthy tremble occurred in the first period). Now,

$$P(h_{\tau-1}|D_1 = D_\tau = 1) = \left(\frac{n-1}{n}\right)^{\tau-1}, \tag{4}$$

since this is the probability of $\tau - 1$ trustworthy returns in a row, given that there is one untrustworthy trustee out of a total of n trustees. Note, here, as well as in (7) and (13) below, that it is important that trustees are assumed to follow a "contagious strategy." That is, if they *ever* defect (e.g., if they are hit by a tremble), then they continue to defect forever. Next,

$$\begin{aligned}
P(h_{T-1}|h_{\tau-1}, D_\tau = 1) &= P(h_{T-1}|h_{\tau-1}, D_T = D_\tau = 1)P(D_T = 1|h_{\tau-1}, D_\tau = 1) \\
&\quad + P(h_{T-1}|h_{\tau-1}, D_T > D_\tau = 1)P(D_T > 1|h_{\tau-1}, D_\tau = 1).
\end{aligned} \tag{5}$$

Now, by Lemma 1,

$$P(D_T > 1|h_{\tau-1}, D_\tau = 1) = \epsilon_1 \quad \text{and} \quad P(D_T = 1|h_{\tau-1}, D_\tau = 1) = 1 - \epsilon_1. \tag{6}$$

Also,

$$P(h_{T-1}|h_{\tau-1}, D_T = D_\tau = 1) = \left(\frac{1}{n}\right)^D \left(\frac{n-1}{n}\right)^{T-D-\tau}. \tag{7}$$

Using (6) and (7) in (5) gives

$$P(h_{T-1}|h_{\tau-1}, D_\tau = 1) = \left(\frac{1}{n}\right)^D \left(\frac{n-1}{n}\right)^{T-D-\tau} + O(\epsilon_1). \tag{8}$$

Combining (4) and (8) in (3) yields (2). ∎

The above lemma shows that the probability of h_{T-1}, given $D_\tau = 1$, does not vanish as the probability of trembles vanishes. The next lemma shows that, if the investor first observes an untrustworthy trustee in period τ, she assumes that there is almost certainly only *one* untrustworthy trustee as of period τ.

Lemma 3 *Let ϵ_2 be the probability of at least two trembles as of period τ, given at least one tremble as of period τ. Then $P(D_\tau = 1|h_{T-1}) = 1 + O(\epsilon_2)$.*

Proof.

$$P(D_\tau = 1|h_{T-1}) = \frac{P(h_{T-1}|D_\tau = 1)P(D_\tau = 1)}{P(h_{T-1}|D_\tau = 1)P(D_\tau = 1) + P(h_{T-1}|D_\tau > 1)P(D_\tau > 1)}. \tag{9}$$

Let ϵ_0 be the probability of at least one tremble as of period τ. Let ϵ_2 be the probability of at least two trembles as of period τ, given at least one tremble as of period τ, as in the statement of the lemma. Then $P(D_\tau = 1) = (1 - \epsilon_2)\epsilon_0$ and $P(D_\tau > 1) = \epsilon_2\epsilon_0$. Using these with (2) from Lemma 2 gives

$$P(D_\tau = 1|h_{T-1}) = \frac{A(1 - \epsilon_2)\epsilon_0}{A(1 - \epsilon_2)\epsilon_0 + P(h_{T-1}|D_\tau > 1)\epsilon_2\epsilon_0} = 1 + O(\epsilon_2), \tag{10}$$

9

since A does not approach zero as ϵ_0, ϵ_1 and ϵ_2 do, by Lemma 2. ∎

The next two lemmas show that, if the investor believes $D_\tau = 1$ in period τ, then, no matter what happens between periods τ and T, he will be nearly certain that $D_T = 1$ as of period T.

Lemma 4 *With notation as in Lemma 2,*

$$B = P(h_{T-1}|D_T = D_\tau = 1) \geq \left(\frac{1}{n}\right)^D \left(\frac{n-1}{n}\right)^{T-D-1}. \tag{11}$$

Proof. Using logic very similar to that in (3),

$$P(h_{T-1}|D_T = D_\tau = 1) \geq P(h_{T-1}|h_{\tau-1},\, D_T = D_\tau = 1)P(h_{\tau-1}|D_1 = D_\tau = 1). \tag{12}$$

Simple probability calculations then show that

$$P(h_{T-1}|D_T = D_\tau = 1) \geq \left[\left(\frac{1}{n}\right)^D \left(\frac{n-1}{n}\right)^{T-D-\tau}\right]\left(\frac{n-1}{n}\right)^{\tau-1}, \tag{13}$$

which immediately yields (11). ∎

Finally, we show that the investor puts most of her weight on $D_T = 1$, given $D_\tau = 1$, if she has ever seen a defection.

Lemma 5 *As in Lemma 2, let ϵ_1 be the probability of at least one additional tremble* after *period τ, given $D_\tau = 1$. Then $P(D_T = 1|h_{T-1},\, D_\tau = 1) = 1 + O(\epsilon_1)$.*

Proof.

$$P(D_T = 1|h_{T-1},\, D_\tau = 1)$$
$$= \frac{P(h_{T-1}|D_T = D_\tau = 1)P(D_T = 1|D_\tau = 1)}{P(h_{T-1}|D_T = D_\tau = 1)P(D_T = 1|D_\tau = 1) + P(h_{T-1}|D_T > D_\tau = 1)P(D_T > 1|D_\tau = 1)}. \tag{14}$$

Using (11), this becomes

$$P(D_T = 1|h_{T-1},\, D_\tau = 1) = \frac{B(1 - \epsilon_1)}{B(1 - \epsilon_1) + O(\epsilon_1)} = 1 + O(\epsilon_1), \tag{15}$$

which proves the result. ∎

Proof of the Theorem. Since h_{T-1} involves a defection in period τ, $D_T = 1$, combined with h_{T-1}, implies $D_\tau = 1$. Thus,

$$P(D_T = 1|h_{T-1}) = P(D_T = D_\tau = 1|h_{T-1})$$
$$= P(D_T = 1|h_{T-1},\, D_\tau = 1)P(D_\tau = 1|h_{T-1}) = (1 + O(\epsilon_1))(1 + O(\epsilon_2)), \tag{16}$$

by Lemma 5 and Lemma 3, respectively. The result now follows from the argument in the first paragraph after the statement of the theorem. ∎

We have shown that the sufficient condition $a \geq (1 - \frac{1}{n})b$ is also a necessary condition to support any sequential equilibrium in which the trustees adopt the contagious strategy. This result implies that it may be difficult to extend the concept of contagious equilibrium from the Prisoner's Dilemma game as in Kandori (1992) to other classes of games, as confirmed by the experimental evidence in Duffy, Xie, and Lee (2012).

One question we want to ask is which aspect of the game structure in the trust game yields this negative result compared to Kandori (1992). The trust game we examine in this paper is different from the Prisoner's Dilemma game in two respects: first, the players move sequentially rather than simultaneously; second, only the trustee (second mover) has a dominant strategy. In order to further investigate this question, consider the case where the stage game is replaced by the following normal-form version of the original sequential trust game.

	Keep	Return
Invest	$0, 1$	$b, 1 - b$
No Invest	$a, 0$	$a, 0$

Given that the investor is the first mover of the sequential trust game, the information available to the investor before she makes an investment decision in each period is the history of previous outcomes in both sequential trust game and the normal form version of the trust game. The investor would learn more in the normal form from periods where she did not trust, but that does not affect the proof. Therefore, it is easy to verify that Theorem 1 still holds if we change the stage game from the sequential trust game to the normal-form trust game. That is, the condition $a \geq (1 - \frac{1}{n})b$ is also a necessary condition to support any sequential equilibrium in which the trustees adopt the contagious strategy in a random matching model where the stage game is the normal-form trust game. This observation implies that the negative result of Theorem 1 comes from the fact that the investor (first mover) of the sequential trust game does not have a dominant strategy rather than from the fact that the trust game has a sequential structure of moves.

5 Conclusion

This paper provides the first analysis of the development of trust and reciprocity among strangers based on the concept of contagious equilibrium. The trust game represents a one-sided incentive problem rather than a two-sided incentive problem such as the Prisoner's Dilemma game. Kandori (1992) discusses cooperation in the one-sided incentive problem under local information processing, in which each player carries a label and the label is revised based on the actions taken

by the matched players in each period. In contrast, we consider the one-sided incentive problem under the most restrictive information structure, that is, players neither observe the outcomes of games in which they are not involved, nor recognize the identity of their opponents.

We find that, in order to sustain the contagious strategy as a sequential equilibrium, it is *necessary* for the investors to have a high outside option. In fact, the investor's payoff from the outside option must converge to that from trust and reciprocity as the community size increases to infinity. Therefore, a contagious equilibrium only supports trust when trust contributes almost nothing to the investors' payoffs. This negative result suggests that when the game structure changes, specifically, when one player does not have a dominant strategy, it is unlikely that the concept of contagious equilibrium will extend from the Prisoner's Dilemma game to other classes of games such as the trust game.

The necessary condition is based on the assumption that the trustees adopt the contagious strategy. For future research, it would be interesting to investigate whether strategies with more forgiveness can sustain trust and reciprocity in equilibrium. On the other hand, we may consider situations with less information restrictions and where different mechanisms coexist and interact with each other. For instance, what happens if some investors can observe the trustees' previous behavior, and use a strategy conditional on reputational information, while other investors cannot observe the information and follow the contagious strategy?

References

[1] Berg, J., Dickhaut, J. and McCabe, K. (1995). "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10, 122-142.

[2] Bolton, G. and Ockenfels, A. (2000). "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review,* 90, 166-193.

[3] Bolton, G., Katok, E. and Ockenfels, A. (2005). "Cooperation among Strangers with Limited Information about Reputation," *Journal of Public Economics,* 89, 1457-1468.

[4] Camera, G. and Casari, M. (2009). "Cooperation among Strangers under the Shadow of the Future," *American Economic Review*, 99 (3), 979-1005.

[5] Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction, Russell Sage Foundation and Princeton University Press.

[6] Duffy, J. and Ochs, J. (2009). "Cooperative Behavior and the Frequency of Social Interaction," *Games and Economic Behavior*, 66(2), 785-812.

[7] Duffy, J., Xie, H. and Lee, Y.-J. (2011). "Social Norms, Information, and Trust among Strangers: Theory and Evidence," *Economic Theory*, forthcoming.

[8] Ellison, G. (1994). "Cooperation in the Prisoner's Dilemma with Anonymous Random Matching," *Review of Economic Studies,* 61, 567-588.

[9] Engle-Warnick, J. and Slonim, R. (2006). "Learning to Trust in Infinitely Repeated Games," *Games and Economic Behavior,* 54, 95-114.

[10] Ghosh, P. and Ray, D. (1996). "Cooperation in Community Interaction without Information Flows," *Review of Economic Studies,* 63, 491-519.

[11] Greif, A. (1993). "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition," *American Economic Review,* 83, 525-548.

[12] Kandori, M. (1992). "Social Norms and Community Enforcement," *Review of Economic Studies,* 59, 63-80.

[13] Kranton, R. (1996). "The Formation of Cooperative Relationships," *Journal of Law, Economics, and Organization*, 12, 214-233.

[14] Kreps, D., Milgrom, P., Roberts, J. and Wilson, R. (1982). "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 27, 245-252.

[15] Kreps, D. and Wilson, R. (1982). "Reputation and Imperfect Information," *Journal of Economic Theory*, 27, 253-279.

[16] Kreps, D. and Wilson, R. (1982). "Sequential Equilibria," *Econometrica*, 50(4), 863-94.

[17] Milgrom, P. and Roberts, J. (1982). "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory*, 27, 280-312.

[18] Okuno-Fujiwara, M. and Postlewaite, A. (1995). "Social Norms and Random Matching Games," *Games and Economic Behavior*, 9, 79-109.

[19] Rigdon, M., McCabe, K. and Smith, V. (2007). "Sustaining Cooperation in Trust Games," *Economic Journal,* Vol. 117, No. 522, 991-1007.

[20] Tirole, J. (1996). "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)," *Review of Economic Studies,* 63, 1-22.

[21] Van Huyck J., Battalio, R. and Walters, M. (1995). "Commitment versus Discretion in the Peasant-Dictator Game," *Games and Economic Behavior*, 10(1), 143-170.

[22] Van Huyck J., Battalio, R. and Walters, M. (2001). "Is Reputation a Substitute for Commitment in the Peasant-Dictator Game?" Unpublished Manuscript.