This paper provides theoretical background for some effects of social networks on trust. We study the implications of a model with rational actors in two settings with three actors. In the first setting, there are two trustees who are involved in transactions with one truster implying that the truster has an exit option. In the second setting, two trusters play with one trustee, which gives the trusters options for voice, i.e., complaining and informing each other about the trustee’s behavior. We compare these models with a baseline model in which there is only one truster and one trustee. It turns out that the opportunities for placing and honoring trust do not change for the exit model compared to the baseline model. The opportunities for trust in the voice model differ from the baseline model only if both trusters inform each other at a rate that is high enough. Only if the possibilities for receiving information and transmitting information are large enough for both trusters, trust will increase due to the information exchange possibilities in the voice model.

JEL-classification: C72

Key Words: trust, social networks, non-cooperative game theory.
1. INTRODUCTION

The relation between social networks and trust is a quite complex one, because social networks may have different types of effects on trust. For example, networks can provide options for exit out of a relation (Lahno, 1995), for obtaining information from previous behavior of other actors (learning, see Buskens, 1999), or for controlling partners through reputational sanctions when they act untrustworthy (Coleman, 1990; Kreps, 1990a; Raub & Weesie, 1990; Buskens, 1999). In this paper, I model these three effects in similar models that are all based on a well-known game-theoretic model with incomplete information to analyze the finitely repeated Prisoner’s Dilemma (see Kreps et al., 1982).

In a trust situation, a truster has to decide first whether or not to trust a trustee. Placing trust allows the trustee to choose between honoring and abusing trust, which would not have been possible if the truster would not have placed trust. The truster regrets placing trust if trust is abused, but benefits from honored trust. The trustee can earn an extra profit from abusing trust in a transaction. Therefore, if a transaction is happening in isolation, i.e., with a trustee that is unknown to the truster and the truster and trustee do not expect to meet at any time after the transaction, the trustee is expected to take this extra profit. Consequently, the truster will not place trust in such a situation. Formally, a trust situation can be represented by a Trust Game (Dasgupta, 1988; Kreps, 1990b), which will be introduced in Section 2.

I am convinced that many exchange relations or transactions resemble a trust situation as given above. For example, an actor who wants to buy a used car knows that the dealer has an incentive to sell the car for a price that is too high. The dealer might conceal essential information about the
history of the car, for example, whether the car has been involved in a major accident that has caused vital damage to the car. The buyer is assumed to be unable to deduce this information by inspection of the car. However, the buyer is also uncertain about the extent to which the dealer has an incentive to sell at a high price. The dealer might be concerned about future business with this buyer or acquaintances of this buyer. Moreover, the dealer might just feel guilty if he would conceal information. Consequently, the buyer is not only uncertain about the quality of the car (which creates the trust problem), but also about the precise incentives of the dealer.

Exchanges or transactions among actors hardly ever happen in isolation. Most transactions are embedded in a larger social setting, for example, because actors have more transactions with each other (temporal embeddedness) or because third parties are connected to the actors in a transaction (network embeddedness). These two types of embeddedness might affect the behavior of the actors involved in a transaction (Raub & Weesie, 1993). In this paper, I want to concentrate on three effects of third parties using the smallest possible networks that exceed the dyadic level: triads. This provides the opportunity to reach some analytic results, which are difficult to obtain for larger systems. Moreover, it provides possibilities for testing the theory in laboratory experiments.

First, I discuss a baseline model in which only one trustor and one trustee are involved in a finite number of transactions. Second, the model will be extended with an exit option for trustors (Hirschman, 1970, Chapter 2). An exit option increases the set of feasible sanctions for the trustor because a trustor cannot only sanction the trustee by withholding trust, but she can switch to another trustee as well. Third, a voice option is incorporated for trustors (Hirschman, 1970, Chapter 3). In this case, there are two
trusters who are involved in transactions with the trustee and they can communicate about the behavior of the trustee. This provides the truster with additional opportunities to control the trustee, because she can inform the other truster about the behavior of the trustee, and the second truster may refrain from placing trust as a result of this information. Moreover, the trusters can learn about the trustee’s incentives to abuse trust from each others experiences with the trustee in the past. Because the learning opportunities of one truster coincide with the control opportunities of the other truster, I can distinguish between these two kinds of effects only by allowing asymmetric information flows between the two trusters. In this way, I hope to disentangle whether trust can be facilitated better by control or learning and what the relative impact of these two effects is if they are combined.

Summarizing, this paper studies whether or not a truster can trust a trustee depending on a number of characteristics of the setting. First, I determine how trust depends on the payoffs in the game and the extent to which one truster and one trustee are involved in a series of transactions. Then, I analyze how these conditions change if trusters have an exit option or a possibility of communication among each other.

The paper is outlined as follows. Section 2 summarizes the theory on the finitely repeated Trust Game with incomplete information, which is called the baseline model. The results resemble results for the finitely repeated Prisoner’s Dilemma’s (Kreps et al., 1982). The result for the finitely repeated Trust Games with two actors is not new, but it is given here as a reference to be compared with the other models. In Section 3, an exit option is introduced. I assume that there is a second trustee and the truster can choose between these trustees and change to the other trustee
in the course of the game. Section 4 analyzes the finitely repeated Trust Game with two trusters. Two trusters play with one trustee and they might communicate about the behavior of the trustee between the stages of play. Section 5 summarizes the main findings and testable hypotheses that follow from the models presented and gives indications for further theoretical developments. Finally, some comments are made about possibilities to test these models.

2. THE BASELINE MODEL

The model developed here closely resembles reputation models in the economic literature, in particular the model develop by Kreps and Wilson (1982a) on the finitely repeated Prisoner’s Dilemma. The constituent game is the Trust Game as shown in Figure 1. The Trust Game can be seen as a one-sided version of the Prisoner’s Dilemma. I assume that there are two types of trustees: “friendly” and “payoff-maximizing.” Both types of actors are utility maximizers, but friendly trustees will always feel guilty to such an extent that $u_2(P_2) < u_2(T_2) < u_2(R_2)$ (see, for example, Guth & Kliemt, 1994; Snijders, 1996). Consequently, friendly trustees will never abuse trust. Therefore, in the discussion of the equilibria, I do not need to consider the behavior of the friendly trusters in detail. It is immediately clear that since friendly trustees have no short-term incentive to abuse trust, they certainly will not have a long-term incentive to abuse trust. One could model these trustees as if they do not have the option to abuse trust, which would essentially lead to the same equilibria. Payoffs represent utility for the payoff-maximizing trustee. I assume that also for the trusters, payoffs represent utility. Therefore, with some abuse of notation, I will denote the utilities for the trusters and payoff-maximizing trustees by the
payoffs as given in the Trust Game. The ex-ante proportions of the two types of trustees are given by $\pi_f$ and $\pi_p$, $\pi_f + \pi_p = 1$. These proportions are common knowledge.

FIGURE 1 ABOUT HERE

The game starts with a move by Nature deciding which type of trustee is going to play. The trustee is “friendly” with a probability $\pi_f$ and “payoff-maximizing” with a probability $\pi_p$. The trustee knows his type, but the truster does not observe the type of the trustee. No discounting is assumed for payoffs received in later stages of the game. The stages are labeled backwards, i.e., the last stage is stage 1 and the first stage is stage $N$. Furthermore, $\pi_{t,n}$ is the belief of the truster that the trustee is of type $t$ at the start of stage $n$; $p_n$ is the probability that a truster places trust in stage $n$; $q_n$ is the probability that a payoff-maximizing trustee honors trust in stage $n$. I define:

$$RISK = \frac{P_1 - S_1}{R_1 - S_1} \quad \text{and} \quad TEMP = \frac{T_2 - R_2}{T_2 - P_2}. \quad (1)$$

RISK represents the risk for trusters to place trust and TEMP represents the temptation for the trustee to abuse trust (see Snijders, 1996; Snijders & Keren, 1999).

In these models, payoff-maximizing actors mimic the behavior of other types if these other types are expected to obtain better outcomes. For example, if there would be only payoff-maximizing trustees in a finitely repeated Trust Game, a backward induction argument would imply that no trust is possible (see, for example, Luce & Raiffa, 1957). This results in a payoff $P_i$ for the players in every stage. However, friendly trustees will receive $R_2 > P_2$ if they can convince the truster that they are actually
friendly trustees. In a mixed population of trustees, payoff-maximizing trustees will mimic friendly trustees until the end of the game comes close and they will try to exploit the trusters at the end. In such a model with more types of trustees, the counterintuitive backward induction result of the finitely repeated Trust Game is likely to be resolved and it seems that behavior of subjects in experiments can be understood quite well with this model (Camerer & Weigelt, 1988; Neral & Ochs, 1992).

From earlier research it is known that the following beliefs and strategies form a sequential equilibrium (Kreps & Wilson, 1982b) for the finitely repeated Trust Game with incomplete information as described above.

Beliefs of the truster

- If the truster does not place trust in stage $n+1$, then $\pi_{t,n} = \pi_{t,n+1}$. 
- If the truster places trust in stage $n+1$ and the trustee honors trust in that stage, the truster updates beliefs: $\pi_{f,n} = \max (\text{RISK}^n, \pi_{f,n+1})$.
- If the truster places trust in stage $n+1$ and the trustee abuses trust, $\pi_{f,n} = 0$.

Strategy of the truster

If $\pi_{f,n} > \text{RISK}^n$, the truster places trust in stage $n$. If equality holds, the truster randomizes with a probability $p_n = \text{TEMP}$. Otherwise, the truster does not place trust.

Strategy of a payoff-maximizing trustee

- If $\pi_{f,n} \geq \text{RISK}^{n-1}$, a payoff-maximizing trustee honors trust.
- If $\pi_{f,n} < \text{RISK}^{n-1}$, a the payoff-maximizing trustee honors trust with a probability $q_n = \frac{\pi_{f,n}}{1 + \pi_{f,n} \left( \frac{1}{\text{RISK}^{n-1}} - 1 \right)}$. 

Theorem 2.1. The strategies and beliefs above constitute a sequential equilibrium in the finitely repeated Trust Game with friendly and payoff-maximizing trustees.

Proof. The proof for $N = 2$ is given by Bower et al. (1997). They also provide the result for $N > 2$, which follows from an induction argument.

The equilibrium can be described as follows. There are three periods in the game. The game starts with a number of stages in which trust is placed and honored. At a certain stage, the trustee starts to randomize. In this second period, both truster and trustee randomize until the truster does not place trust or the trustee abuses trust. After any instance of no trust or abuse of trust the third and last period starts in which there will be no more trust until the end of the game. The equilibrium is visualized in Figure 2. Table 1 provides the explanation for this figure as well as for the figures that will illustrate the more complex equilibria in Section 4. Figure 2 shows that the actors remain in the randomization period as long as the trustee honors trust and the truster places trust.

FIGURE 2 ABOUT HERE

TABLE 1 ABOUT HERE

The equilibrium demonstrates that whether or not a truster trusts the trustee depends mainly on RISK, the number of stages to be played until the end of the game, and the proportion of friendly trustees in the total population. The higher the risk of placing trust for the truster, the smaller
the number of stages to be played, and the smaller the percentage of friendly trustees, the shorter the first (trust) period of the game will be. When the payoff-maximizing trustee starts to randomize, the truster learns in the sense that she updates her belief about the probability that she is playing with a friendly trustee. This probability increases gradually as long as trust is honored and becomes zero as soon as trust is abused.

It is striking that the payoffs of the truster incorporated in RISK are of major importance to determine how close the game can approach the end before trust breaks down. This contrasts with well-known results for the infinitely repeated Trust Games with complete information and discounting. In the latter case, trust is explained completely by the discount factor $w$ of the trustee and the payoffs of the trustee. Namely, $w$ should be larger than TEMP (see, for example, Kreps, 1990a). TEMP plays only a role in the randomization of the trusters in the game analyzed here. If TEMP is larger, the probability that the truster places trust in the randomization period is larger. This is in a sense counterintuitive, since this implies that trusters place more trust if the temptation of the trustee for abusing trust is larger. The reason is that the truster’s randomization probability is chosen such that the trustee is indifferent in the stage before. Therefore, the larger the temptation for the trustee, the higher the probability needs to be that trust is placed in the following stage again, to make him indifferent between honoring and abusing trust. In the following two sections, I will discuss changes in the predictions if the truster has an exit option or if there are two trusters who can communicate about the trustee’s behavior.

3. THE EXIT MODEL
The first extension is an exit option. Now, one truster plays a finitely repeated Trust Game with two (or more) trustees. Between every two stages of the game, the truster has an additional choice, namely, which trustee she wants to encounter in the next stage. It is assumed that a trustee who is not playing with the truster in a stage receives a payoff $P_2$, the non-playing payoff that corresponds with the no-trust payoff. Again there are friendly and payoff-maximizing trustees. The initial beliefs of the truster about the distribution of trustees are given. The costs of switching trustees for the truster are small compared to the payoffs in the game, such that switching costs only affect the truster’s behavior if she would be indifferent in case there would be no costs of switching. The truster starts with choosing the trustee she wants to play with and, thereafter, she chooses whether she does or does not trust this trustee.

The main equilibrium for this game can be derived easily by reconsidering the equilibrium for the baseline model. It is known that payoff-maximizing trustees mimic friendly trustees as long as that is better for them. At a certain moment, the payoff-maximizing trustee switches to randomization to convince the truster that he is really friendly. If this randomization leads to an abuse of trust, the truster probably would switch to another trustee, but at that moment it is too late for the truster to experiment with a new trustee because the future is too short, and payoff-maximizing trustees will not honor trust anymore. Formally, since the trustee abused trust in the foregoing stage, it holds in the present stage $n$ that $\pi_f < \text{RISK}_n$, which implies that the truster will not trust a trustee she did not encounter before. Thus, as long as all trustees always honor trust, the truster has no reason to change trustees. As soon as she discovers that she is playing with a payoff-maximizing trustee, it is too late to trust any trustee. One might expect
that if the trustee’s non-playing payoff is smaller than $P_2$, the sanction from changing trustees increases for the untrustworthy trustee. However, such a payoff change for the trustee does only affect the randomization probability of the truster. Moreover, the probability that the truster places trust after trust is honored in the randomization period even decreases, because the expected payoff for the trustee after honoring trust should be the same as his expected payoff after abusing trust. Thus, since the expected payoff after abusing trust decreases as a result of the additional sanction, this implies that the expected payoff after honoring trust should decrease as well. The truster will not be better off as a result of this change, because the expected payoff for her in the randomization period remains $P_1$. Consequently, the equilibrium strategy for the truster is choosing one of the trustees to begin with and play the equilibrium strategy of the baseline model with this trustee. The chosen trustee will also follow the equilibrium strategy of the baseline model.

If there would be no costs of switching trustees for the truster, the truster would be indifferent between switching or staying with the same trustee in the early stages of the game (because she did not learn anything yet) or after trust has been abused (because she will not trust any truster). This would cause uncertainty on the side of the trustees about whether the truster will continue to play with him. By imposing some costs of switching, which seems a reasonable assumption, I circumvent this problem.

It might be clear that this outcome for the inclusion of an exit option is not a satisfying outcome. An exit option increases the sanction opportunities for the truster and decreases the dependence of the truster on the trustee. It should be noted that also in the repeated Trust Game with complete information, an exit option would not have had an effect on the
solution. This is often attributed to the fact that all the trustees are the same and there is complete information about their characteristics. I have shown here that assuming incomplete information does not automatically solve this problem. Still, there are many other ways to model exit opportunities, although this easily leads to rather complicated models. One option would be to model a repeated Trust Game with monitoring problems in which the trustee may abuse trust unintentionally, while different trustees have different capabilities, i.e., some trustees abuse trust unintentionally with a higher probability than others. Other authors have developed models for exit in different settings related to trust and other cooperation problems (see Schüßler, 1989; Vanberg & Congleton, 1992; Lahno, 1995; Weesie, 1996; Macy & Skvoretz, 1998).

4. THE VOICE MODEL

Now, the baseline model is extended with a second truster. Both trusters play $N$ Trust Games with the trustee. One truster starts and in the following stage the other truster plays. Consequently, there will be $2N$ stages of play and the game starts with stage $2N$. Between two stages, it is decided by a probabilistic mechanism whether the former truster can transmit information to the other truster about former play. For now assume that, if possible, the former truster informs the latter truster truthfully about what happened in the last stage. Later, I will discuss this assumption in some more detail. A nice result will be that it does not matter for the equilibrium discussed here whether or not the trustee observes communication between the trusters as long as he knows the probabilities for information transmission. The probabilities for information transmission are $c_{ij}, i, j = 1, 2, i \neq j$. The transmission probabilities do not need to be
the same, so it does not need to be the case that $c_{ij} = c_{ji}$. This results in a small, asymmetric communication network. The communication probability from truster 1 to truster 2 $c_{12}$ is called the outdegree of truster 1 or the indegree of truster 2. Similarly, $c_{21}$ is the indegree of truster 1 and the outdegree of truster 2. Truster 1 controls the trustee through her outdegree $c_{12}$, while she learns about the trustee through her indegree $c_{21}$. If trust would be based primarily on what trusters hear about the trustee, truster 1’s trust would be affected more by her indegree $c_{21}$, while truster 2’s trust would be more affected by $c_{12}$. If trust would be based more on the potential sanctions imposed on the trustee after abusing trust, outdegree should be more important (i.e., $c_{12}$ should be more important for trust of truster 1 and $c_{21}$ should be more important for trust of truster 2).

In principle, it can occur that trusters will get mixed information about the trustee. One truster might be doing well herself with the trustee while the other truster’s trust is abused. Still, if a truster ever obtains information about any abuse of trust by the trustee, she knows that she is playing with a payoff-maximizing trustee and the backward induction argument comes into play, which implies that no more trust will be placed.

Before I describe the equilibrium, note that because truster 1 starts to play, truster 1 plays all the stages with the even numbers and truster 2 plays the stages with the odd numbers. First, I describe the beliefs of the trusters and, thereafter, the strategies of the players. I will not use different indices for the two trusters. The indices refer to the stage in the game which implies that even indices are related to truster 1, and odd indices are related to truster 2. Two cases need to be distinguished:

- Case 1: $c_{12} \geq$ TEMP and $c_{21} \geq$ TEMP;
- Case 2: $c_{12} <$ TEMP.
For the cases in which $c_{12} \geq \text{TEMP}$ and $c_{21} < \text{TEMP}$, the equilibrium resembles the equilibrium of Case 2, and the qualitative implications are the same as for Case 2. I will not discuss them in detail, but all analyses are available from the author.

**Beliefs of the trusters**

- If a truster does not place trust, does not obtain information from the other truster, or is informed that the other truster did not place trust, beliefs about the trustee do not change.
- If a truster knows about any abuse of trust by the trustee, $\pi_f = 0$ for all subsequent stages of this truster (the truster knows that the trustee is not friendly).
- In any case, if a truster receives information from the other truster about behavior of the trustee, she updates her belief about the probability that the trustee is a friendly trustee to the same value as the belief of the truster who transmits the information to her.
- Case 1: If the truster places trust in stage $n + 1$ and the trustee honors trust in that stage, the truster updates her belief to the value $\pi_{f,n} = \max(\text{RISK}^n, \pi_{f,n+1})$.
- Case 2: If truster 1 places trust and the trustee honors trust in stage $n + 1$, she updates her belief such that $\pi_{f,n} = \max(\text{RISK}^{(n-1)/2}, \pi_{f,n+1})$. If truster 1 did not place trust in stage $n + 1$, but she received information about honored trust in stage $n$ from truster 2, she will update her belief to $\pi_{f,n-1} = \max\left(\text{RISK}^{(n-1)/2}, \pi_{f,n}\right)$. If truster 2 received information from truster 1 before her stage, her belief will not change after her own stage. If she did not receive information from truster 1, she updates her belief after honored trust in stage $n + 1$ to $\pi_{f,n} = \max\left(\text{RISK}^{n/2}, \pi_{f,n+1}\right)$. 

Strategies of the trusters

- Case 1: If $\pi_{f,n} > \text{RISK}^n$, the truster places trust in stage $n$. If $\pi_{f,n} = \text{RISK}^n$, trusters 1 and 2 place trust with a probability $\frac{\text{TEMP}}{c_{12}}$ and $\frac{\text{TEMP}}{c_{12}}$ respectively. Otherwise, the trusters do not place trust.

- Case 2: If $\pi_{f,n} > \text{RISK}^{\left(\frac{n}{2}\right)}$, the truster places trust in stage $n$. If $\pi_{f,n} = \text{RISK}^{\left(\frac{n}{2}\right)}$, truster 1 places trust with a probability $\frac{\text{TEMP}(1-c_{12})-c_{12}\text{TEMP}+c_{12}^2}{(1-c_{12})^2-c_{12}\text{TEMP}+c_{12}^2}$ if she placed trust in her previous stage, but she will not place trust if she did not place trust in her previous stage. If $\pi_{f,n} = \text{RISK}^{\left(\frac{n}{2}\right)}$, truster 2 places trust if she just obtained information that truster 1 did not place trust and she places trust with a probability $\frac{\text{TEMP}-c_{12}}{1-c_{12}}$ if she did not receive any information from truster 1. Otherwise, the trusters do not place trust.

Strategy of a payoff-maximizing trustee

- Case 1: If $\pi_{f,n} < \text{RISK}^{n-1}$, the trustee honors trust with a probability $q_n = \frac{\pi_{f,n}}{1-\pi_{f,n}} \left(1-\frac{1}{\text{RISK}^{n-1}}\right)$. If $\pi_{f,n} \geq \text{RISK}^{n-1}$, a payoff-maximizing trustee honors trust.

- Case 2: If $\pi_{f,n} \geq \text{RISK}^{(n-2)/2}$, the trustee honors trust placed by truster 1. If $\pi_{f,n} < \text{RISK}^{(n-2)/2}$, the trustee honors trust placed by truster 1 with probability $q_n = \frac{\pi_{f,n}}{1-\pi_{f,n}} \left(1-\frac{1}{\text{RISK}^{(n-2)/2}}\right)$. The trustee repeats his move from stage $n$ with truster 1 in stage $n-1$ with truster 2. If truster 1 did not place trust in stage $n$, the trustee plays the move he would have played in stage $n$ in stage $n-1$ with truster 2.

Theorem 4.1. Considering the beliefs and strategies described above:

- The strategies and beliefs are a sequential equilibrium in the finitely repeated Trust Game with two trusters.
• In Case 1, if $\pi_f < \text{RISK}^{2N-1}$, there is one other sequential equilibrium, namely, never placing trust by both trusters and always abusing trust by the trustee.

• Otherwise, if $\pi_f \neq \text{RISK}^n$ for $n \leq 2N$, every sequential equilibrium for Case 1 has on-the-equilibrium-path strategies as described previously.

• For Case 2, there exists no equilibrium for which the trustee starts randomizing more than one stage later than in the equilibrium described here.

Proof. The proof of this theorem is presented in the appendix.

The most important substantive finding of the last theorem is that trust increases with the communication opportunities if and only if both trusters transmit information at a high rate, i.e., if $c_{12} \geq \text{TEMP}$ and $c_{21} \geq \text{TEMP}$ (Case 1). As in the baseline model, the equilibrium consists for both cases of three periods. In the first period, trust is placed and honored. After some time, the trustee starts to randomize in a stage with truster 1. Thereafter, the actors remain in the randomization period depending on whether or not trust honored, trust is placed, and information is communicated between the trusters.

FIGURE 3 ABOUT HERE

Now, I discuss the equilibrium for the two cases in detail. Case 1 resembles the game with one truster who plays $2N$ stages with the trustee (see Figure 3). It can easily be checked that if $c_{12} = c_{21} = 1$, the equilibrium exactly corresponds with the equilibrium in the game with one truster. This implies that the trustee remains trustworthy with as many stages left
as in the baseline model. Consequently, both trusters can trust until they have each only half of the stages left compared to the baseline model. This implies that, compared to the game in which both trusters play in isolation with the trustee, there is considerably more trust in this game. The timing of the start of the randomization period still depends on the payoffs of the trustee and the proportion of friendly trustees in the total population. The higher RISK and the smaller the proportion of friendly trustees, the earlier the randomization period will start. The randomization probabilities of the trusters depend on the temptation for the trustee and the communication probabilities. The probability that a truster places trust in the randomization period increases with TEMP and decreases with the probability that she receives information from the other truster \( (c_{12}, c_{21}) \). These outcomes are in line with the results for the baseline model. A first intuition would probably predict that trust decreases with TEMP and increases with communication probabilities. However, since the randomization probabilities of the trusters are chosen such that the trustee is indifferent, the truster has to choose higher probabilities of placing trust if the situation for the trustee is better, i.e., if TEMP is larger and if communication between the trusters is less frequent. Therefore, as soon as the randomization period starts, trust breaks down as easily as in the baseline model, because information has to be transmitted between every stage for sustaining trust. If trust is not placed in any stage or information about honored trust is not passed to the following truster, no trust will be placed in the remainder of the game. Consequently, if the probabilities of communication are smaller, trust breaks down more easily, and the trustee needs to be compensated with higher probabilities of placing trust by the truster. The net result is
that the probabilities to go from the randomization period to the “no-trust” period are exactly the same as in the baseline model.

With respect to the information assumption, I can argue now that truthful information transmission is beneficial for both trusters in this case. Withholding information or transmitting wrong information is harmful for both trusters. The reason is that the equilibrium is built on truthful information exchange among the trusters. The trustee would stop honoring trust earlier if he cannot be sure that the trusters will exchange information truthfully, especially in the randomization period. The trustee does not need to observe actual communication between the trusters, because trusters do not place trust anymore as soon as information is not transmitted in the randomization period. Therefore, the trustee can infer whether information has been transferred from one truster to the other between two stages from the actions of the trusters.\(^5\)

In Case 2, truster 1 transmits information at a low rate. Note that it does not matter whether truster 2 transmits information at a high or a low rate. Now, the equilibrium basically breaks down to the equilibrium of the baseline model. The randomization period starts for both trusters with the same number of stages left as it would start in the baseline model. If \(c_{12} = 0\), the equilibrium reduces to the equilibrium of the baseline model for each of the trusters, whatever the value of \(c_{21}\) is. The strategy of the trustee prescribes that he starts randomizing in a stage with truster 1 who has a low transmission rate, and he repeats the move that is the outcome of this randomization in the following stage with the other truster. In Figure 4, it can be seen that the trustee does not randomize in the encounters with truster 2. By starting randomizing in a stage with truster 1, he has a relatively large probability to abuse trust twice, and obtain the \(T_2\) payoff.
twice. As before it can be seen that trust can be placed longer if RISK is smaller. The most important condition for the randomization probabilities of the trusters is that they have to be (on average) slightly smaller than TEMP, and the extent to which they have to be smaller depends on $c_{12}$. Consequently, again the randomization probabilities of the trusters increase with TEMP. The randomization probability of truster 2 decrease with $c_{12}$, while the randomization probability of truster 1 equals TEMP if $c_{12} = 0$ and if $c_{12} = \text{TEMP}$. In between these two extreme values for $c_{12}$, the randomization probability first decreases and then increases again. Consequently, the probability that truster 1 places trust in the randomization period is always smaller than TEMP. The randomization probabilities of both trusters do not depend on $c_{21}$.

FIGURE 4 ABOUT HERE

In Case 2, communication between the trusters after every stage is not necessary to sustain trust in the randomization period. Information transmission from truster 2 to truster 1 is worthless. Due to the strategy of the trustee, the information of truster 2 is not new for truster 1 or she will not use the information to adapt her behavior. Figure 4 demonstrates that after trust of truster 1 is honored, the behavior of truster 1 in her next stage does not depend on whether or not truster 2 placed trust neither on whether or not information is communicated between the trusters. However, truster 2 might profit from the information she receives, because she does not need to randomize if she obtains information about honored trust, and she certainly will receive $R_1$ in these stages. In the game with one truster, any defection automatically implies no more trust. Here, truster 2 can randomize or trust again even after she did not place trust before, if
she receives information that the trustee has honored trust placed by the other truster in the last stage. Note that this implies that truster 1 has been placing trust and her trust has been honored in all previous stages. Finally, truster 2 profits from the fact that the other truster might inform her about the first time that trust is abused, and she can avoid that the trustee takes advantage of her as well. All these profits are rather small because they depend on a relatively small transmission probability. Still, they cause some of the more complex formulations of the equilibrium that incorporate these profits.6

There exist equilibria comparable to the one for Case 2 if \( c_{12} \geq \text{TEMP} \) and \( c_{21} < \text{TEMP} \), although this case has to be treated again in distinct subcases. The main difference is that the trustee now starts randomizing in a stage played with truster 2 and repeats his move in the next stage with truster 1 in order to avoid the large information transmission probability of truster 1. The strategies of the trusters are reversed. Analyses become more complex because the game starts with a single stage by truster 1 followed by \( N - 1 \) pairs of stages in which the trustee plays the same move, and the game ends with one stage played with truster 2. Therefore, the first and last stage cause some additional considerations and complicate the formulation of the equilibria without changing the substantive implication that the randomization period starts at the moment that each truster has about the same number of stages left as she would have left in the baseline model.

Concerning the communication, it is worthwhile to note that trusters only need to communicate what happened in the last stage, because what happen in earlier stages can be derived from the last stage or the information does not influence the behavior of the truster. If communication is
assumed to be an actual choice, there may be other equilibria in Case 2, but these other equilibria hardly affect the substantial outcome of the game, because the trusters have no possibility to exchange information in such a manner that trust considerably increases compared to the baseline model. As far as the trustee is concerned, if trust is placed and the trustee has to move, his behavior is never affected by whether or not the trusters have communicated before. Therefore, it is not necessary to assume that the trustee observes communication among the trusters.

Concluding, I have shown that the effect of the network is large only if the information transmission rate is large in two directions. If there is an asymmetric relation, the network does not have an effect on the timing of the randomization period compared to the baseline model. In the concluding section, I will discuss what this implies for learning and control issues.

5. IMPLICATIONS AND CONCLUSIONS

This paper has analyzed effects of adding a third actor to a finitely repeated Trust Game with incomplete information. The main findings of the two-person case remain valid in the three-person games. Trust will be placed in more stages of the game if the truster has less to lose in each stage (RISK is smaller). Trust will be placed and honored in the early stages of the game. Thereafter, there is a randomization period in which eventually trust breaks down and after that no trust will be placed anymore. The probability that trusters place trust in the randomization period increases with the temptation of the trustee to abuse trust in a given stage (TEMP).

In addition to these known results from the two-person case, the paper provides new results for the three-person cases from which some are coun-
terintuitive. First, adding more trustees to the model providing the truster with an exit option does not have an effect on trust. This is a prediction that is compatible with the prediction of many economic models that assume complete information. Here, I have shown that incomplete information and having more types of trustees is not a sufficient condition for exit to be an essential element within the model. Consequently, if effects of exit opportunities are found in experiments, this would imply that the model still lacks some key elements. Second, adding a voice option for the trusters by including two trusters instead of one truster provides more trust only if both trusters inform each other at a high rate about the trustee’s behavior. What is considered to be high depends on the temptation for the trustee to abuse trust.

Linking these results to the discussion about learning and control in the beginning of the paper, it can be concluded that an exit option actually does not provide the truster with additional control opportunities in this model. This result is related to the fact that the timing of the start of the randomization period is determined by the payoffs of the truster rather than the payoffs of the trustee. Adding a voice opportunity by introducing a second truster to the model provides the trusters with more learning as well as control possibilities. I hinted in the beginning of this paper at the possibility to determine whether it is more important for trust to learn from the other truster about the behavior of the trustee or to sanction the trustee by informing the other truster. It turns out that the two aspects of voice need to be combined by the trusters to enable more trust in the trustee. If a truster only exercises control through voice but does not learn from information received from the other truster, control does not have an effect on trust and a truster’s trust is based only on her own stages with
the trustee. The same is true if the truster only learns from information transmitted by the other truster, but does not control the trustee herself by transmitting information about his behavior.

This last implication of the model provides an opportunity to test the model against models that assume other types of learning than Bayes’ updating such as learning by reinforcement (Roth & Erev, 1995; Erev & Roth, 1998). Models based on reinforcement predict that learning has an effect even in the absence of control options for a truster. Although it cannot be excluded that a separate learning effect could also follow from a model with Bayesian updating in which incomplete information is introduced in a different way, such a result would be in favor of a reinforcement model compared to the model developed in this paper. A second type of models that would likely predict effects of exit and more pronounced effects of learning are models in which trusters sometimes experience bad outcomes although the trustee did not intentionally abuse trust. Such situations are expected to be described better by models on monitoring problems (see, for example, Radner, 1981; Porter, 1983; Green & Porter, 1984).

A disadvantage of the type of model presented in this paper is that after any abuse of trust, there will never be any doubt about the type of trustee the truster is playing with. As a result, the period of the game in which there is any learning is limited. Learning is expected to be more important in models in which trustees do not have a fixed type, but there is a small probability that the type of a trustee changes and the trusters cannot observe these changes. Then, trusters can never be completely sure about the type of a trustee they continue updating their beliefs throughout the game. Moreover, trustees are not perfectly able to reveal or conceal their type, so every experience is worthwhile to the trusters (cf. Tadelis,
However, these two models have also two important drawbacks. First, the trusters are myopic or play just one time and, second, the outcomes of any encounter are public knowledge. The model in this paper is a first attempt to understand the importance of communication of outcomes among different players in a game. I doubt whether random type changes of the trustee would cause essential changes to the outcomes, since all (types of) players in the game want to place and honor trust in a considerable part of the game. The randomization period is the only period in which different types of trustees act differently, and this is generally only a short period in the game.

Clearly, the discussion about effects of changes in assumptions can be extended much further. However, I think that we lack considerable knowledge about what actually reasonable assumptions are especially related to information availability of actors, information exchange among actors, and how actors actually use this information (belief updating, sanctioning). Especially, in the light of the finding that implications of models such as the finitely repeated Trust Games change dramatically if we move from complete information to incomplete information, it seems necessary to search for (sets of) assumptions under which the findings are more robust to changes that should not have large effects. Although I do not want to start a philosophical discussion at the end of the paper, I would favor a careful design of experiments that do not only allow for testing the implications of theoretical models, but also allow for testing some assumptions especially about how actors use and react to information they obtain in the games they play. Camerer et al. (1993) provide an example for how experiments can be designed in which it is possible to follow more or less the decision
making process of the trusters rather than only the final decisions of the actors.

**APPENDIX: PROOF OF THEOREM 4.1**

Proof. The theorem is proved by induction on the number of stages each truster plays. Therefore, we will first reformulate the theorem for $N = 1$.

**Beliefs of the trusters**

The only occasion that a truster who still has to play updates her beliefs is when truster 2 receives information about the behavior of the trustee in stage 2. If this information indicates that trust is honored by the trustee and $c_{12} < \text{TEMP}$, the trustee is definitely friendly ($\pi_{f,1} = 1$). If she receives information about trust honored by the trustee and $c_{12} \geq \text{TEMP}$, she will update her belief $\pi_{f,1} = \max(\text{RISK}, \pi_f)$. If she obtains information about abused trust $\pi_{f,1} = 0$.

**Strategies of the trusters**

If $c_{12} < \text{TEMP}$, truster 1 places trust if $\pi_f \geq \text{RISK}$. If $c_{12} \geq \text{TEMP}$, truster 1 places trust if $\pi_f \geq \text{RISK}^2$. Otherwise, truster 1 does not place trust. Truster 2, subsequently, places trust if $\pi_{f,1} > \text{RISK}$. If $\pi_{f,1} = \text{RISK}$, truster 2 places trust with a probability $\min(1, \frac{\text{TEMP}}{c_{12}})$. Otherwise, truster 2 does not place trust.

**Strategy of a payoff-maximizing trustee**

If $c_{12} < \text{TEMP}$, a payoff-maximizing trustee always abuses trust. If $c_{12} \geq \text{TEMP}$, a payoff-maximizing trustee honors trust in the stage 2 if $\pi_f \geq \text{RISK}$ and honors trust with a probability $\frac{\pi_f}{1 - \pi_f} \left(\frac{1}{\text{RISK}} - 1\right)$ if $\pi_f < \text{RISK}$. The trustee abuses trust in stage 1.

**Theorem A.1.** Consider the beliefs and strategies described above.
These beliefs and strategies constitute a sequential equilibrium in the two-stage finitely repeated Trust Game with two trusters (N = 1).

If \( \pi_f < \text{RISK} \) and \( c_{12} > \text{TEMP} \), there is one other sequential equilibrium, namely, never placing trust by both trusters, and always abusing trust by the trustee.

Otherwise, if \( \pi_f \neq \text{RISK}^n \) for \( n = 1, 2 \), every sequential equilibrium has the same on-the-equilibrium-path strategies as described before.

Checking that these beliefs are consistent with Bayes’ rationality is straightforward. If the trustee honors trust in stage 2 and truster 2 receives this information, then

\[
\pi_{f,2} = \frac{\Pr(C_2|\text{friendly})\Pr(\text{friendly})}{\Pr(C_2|\text{friendly})\Pr(\text{friendly}) + \Pr(C_2|\text{payoff-max.})\Pr(\text{payoff-max.})},
\]

which results in the given probabilities for all the relevant cases. The only case for which Bayes’ rule does not apply is if trust is abused in stage 2, while the payoff-maximizing trustee should honor trust with probability 1. The theorem poses that \( \pi_f = 0 \) in this out-of-equilibrium instance of updating.

Assume \( c_{12} < \text{TEMP} \). If truster 2 does not receive information about stage 2 from truster 1, she does not update her beliefs (\( \pi_{f,1} = \pi_f \)), which implies that she places trust if and only if \( \pi_f R_1 + (1 - \pi_f) S_1 \geq P_1 \Leftrightarrow \pi_f \geq \text{RISK} \), because the payoff-maximizing trustee will abuse trust. Truster 2 is indifferent if equality holds. If truster 2 obtains information about honored trust, the trustee must be a friendly trustee and, therefore, she will place trust in her stage. If truster 2 receives information about abused trust, then \( \pi_{f,1} = 0 \), which implies that she has no incentive to place trust. Because a payoff-maximizing trustee abuses trust in the stage 2 as well, truster 1 will place trust if and only if \( \pi_f \geq \text{RISK} \).
The payoff-maximizing trustee only has a choice in stage 2 if \( \pi_f \geq \text{RISK} \). Then, the payoff-maximizing trustee’s expected payoff from honoring trust is \( u_2(C_2) = R_2 + c_{12}T_2 + (1 - c_{12})T_2 \). On the other hand, \( u_2(D_2) = T_2 + c_{12}P_2 + (1 - c_{12})T_2 \). Consequently, \( u_2(C_2) > u_2(D_2) \Leftrightarrow c_{12} > \text{TEMP} \), which implies that a payoff-maximizing trustee will abuse trust, because \( c_{12} < \text{TEMP} \). Trivially, the best move for the payoff-maximizing trustee in the last stage is abusing trust.

Now, assume that \( c_{12} \geq \text{TEMP} \) and \( \pi_f \geq \text{RISK} \). If truster 2 does not receive any information about behavior of the trustee, she will again place trust if and only if \( \pi_f \geq \text{RISK} \) for the same reason as given above. If she receives information, she will place trust if \( \pi_{f,1} \geq \text{RISK} \). This is the case if she receives information about honored trust, because then \( \pi_{f,1} \geq \pi_f \geq \text{RISK} \). If she receives information about abused trust, she will not place trust because \( \pi_{f,1} = 0 \). Because the payoff-maximizing trustee honors trust in stage 2, Truster 1 is certainly better off placing trust compared to not placing trust.

The trustee’s expected payoff from honoring trust equals \( u_2(C_2) = R_2 + c_{12}T_2 + (1 - c_{12})(p_1T_2 + (1 - p_1)P_2) \) and his payoff from abusing trust is \( u_2(D_2) = T_2 + c_{12}P_2 + (1 - c_{12})(p_1T_2 + (1 - p_1)P_2) \). Therefore, \( u_2(C_2) > u_2(D_2) \Leftrightarrow c_{12} > \text{TEMP} \) and the trustee should indeed honor trust in stage 2.

Finally, assume that \( c_{12} \geq \text{TEMP} \) and \( \pi_f < \text{RISK} \). Now, truster 2 is indifferent between honoring and abusing trust if she receives information about honored trust, because \( \pi_{f,1} = \text{RISK} \). Therefore, randomizing is optimal for truster 2. The expected payoffs for truster 1 are \( u_1(C_1) = \pi_fR_2 + (1 - \pi_f)(q_2R_1 + (1 - q_2)S_1) \) and \( u_1(D_1) = P_1 \), where
28

\( q_2 = \frac{\pi_f}{1-\pi_f}(\frac{1}{\text{RISK}} - 1) \). Straightforward manipulations yields that \( u_1(C_1) > u_1(D_1) \Leftrightarrow \pi_f > \text{RISK}^2 \).

The trustee’s expected payoff from honoring trust equals \( u_2(C_2) = R_2 + c_{12}p_1T_2 + c_{12}(1-p_1)P_2 + (1-c_{12})P_2 \), where \( p_1 = \frac{\text{TEMP}}{c_{12}} \) and his payoff from abusing trust is \( u_2(D_2) = T_2 + P_2 \). Therefore, \( u_2(C_2) = u_2(D_2) \), which implies that the trustee is indifferent in stage 2.

Checking that the second equilibrium for \( \pi_f < \text{RISK} \) and \( c_{12} \geq \text{TEMP} \) is an equilibrium is straightforward. Other randomization probabilities are not possible because if the probability is larger, the trustee would always honor trust in the stage before, which leads to a contradiction, since truster 2 will never be able to update her belief and cannot randomize in her stage. If the probability is smaller the trustee will abuse trust in stage 2, which implies that truster 1 cannot place trust and, therefore, truster 2 cannot randomize. That there are no other sequential equilibria follows from the fact that moving backward in the game tree, all the other moves are uniquely determined.

Now, I continue with the proof of the general theorem. It can be shown easily that the beliefs of the trusters are consistent with Bayesian updating. An important observation is that the trusters in Case 2, can update their beliefs at most once within every pair of moves, because the trustee is randomizing only once and uses the outcome of the randomization in any of the two stages in which the truster places trust. Therefore, if truster 1 has placed trust in such a pair of moves, she will not obtain any new information if truster 2 informs her after truster 2’s move. Moreover, if truster 2 received in this pair of moves information about the behavior of the trustee in the previous stage, she will not update again her beliefs after her own stage. Note that if truster 1 does not place trust in a stage in which
she was indifferent, but she receives information from truster 2 about trust honed by the trustee, she will be indifferent again in her following stage. (This combination of moves can only occur in the randomization period.) Nevertheless, she will not place trust. Consequently, the information she obtains is irrelevant for further play.

I proceed to prove the theorem for these cases using the induction argument. It has to be proved for both cases that the first move of each truster and the first two moves by the trustee are their optimal moves considering the strategies of the other players as given. This implies that I will consider stages \(2N\) and \(2N - 1\). If both trusters have no incentive to withhold trust in stage \(2N - 2\) and \(2N - 3\), the trustee will honor trust in the first two stages. The reason is that he obtains two times \(R_2\) and, thereafter, he is in a similar position as before and will still be able to obtain short-term gains from abusing trust. If he abuses trust immediately, he will shift to \(P_2\) payoffs thereafter, losing the opportunity to receive some \(R_2\) payoffs. Consequently, the trusters will place trust in their first stages of play if they are going to place trust in stages \(2N - 2\) and \(2N - 3\), because they can never do better than obtaining \(R_1\). Thus, the key cases to be checked are those where the trusters probably do not place trust in their first or second stage of play.

Let's start with Case 1. Because the trustee’s strategy is exactly the same as in Theorem 2.1, he acts as if he is playing \(2N\) stages with the same truster. Consequently, it follows from Theorem 2.1 that the trusters play an optimal response, because they also play the game as if only one truster is involved with the only exception that they use other randomization probabilities. Now it only has to be checked whether the behavior of the trustee is optimal.
Assume $\text{RISK}^{2N-2} < \pi_f < \text{RISK}^{2N-3}$. From the induction assumption, it is known that the trustee is indifferent in stage $2N - 2$ and will be randomizing in this stage. For the calculation of the expected payoffs, it can be assumed that he abuses trust in this stage and in stage $2N - 3$. Throughout this proof I use this argument to calculate the payoffs in the stages that are covered by the induction assumption, which mostly implies that payoffs are calculated for the trustee abusing trust and truster not placing trust, because then the payoffs for the remaining stages can be determined straightforwardly. Now, consider the first two moves of the trustee.

$$u_2(C_2C_2) = R_2 + R_2 + T_2 + c_{12}P_2$$
$$+ (1 - c_{12})\left(p_{2N-3}T_2 + (1 - p_{2N-3})P_2\right) + (2N - 4)P_2$$
$$= 2R_2 + T_2 + (2N - 3)P_2;$$

$$u_2(C_2D_2) = R_2 + T_2 + c_{21}P_2 + (1 - c_{21})T_2 + (2N - 3)P_2$$
$$< 2R_2 + T_2 + (2N - 3)P_2;$$

$$u_2(D_2C_2) = T_2 + c_{12}P_2 + (1 - c_{12})R_2 + (2N - 2)P_2 < u_2(D_2D_2);$$

$$u_2(D_2D_2) = T_2 + c_{12}P_2 + (1 - c_{12})T_2 + (2N - 2)P_2 < R_2 + T_2$$
$$+ (2N - 2)P_2.$$

Consequently, two times honoring trust is equilibrium play for the trustee here, which is the behavior specified in the theorem. Now, assume that $\text{RISK}^{2N-1} < \pi_f < \text{RISK}^{2N-2}$.

$$u_2(C_2C_2) = R_2 + R_2 + c_{21}\left(p_{2N-3}T_2 + (1 - p_{2N-3})P_2\right)$$
$$+ (1 - c_{12})P_2 + (2N - 3)P_2;$$

$$u_2(C_2D_2) = R_2 + T_2 + (2N - 2)P_2;$$
TRUST IN TRIADS

\[ u_2(D_2C_2) = T_2 + c_{12}P_2 + (1 - c_{12})R_2 + (2N - 2)P_2 < u_2(D_2D_2); \]
\[ u_2(D_2D_2) = T_2 + c_{12}P_2 + (1 - c_{12})T_2 + (2N - 2)P_2 \]
\[ < R_2 + T_2 + (2N - 2)P_2. \]

Therefore, only \( C_2C_2 \) and \( C_2D_2 \) can be equilibrium strategies for the first two moves of the trustee. Moreover, \( u_2(C_2C_2) = u_2(C_2D_2) \leftrightarrow p_{2N-2} = \frac{\text{TEMP}_{c_{21}}}{T_2} \), which is the randomization probability the truster uses in this stage. This proves that the randomizing strategy of the trustee is indeed an equilibrium strategy.

Finally, I consider \( \pi_f < \text{RISK}^{2N-1} \), which implies that the trustee randomizes in the first two stages according to the theorem. Consequently, it needs to be checked whether the trustee is indifferent between playing any pair of moves in the first two stage assuming that the trusters place trust. Otherwise the trustee does not have a choice anyway.

\[ u_2(C_2C_2) = R_2 + c_{12}\left(p_{2N-1}R_2 + (1 - p_{2N-1})P_2\right) \]
\[ + (1 - c_{12})P_2 + c_{12}c_{21}p_{2N-1}\left(p_{2N-2}T_2 + (1 - p_{2N-2})P_2\right) \]
\[ + (1 - c_{12}c_{21}p_{2N-1})P_2 + (2N - 3)P_2; \]
\[ = R_2 + \text{TEMP}(R_2 - P_2) + \text{TEMP}^2(T_2 - P_2) + (2N - 1)P_2 \]
\[ = T_2 + (2N - 1)P_2; \]
\[ u_2(C_2D_2) = R_2 + c_{12}\left(p_{2N-1}T_2 + (1 - p_{2N-1})P_2\right) + (1 - c_{12})P_2 \]
\[ + (2N - 2)P_2 = T_2 + (2N - 1)P_2; \]
\[ u_2(D_2C_2) = T_2 + (2N - 1)P_2; \]
\[ u_2(D_2D_2) = T_2 + (2N - 1)P_2. \]
Above I substituted $p_{2N-1} = \frac{\text{TEMP}_{c_{12}}}{c_{12}}$ and $p_{2N-2} = \frac{\text{TEMP}_{c_{21}}}{c_{21}}$, which represent the strategies of the trusters. It can be concluded that a payoff-maximizing trustee is indifferent among any combination of moves in the first two stages and, consequently, may randomize in both stages. The existence of a second equilibrium in this situation follows from the same considerations as given for $N = 1$. That there are no other equilibria follows from the uniqueness of the corresponding equilibrium in the baseline model.

For Case 2, I first consider the strategies of the trusters in stage $2N$ and $2N - 1$. If the trustee honors trust with certainty, the trusters place trust. This should be the case because $R_1$ is their best possible payoff. Therefore, I only need to consider the case were the trustee randomizes in stage $2N$, i.e., if $\pi_{f,2N} < \text{RISK}^{N-1}$ in stage $2N$ with truster 1.

Then, it holds for truster 1 that

\[
\begin{align*}
\quad u_1(C_1) &= \pi_f R_1 + (1 - \pi_f) \left(q_{2N} R_1 + (1 - q_{2N}) S_1\right) + (N - 1) P_1; \\
\quad u_1(D_1) &= NP_1 \\
\quad u_1(C_1) &> u_1(D_1) \Leftrightarrow \pi_{f,2N} > \text{RISK}^N.
\end{align*}
\]

Thus, this shows that truster 1 is acting optimal. The truster is indifferent if equality holds. The expected payoffs for truster 2 if truster 1 did not place trust or if truster 2 did not receive information about the trustee’s behavior in stage $2N$ are

\[
\begin{align*}
\quad u_1(C_1) &= \pi_f R_1 + (1 - \pi_f) \left(q_{2N} R_1 + (1 - q_{2N}) S_1\right) + u_1(\text{after } C_1), \\
\quad u_1(D_1) &= P_1 + u_1(\text{after } D_1),
\end{align*}
\]

where $u_1(\text{after } C_1) = u_1(\text{after } D_1)$, because the behavior of truster 1 does not depend on whether truster 2 informs her about the trustworthiness of
TRUST IN TRIADS

the trustee. If truster 1 did not place trust once, she will never place trust again although the trustee might still be trustworthy toward truster 2. If truster 1 did place trust, she knows what the trustee will play in the following stage with truster 2, so nothing will change for her whether or not truster 2 informs her before her next stage. Therefore, the probability that truster 2 obtains information about the trustee in a foregoing stage does not depend on her own behavior. Consequently, $u_1(C_1) > u_1(D_1) \Rightarrow \pi_{f,2N} > \text{RISK}^N$, and truster 2 is indifferent if equality holds. Note that if truster 2 is informed about the behavior of the trustee in the previous stage, she knows what the trustee will play against her and she should play the best response against that move. This will increase her payoff as long as there is a probability that she obtains information from the other truster about trust honored by the trustee.

Now consider the four possible strategies for the trustee for the first two stages. Again I only need to consider the situation in which the trusters probably do not place trust in their second stage, i.e., if $\pi_{f,2N} < \text{RISK}^{N-1}$. Otherwise, the best option for the trustee is to honor trust in the first two stages.

Before the calculation, note that for the randomization probabilities of the trusters holds

$$p_{2N-2} + p_{2N-3} - c_{12}p_{2N-2}p_{2N-3} = 2\text{TEMP} - c_{12}, \quad (A.1)$$

$$0 < p_{2N-3} = \frac{\text{TEMP}(1 - c_{12}) - c_{12}\text{TEMP} + c_{12}^2}{(1 - c_{12}) - c_{12}\text{TEMP} + c_{12}^2} < \frac{\text{TEMP}}{1 - c_{12}} \quad (A.2)$$

$$0 < p_{2N-2} = \frac{\text{TEMP} - c_{12}}{1 - c_{12}} < \frac{\text{TEMP} - c_{12}}{1 - c_{21}}. \quad (A.3)$$

Condition (A.1) follows from straightforward manipulation. For Condition (A.2), one has to realize that the numerator equals $\text{TEMP} - \text{TEMP}^2 +$
(\text{TEMP} - c_{12})^2 > 0 \text{ and that } -c_{12}\text{TEMP} + c_{12}^2 < 0. \text{ Condition (A.3) follows from } c_{12} < c_{21}.

Now, it can be shown that the trustee is indifferent between playing $D_2D_2$ and $C_2C_2$, and the other combinations provide him with a lower payoff.

$$u_2(C_2C_2) = R_2 + R_2 + p_{2N-2}T_2 + (1 - p_{2N-2})P_2 + p_{2N-2}c_{12}P_2$$
$$+ (1 - p_{2N-2}c_{12})\left(p_{2N-3}T_2 + (1 - p_{2N-3})P_2\right)$$
$$+ (2N - 4)P_2$$
$$= 2R_2 + \left(p_{2N-2} + p_{2N-3} - c_{12}p_{2N-2}p_{2N-3}\right)(T_2 - P_2)$$
$$+ (2N - 2)P_2 = 2T_2 - c_{12}(T_2 - P_2) + (2N - 2)P_2;$$

$$u_2(C_2D_2) = R_2 + T_2 + c_{21}P_2 + (1 - c_{21})\left(p_{2N-2}T_2 + (1 - p_{2N-2})P_2\right)$$
$$+ (2N - 4)P_2$$
$$= R_2 + T_2 + p_{2N-2}(1 - c_{21})(T_2 - P_2) + (2N - 2)P_2$$
$$< R_2 + T_2 + (\text{TEMP} - c_{12})(T_2 - P_2) + (2N - 2)P_2$$
$$= 2T_2 - c_{12}(T_2 - P_2) + (2N - 2)P_2;$$

$$u_2(D_2C_2) = T_2 + c_{12}P_2 + (1 - c_{12})R_2 + P_2$$
$$+ (1 - c_{12})^2\left(p_{2N-3}T_2 + (1 - p_{2N-3})P_2\right)$$
$$+ \left(1 - (1 - c_{12})^2\right)P_2 + (2N - 4)P_2$$
$$= T_2 + R_2 - c_{12}(R_2 - P_2) + (1 - c_{12})^2p_{2n-3}(T_2 - P_2)$$
$$+ (2N - 2)P_2$$
$$< T_2 + R_2 - c_{12}(T_2 - P_2) + c_{12}(T_2 - R_2)$$
$$+ (1 - c_{12})(T_2 - R_2) + (2N - 2)P_2$$
$$= 2T_2 - c_{12}(T_2 - P_2) + (2N - 2)P_2;$$

$$u_2(D_2D_2) = T_2 + c_{12}P_2 + (1 - c_{12})T_2 + (2N - 2)P_2$$
\[ 2T_2 - c_{12}(T_2 - P_2) + (2N - 2)P_2. \]

Consequently, \( u_2(C_2C_2) = u_2(D_2D_2) \) and it is indeed optimal for the trustee to randomize between these two pairs of moves here. Finally, it has to be shown for the induction assumption to be applicable and to ensure perfectness that if the trustee randomizes and truster 1 does not place trust that the trustee also in that situation does not have an incentive to deviate in his subsequent stage with truster 2. Realize first that if truster 1 does not place trust, she will never place trust again. Therefore, the game essentially reduces to a two-person game for which it is known that in equilibrium the truster places trust with a probability TEMP in the randomization phase. The probability that truster 2 places trust equals \( c_{12} + (1 - c_{12}) \frac{\text{TEMP} - c_{12}}{1 - c_{12}} = \text{TEMP} \), which proves that the trustee is also indifferent between honoring and abusing trust if truster 1 did not place trust.

There are no equilibria in which the trustee starts randomizing later because the trustee can never rely on the probability \( c_{12} \) to make truster 2 indifferent. If that would be necessary, he would always prefer to abuse trust placed by truster 1. It can also be checked easily that for \( N = 2 \) and \( \pi_f < \text{RISK} \), the trustee has to start randomizing or abusing trust at least in his first encounter with truster 2, which is stage 3. Using again an induction argument it follows that the trustee cannot continue to place trust with certainty beyond an encounter with truster 2 for which \( \pi_f < \text{RISK}^n \) in stage \( \frac{n-1}{2} \).
ACKNOWLEDGEMENT

Stimulating comments of and discussions with Werner Raub, Chris Snijders, Jeroen Weesie, and an anonymous referee are gratefully acknowledged. Financial support was provided by the Netherlands Organization for Scientific Research (NWO) under grant S 96-168 and PGS 50-370. This paper was largely written during research visits at the Stanford Graduate School of Business and the Department of Sociology at the University of Chicago. I thank both institutes for their support.

References


1 This result is discussed in detail by Bower et al. (1997) and is mentioned earlier in a more informal way by Camerer and Weigelt (1988), Dasgupta (1988), and Neral and Ochs (1992).

2 It is known from earlier experiments that there is a considerable number of trustees who actually honor trust also if the Trust Game or similar games are played only once (McKelvey & Palfrey, 1992; Berg et al., 1995; Snijders, 1996; Guth et al., 1997; Snijders & Keren, 1999).

3 It is assumed that the payoffs are the same in all the stages. The analysis can be generalized for arbitrary payoffs in all stages, but this will considerably complicate the notation.

4 If \( N \) is large enough and there are at least some friendly trustees, there will always be a (maybe short) period of trust in this finitely repeated game, because \( \lim_{N \to \infty} \text{RISK}^N = 0 \).

5 The reason that there is another equilibrium in one instance is that truster 2 has to start randomizing in her first move. In the baseline model, the truster has to incorporate this randomizing period to ensure that she can place trust in earlier stages. Thus, the randomization is optimal because of gains to be earned before the randomization starts. However, if truster 2 has to start randomizing in her first stage, her expected payoff is the same if she does not place trust or uses another randomization probability. It can easily be checked that another positive randomization probability cannot lead to an equilibrium. If truster 2 does not place trust, truster 1 cannot place trust as well. This other equilibrium is weakly Pareto inferior to the first one, because both truster 1 and the trustee are worse off, while truster 2 has the same expected payoff. Uniqueness in all other situations follows from the fact that the game is similar to the baseline
model and that both trusters need the randomization period to sustain trust in the earlier stages of the game.

6 The equilibrium in Case 2 is not unique. The two trusters have to coordinate their randomization probabilities such that the conditions as they are mentioned in the proof are met. However, they still have some freedom how to choose these randomization probabilities. If communicating would be a choice itself, there is another equilibrium in which truster 1 chooses not to communicate information to truster 2. Truster 1 has no incentive to transmit any information to truster 2. Then, both trusters would randomize with a probability TEMP and truster 2 loses the small profits from communication. However, essentially the equilibrium remains the same, and as the last statement in the theorem says, there are no equilibria for which the randomization period starts considerable later than for the equilibrium described here.
- T indicates that the truster places trust; N that the truster does not place trust; H indicates that the trustee honors trust; and A that the trustee abuses trust.

- Normal arrows indicate that the actor who is going to play knows what happened in the last move. Dotted arrows indicate that he or she does not know what happened in the last move. Of course, the trustee always knows what the truster did in her last move. In the baseline model, the truster is always informed about the last move of the trustee as well. Information is not an issue for arrows that enter or leave a boxes.

- Arrows that split in two parts indicate randomization moves.

- In all positions within a figure indicated by T*, the game continues in the same way. This is also true for positions indicated with N*.

TABLE 1.
Legend belonging to the figures

FIG. 1. Extensive form of a Trust Game with incomplete information, where $R_i > P_i, (i = 1, 2)$, $P_1 > S_1$, $T_2 > R_2$, and $F$ is the distribution over the types of trustees.
FIG. 2. Equilibrium play in the finitely repeated Trust Game with one truster and one trustee.

FIG. 3. Equilibrium play in the finitely repeated Trust Game with two trusters and one trustee (Case 1; note that the two trusters are completely interchangeable in this case; further explanation can be found in Table 1 and the main text).
First period: Always placing and honoring trust until $\pi_f < \text{RISK}^{(n-2)/2}$

Third period: No more trust

Second period: Randomization

Truster 1 Trustee Trustee Trust 1 Trustee Truster 1 Trustee Truster 2 Trustee

**FIG. 4.** Equilibrium play in the finitely repeated Trust Game with two trusters and one trustee (Case 2; further explanation can be found in Table 1 and the main text).