Trust in Triads: Effects of Exit, Control, and Learning

Vincent Buskens*
Department of Sociology
Utrecht University
Utrecht, The Netherlands

December 15, 1999

---

*Stimulating comments of and discussions with Werner Raub, Chris Snijders, and Jeroen Weesie 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.

Running Head: Trust in Triads

Abstract

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 trustor implying that the trustor has an *exit* option. In the second setting, two trustors play with one trustee, which gives the trustors 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 trustor 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 trustors 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 trustors, trust will increase due to the information exchange possibilities in the voice model.

JEL-classification: C72

Keywords: Trust, social networks, non-cooperative game theory.

Address:
Department of Sociology,
Utrecht University,
Heidelberglaan 1,
3584 CS Utrecht, the Netherlands.
Email: v.buskens@fss.uu.nl

Trust in Triads

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 and Weesie 1990; Buskens 1999). In this paper, we 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, Milgrom, Roberts, and Wilson 1982). However, before proceeding we want to clarify what we mean by trust.

The way we define a trust situation resembles Coleman's definition (1990: Chapter 5). In a trust situation, a trustor first has to decide 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 trustor would not have placed trust. The trustor 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 trustor and the trustor and trustee do not expect to meet at any time after the transaction, the trustee is expected to take this extra profit. Consequently, the trustor will not place trust in such a situation. Formally, a trust situation can be represented by a Trust Game (Dasgupta 1988; Kreps 1990b) as is shown in Figure 1.

We are 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. This last element is introduced in the Trust Game by assuming the the trustee is chosen by Nature from a distribution F and the utility of an outcome $u_2(X_2), X_2 = P_2, R_2, T_2$ for the trustee depends on characteristics of the trustee unknown to the trustor.

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 transactions (*network embeddedness*). These two types of embeddedness might affect the behavior of the actors involved in a transaction (Raub and Weesie 1993). In this paper, we want to concentrate on three effects of third parties using the smallest possible networks that exceed the dyadic level: *triads*. This provides the op-

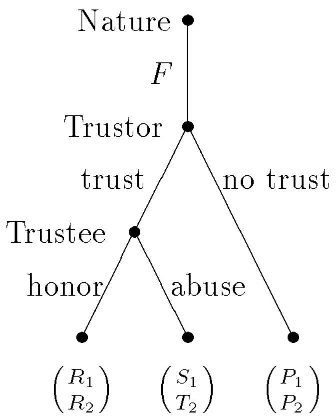

Figure 1 *Extensive form of a Trust Game with incomplete information, where* $R_i > P_i, (i = 1, 2)$, $P_1 > S_1$, $T_2 > R_2$, *and* F *is the distribution over the types of trustees.*

portunity 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, we discuss a baseline model in which only one trustor and one trustee are involved in a finite number of transactions. Second, we extend the model 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. If a trustor can only withhold trust, she cannot obtain more than P_1 herself. However, if the trustor has an alternative partner, she can manage to obtain R_1 if she trusts the other partner, while the first trustee remains without the possibility of having a transaction. Third, we incorporate a *voice* option for trustors (Hirschman 1970: Chapter 3). In this case, there are two trustors who are involved in transactions with the trustee and they can communicate about the behavior of the trustee. This provides the trustor with additional opportunities to *control* the trustee, because she can inform the other trustor about the behavior of the trustee, and the second trustor may refrain from placing trust as a result of this information. Moreover, the trustors 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 trustor coincide with the control opportunities of the other trustor, we can distinguish between these two kinds of effects only by allowing asymmetric information flows between the two trustors. In this way, we 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, we study in this paper whether or not a trustor can trust a trustee depending on a number of characteristics of the setting. First, we determine how trust depends on the payoffs in the game and the extent to which one trustor and one trustee are involved in a series of transactions. Then, we analyze how these conditions change if trustors have an exit option or a possibility of communication among each other.

Although an earlier model in which a whole network of trustors played Trust Games with one trustee is seemingly more complex (Buskens and Weesie 1999; Buskens 1999:

Chapter 3), we investigate triads here to address some of the problems that are part of the assumptions in the earlier model. In the earlier model, complete information is assumed about all elements of the model including the incentives for the trustee. This implies that there is no reason for exit, because all trustees are assumed to be the same and, consequently, act the same and are completely interchangeable. Moreover, the assumptions imply that trustors cannot learn about the trustee. As a result, the model only predicts control effects depending on the extent to which trustors can inform other trustors about the behavior of the trustee. Moreover, trustors act sequentially in this former model, i.e., a trustor whose transactions with the trustee ends may transmit her information about the behavior to the following trustor, but she will not return as a partner of the trustee or she returns without any knowledge about her earlier transactions with the trustee. These two assumptions are relaxed in the new models paying the price of reducing the number of actors to three.

Assuming incomplete information on the side of the trustors allows the trustors to learn about the incentives of the trustee during the game. If they encounter an unfavorable type of trustee, they may want to exit and search for another trustee. We assume that there are two types of trustees. A good trustee feels so guilty about abusing trust that his actual utility from abusing trust is smaller than his utility from honoring trust $u_2(T_2) < u_2(R_2)$ (see, for example, Güth and Kliemt 1994; Snijders 1996). Such a trustee will never abuse trust in a Trust Game even in isolation. 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 and Palfrey 1992; Berg, Dickhaut, and McCabe 1995; Snijders 1996; Güth, Ockenfels, and Wendel 1997; Snijders and Keren 1999). We assume that these good trustees exist next to payoff-maximizing trustees for whom the payoffs represent the utility they derive from the outcomes of the game.

The paper is outlined as follows. Section 2 provides 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.[1] In Section 3, we introduce an exit option. We assume that there is a second trustee and the trustor 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 trustors. Two trustors 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 and possibilities to test the models.

2 Theoretical Background and the Baseline Model

The model we develop 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. 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 and Raiffa 1957). This results in a payoff P_i for the players in every stage. However, good trustees will receive $R_2 > P_2$ if they can convince the trustor that they are actually good trustees. In a mixed population of trustees, as we will see, payoff-maximizing trustees will mimic good trustees until the end of the game comes close and they will try to exploit the trustors at the end. In such a model with more types of trustees, the counter-intuitive backward induction result of the finitely repeated Trust Game is likely to be resolved and it seems that we can understand behavior of subjects in experiments quite well with this model (Camerer and Weigelt 1988; Neral and Ochs 1992).

As mentioned before, we assume that there are two types of trustees: good and payoff-maximizing. Both types of actors are utility maximizers, but good trustee will always feel

---

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

guilty to such an extent that $u_2(P_2) < u_2(T_2) < u_2(R_2)$. Consequently, good trustees will never abuse trust. Therefore, in the discussion of the equilibria, we do not need to consider the behavior of the good trustors in detail. It is immediately clear that since good 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. We assume that also for the trustors, payoffs represent utility. Therefore, with some abuse of notation, we will denote the utilities for the trustors and payoff-maximizing trustees by the payoffs as given in the Trust Game. We assume that the payoffs are the same in all the stages. The analyses can be generalized for arbitrary payoffs in all stages, but this will considerably complicate the notation. The ex-ante proportions of the two types of trustees are given by π_g and π_p, $\pi_g + \pi_p = 1$. These proportions are common knowledge. The game starts with a move by Nature deciding which type of trustee is going to play. The trustee is good with a probability π_g and payoff-maximizing with a probability π_p. The trustee knows his type, but the trustors do not observe the type of the trustee. We assume no discounting 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 trustor that the trustee is of type t at the start of stage n. Then, p_n is the probability that a trustor places trust in stage n and q_n is the probability that a payoff-maximizing trustee honors trust in stage n. We define:

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

RISK represents the risk for trustors to place trust and TEMP represents the temptation for the trustee to abuse trust (see Snijders 1996).

We show that the following beliefs and strategies form a sequential equilibrium (Kreps and Wilson 1982b). By definition, $\pi_{t,N} = \pi_t$ with $t = g, p$. Clearly, $\pi_{g,n} = 0$ for all following stages as soon as placing trust is followed by abusing trust. The trustor then knows that she is not dealing with a good trustee. If trust by the trustor has thusfar been answered by honoring trust and if the trustor trusted for the last time in game k, $\pi_{g,n} = \max(\text{RISK}^{k-1}, \pi_k)$. As long as the trustor does not trust the trustee, $\pi_{t,n} = \pi_t$ for $t = g, p$. This description corresponds to the following recursive definition regarding the beliefs of the trustor during the game:

BELIEFS OF THE TRUSTOR

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

STRATEGY OF THE TRUSTOR

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

STRATEGY OF A PAYOFF-MAXIMIZING TRUSTEE

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

Theorem 1 *The strategies and beliefs above constitute a sequential equilibrium in the finitely repeated Trust Game with good and payoff-maximizing trustees.*

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

The equilibrium can be described as follows. The game starts with a number of stages in which trust is placed and honored. At a certain stage the trustee starts to randomize. Thereafter, both trustor and trustee randomize until one of them does not place trust or abuses trust. Then, there will be no more trust until the end of the game. It does not matter how many kinds of trustees there are, if N is large enough and we have at least some good trustees, there will always be a (maybe short) period of trust in this finitely repeated game, because $\lim_{N\to\infty} \mathrm{RISK}^N = 0$.[2]

Whether or not a trustor trusts the trustee depends mainly on RISK, the number of stage to be played until the end of the game, and the proportion of good trustees in the total population. The more the trustor has to lose in each stage the less likely it is that she places trust and the earlier in the game trust is expected to break down. The more stages to be played to more likely it is that the trustor trust the trustee. And, the larger the proportion of good trustees, the more stages can be played before the randomization period starts. The equilibrium implies that the trustor will never place trust if the trustee has ever abused trust, which means that the trustor uses a *conditionally cooperative* strategy. Third, when the payoff-maximizing trustee starts to randomize, the trustor is learning in the sense that her belief that she is dealing with a good trustee increases as long as the trustee honors trust. It is questionable whether this can be interpreted as increasing trust in the trustee, because on the other hand, the trustor starts randomizing as well and is, therefore, less likely to place trust. Fourth, the equilibrium moves to a situation in which no trust is placed any more as soon as either the trustee abuses trust or the trustor does not place trust. This implies that "trust" breaks down quickly, namely, after one defection, while the belief increased gradually. Lahno (1998) argues that the randomization period will be very short for "normal" parameter values, which indicates that this model predicts only a limited extent of "dynamics of trust."

It is striking that the payoffs of the trustor 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

---

[2] In an earlier version we included bad trustees who take every (short-term) opportunity to abuse trust placed by the trustees. However, bad trustees will always reveal themselves in their first encounter with a trustor. Thereafter, the trustor will not trust them anymore. Consequently, such trustees make the formulation of the equilibria in the stage N of a game more complicated, but do not really contribute to the insights in the mechanisms that influence trust.

discount factor w of the trustee and the payoffs of the trustee. Namely w should be larger than TEMP (see, for example, Kreps 1990a; Snijders 1996). TEMP plays only a role in the randomization of the trustors in the game analyzed here. This might also explain why the effect of RISK is larger than the effect of TEMP in the one-shot Trust Game, as is found by Snijders (1996).[3]

The following theorem states that the equilibrium we discussed above is actually the only relevant equilibrium for this game, which implies that we have a rather strong prediction for behavior in this game.

Theorem 2 *If* $\pi_g \neq \mathrm{RISK}^n$ *for* $n \leq N$*, every sequential equilibrium has on-the-equilibrium-path strategies as described previously.*

Proof. Again, the proof for $n = 2$ is given by Bower, Garber, and Watson (1997). The generalization for $n > 2$ follows from an induction argument. □

There are some similar variations possible in the equilibrium as mentioned by Kreps and Wilson (1982a). Updating of beliefs when Bayes' updating is not possible is not uniquely determined for this equilibrium. If a trustee abuses trust before a payoff-maximizing trustee should abuse trust according to the equilibrium, the probability that the trustor is playing with a good trustee does not need to be set to zero. As long as π_g is set smaller than RISK^n, where n is the number of the stage in which the anomaly happened, the strategies are still in equilibrium and consistent with the beliefs. If one explicitly excludes an option to abuse trust for the good trustee, setting $\pi_g = 0$ is the only plausible adaptation of the beliefs in this case. However, if we define a good trustee only as a trustee who does not have an incentive to abuse trust in the constituent game although

---

[3] In an infinitely repeated Trust Games with incomplete information operationalized in the same way as above and a discount factor w, we obtain an equilibrium in which trust is always placed and honored if $w >$ TEMP. If $w <$ TEMP, the trustor will place trust in the first stage if $\pi_g > \frac{P_1 - S_1}{R_1 - S_1 + \frac{w}{1-w}(R_1 - P_1)}$, but a payoff-maximizing trustee will abuse trust in the first stage. Therefore, also in such a game, the payoffs of the trustee are of major importance to decide whether a cooperative equilibrium exists, and if this is not the case there should be enough good trustees for the trustor to find out in the first stage whether she is dealing with a good or a payoff-maximizing trustee.

she has a behavioral option to abuse trust, the updating is less clear because the act of the trustee is off-equilibrium for any trustee, which gives no obvious reason why trust should be abused by the good or the payoff-maximizing trustee. Still, we are discussing here hypothetical situations that should not occur in equilibrium.

If we want to test this model in an experiment, one could explicitly inform the trustors about the distribution of trustees and tell them that they are playing with "computerized" opponents that are chosen from a given distribution and have the described properties. However, such an experiment will hardly have the flavor of a social interaction. The other option is assuming that trustors have a distribution in mind that resembles the distribution given in the model. One might relax the assumption that the distribution the trustors have in mind is common knowledge. The solution of the model does not change fundamentally if we assume that there exist optimistic and pessimistic trustors, and the optimistic trustors think that the probability that they are dealing with a good trustee is larger than for the pessimistic trustors. Assuming that the trustee does not know whether he is dealing with an optimistic or a pessimistic trustor, he will play in such a way that his behavior is adapted for the beliefs of a pessimistic trustor, which causes the optimistic trustor to place trust as long as trust is not abused.

We will always have difficulties to explain off-equilibrium behavior such as trustors who place trust again after having been deceived ones. Qualitatively, we expect the probability that trustors place trust and trustees honor trust to decrease if the end of the game comes nearer. This finding is confirmed in earlier papers by Camerer and Weigelt (1988) and Neral and Ochs (1992). Moreover, we expect that these probabilities decreases with RISK. An unexpected implication is that during the randomization phase, the probabilities that trustors place trust are larger if TEMP is larger. This can be understood as follows. The randomization probability is chosen such that the trustee is indifferent in the stage before. If the trustee has a large incentive to abuse trust, a larger probability that trust will be placed in the following stage is necessary to make him indifferent. Neral and Ochs (1992) claim that in their experiment the hypothesis about TEMP is rejected. However, since the prior beliefs of the subjects about trustworthiness of the trustees are not known, it is difficult to determine in which stages the effect of TEMP should be found.

Except for the prediction related to TEMP, predictions similar to the predictions from this model will result from models that assume simple heuristics for the subjects to determine their behavior as long as these heuristics take into account some endgame effects and concerns about the payoffs. Therefore, these predictions do not necessarily need to be results of completely rational behavior but can also be explained by, for example, choice-reinforcement learning models (see Roth and Erev 1995; Erev and Roth 1998). By extending the model with a third actor, it is not only possible to obtain insights in effects of social networks, but we might also be able to find more predictions that are unlikely to result from other learning models.

3 The Exit Model

The first extension we will discuss is an exit option. Now, the finitely repeated Trust Game is played with two (or more) trustees. Between every two stages of the game, the trustor has an additional choice, namely, which trustee she wants to encounter in the next stage. We assume that a trustee who is not playing with the trustor in a stage receives a payoff P_2, the reservation payoff that corresponds with the no-trust payoff. We include the same types of trustees. The initial beliefs of the trustor are for each trustee the same. There are some costs of switching trustees for the trustor. The trustor starts with choosing the trustee she wants to play with and, thereafter, she chooses whether she does or does not trust this trustee.

Analyzing this game in detail is not necessary, since the equilibrium can be derived easily by reconsidering the equilibrium for the baseline model. We know that payoff-maximizing trustees mimic good trustees as long as that is better for them. At a certain moment, the payoff-maximizing trustee switches to randomization to convince the trustor that he is really good. If this randomization leads to an abuse of trust, the trustor probably would switch to another trustee, but at that moment it is too late for the trustor 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 $\pi_g < \text{RISK}^n$, which implies that the trustor will not trust a trustee she did not encounter before. Thus, the trustor's trust will not increase by the exit option.

If there would be no costs of switching trustees for the trustor, the trustor 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. This would cause uncertainty on the side of the trustees about whether the trustor will continue the play with him. By imposing some costs of switching, which seems a reasonable assumption, we 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 trustor and decreases the dependence of the trustor 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. We 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 lead easily 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 for trust and other cooperation problems (see Schüßler 1989; Vanberg and Congleton 1992; Lahno 1995; Weesie 1996; Macy and Skvoretz 1998).

4 The Voice Model

Now we extend the baseline model with a second trustor. Both trustors play N Trust Games with the trustee. One trustor starts and in the following stage the other trustor plays. Consequently, there will be $2N$ stages of play. Between two stages, it is decided by a probabilistic mechanism whether the former trustor can transmit information to the other trustor about former play. For now we assume that, if possible, the former trustor informs the latter trustor truthfully about what happened in the last stage. Later, we will discuss this assumption in some more detail. Furthermore, we will see that it does not matter for the equilibrium discussed here whether the trustee observes communication between the trustors as long as he knows the probabilities for information transmission. The probabilities for information transmission are $\pi_{ij}, i, j = 1, 2, i \neq j$. The transmission probabilities do not need to be the same, so it does not need to be the case that $\pi_{ij} = \pi_{ji}$. In this way, we have a small communication network. We call π_{12} the *outdegree* of trustor 1 or the *indegree* of trustor 2. Similarly, π_{21} is the indegree of trustor 1 and the outdegree of trustor 2. The average $\frac{\pi_{12}+\pi_{21}}{2}$ is the density of the network. If trust would be based primarily on what trustors hear about the trustee, indegree would have the most important effect on trust. While if trust would be based more on the potential sanctions imposed on the trustee after abusing trust, outdegree should be more important.

In principle, it can occur that trustors will get mixed information about the trustee. One trustor might be doing well herself with the trustee while the other trustor turns out to be defeated after some time. Still, a trustor 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.

For $N = 1$ we obtain already a remarkable result. It turns out that for $\pi_{12} \geq$ TEMP, the equilibrium of the game corresponds with the two times played Trust Game with one trustor with the exception that the second trustor may not be able to update her beliefs and has to make her decision based on the original distribution of trustees. For $\pi_{12} <$ TEMP, the network is not useful for the first trustor and this trustor can only place trust if she would have placed trust in the one-shot Trust Game. Consequently, also the second trustor can only place trust if she would have had that opportunity in the one-shot game, because she will not obtain information from the first trustor and, therefore, she cannot place trust herself. This implies that trust increases with the network in a very discrete manner. If the network connection is strong enough, the game resembles a finitely repeated game with two stages. Otherwise, none of the trustors is better off than playing alone with the trustee and this does not depend on whether the trustor is on the transmitting or receiving end of the tie that is relevant for $N = 1$. We now formulate the equilibrium beliefs and strategies for $N = 1$. Note that the game starts with stage 2.

Beliefs of the trustors

The only occasion that a trustor who still has to play updates her beliefs is when trustor 2 receives information about the behavior of the trustee in stage 2. If this information is positive and $\pi_{12} < \text{TEMP}$, the trustee has to be good ($\pi_{g,1} = 1$). If she receives positive information and $\pi_{12} \geq \text{TEMP}$, she will update her belief $\pi_{g,1} = \max\left(\frac{\pi_g}{1-\pi_g}, \text{RISK}\right)$. If she obtains information about abused trust $\pi_{g,1} = 0$.

Strategies of the trustors

If $\pi_{12} < \text{TEMP}$, trustor 1 places trust if $\pi_g \geq \text{RISK}$. If $\pi_{12} \geq \text{TEMP}$, trustor 1 places trust if $\pi_g \geq \text{RISK}^2$. Otherwise, trustor 1 does not place trust. Trustor 2 places trust if $\pi_{g,1} > \text{RISK}$. If $\pi_{g,1} = \text{RISK}$, trustor 2 places trust with a probability $\min(1, \frac{\text{TEMP}}{\pi_{12}})$. Otherwise, trustor 2 does not place trust.

Strategy of a payoff-maximizing trustee

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

Theorem 3 *With respect to the beliefs and strategies described above holds that:*

- *These beliefs and strategies constitute a sequential equilibrium in the two-stage finitely repeated Trust Game with two trustors ($N = 1$).*
- *If $\pi_g < \text{RISK}$ and $\pi_{12} > \text{TEMP}$, there is one other sequential equilibrium, namely, never placing trust by both trustors, and always abusing trust by the trustee.*

- *Otherwise, if* $\pi_g \neq \mathrm{RISK}^n$ *for* $n = 1, 2$, *every sequential equilibrium has the same on-the-equilibrium-path strategies as described before.*

Proof. The proofs of this theorem and the following theorem are presented in the appendix. □

The reason that there is another equilibrium for one instance, is that trustor 2 has to start randomizing in her first move. In the baseline model, the trustor 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 trustor 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 trustor 2 does not place trust, trustor 1 can not place trust as well. This other equilibrium is weakly Pareto inferior to the first one, because both trustor 1 and the trustee are worse off, while trustor 2 has the same expected payoff.

Now, we continue with the general case of the finitely repeated Trust Game with two trustors who play N stages each. The total game has $2N$ stages and the game starts with stage $2N$. Trustor 1 plays all the stages with the even numbers and trustor 2 plays the stages with the odd numbers. First, we describe the beliefs of the trustors and, thereafter, the strategies of the players. Note that we will not use different indices for the two trustors. The indices refer to the stage in the game which implies that even indices are related to trustor 1, and odd indices are related to trustor 2. We distinguish two cases:

- Case 1: $\pi_{12} \geq \mathrm{TEMP}$ and $\pi_{21} \geq \mathrm{TEMP}$;
- Case 2: $\pi_{12} < \mathrm{TEMP}$.

For the cases in which $\pi_{12} \geq \mathrm{TEMP}$ and $\pi_{21} < \mathrm{TEMP}$, the equilibrium resembles the equilibrium of case 2, and the qualitative implications are the same as for case 2. We will not discuss them in detail, but all detailed analyses are available from the author.

BELIEFS OF THE TRUSTORS

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

STRATEGIES OF THE TRUSTORS

- Case 1: If $\pi_{g,n} > \mathrm{RISK}^n$, the trustor places trust in stage n. If $\pi_{g,n} = \mathrm{RISK}^n$, trustor 1 and 2 places trust with a probability $\frac{\mathrm{TEMP}}{\pi_{21}}$ and $\frac{\mathrm{TEMP}}{\pi_{12}}$ respectively. Otherwise, the trustors do not place trust.
- Case 2: If $\pi_{g,n} > \mathrm{RISK}^{\lceil \frac{n}{2} \rceil}$, the trustor places trust in stage n. If $\pi_{g,n} = \mathrm{RISK}^{\lceil \frac{n}{2} \rceil}$, trustor 1 places trust with a probability $\frac{\mathrm{TEMP}(1-\pi_{12})-\pi_{12}\mathrm{TEMP}+\pi_{12}^2}{(1-\pi_{12})-\pi_{12}\mathrm{TEMP}+\pi_{12}^2}$ if she placed trust in her previous stage, but she will not place trust if she did not place trust in the her previous stage. If $\pi_{g,n} = \mathrm{RISK}^{\lceil \frac{n}{2} \rceil}$, trustor 2 places trust if she just obtained information that trustor 1 did not place trust and she places trust with a probability $\frac{\mathrm{TEMP}-\pi_{12}}{1-\pi_{12}}$ if she did not receive any information from trustor 1. Otherwise, the trustor does not place trust.

- Case 1: If $\pi_{g,n} < \mathrm{RISK}^{n-1}$, the trustee honors trust with a probability $q_n = \frac{\pi_{g,n}}{1-\pi_{g,n}}(\frac{1}{\mathrm{RISK}^{n-1}} - 1)$. If $\pi_{g,n} \geq \mathrm{RISK}^{n-1}$, a payoff-maximizing trustee honors trust.
- Case 2: If $\pi_{g,n} \geq \mathrm{RISK}^{(n-2)/2}$, the trustee honors trust placed by trustor 1. If $\pi_{g,n} < \mathrm{RISK}^{(n-2)/2}$, the trustee honors trust placed by the trustor 1 with probability $q_n = \frac{\pi_{g,n}}{1-\pi_{g,n}}(\frac{1}{\mathrm{RISK}^{(n-2)/2}} - 1)$. The trustee repeats his move from stage n with trustor 1 in stage $n-1$ with trustor 2. If trustor 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 trustor 2.

Theorem 4 *With respect to the beliefs and strategies described above holds that:*

- *The strategies and beliefs are a sequential equilibrium in the finitely repeated Trust Game with two trustors.*
- *In case 1, if* $\pi_g < \mathrm{RISK}^{2N-1}$*, there is one other sequential equilibrium, namely, never placing trust by both trustors and always abusing trust by the trustee.*
- *Otherwise, if* $\pi_g \neq \mathrm{RISK}^n$ *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.*

The most important substantial finding of the last theorem is that in case 1, the randomization period starts considerably later than in case 2. Case 1 resembles the situation in which their would be only one trustor playing $2N$ stages with the trustee. This implies that trust increases considerably with the communication opportunities if both trustors transmit information at a high rate, since trustors can place trust until they have each only half as many stages left compared to the baseline model. Still, 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 trustor, no trust will be placed in the remainder of the game. Therefore, truthful information transmission is beneficial for both trustors in this case. Withholding information or transmitting wrong information is harmful for both trustors. The role of RISK and TEMP are virtually the same as in the baseline model. Furthermore, it turns out that the probability that the trustors place trust in the randomization period decreases in their indegree. This is again a somewhat counterintuitive result, but it can be explained by the fact that these probabilities are chosen in such a way to make the trustee indifferent. Therefore, if the probabilities of communication are smaller, trust breaks down more easily, and the trustee needs to be compensated by this with higher probabilities of placing trust by the trustor. Since the trustors do never expect more than P_1 during the randomization process, their own outcomes do not depend on these probabilities.

A similar second equilibrium exists for case 1 as we indicated for $N = 1$, but note that this equilibrium occurs only if the randomization period would start in stage $2N - 1$ in the equilibrium described initially. Uniqueness in all other situations follows from the fact that the game is similar to the baseline model and that both trustors need the randomization period to sustain trust in the earlier stages of the game.

In case 2, trustor 1 transmits information at a low rate. Note that it does not matter whether trustor 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 trustors with the same number of stages left as it would start in the baseline model. The strategy of the trustee prescribes that he starts randomizing in a stage with trustor 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 trustor. By starting randomizing in a stage with trustor 1, he has a relatively large probability to abuse trust twice, and obtain the T_2 payoff twice. As before we see that trust can be placed longer if RISK is smaller. The most important condition for the randomization probabilities of the trustors is that they have to be (on average) slightly smaller than TEMP, and the extent to which they have to be smaller depends on π_{12}. Consequently, we see again that the randomization probabilities of the trustors increase with TEMP and decrease with π_{12}, while they do not depend on π_{21}.

In case 2, communication among the trustors is not necessary to sustain the trust. Information transmission from trustor 2 to trustor 1 is worthless due to the strategy of the trustee, because the information is not new for the trustor 1 or she will not use the information to adapt her behavior. However, the trustor with the high information transmission rate 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 trustor any defection automatically implies no more trust. However, the trustor with the high transmission rate 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 trustor in the last stage. Note that this implies that the informing trustor has been placing trust and her trust has been honored in all previous stages. Finally, the trustor with the high transmission rates profits from the fact that the other trustor 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 the relatively small transmission probability. Still, they cause some of the more complex formulations of the equilibrium that incorporate these profits. If the small information transmission probability would be equal to 0, the equilibrium condition reduce to the conditions for the baseline model.

The equilibrium in case 2 is not unique. The two trustors 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 trustor 1 chooses not to communicate information to trustor 2. Trustor 1 has no incentive to transmit any information to trustor 2. Then, both trustors would randomize with a probability TEMP and trustor 2 loses the small profits from the communication. However, essentially the equilibrium remains the same, and as the last statement in the theorem say, there are no equilibria for which the randomization period starts considerable later than for the equilibrium described here.

There exists a similar equilibrium as the equilibrium of case 2 if $\pi_{21} < TEMP$, 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 trustor 2 and repeats his move in the next stage with trustor 1 in order to avoid the large information transmission probability of trustor 1. The strategies of the trustors are reversed. Analyses become more complex because the game starts with a single stage by trustor 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 trustor 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 with for each trustors as many stages left as it would be in the baseline model.

Concerning the communication, it is worthwhile to note that trustors 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 trustor as is the case with trustor 1 in case 2. If communication is assumed to be an actual choice by the trustor given that the possibility occurs, we have seen that this leads to the same equilibrium in case 1, and that there may be other equilibria in case 2, but that these other equilibria hardly affect the substantial outcome of the game. As far as the trustee is concerned. In case 1, he will know whether information is transferred in the randomization state, because otherwise the trustor would not place trust. In case 2, his behavior is never effected by whether or not the trustors have communicated before a stage. Therefore, we do not need to assume that the trustee observes communication among the trustors.

Concluding, we 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 randomization period compared to the baseline model. This implies that if a trustor only exercises control but does not learn from information received from the other trustor, control does not have an effect on trust and a trustor's trust is based only on her own stages with the trustee. The same is true if the trustor only learns from information transmitted by the other trustor, but does not control the trustee herself by transmitting information about his behavior.

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 trustor 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 trustors 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 counter-intuitive compared to the more straightforward heuristics that we mentioned at the beginning of the paper. First, adding more trustees to the model providing the trustor 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 we 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 we would find effects of exit opportunities in experiments, this would imply that the model still fails some key elements. Second, adding a voice option for the trustors by including two trustors instead of one trustor provides more trust only if both trustors inform the other trustor about the trustee's behavior at a high rate. What is considered to be high depends on the temptation for the trustee to abuse trust. This implies that this small three-actor model does not predict effects of learning of control separately. The model predicts an interaction effect of control and learning, but no main effects of the two mechanisms themselves.

This last implication of our model gives us an opportunity to test our model against models that assume other types of learning than Bayes' updating such as learning by reinforcement (Roth and Erev 1995; Erev and Roth 1998). Models based on reinforcement would predict that learning has an effect even in the absence of control options for a trustor. Although it cannot be excluded that a seperate 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 trustors sometimes experience bad outcomes although the trustee did not intentionally abuse trust. Such situations are expected to be describedd better by models on monitoring problems (see, for example, Radner 1981; Porter 1983; Green and Porter 1984). Learning is also 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. Then, trustors 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 trustors (cf. Mailath and Samuelson 1998b, 1998b; Tadelis 1999).

Clearly, many theoretical issues have to be resolved. However, some three-person experiments in which subjects play finitely repeated Trust Games have to be done for obtaining better insight in the merits and problems of the models presented here. Considering that the implications of different models depend to a large extent on what the subjects take into account to make their decisions, it is advisable to design an experiment in which it is possible to follow more or less the decision making process of the trustors as is done, for example, by Camerer, Johnson, Rymon, and Sen (1993).

A Proofs of the Theorems

Proof of Theorem 3. Checking that the beliefs are consistent with Bayes' rationality is straightforward. If the trustee honors trust in stage 2 and trustor 2 receives this information, then

$$\pi_{g,2} = \frac{Pr(C_2|\text{good})Pr(\text{good})}{Pr(C_2|\text{good})Pr(\text{good}) + Pr(C_2|\text{payoff-maximizing})Pr(\text{payoff-maximizing})}, \tag{2}$$

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 $\pi_g = 0$ in this out-of-equilibrium instance of updating.

Assume $\pi_{12} <$ TEMP. If trustor 2 does not receive information about stage 2 from trustor 1, she does not update her beliefs ($\pi_{g,1} = \pi_g$), which implies that she places trust if and only if $\pi_g R_1 + (1 \Leftrightarrow \pi_g)S_1 \geq P_1 \Leftrightarrow \pi_g \geq$ RISK, because the payoff-maximizing trustee will abuse trust. Trustor 2 is indifferent if equality holds. If trustor 2 obtains information about honored trust, the trustee must be a good trustee and, therefore, she will place trust in her stage. If trustor 2 receives information about abused trust, then $\pi_{g,1} = 0$, which implies that she has no incentive to place trust. Because a payoff-maximizing trustee abuses trust in the stage 2 as well, trustor 1 will also place trust if and only if $\pi_g \geq$ RISK.

The payoff-maximizing trustee only has a choice in the stage 2 if $\pi_g >$ RISK. Then, the payoff-maximizing trustee's expected payoff from honoring trust is $u_2(C_2) = R_2 + \pi_{12}T_2 + (1 \Leftrightarrow \pi_{12})T_2$. On the other hand, $u_2(D_2) = T_2 + \pi_{12}P_2 + (1 \Leftrightarrow \pi_{12})T_2$. Consequently, $u_2(C_2) > u_2(D_2) \Leftrightarrow \pi_{12} >$ TEMP, which implies that a payoff-maximizing trustee will abuse trust, because $\pi_{12} <$ TEMP. Trivially, the best move for the payoff-maximizing trustee in the last stage is abusing trust.

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

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

Finally, assume that $\pi_{12} \geq \text{TEMP}$ and $\pi_g < \text{RISK}$. Now, trustor 2 is indifferent between honoring and abusing trust if she receives information about honored trust, because $\pi_{g,1} = \text{RISK}$. Therefore, randomizing is optimal for trustor 2. The expected payoffs for trustor 1 are $u_1(C_1) = \pi_g R_2 + (1 \Leftrightarrow \pi_g)(q_2R_1 + (1 \Leftrightarrow q_2)S_1)$ and $u_1(D_1) = P_1$, where $q_2 = \frac{\pi_g}{1-\pi_g}(\frac{1}{\text{RISK}} \Leftrightarrow 1)$. Straightforward manipulations yields that $u_1(C_1) > u_1(D_1) \Leftrightarrow \pi_g > \text{RISK}^2$.

The trustee's expected payoff from honoring trust equals $u_2(C_2) = R_2 + \pi_{12}p_1T_2 + \pi_{12}(1 \Leftrightarrow p_1)P_2 + (1 \Leftrightarrow \pi_{12})P_2$, where $p_1 = \frac{\text{TEMP}}{\pi_{12}}$ 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 $\pi_g < \text{RISK}$ and $\pi_{12} \geq \text{TEMP}$ is an equilibrium is straightforward. Other randomization probabilities are not possible because if this probability is larger, the trustee would always honor trust in the stage before, which leads to a contradiction, since trustor 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 the trustor 1 cannot place trust and, therefore, the trustor 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. □

Proof of Theorem 4. It can be shown easily that the beliefs of the trustors are consistent with Bayesian updating. An important observation is that the trustors 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 trustor places trust. Therefore, if the first trustor in such a pair of moves has placed trust, she will not obtain any new information if the other trustor after her move informs the first trustor. Moreover, if the second trustor in a pair of moves received 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 trustor 1 does not place trust in a stage in which she was indifferent, but she receives information from trustor 2 about trust honored 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.

If $N = 1$, the theorem reduces to Theorem 3. Now, we proceed to prove the theorem for these cases using an induction argument. For each case, we have to prove that the first move of each trustor and the first two moves by the trustee are their optimal moves considering the strategies of the other players as given. This implies that we will consider stages $2N$ and $2N-1$. If both trustors 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 in this way two times R_2 and is thereafter 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 trustors 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 we have to check are those where the trustors probably do not place trust in their first or second stage of play.

We start with case 1. Because the trustee's strategy is exactly the same as in Theorem 1, he acts as if he is playing $2N$ stages with the same trustor. Consequently, it follows from Theorem 1 that the trustors play an optimal response, because they also play the game as if only one trustor is involved with the only exception that they use other randomization probabilities. Now we only have to check whether the behavior of the trustee is optimal.

Assume $\text{RISK}^{2N-2} < \pi_g < \text{RISK}^{2N-3}$. From the induction assumption we know that the trustee is indifferent in stage $2N-2$ and will be randomizing in this stage. For the calculation of the expected payoffs we can assume that he abuses trust in this stage and in stage $2N-3$. Throughout this proof we use this argument to calculate the payoffs in the stages that are covered by the induction assumption, which mostly implies that we calculate payoffs for the trustee abusing trust and trustor not placing trust, because then the payoffs for the remaining stages can be determined straightforwardly. Now, we consider the first two moves of the trustee.

$$\begin{aligned} u_2(C_2C_2) &= R_2 + R_2 + T_2 + \pi_{12}P_2 + (1-\pi_{12})\Big(p_{2N-3}T_2 + (1-p_{2N-3})P_2\Big) + (2N-4)P_2 \\ &= 2R_2 + T_2 + (2N-3)P_2; \end{aligned}$$

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

$$u_2(D_2C_2) = T_2 + \pi_{12}P_2 + (1 \Leftrightarrow \pi_{12})R_2 + (2N \Leftrightarrow 2)P_2 < u_2(D_2D_2);$$

$$u_2(D_2D_2) = T_2 + \pi_{12}P_2 + (1 \Leftrightarrow \pi_{12})T_2 + (2N \Leftrightarrow 2)P_2 < R_2 + T_2 + (2N \Leftrightarrow 2)P_2.$$

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

$$u_2(C_2C_2) = R_2 + R_2 + \pi_{21}\Big(p_{2N-2}T_2 + (1 \Leftrightarrow p_{2N-2})P_2\Big) + (1 \Leftrightarrow \pi_{12})P_2 + (2N \Leftrightarrow 3)P_2;$$

$$u_2(C_2D_2) = R_2 + T_2 + (2N \Leftrightarrow 2)P_2;$$

$$u_2(D_2C_2) = T_2 + \pi_{12}P_2 + (1 \Leftrightarrow \pi_{12})R_2 + (2N \Leftrightarrow 2)P_2 < u_2(D_2D_2);$$

$$u_2(D_2D_2) = T_2 + \pi_{12}P_2 + (1 \Leftrightarrow \pi_{12})T_2 + (2N \Leftrightarrow 2)P_2 < R_2 + T_2 + (2N \Leftrightarrow 2)P_2.$$

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

Finally, we consider $\pi_g < \mathrm{RISK}^{2N-1}$, which implies that the trustee randomizes in the first two stages according to the theorem. Consequently, we need to check whether the trustee is indifferent between playing any pair of moves in the first two stage assuming that the trustors place trust. Otherwise the trustee does not have a choice anyway.

$$\begin{aligned}
u_2(C_2C_2) &= R_2 + \pi_{12}\Big(p_{2N-1}R_2 + (1 \Leftrightarrow p_{2N-1})P_2\Big) + (1 \Leftrightarrow \pi_{12})P_2 \\
&\qquad + \pi_{12}\pi_{21}p_{2N-1}\Big(p_{2N-2}T_2 + (1 \Leftrightarrow p_{2N-2})P_2\Big) + (1 \Leftrightarrow \pi_{12}\pi_{21}p_{2N-1})P_2 + (2N \Leftrightarrow 3)P_2; \\
&= R_2 + \mathrm{TEMP}(R_2 \Leftrightarrow P_2) + \mathrm{TEMP}^2(T_2 \Leftrightarrow P_2) + (2N \Leftrightarrow 1)P_2 = T_2 + (2N \Leftrightarrow 1)P_2;
\end{aligned}$$

$$\begin{aligned}
u_2(C_2D_2) &= R_2 + \pi_{12}\Big(p_{2N-1}T_2 + (1 \Leftrightarrow p_{2N-1})P_2\Big) + (1 \Leftrightarrow \pi_{12})P_2 + (2N \Leftrightarrow 2)P_2 \\
&= T_2 + (2N \Leftrightarrow 1)P_2;
\end{aligned}$$

$$u_2(D_2C_2) = T_2 + (2N \Leftrightarrow 1)P_2;$$

$$u_2(D_2D_2) = T_2 + (2N \Leftrightarrow 1)P_2.$$

Above we substituted $p_{2N-1} = \frac{\mathrm{TEMP}}{\pi_{12}}$ and $p_{2N-2} = \frac{\mathrm{TEMP}}{\pi_{21}}$, which represent the strategies of the trustors. 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 the other equilibrium follows from the same considerations as given for $N = 1$. That there are no other equilibria follows from the uniqueness of the similar equilibrium in the baseline model.

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

Then, it holds for trustor 1 that

$$u_1(C_1) = \pi_g R_1 + (1 \Leftrightarrow \pi_g)\Big(q_{2N} R_1 + (1 \Leftrightarrow q_{2N}) S_1\Big) + (N \Leftrightarrow 1) P_1;$$

$$u_1(D_1) = N P_1$$

$$u_1(C_1) > u_1(D_1) \Leftrightarrow \pi_{g,2N} > \mathrm{RISK}^N.$$

Thus, this shows that trustor 1 is acting optimal. The trustor is indifferent if equality holds. The expected payoffs for trustor 2 are

$$u_1(C_1) = \pi_g R_1 + (1 \Leftrightarrow \pi_g)\Big(q_{2N} R_1 + (1 \Leftrightarrow q_{2N}) S_1\Big) + u_1(\text{after } C_1),$$

$$u_1(D_1) = P_1 + u_1(\text{after } D_1),$$

where $u_1(\text{after } C_1) = u_1(\text{after } D_1)$, because the behavior of trustor 1 does not depend on whether trustor 2 informs her about the trustworthiness of the trustee. If trustor 1 did not place trust once, she will never place trust again although the trustee might still be trustworthy toward trustor 2. If trustor 1 did place trust she knows what the trustee will play in the following stage with trustor 2, so nothing will change for her whether or not trustor 2 informs her before her next stage. Therefore, the probability that trustor 2 obtains information about the trustee in a foregoing stage does not depend on her own behavior. Consequently, $u_1(C_1) > u_1(D_1) \Leftrightarrow \pi_{g,2N} > \mathrm{RISK}^N$, and trustor 2 is indifferent if equality holds. Note that if trustor 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 trustor about trust honored by the trustee.

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

Before the calculation we note that for the randomization probabilities of the trustors holds

$$p_{2N-2} + p_{2N-3} \Leftrightarrow \pi_{12} p_{2N-2} p_{2N-3} = 2\mathrm{TEMP} \Leftrightarrow \pi_{12}, \tag{3}$$

$$0 < p_{2N-3} = \frac{\mathrm{TEMP}(1 \Leftrightarrow \pi_{12}) \Leftrightarrow \pi_{12}\mathrm{TEMP} + \pi_{12}^2}{(1 \Leftrightarrow \pi_{12}) \Leftrightarrow \pi_{12}\mathrm{TEMP} + \pi_{12}^2} < \frac{\mathrm{TEMP}}{1 \Leftrightarrow \pi_{12}}, \text{ and} \tag{4}$$

$$0 < p_{2N-2} = \frac{\mathrm{TEMP} \Leftrightarrow \pi_{12}}{1 \Leftrightarrow \pi_{12}} < \frac{\mathrm{TEMP} \Leftrightarrow \pi_{12}}{1 \Leftrightarrow \pi_{21}}. \tag{5}$$

Condition (3) follows from straightforward manipulation. For Condition (4), we need to realize that the enumerator equals $\mathrm{TEMP} \Leftrightarrow \mathrm{TEMP}^2 + (\mathrm{TEMP} \Leftrightarrow \pi_{12})^2 > 0$ and that $\Leftrightarrow \pi_{12}\mathrm{TEMP} + \pi_{12}^2 < 0$. Condition (5) follows from $\pi_{12} < \pi_{21}$.

Now, we can show that the trustee is indifferent between playing D_2D_2 and C_2C_2, and the other combinations provide him with a lower payoff.

$$\begin{aligned}
u_2(C_2C_2) &= R_2 + R_2 + p_{2N-2}T_2 + (1 \Leftrightarrow p_{2N-2})P_2 + p_{2N-2}\pi_{12}P_2 \\
&\qquad + (1 \Leftrightarrow p_{2N-2}\pi_{12})\Big(p_{2N-3}T_2 + (1 \Leftrightarrow p_{2N-3})P_2\Big) + (2N \Leftrightarrow 4)P_2 \\
&= 2R_2 + \Big(p_{2N-2} + p_{2N-3} \Leftrightarrow \pi_{12}p_{2N-2}p_{2N-3}\Big)(T_2 \Leftrightarrow P_2) + (2N \Leftrightarrow 2)P_2; \\
&= 2T_2 \Leftrightarrow \pi_{12}(T_2 \Leftrightarrow P_2) + (2N \Leftrightarrow 2)P_2; \\
u_2(C_2D_2) &= R_2 + T_2 + \pi_{21}P_2 + (1 \Leftrightarrow \pi_{21})\Big(p_{2N-2}T_2 + (1 \Leftrightarrow p_{2N-2})P_2\Big) + (2N \Leftrightarrow 3)P_2 \\
&= R_2 + T_2 + p_{2N-2}(1 \Leftrightarrow \pi_{21})(T_2 \Leftrightarrow P_2) + (2N \Leftrightarrow 2)P_2 \\
&< R_2 + T_2 + (\mathrm{TEMP} \Leftrightarrow \pi_{12})(T_2 \Leftrightarrow P_2) + (2N \Leftrightarrow 2)P_2 \\
&= 2T_2 \Leftrightarrow \pi_{12}(T_2 \Leftrightarrow P_2) + (2N \Leftrightarrow 2)P_2;
\end{aligned}$$

$$\begin{aligned}
u_2(D_2C_2) &= T_2 + \pi_{12}P_2 + (1 - \pi_{12})R_2 + P_2 + (1 - \pi_{12})^2\Big(p_{2N-3}T_2 + (1 - p_{2N-3})P_2\Big) \\
&\qquad + \Big(1 - (1 - \pi_{12})^2\Big)P_2 + (2N - 4)P_2 \\
&= T_2 + R_2 - \pi_{12}(R_2 - P_2) + (1 - \pi_{12})^2\pi_{2n-3}(T_2 - P_2) + (2N - 2)P_2 \\
&< T_2 + R_2 - \pi_{12}(T_2 - P_2) + \pi_{12}(T_2 - R_2) + (1 - \pi_{12})(T_2 - R_2) + (2N - 2)P_2 \\
&= 2T_2 - \pi_{12}(T_2 - P_2) + (2N - 2)P_2; \\
u_2(D_2D_2) &= T_2 + \pi_{12}P_2 + (1 - \pi_{12})T_2 + (2N - 2)P_2 \\
&= 2T_2 - \pi_{12}(T_2 - P_2) + (2N - 2)P_2.
\end{aligned}$$

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, we need to show for the induction assumption to be applicable and to ensure perfectness that if the trustee randomizes and trustor 1 does not place trust that the trustee also in that situation does not have an incentive to deviate in his subsequent stage with trustor 2. Realize first that if trustor 1 does not place trust, she will never place trust again. Therefore, the game essentially reduces to a two-person game for which we know that in equilibrium the trustor places trust with a probability TEMP in the randomization phase. The probability that trustor 2 places trust equals $\pi_{12} + (1 - \pi_{12})\frac{\text{TEMP}-\pi_{12}}{1-\pi_{12}} = \text{TEMP}$, which proves that the trustee is also indifferent between honoring and abusing trust if trustor 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 π_{12} to make trustor 2 indifferent. If that would be necessary, he would always prefer to abuse trust placed by trustor 1. It can also be checked easily that for $N = 2$ and $\pi_g < \text{RISK}$, the trustee has to start randomizing or abusing trust at least in his first encounter with trustor 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 trustor 2 for which $\pi_g < \text{RISK}^n$ in stage $\frac{n-1}{2}$.

References

Berg, J., J. Dickhaut, and K. McCabe (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior 10*, 122–142.

Bower, A. G., S. Garber, and J. C. Watson (1997). Learning about a Population of Agents and the Evolution of Trust and Cooperation. *International Journal of Industrial Organization 15*, 165–190.

Buskens, V. (1999). *Social Networks and Trust.* Amsterdam: Thela Thesis.

Buskens, V. and J. Weesie (1999). Cooperation via Networks. *Analyse und Kritik*. (forthcoming).

Camerer, C. and K. Weigelt (1988). Experimental Tests of a Sequential Equilibrium Reputation Model. *Econometrica 56*, 1–36.

Camerer, C. F., E. J. Johnson, T. Rymon, and S. Sen (1993). Cognition and Framing in Sequential Bargaining for Gains and Losses. In K. Binmore, A. Kirman, and P. Tani (Eds.), *Frontiers of Game Theory*, pp. 27–47. Cambridge, MA: MIT Press.

Coleman, J. S. (1990). *Foundations of Social Theory.* Cambridge, MA: The Belknap Press of Harvard University Press.

Dasgupta, P. (1988). Trust as a Commodity. In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations*, pp. 49–72. Oxford: Blackwell.

Erev, I. and A. E. Roth (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review 88*, 848–881.

Green, E. J. and R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica 52*, 87–100.

Güth, W. and H. Kliemt (1994). Competition or Co-operation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes. *Metroeconomica 45*, 155–178.

Güth, W., P. Ockenfels, and M. Wendel (1997). Cooperation Based on Trust: An Experimental Investigation. *Journal of Economic Perspectives 18*, 18–43.

HIRSCHMAN, A. O. (1970). *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Cambridge, MA: Harvard University Press.

KREPS, D. M. (1990a). Corporate Culture and Economic Theory. In J. ALT AND K. SHEPSLE (Eds.), *Perspectives on Positive Political Economy*, pp. 90–143. Cambridge: Cambridge University Press.

KREPS, D. M. (1990b). *A Course in Microeconomic Theory.* New York: Harvester Wheatsheaf.

KREPS, D. M., P. MILGROM, J. ROBERTS, AND R. WILSON (1982). Rational Cooperation in the Finitely Repeated Prisoners' Dilemma. *Journal of Economic Theory 27*, 245–252.

KREPS, D. M. AND R. WILSON (1982a). Reputation and Imperfect Information. *Journal of Economic Theory 27*, 253–279.

KREPS, D. M. AND R. WILSON (1982b). Sequential Equilibria. *Econometrica 50*, 863–894.

LAHNO, B. (1995). Trust, Reputation, and Exit in Exchange Relationships. *Journal of Conflict Resolution 39*, 495–510.

LAHNO, B. (1998). *Is Trust the Result of Bayesian Learning?* mimeo, Heinrich Heine Universität Düsseldorf.

LUCE, R. D. AND H. RAIFFA (1957). *Games and Decisions: Introduction and Critical Survey.* New York: Wiley.

MACY, M. W. AND J. SKVORETZ (1998). The Evolution of Trust and Cooperation between Strangers: A Computational Model. *American Sociological Review 63*, 638–660.

MAILATH, G. J. AND L. SAMUELSON (1998a). *Who Wants a Good Reputation?* CARESS Working Paper 98-12, University of Pennsylvania.

MAILATH, G. J. AND L. SAMUELSON (1998b). *Your Reputation is Who You're Not, Not Who You'd Like to be.* CARESS Working Paper 98-11, University of Pennsylvania.

MCKELVEY, R. D. AND T. R. PALFREY (1992). An Experimental Study of The Centipede Game. *Econometrica 60*, 803–836.

NERAL, J. AND J. OCHS (1992). The Sequential Equilibrium Theory of Reputation Building: A Further Test. *Econometrica 60*, 1151–1169.

PORTER, R. H. (1983). Optimal Cartel Trigger-Price Strategies. *Journal of Economic Theory 29*, 313–338.

RADNER, R. (1981). Monitoring Cooperative Agreements in a Principal-Agent Relationship. *Econometrica 53*, 1173–1198.

RAUB, W. AND J. WEESIE (1990). Reputation and Efficiency in Social Interactions: An Example of Network Effects. *American Journal of Sociology 96*, 626–654.

RAUB, W. AND J. WEESIE (1993). *The Management of Matches: Decentralized Mechanism for Cooperative Relations with Applications to Organizations and Households (revised version).* ISCORE Paper 1, Utrecht University.

ROTH, A. E. AND I. EREV (1995). Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior 8*, 164–212.

SCHÜSSLER, R. A. (1989). Exit Threats and Cooperation under Anonymity. *Journal of Conflict Resolution 33*, 728–749.

SNIJDERS, C. (1996). *Trust and Commitments.* Amsterdam: Thesis Publishers.

SNIJDERS, C. AND G. KEREN (1999). Determinants of trust. In D. V. BUDESCU, I. EREV, AND R. ZWICK (Eds.), *Games and Human Behavior*, pp. 355–385. Mahwah, NJ: Lawrence Erlbaum.

TADELIS, S. (1999). What's in a Name? Reputation as a Tradeable Asset. *American Economic Review*, forthcoming.

VANBERG, V. AND R. D. CONGLETON (1992). Rationality, Morality, and Exit. *American Political Science Review 86*, 418–431.

WEESIE, J. (1996). *Disciplining via Exit and Voice.* ISCORE Paper 88, Utrecht University.