Analyse & Kritik 22/2000 (© Lucius & Lucius, Stuttgart) p. 44–74

Vincent Buskens/Jeroen Weesie

Cooperation via Social Networks*

Abstract: Sufficiently frequent interaction between partners has been identified by, a.o., Axelrod as a more-or-less sufficient condition for stable cooperation. The underlying argument is that rational cooperation is ensured if short-term benefits from opportunistic behavior are offset by the long-term costs of sanctions imposed on the culprit. In this paper, we develop a model for 'embedded trust' in which a trustee interacts with a number of trustors who may communicate via a social network with each other about the behavior of the trustor. The analysis reconfirms the standard predictions about how the level of trust depends on the payoffs and shadow of the future. We provide new predictions both on between-network effects ("which network is more favorable for cooperation?") and on within-network effects ("in what network position can you trust more?").

1. Introduction

The study of the conditions under which cooperation among rational egoists is possible has had a great impetus from the important work of Axelrod (1984). While not the originator of the argument that cooperative behavior may be in an actor's enlightened self-interest if the short-term benefits from non-cooperative behavior are exceeded by the long-term costs of sanctions imposed upon him, Axelrod's seminal work contributed to the wide diffusion of the argument in the social science literature, and has triggered a huge pile of research. The results from Axelrod's computer tournament about the powerful effect of the shadow of the future on the prospects for cooperation and the apparent success of one particularly simple strategy, Tit-for-tat, are highly fascinating, even though not beyond criticism (see, e.g., Binmore 1998). In this paper, we abstain from a critical review of tiny parts or even larger fragments of Axelrod's approaches and results. Rather, we seek to elaborate on

* Stimulating comments of and discussions with Werner Raub are gratefully acknowledged. Financial support was provided by the Netherlands Organization for Scientific Research (NWO) under grant PGS 50-370. Please direct all correspondence to the first author at the address: Department of Sociology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands. Email: V.Buskens@fss.uu.nl.

his results by focusing on the conditions for cooperation in order to contribute to the research agenda set implicitly by Axelrod's work.

Axelrod focused on repeated two-person games, in which sanctions can only be imposed by the victim, and not by other actors in the tournament. We argue that Axelrod's conditions may have been *too pessimistic* about the prospects of cooperation, since he restricts the long-term costs of non-cooperative behavior to the sanctions imposed by the direct victim. In most real settings, this is overly restrictive. Usually, dyadic relations are embedded in a 'network' of relations with overlapping sets of actors. These overlaps create dependencies between the relations that are ignored in Axelrod's work. A prime example is provided by communication networks that allow for reputation effects: An actor has to take into account not only the partner's responses to his actions, but also how his behavior affects his *reputation* with other actors that are relevant to him. We hasten to admit that Axelrod was well aware that he made a substantial simplification by ignoring these reputation effects as analyzed here, and we expect that he probably shared our intuition about their importance as well (see, e.g., Axelrod 1984, chapter 1).[1] One of Axelrod's pieces of advice on promoting cooperation, however, is more ambiguous (1984, chapter 7). Axelrod's advice is "keep others away" (130). Keeping others away makes good sense if this implies that the shadow of the future with one's partners is increased. In this paper, however, it is argued that cooperation is also facilitated if players inform third parties about interactions with a partner, and actually stimulate interactions of these third parties with the partner.

In recent years, many empirical and theoretical studies have addressed the effects of reputation on the behavior of actors in cooperation problems (for example, Granovetter 1985; Kreps/Wilson 1982; Wilson 1985; Kreps 1990b; Klein 1997). In nearly all studies that we are aware of, the authors do not really try to assess the consequences of the observation that the reputational mechanism usually depends on *informal contacts* between actors, and hence the efficiency of the reputation mechanism should vary with the network structure.[2] Raub and Weesie (1990) have analyzed the first game-theoretic model in which the actors who play iterated prisoner's dilemmas are embedded in a network of relations interpreted as information channels. Their model allowed to make predictions how cooperation rates vary between networks, but did not allow predictions about differences within networks. In this paper, we

[1] Axelrod's discussion of reputations (1984, 150–154) is more in-line with literature on repeated games with incomplete information, in which reputation refers to a hidden characteristics of a player, that you can learn about, both from own experience in interacting with that player, and from observing his behavior vis-à-vis third parties.

[2] Interesting exceptions concern the analyses of *institutions* such as law merchants and credit rating firms, that provide formal rather than informal means to affect reputations. See, a.o., Milgrom/Roberts 1992, 266–269.

seek to address this problem. Thus, the information networks analyzed here are not homogenous, i.e., we do not assume that all actors occupy a 'similar' position in the network.

The remainder of this paper is organized as follows. Section 2 gives a detailed description of the model. In Section 3, some of the mathematical properties of the model will be analyzed and equilibria are identified. Section 4 examines how the equilibrium that is chosen as the solution for the game depends on the model parameters. Section 5 addresses how the level of trust depends on the network among trustors. In particular, we apply an analytical approximation method (linearization) to obtain predictions about the effects of global and local aspects of the network between trustors. Section 6 summarizes the main findings, limitations of the model, and possibilities for further research. An appendix collects the proofs of the theorems and contains some further technical details.

2. Construction of the Model

This section describes an analytically tractable model for predicting network effects in trust situations. To model network effects, the minimal requirement is a model with repeated games between (subsets of) actors who may inform other actors about their experiences. The first element needed is a constituent game that is played in the different periods of the game. For reasons outlined below, we use as the constituent game a Heterogeneous Trust Game (HTG) Γ_F, a variant of the Trust Game for modeling simple trust relations (Dasgupta 1988; Kreps 1990a). The extensive form of the HTG is shown in Figure 1.

2.1 The Constituent Game

Nature generates $\theta > 0$ randomly from a probability distribution F in the first move of Γ_F. θ is the incentive for the trustee to abuse trust placed by the trustor. For technical reasons it is assumed that F is an atomless probability distribution with full support on $[0, \infty]$. An example of such a distribution with favorable analytical properties is $F_a(\vartheta) = Pr(\theta \leq \vartheta) = \vartheta/(a + \vartheta)$ (cf. Raub/Weesie 1993; Weesie et al. 1998). Here, a is the median of F_a and so a can be interpreted as the trustee's 'average' incentive for opportunistic behavior. In the HTG, both trustor and trustee are informed on θ.[3] In the second move, the trustor chooses whether or not she places trust. If the trustor does not place trust, the constituent game is over and the actors receive a payoff P_i. If the trustor places trust, the constituent game continues with

[3] In an interesting variant of the model, it can be assumed that the trustor is *not* informed about the incentive θ for the trustee to abuse trust. We plan to provide an analysis of this model with incomplete information in a future paper.

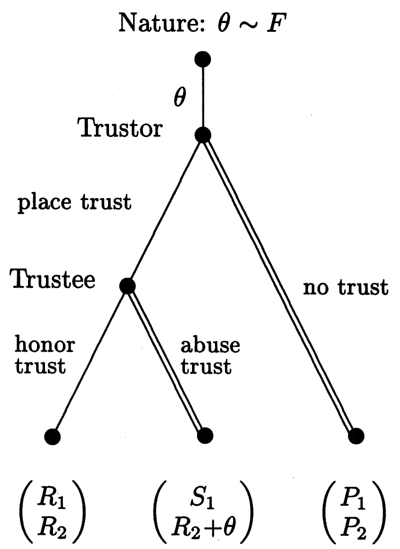

Figure 1: Extensive form of the Heterogeneous Trust Game Γ_F, where $R_i > P_i$, $(i = 1, 2)$, $P_1 > S_1$, $\theta > 0$.

a third move. Here, the trustee can honor trust or abuse trust. If he honors trust, the actors receive $R_i > P_i$. If he abuses trust, the trustee receives $R_2 + \theta > R_2$ and the trustor receives $S_1 < P_1$. Clearly, the trustee will always abuse trust, and the trustor will not place trust for any $\theta > 0$ in the unique subgame-perfect equilibrium of the constituent game Γ_F.[4] The double lines in Figure 1 indicate the predicted move for each actor in the corresponding node of the constituent game. This equilibrium is Pareto inefficient, because both actors prefer placing trust and honoring trust over the situation in which no trust is placed. The difference $R_i - P_i$ is the efficiency loss for the trustor $(i = 1)$ and the trustee $(i = 2)$ due to the trust problem.

2.2 Repeating the HTG in a Network

Now, we define the Iterated Heterogeneous Trust Game IHTG $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A})$. The constituent game Γ_F is played at discrete moments in time, $t = 0, 1, 2, \ldots$ between a trustee and one trustor in a network of n types of trustors $(\boldsymbol{\pi}, \mathbf{A})$.[5]

[4] Formally, it should be added that if the trustor would place trust, the trustee abuses trust in equilibrium.

[5] To avoid notational complexity the homogeneity assumption is made that the payoffs of the trustee and the distribution F are independent of the trustor with whom the trustee plays. The analyses easily generalize for the case that the payoffs depend on the trustor.

Thus, there are two main differences between the ordinary Iterated Trust Game and the IHTG. First, the ordinary Trust Game is replaced by the HTG as the constituent game. In the discussion about the assumptions of the model below, we will explain the advantages of this modification. Second, we distinguish n types of trustors. The only difference between types is their relation with other trustors, i.e., their position in the communication network. The trustee plays with a network of n types of trustors ($\boldsymbol{\pi}$, $\mathbf{A}$) instead of one trustor. The vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ reflects the proportion of the different types of trustor in the 'population', with $\pi_i > 0$ and $\sum_{i=1}^{n} \pi_i = 1$. The proportion π_i is the probability that a trustor of type i is selected for transactions with the trustee. The entries α_{ij} of the network matrix $\mathbf{A}$ are the densities of the network between trustors of type i and trustors of type j. The α_{ij} is the probability that one trustor of type i has a tie to a given trustor of type j. It may be the case that $i = j$. We assume that the networks between two types of trustors and among trustors of one type are 'homogeneous' in the sense that all trustors of a certain type have ties to the same proportion of trustors of another type and to the same proportion of trustors within their own type.

Now, we will explain how the sequences of transactions are modeled and when communication is possible. Again, in the discussion of the assumptions, we will pay attention to the reasons for these assumptions. To reduce complexity while maintaining the essential character of network information diffusion, we constructed a scenario in which the information diffusion process is somehow restricted (see also Weesie et al. 1998). The first assumption is that the trustee has transactions with trustors 'forever'. Trustors play series of constituent games to allow for effects of temporal embeddedness as described in the introduction, in addition to the network effects. The second assumption is that the trustors play series of transactions with the trustee *sequentially* (see Figure 1). Only after a trustor drops out, a new trustor is chosen for a next series of transactions. A series of transactions with a trustor of type i ends with a probability δ_i, $0 \leq \delta_i \leq 1$, called the *drop-out rate.* The termination of a series of transactions is stochastically independent of what happened in earlier transactions.[6] The probability that the new trustor is a trustor of type j is π_j independent of the identity of the trustor who drops out. It may be the case that the new trustor is of the same type as the trustor who drops out. It is assumed that each type of trustors consists of infinitely many trustors to prevent that the same trustor is chosen again for a series of transactions.[7]

---

[6] This implies, for example, that the continuation of a relation does not depend on decisions of players whether or not to continue. Endogenous probabilities are used in models about exit out of a relation (Schuessler 1989; Vanberg/Congleton 1992; Lahno 1995; Weesie 1996; Blumberg 1997). Introducing exit endogenously in the model would complicate the analysis considerably.

[7] If one wants to assume that there are only a finite number of trustors for each type,

Type of trustor who has transactions

$1,1,\ldots,1 \downarrow 3,3,\ldots,3 \downarrow 2,2,\ldots,2 \downarrow 2,2\ldots,2 \downarrow 1,1\ldots,1$

opportunity for communication from a type 1 trustor to a type 3 trustor	opportunity for communication from a type 3 trustor to a type 2 trustor	opportunity for communication from a type 2 trustor to another type 2 trustor	opportunity for communication from a type 2 trustor to a type 1 trustor

Figure 2: A sequence of transactions

The third assumption is that *exclusively* at the moment that one trustor drops out and another trustor enters the game, information can be communicated and *only* between the old and new trustor (see Figure 1). The trustor who drops out transmits information to the next trustor if a tie exists between the old and the new trustor. Thus, since the identity of a new trustor is selected at random, the probability that a trustor of type i informs a trustor of type j about the behavior of the trustee after a series of transactions equals α_{ij}. If a trustor informs the next trustor, she does not only communicate her own experiences with the trustor, but also all information she obtained from trustors before her. In this way, it is still possible that information is transmitted from one trustor to another to yet another and so on. If a trustor does not communicate with a subsequent trustor, all information of previous transactions is lost because this trustor will not have another series of transactions with the trustee. The new trustor has to start her series of transactions without information about previous transactions. Furthermore, it is assumed that all information is accurate; no incentives to withhold or misrepresent information strategically are analyzed. Finally, it is assumed that the trustee knows whether or not the next trustor is informed by her predecessor. The argument for the last assumption will become apparent in the discussion about the equilibria that will be studied.

Summarizing, three things might happen 'between' time t and time $t+1$. First, the same trustor continues a series of transactions. Second, a new trustor is chosen and the old trustor has a tie to the new trustor, and so informs the new trustor about the behavior of the trustee. Third, a new trustor is chosen who is not tied to the new trustor, and she does not inform the new trustor. Now, a transition matrix $\mathbf{T}$ can be derived, where t_{ij} equals

one should assume that trustors do not remember information from earlier transactions if they are chosen for a new series of transactions (Buskens 1999).

the probability that the former transaction was a transaction of the trustee with a trustor of type i, *and* the following transaction is with a trustor of type j, *and* the trustor of type i communicates the information she has to the trustor of type j. Thus, $t_{ij} = \delta_i \pi_j \alpha_{ij}$ if $i \neq j$. For the diagonal elements, it holds that $t_{ii} = 1 - \delta_i + \delta_i \pi_i \alpha_{ii}$, because $1 - \delta_i$ is the probability that a series of transactions of a trustor of type i continues, and $\delta_i \pi_i \alpha_{ii}$ is the probability that a series of transactions of a trustor of type i ends and another trustor of type i is chosen for a new series of transactions. In matrix form, the transition matrix is

$$\mathbf{T} = \begin{pmatrix} 1-\delta_1+\delta_1\pi_1\alpha_{11} & \delta_1\pi_2\alpha_{12} & \cdots & \delta_1\pi_n\alpha_{1n} \\ \delta_2\pi_1\alpha_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta_{n-1}\pi_n\alpha_{n-1,n} \\ \delta_n\pi_1\alpha_{n1} & \cdots & \delta_n\pi_{n-1}\alpha_{n,n-1} & 1-\delta_n+\delta_n\pi_n\alpha_{nn} \end{pmatrix}. \quad (1)$$

The rows of $\mathbf{T}$ do not add up to 1, because the transitions in which information is not communicated between two consecutive trustors are not included in $\mathbf{T}$. Hence, $\mathbf{T}$ is not a real transition matrix.

2.3 Payoffs

The payoff function is defined similarly as in the ordinary Iterated Trust Game. If a trustor is not involved in a transaction at time t, she receives a payoff 0. A trustor and trustee who are involved in the game at time t obtain the payoffs related to the outcome of the game in that time period. Payoffs are discounted exponentially with discount factor w, $0 \leq w \leq 1$, for the trustee and all types of trustors.[8] This discount factor reflects pure time preferences. Thus, the total payoff of an actor i associated with a stream of payoffs $(u_{i0}, u_{i1}, \ldots)$ equals $\sum_{t=0}^{\infty} w^t u_{it}$. An important advantage of the approach taken here is that the payoffs, discount factor, drop-out rates, and the social network are included simultaneously in the one model. Therefore, it is possible to deduce not only *main effects* of all these different elements of the model, but also *interaction effects*, in particular between network parameters and other elements of the game.

[8] In the analyses, only the discount factor of the trustee is relevant. Thus, this homogeneity assumption is made only to avoid useless notational complexity.

2.4 Discussion of Assumptions

First, we will discuss why the Heterogeneous Trust Game rather than the ordinary Trust Game was chosen as the constituent game. In the solution of the ordinary Iterated Trust Game, the trustor always places trust or never places trust, depending on the parameters of the game. The advantage of the IHTG is that trustors will not place trust in a period with very large θ. These are 'golden opportunities' for the trustee in which the incentive to abuse trust is so large that it exceeds possible long-term losses. This implies that, in the IHTG, the trustor who has transactions with the trustee and the trustee will not always obtain R_i. Thus, there will always be some inefficiency. The extent of inefficiency depends on the size of the incentive to abuse trust for which trustors cannot trust the trustee anymore. It will be shown that in the solution of the IHTG the trustor will only place trust if θ is not too large. The more the trustor trusts the trustee, the larger the values θ for which the trustor still places trust, and, thus, the higher the efficiency level that is reached. While for the ordinary Iterated Trust Game hypotheses can only be derived for the condition under which trust can always be placed, for the IHTG, hypotheses can be derived that are direct consequences of the comparative statics on the *extent* to which trust can be placed by the trustor. One thing that does not change in comparison with the ordinary Iterated Trust Game the is that the trustee never abuses trust in the solution presented in the next section.

Second, we want to comment on the scenario for the order of transactions and opportunities for communication among an infinite number of trustors. In an 'ideal' and more realistic model, a (finite) number of trustors would have transactions with the trustee simultaneously as well as sequentially. Between transactions, the trustors may have opportunities to communicate information about the behavior of the trustee. However, the analysis of a model in which a finite number of trustors have simultaneous transactions with the trustee and communicate information is already very demanding. In particular, it is complex to determine what the optimal behavior for the trustee is after one abuse of trust. This can be understood as follows. Even under the assumption that trustors never forget information about abused trust and the deceived trustor transmits this information as soon as possible to other trustors, it is not trivial whether the trustee should abuse trust again. On the one hand, the trustee has to take into account that the trustor can obtain information through the network that the trustee deceived another trustor and, consequently, that the trustor might not trust the trustee anymore in the near future. Therefore, it might be profitable for the trustee to take the short-term profit from abusing trust. On the other hand, if the trustee abuses trust again, the deceived trustor also starts to transmit information about this deceit through the network. This causes that other trustors will be informed faster about the untrustworthiness of the trustee. Thus, the trustee has to take into account all possible information diffusion patterns among the trustors including his own role in this process if he would abuse trust again. Such a scenario could be studied using simulation. We do not apply such a simulation method, because hardly any analytic results exist for effects of detailed network parameters on information diffusion in a game-theoretic context as presented here. For such a complex model, it is then quite difficult to obtain *robust* results in a simulation study. Therefore, it is preferable to obtain some analytic results first, even if the model is more restrictive. After some analytic results are obtained, the model may be extended and a simulation could be guided by the analytic results (see Buskens 1999).

3. The Solution of the Model

We will analyze equilibria in a particular type of strategies, namely, trigger strategies (see Friedman 1971) for $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A})$. The constituent game can be characterized by the pair (i, θ), where i is the type of trustor involved and θ is the incentive for the trustee to abuse trust. Trigger strategies for trustors and trustee are defined via thresholds $\boldsymbol{\vartheta}$. In a constituent game (i, θ), a trustor of type i will place trust if $\theta \leq \vartheta_{i1}$ and she has no information that the trustee abused trust in the past. Otherwise, the trustor does not place trust. Thus, a trustor does not place trust if the trustee's incentive for abusing trust in a particular period is too large, i.e., if $\theta > \vartheta_{i1}$, or if she has any information that the trustee abused trust in the past. Consequently, as soon as the trustor obtains information about trust abused by the trustee, either from own experience or from another trustor, she will *never* place trust again.

Similarly, the trustee uses a trigger strategy with threshold ϑ_{i2} for his decision node in the game (i, θ). That means that in a period with a trustor of type i and incentive to abuse trust θ, the trustee will honor trust if trust is placed and $\theta \leq \vartheta_{i2}$ and abuse trust otherwise. Moreover, the trustee will abuse trust if the trustor has information about any abuse of trust by the trustee.[9]

Important advantages of trigger strategies are that they are analytically tractable and cognitively simple, i.e., they do not make excessive demands on the cognitive skills or memory of the actors. Equilibria will be found in which trust is only placed if the incentive for abusing trust is not too large. These equilibria are suboptimal, because all actors would receive a higher

[9] In the appendix, it is shown that this last addition to the strategy description of the trustee is only necessary to obtain *subgame-perfect* equilibria. In equilibrium, the trustor will never place trust if she has information about an abuse of trust. Therefore, the behavior of the trustee in such situations only guarantees equilibrium behavior in subgames that are not part of the equilibrium path.

payoff, if trust would always be placed and honored. This is a consequence of choosing a Heterogeneous Trust Game instead of a standard Trust Game as the constituent game, as we explained before. This leads to the favorable property of the model that the thresholds are a *measure for efficiency* that permits comparison of the trigger equilibria with the efficient situation that trust is always placed and honored. The higher the threshold ϑ, the higher the proportion of transactions $F(\vartheta)$ in which the trustor will trust the trustee, the higher the efficiency.

The following theorem states that in a subgame-perfect equilibrium in trigger strategies, for each type of trustors, the associated thresholds for the trustee and the trustor are the same ($\vartheta_{i1} = \vartheta_{i2}$).[10] Hence, on the equilibrium path, trust is placed and honored, or trust is not placed, but it never happens that trust is placed and the trustee abuses trust. In other words, trust will not always be placed, but if trust is placed, it will never be abused in equilibrium.[11]

Theorem 1 *The* IHTG $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \boldsymbol{A})$ *has the following properties.*

i) At least one subgame-perfect equilibrium in trigger strategies exists: $\vartheta_{i1} = \vartheta_{i2} = 0$ *for all* i*.*

ii) If a vector of trigger strategies is a subgame-perfect equilibrium, $\vartheta_{i1} = \vartheta_{i2}$ *for all* i*.*

Proof The proofs of this theorem and all other theorems in this paper can be found in the technical appendix.

As a result of this theorem, trigger-strategy equilibrium vectors are denoted, with some abuse of notation, by $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$. In the following theorem, a solution of the game $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A})$ is specified. A condition for subgame-perfect equilibrium in trigger strategies is given. Because usually multiple

---

[10] Strictly speaking, the game does not have proper subgames as soon as a trustor does not communicate with the subsequent trustor, because everything what has happened before is unknown by the new trustor. Still, it is a subgame if a period starting with an uninformed new trustor is considered as the start of a new game or a collapse of all possible prior states to one new state that resembles the initial beginning of the game. For the trustor, the situation is exactly the same as in the beginning of the game. The only concern is that the trustee remembers what happened before in the game. This is not problematic because the trigger strategies of the trustee are not conditioned on what happened earlier in the game, but only on what the trustor knows about past periods (see also the appendix).

[11] There exist many equilibria that involve other strategies as follows from the Folk Theorem (see, for example, Fudenberg/Tirole 1991, Section 5.1). Other strategies, for example, strategies in which trustors refrain from placing trust for only a finite number of time periods might be in equilibrium. Threatening with 'eternal' punishment, however, is the most effective in our model in which actors perfectly monitor each other. The reason is that the trustee's loss after abused trust is maximized. Therefore, trust of the trustor will be as large as possible using trigger strategies if there exist equilibria in these trigger strategies.

equilibria in trigger strategies exist, the condition of the first part of the theorem does not provide explicit predictions for the behavior of the actors. For example, according to Theorem 1, never placing trust by the trustor is an equilibrium ($\boldsymbol{\vartheta} = \mathbf{0}$). For equilibrium selection within the class of trigger strategies we use payoff dominance (Harsanyi/Selten 1988, 80–81), i.e, one equilibrium is selected over another equilibrium if the expected payoff is at least as large for all actors. The second part of the next theorem states that the payoff dominance selection criterion yields a unique subgame-perfect equilibrium in the class of trigger strategies. The $\boldsymbol{\vartheta}^*$ that belong to this subgame-perfect equilibrium is called the *solution* of the game.

Theorem 2 *Consider the* IHTG $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \boldsymbol{A})$ *with transition matrix* $\mathbf{T}$. *Let* $\tilde{\mathbf{T}}_w = (\mathbf{I} - w\mathbf{T})^{-1}$.[12]

i) The vector $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_n)$ *of trigger strategies is in subgame-perfect equilibrium if and only if*

$$\vartheta_i \leq (R_2 - P_2)\mathbf{e}_i'(\tilde{\mathbf{T}}_w - \mathbf{I})F(\boldsymbol{\vartheta}) \text{ for all } i. \quad (2)$$

ii) There exists a unique Pareto-optimal subgame-perfect equilibrium in the class of trigger strategies, with thresholds $\boldsymbol{\vartheta}^*$. *This* solution $\boldsymbol{\vartheta}^*$ *can be characterized as the maximal solution in* $\boldsymbol{\vartheta}$ *of*

$$\vartheta_i = (R_2 - P_2)\mathbf{e}_i'(\tilde{\mathbf{T}}_w - \mathbf{I})F(\boldsymbol{\vartheta}) \text{ for all } i, \quad (3)$$

where

$$\mathbf{e}_i \text{ is the } i\text{-th unit vector of length } n,$$

$$\mathbf{I} \text{ is an identity matrix of size } n, \text{and}$$

$$F(\boldsymbol{\vartheta}) = (F(\vartheta_1), \ldots, F(\vartheta_n)).$$

The solution $\boldsymbol{\vartheta}^*$ specifies for every trustor the Pareto-optimal equilibrium threshold and indicates 'how much' the trustor can trust the trustee. We will call the threshold that belongs to the solution of the game the *trust threshold.* The essential property of the solution is that the trustor places trust if the incentive to abuse trust for the trustee is 'compensated' by the expected number of times he will be sanctioned by the trustors, weighted by time preferences. This number depends on the 'network effect' of a trustor, in particular, it depends on how long the information about abuse of trust will be in the network. In the following section, I elaborate on how the solution of the model depends on the parameters of the model.

[12] We define generally $\tilde{\mathbf{X}}_w = (\mathbf{I} - w\mathbf{X})^{-1}$. Invertability is ensured by the theory of non-negative matrices (Berman/Plemmons 1979, 133).

4. Properties of the Solution

Although equation (3) seems to be a rather complicated formula, the different elements of the model are clearly separated. The first part on the right-hand side contains the sanction costs for the trustee $(R_2 - P_2)\mathbf{e}'_i$. The second term involves the matrix $(\tilde{\mathbf{T}}_w - \mathbf{I})$; the element $(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}$ is the expected discounted number of times a trustor of type j will not place trust after the trustee has abused trust placed by a trustor of type i. Note that in this expression the discount factor w is the discount factor of the trustee. The discount factors of the trustors do *not* affect the trust thresholds. Moreover, the payoffs of the trustors do *not* influence the trust threshold. In particular, the payoff S_1 for the trustors when trust would be abused, does not affect the extent to which the trustors trust the trustee, although it can be argued that the higher the payoff for the trustor if trust would be abused, the less problematic placing trust is.[13] S_1 does not affect the solution because in the equilibria in the class of trigger strategies, trust is never abused. Namely, if trust would be abused, the trustor would have 'known' that in advance and she would not have placed trust. Consequently, the model does not provide predictions about effects of payoffs and the discount factors of trustors on the extent to which trustors can place trust. Finally, it can be seen from equation (3) that the distribution F of incentives to abuse trust is important. Before the comparative statics of the trust thresholds are studied, we will discuss some special cases.

4.1 Special Cases

NO NETWORK EMBEDDEDNESS AND NO TEMPORAL EMBEDDEDNESS. Without social network and without temporal embeddedness, $\alpha_{ij} = 0$ and $\delta_i = 1$ for all i and j, the game is a one-shot game. No information of abusing trust by the trustee is transferred to the next period. Thus, there is a unique equilibrium, in which trust is never placed ($\boldsymbol{\vartheta} = \mathbf{0}$).

NO NETWORK EMBEDDEDNESS. If $\alpha_{ij} = 0$ for all i and j, no information transfer between different types of trustors occurs (see also Raub/Weesie 1993). The extent to which trust is placed, is the maximal solution in ϑ_i of

$$\vartheta_i = \frac{w(1-\delta_i)}{1 - w(1-\delta_i)}(R_2 - P_2)F(\vartheta_i) \text{ for all } i. \tag{4}$$

Thus, what happens between a trustor of type i and the trustee does not depend on the trustors of other types. The constructive proof of assertion *ii*) of Theorem 2 (see the technical appendix) implies that, although the solution is

[13] See Snijders 1996 for experimental evidence in one-shot games.

only characterized implicitly, if the right-hand side of (4) increases, the equality becomes a strict inequality and, consequently, that the new solution is a (strict) Pareto improvement of the former solution. Thus, the trust thresholds are higher if temporal embeddedness increases (δ_i decreases), if the discount factor (w) is higher, and if the sanction costs for the trustee ($R_2 - P_2$) are higher.

PERFECT NETWORK EMBEDDEDNESS. If $\alpha_{ij} = 1$ for all i and j, information of abusing trust by the trustee will be known for ever by the trustors who have to play the game. Therefore, trust will never be placed after the first defection of the trustee. Thus, $\boldsymbol{\delta}$ does not matter anymore; the identities of the trustors have become irrelevant. The restriction for payoff dominant subgame-perfect equilibrium in this case is given by the maximal solution in ϑ of

$$\vartheta = \frac{w}{1-w}(R_2 - P_2)F(\vartheta) \text{ for all } i, \tag{5}$$

which is a special case of (4) with $\boldsymbol{\delta} = 0$. Thus, here trust thresholds increase with the discount factor and the sanction costs.

HOMOGENEOUS NETWORK. Let $\alpha_{ij} = \alpha_2$ for all $i \neq j$, $\alpha_{ii} = \alpha_1$, $\delta_i = \delta$, and $\pi_i = \frac{1}{n}$ for all i. Then, the trust thresholds are the same for all types of trustors. The assumption implies that for all pairs of types of trustors, the between-type network has density α_1, while for all types of trustors the within-type network has density α_2. The 'within-type' density may be different from the 'between-type' density. The result associated with these parameters is similar to the result from Weesie et al. (1998) for homogeneous networks, and equals the maximal solution in ϑ of

$$\vartheta = \frac{w\eta_1 + w\eta_2}{1 - w\eta_1 - nw\eta_2}(R_2 - P_2)F(\vartheta), \tag{6}$$

where $\eta_1 = 1 - \delta + \frac{\delta\alpha_1}{n}$ and $\eta_2 = \frac{\delta\alpha_2}{n}$. This equation implies that the trust thresholds increase in the within-type density α_1 and in the between-type density α_2.

IDENTICAL OUTDEGREES. The last special case is the case that $\delta_i = \delta$ and $\pi_i = \frac{1}{n}$ for all i, and all types of trustors have the same probability to transmit their information to the next trustor, i.e., all types of trustors have the same outdegree:

$$D_{\text{out}}(i) = \frac{\sum_{j=1}^{n} \pi_j \alpha_{ij}}{\sum_{j=1}^{n} \pi_j} = \alpha \text{ for all } i. \tag{7}$$

The fact that all outdegrees are the same implies that all types of trustors transmit information to the next trustor with the same probability. This implies that the expected time that information about a deceit by the trustee remains in the network is always the same. Therefore, trust thresholds are the same for all types of trustors if all types of trustors have the same $D_{\text{out}}(i)$. The trust thresholds satisfy the solution for homogeneous networks in equation (6). Thus, the structure of the network does not matter if the $D_{\text{out}}(i)$ are the same for all types of trustors.

4.2 General Results

The thresholds $\boldsymbol{\vartheta}^*$ indicate the maximal incentives for the trustee to abuse trust for which the trustors still trust the trustee. In this sense, $\boldsymbol{\vartheta}^*$ is an indicator for the extent to which types of trustors can trust the trustee. $F(\boldsymbol{\vartheta}^*) = (F(\vartheta_1), \ldots, F(\vartheta_n))$ are the proportions of transactions in which the different types of trustors will place trust. Because $F(\boldsymbol{\vartheta}^*)$ is the expected proportion of periods a trustor expects to obtain R_1 compared to P_1, this might be an even better indicator than $\boldsymbol{\vartheta}^*$ for how efficiently trustors can arrange their transactions. We derive how both indicators for trust depend on the parameters of the model.

Theorem 3 *The payoff dominant subgame-perfect equilibrium in trigger strategies with trust thresholds $\boldsymbol{\vartheta}^*$ has the following properties.*

i) ϑ_i^* and $F(\vartheta_i^*)$ increase in the sanction costs for the trustee $R_2 - P_2$ for all i.

ii) ϑ_i^* and $F(\vartheta_i^*)$ increase in the discount factor w of the trustee for all i and are independent of the discount factors of the trustors.

iii) ϑ_i^* and $F(\vartheta_i^*)$ decrease in the drop-out rate δ_j if and only if a path exists from any trustor of type i to some trustors of type j in the network $(\boldsymbol{\pi}, \mathbf{A})$.

iv) ϑ_i^* and $F(\vartheta_i^*)$ increase in α_{jk} if and only if a path exists from any trustor of type i to some trustors of type j in the network $(\boldsymbol{\pi}, \mathbf{A})$.

v) ϑ_i^* and $F(\vartheta_i^*)$ decrease in F in the sense of stochastic ordering.[14]

By Theorem 3, a trustor will trust the trustee more often and efficiency will be higher, if the sanction costs for the trustee $(R_2 - P_2)$ are higher. In other words, if the trustee suffers more if a trustor does not place trust, this trustor will more frequently trust the trustee. Moreover, if the future is more important to the trustee (w is larger), punishment by the trustor will be

[14] $F_1 < F_2$ in the sense of stochastic ordering means that $F_1(\theta) > F_2(\theta)$ for all $\theta > 0$.

more severe for the trustee and so the trust threshold ϑ_i^* and the degree of efficiency $F(\vartheta_i^*)$ will be higher. If the incentives for abusing trust become smaller in the sense of stochastic ordering, the trust threshold will be higher, and the proportion of periods in which trust can be placed increases. If the trustee deals with a trustor of type i for a longer time (δ_i is smaller), the trustor of type i has better punishment possibilities and so ϑ_i^* becomes larger with temporal embeddedness. In addition, trust thresholds of those trustors increase who, directly or indirectly, have ties to trustors of type i. The reason for this is that if a trustor of type j is deceived by the trustee, there is a probability that a trustor of type i receives information about that deceit. The corresponding punishment of the trustee by trustors of type i is expected to be larger if δ_i increases. Thus, effects of temporal embeddedness increase with a longer bilateral shadow of the future. And, network effects increase with the density of the network between types of trustors (α_{jk} increases) and with the shadow of the future for connected others. These effects imply that sanctions for the trustee are larger if information about abused trust stays longer among trustors, not necessarily the deceived trustor, and if information is transferred with a higher probability to other trustors. Again, it is seen here that, in particular, possibilities for transmitting information to others increase trust.

The comparative statics of the model for non-network parameters are in correspondence with earlier models in which the network structure was not modeled explicitly (for example, Raub/Weesie 1993; Weesie et al. 1998). Thus, the additional assumptions made to incorporate the social network in a way that analytical tractability is maintained, do not distort findings about the effects of other parameters. However, concerning network embeddedness, the only result for the new model is that trust increases in the α_{ij}. This does not even imply that trust increases in the overall density of the network.

Unfortunately, the analytic results do not yet provide insights in the effects of network parameters on the trust thresholds. The results do not show whether trust thresholds increase or decrease in the outdegree of a particular trustor, or what the effect is of centralization of the network. We were not able to derive such results analytically, at least partly because the trust thresholds are not given as explicit functions of the relevant network parameters. Section 5 describes the linearization if the trust thresholds. Using this approximation, hypotheses are developed about the effects of the network parameters on trust thresholds as 'approximate implications' of the model. Before we continue with the approximation, we discuss two theorems on the relation between the trust thresholds and the global network structure.

4.3 Global Network Implications

We present two theorems stating what the 'optimal' global network configuration would be if the outdegrees $D_{\text{out}}(i)$ of all types of trustors are given. If no other restrictions are made with respect to the contacts of the trustors, the set of equilibria have a lattice structure: an optimal structure exists that is Pareto superior—in the sense that the trust thresholds are higher for all types of trustors—to other all network structures with the same outdegrees. If the extra restriction is imposed that ties should be symmetric, a unique Pareto-optimal network structure does generally not exist. Nevertheless, we can indicate the optimal structure under the assumption that all trustors spread their ties in an individually rational manner, but that the ties cannot be forced unilaterally. First, we discuss the theorem for $n = 2$ to find indications about structures that lead to the highest levels of trust. Thereafter, we generalize this theorem with an inductive argument to $n > 2$.

Theorem 4 *Consider the* IHTG $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A})$ *with* $k = 2$ *and* $\delta_1 = \delta_2 = \delta$.

i) If $D_{\text{out}}(1) > D_{\text{out}}(2)$*, the* ϑ_1^* *and* ϑ_2^* *are optimal with relation to the network structure for a transition matrix* $\mathbf{T}^*$,

$$\mathbf{T}^* = \begin{pmatrix} 1 - \delta + \delta \min(\pi_1, D_{\text{out}}(1)) & \delta \max(0, D_{\text{out}}(1) - \pi_1) \\ \delta \min(\pi_1, D_{\text{out}}(2)) & 1 - \delta + \delta \max(0, D_{\text{out}}(2) - \pi_1) \end{pmatrix}. \quad (8)$$

ii) If $D_{\text{out}}(1) > D_{\text{out}}(2)$, $\alpha_{12} = \alpha_{21}$ *(symmetric relations), and we assume that bilateral relations cannot be forced unilaterally, then* ϑ_1^* *and* ϑ_2^* *are (individually) optimal with relation to the network structure for a transition matrix* $\mathbf{T}^*$

$$\mathbf{T}^* = \begin{pmatrix} 1 - \delta + \delta \min(\pi_1, D_{\text{out}}(1)) & \delta \max(0, D_{\text{out}}(1) - \pi_1) \\ \delta \max(0, D_{\text{out}}(1) - \pi_1) & 1 - \delta + \delta(D_{\text{out}}(2) - \max(0, D_{\text{out}}(1) - \pi_1) \end{pmatrix}. \quad (9)$$

Theorem 4 shows that the network that is *centralized* around the trustors with the highest outdegree is the 'best' network that can be constructed for fixed outdegrees. Part *ii*) of the theorem considers only symmetric relations ($\alpha_{ij} = \alpha_{ji}$). Because trustors of type 1 have the most ties, the optimal distribution of ties as given in the theorem prescribes that there are as much as possible ties among trustors of type 1. If the network between trustors of type 1 is complete and trustors of type 1 have ties left, these are ties between trustors of type 1 and trustors of type 2. The remainder of the ties of trustors of type 2 are used between themselves. The following theorem generalizes the results from Theorem 4 for $n > 2$.

Theorem 5 *Consider the* IHTG $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A})$ *with* $\delta_1 = \delta_2 = \cdots = \delta_k = \delta$. *Then the following properties hold.*

i) If $D_{\text{out}}(1) > D_{\text{out}}(2) > \cdots > D_{\text{out}}(k)$ *then the* $\vartheta_1^*, \ldots, \vartheta_k^*$ *are optimal if the transition matrix* $\mathbf{T}$ *is chosen such that everybody communicates as much as possible to trustors of type 1, after that communicates as much as possible to trustors of type 2, and so on.*

ii) If $D_{\text{out}}(1) > D_{\text{out}}(2) > \cdots > D_{\text{out}}(k)$ *and* $\alpha_{ij} = \alpha_{ji}, i \neq j$ *(symmetric relations) and we assume that bilateral relations cannot be forced unilateral, then the* ϑ_i^* *are (individually) optimal if in the transition matrix* $\mathbf{T}$ α_{11} *is chosen maximal first. After that we choose the* α_{ij} *maximal for which* $i + j = 3$. *Then the* α_{ij} *for which* $i + j = 4$ *are set maximal, and so on.*

The last two theorems indicate how trustors can organize their (limited) number of ties such that information is optimally communicated through the network. It should be noted that the emphasis in the model is on the *transmission* of information to others. For optimal sanctions, information should be transmitted as long as possible to other trustors. For this aim, the network structures specified in the theorems above are optimal. The trustors are not interested in receiving information because in equilibrium trust is never abused and, consequently, there will never be information about abused trust in the network. In the asymmetric case, the trustors with relatively few ties will hardly ever or never receive information, but they do not care. In the symmetric case, the optimal structure is as well suited for transmitting as for receiving information as a result of the symmetry. However, although the symmetric structures are individually optimal based on bilateral choices from trustors, they are in fact worst cases for the trustors with the lowest outdegrees because these trustors are largely condemned to communication among each other which implies that information does not reach very far.

5. Approximation Using Linearization

In Section 3, we characterized implicitly the trust thresholds of the trustors in the IHTG in terms of the parameters of $\Gamma(\Gamma_F, w, \boldsymbol{\delta}, \boldsymbol{\pi}, \mathbf{A})$. To obtain a better understanding in the properties of the solution (3) of the IHTG, we apply linearization. Explicit linearization can be applied around function values where explicit expressions of the trust thresholds can be obtained. As can be seen from the 'special cases', such an explicit expression is possible for a homogeneous network

$$\mathbf{A}_0 = \alpha_1 \mathbf{I} + \alpha_2 \mathbf{J}, \text{ where } \mathbf{J} = \mathbf{1}\mathbf{1}', \tag{10}$$

and if we use a specific distribution of incentives to abuse trust, $F_a(\theta) = \frac{\theta}{a+\theta}, \theta \geq 0, a > 0$. Below, we will also use the first derivative of $F_a(\theta)$ with respect to θ, $f_a(\theta) = \frac{a}{(a+\theta)^2}$. To be able to focus on the network effects on the trust thresholds, we use two additional homogeneity assumptions:

$$\pi_i = \frac{1}{n} \quad \text{and} \quad \delta_i = \delta \quad \text{for all } i.$$

A first-order approximation is deduced for the trust threshold corresponding to a network matrix $\mathbf{A}$ in the 'neighborhood' of $\mathbf{A}_0$ under the assumption that positive equilibria exist. The approximation symbol ($\approx$) in the Theorem 6 indicates that the difference between the approximation and the true value is in the order of magnitude of the difference between $\mathbf{A}$ and $\mathbf{A}_0$ (see the appendix for details). Theorem 6 gives the first-order approximation of the trust thresholds $\boldsymbol{\vartheta}^*(\mathbf{A})$ for $\mathbf{A}$.

Theorem 6 *Let $\mathbf{A} \approx \mathbf{A}_0 = \alpha_1\mathbf{I} + \alpha_2\mathbf{J}$ and $\vartheta_i^*(\mathbf{A}_0) > 0$ for all i. Then the solution of equation (3) for trust thresholds ϑ^* satisfies*

$$\begin{aligned} \vartheta^*(\mathbf{A}) &\approx \vartheta^*(\mathbf{A}_0) + \rho_1(\mathbf{I} + \rho_2\mathbf{J})(\mathbf{A} - \mathbf{A}_0)\mathbf{1} \\ &= \vartheta^*(\mathbf{A}_0) + n\rho_1\Big(\mathbf{D}_{\text{out}}(\mathbf{A}) - \mathbf{D}_{\text{out}}(\mathbf{A}_0)\Big) \\ &\quad + n^2\rho_1\rho_2\Big(\Delta(\mathbf{A}) - \Delta(\mathbf{A}_0)\Big)\mathbf{1}, \end{aligned} \tag{11}$$

where

$$F_a(\theta) = \frac{\theta}{a+\theta}, f = F', \tag{12}$$

$$\rho_1 = \frac{\delta w(R_2 - P_2)F_a(\vartheta_0^*)}{n(1 - w\eta_1 - nw\eta_2))(1 - w\eta_1(1+\mu))} > 0, \tag{13}$$

$$\rho_2 = \frac{(1 + f_a(\vartheta_0^*)(R_2 - P_2))w\eta_2}{1 - (w\eta_1 + nw\eta_2)(1+\mu)} > 0, \tag{14}$$

$$\mathbf{D}_{\text{out}} \textit{ the vector of all outdegrees, and} \tag{15}$$

$$\mu = f_a(\vartheta_0^*)(R_2 - P_2), \tag{16}$$

$$\eta_1 = 1 - \delta + \frac{\delta\alpha_1}{k}, \textit{ and } \eta_2 = \frac{\delta\alpha_2}{k}. \tag{17}$$

It follows immediately from Theorem 6 that $n\rho_1$ indicates the magnitude of the change in the trust thresholds for a change in the outdegree for a certain trustor, while $n^2\rho_1\rho_2$ is the 'weight' of a small increase in total density of the network. Because ρ_1 and ρ_2 are both positive, positive effects are predicted of outdegree and density on the trust thresholds $\boldsymbol{\vartheta}^*$. The magnitudes of the effects of outdegree and density depend on the other parameters of the model ($R_2 - P_2$, a, w, δ, and n). If ρ_1 and ρ_2 would be clearly monotonic in some of the parameters, interaction effects could be derived from the theorem straightforwardly. However, we did not find arguments that ρ_1 is monotonic in any of the parameters in the relevant range. Consequently, we do not find hypotheses about interaction effects between outdegree and density, and the other parameters directly from this linearization.

However, the linearization result does imply that the parameters in the game-theoretic model determine how large the *relative* size of the effects of changes in outdegree and density will be. The parameter ρ_2, or more precisely $n\rho_2$, can be interpreted as the relative effect of density compared to outdegree. Again, monotonicity of ρ_2 is not guaranteed for w and δ. However, two monotonic effects are found, namely, ρ_2 increases in a and decreases in $R_2 - P_2$.[15] Therefore, it is predicted that density becomes more important compared to outdegree if the sanction costs for the trustee ($R_2 - P_2$) decrease and if the median incentive for the trustee to abuse trust (a) increases. These two results are intuitively appealing because they indicate that the 'whole network' (density) becomes more important compared to 'ego-centered network' (outdegree) if the trust problem in the constituent game increases. Thus, if trust problems are relatively small, it is important that trustors know a number of others who may sanction an untrustworthy trustee. However, if trust problems increase it becomes more and more important that these others also know each other.

The findings based on linearization are limited to networks 'close' to a homogeneous situation. They cannot be generalized to more heterogeneous networks without further examination. Therefore, in another paper, we use a simulation method in which we also investigate effects of network parameters other than density and outdegree on trust thresholds in more heterogeneous networks (Buskens 1998). The following conclusions were based on this simulation. First, outdegree and density are also the most important predictors for the trust thresholds in more heterogeneous networks. The effects for other network parameters are (much) smaller. Second, it is demonstrated that the trust thresholds increase with the centralization of the network only if trustors with higher outdegrees have also high indegrees. This corresponds with our results in Theorem 4 and Theorem 5. Finally, the simulation provides several predictions for interaction effects between network parameters and other aspects of the IHTG. Most notably, the simulation predicts that density becomes more important compared to outdegree if a trustor is more vulnerable, i.e., if sanctions for the trustee of withholding trust are smaller, the duration of a series of transactions is shorter, and the trustee has larger incentives to abuse trust.

---

[15] These results can be seen directly from (14), realizing that $\mu(R_2 - P_2) f_a(\vartheta_0^*) = \frac{(R_2-P_2)a}{(\vartheta_0^*+a)^2} = \frac{a(1-w\eta_1-nw\eta_2)^2}{(R_2-P_2)w^2(w\eta_1+w\eta_2)^2}$, which decreases in $R_2 - P_2$ and increases in a.

6. Conclusions and Discussion

Axelrod (1984) has argued, a.o., that cooperation can be based on the weighting of short-term benefits, and long-term costs that increase in the 'shadow of the future'. The analysis of the model discussed in this paper demonstrates that this argument can be extended from dyadic encounters to encounters that are embedded in a social network. Sanctions against untrustworthy partners can be executed not only by the direct partner but also by other actors in the social network. This aggravates the consequences of the sanctions and, thus, the possibilities for cooperative behavior increase with network embeddedness. In accordance with existing informal and formal literature (Granovetter 1985; Coleman 1990; Raub/Weesie 1993; Weesie et al. 1998), more trust is possible if the costs of sanctions for the trustee are higher ($R_2 - P_2$), if the trustee is more patient (i.e., w is higher), if the trustor expects to be involved with the trustee for a longer time (δ_i), and if the average incentive for the trustee to abuse trust (F) becomes smaller. In addition we predict that trust is higher in more dense networks, and that trustors with a higher outdegree can place more trust. Furthermore, we obtained some results for specific network structures. First, if all types of trustors in the network have the same outdegree, the network structure has no affect on the possibility to arrange transactions with incomplete contracts. Second, if the outdegrees differ, the network that is centralized around the trustors with the highest outdegrees produces the highest levels of trust.

Our results are derived in a model that makes strong—maybe too strong—assumptions that we may relax in future models. A restrictive feature of our model is that trustors act sequentially, not simultaneously. Also, only after a trustor's series of transaction ends and a new trustor starts her interactions with the trustee, trustors can exchange information, and they can do so only once. Obtaining analytical results for models with trustors, operating in parallel, who are incompletely connected via information channels, will probably prove to be very difficult, if not impossible for all but the simplest of cases. Various types of computer simulation could be used here. Simulations in the spirit of evolutionary game theory (i.e., with a fixed strategy pool as in the famous computer tournaments of Axelrod) or of genetic algorithms (with a strategy pool, evolving in a quasi-biological way; see Holland 1975; 1998; Axelrod 1986; 1997) can be useful here. We feel, however, that it is hard not to be cautious enough in generalizing simulation results for so complicated models if a firm analytical understanding of at least certain special cases is not available (see Binmore 1998 for similar criticism on Axelrod's original simulations).

Another possibility is to explore Von Neumann's cellular automata in which cells correspond to actors who play some (iterated) game with their neighbors on the board. This approach was also pioneered by Axelrod (1984, chapter 8), and can be seen as a study of network effects as well. The network effect in the cellular automaton approach is of a totally different nature than in this paper, namely the network is seen as a source of information about what constitutes successful behavior (of course, the conceptual step from observing behavior to observing strategies is a very long one!) that one simply imitates. Thus, the network may give information about successful strategies, irrespective of the partner, and not only about the behavior or characteristics or specific partners.

One can also turn to developing models that make simplifying assumptions in another direction. An example is the stochastic diffusion model considered by Buskens/Yamaguchi (1999). Here, the micro-behavioral assumption of incentive-guided behavior is replaced by a much simpler stochastic model. This allows to study more complicated structures in terms of scheduling of actions of actors.

A second problem with our model is that we cannot make predictions about the effect of the payoffs and time preferences of the trustors on the levels of trust. Technically, this is mainly due to our assumption that the trustors can perfectly observe the behavior of the trustee. If this would not be the case, trustors would have to be more careful with their punishment strategy—recall, they are now assumed to be non-forgiving, the most severe punishment possible—because they may think that the trustee defected, while in fact the trustee did not. Punishment is also costly for the trustor and the more costly it is the more careful she should be (see, e.g., Green/Porter 1984). Another complete information assumption may even be less realistic: We have assumed that the social network among the trustors is common knowledge to all players. There is ample empirical evidence that people have very scarce and even systematically biased information about the network they are embedded in. For instance, people often think of themselves as more more central in networks then they really are (Kumbasar et al. 1994). Intuitively, it is not very clear how our predictions are affected if we relax this assumption. We would not be surprised if a theoretical analysis may yield that, under some conditions yet unknown, a trustee will rationally be very careful and not abuse trust unless he is quite certain he can do so with limited consequences. In other cases, it may simply be 'boundedly rational' to play it safe if one intends 'to stay in business' (Macaulay 1963).

Appendix Mathematical Details

Proof of Theorem 1 (page 53) If all actors play the one-shot equilibrium strategies D_i all the time, this constitutes a subgame-perfect equilibrium. This proves assertion *i)*. Now consider a strategy vector $(\vartheta_{11}, \ldots, \vartheta_{n1}, \vartheta_{12}, \ldots, \vartheta_{n2})$. First, assume $\vartheta_{i1} < \vartheta_{i2}$ for at least one i. Then, a trustor of type i does not maximize her payoff. If she chooses $\vartheta_{i1} = \vartheta_{i2}$, her payoff increases with $R_1 - P_1$ every time that $\vartheta_{i1} < \theta < \vartheta_{i2}$. This occurs with positive probability since F has full support. In all other cases the payoff is the same. On the other hand, assume $\vartheta_{i1} > \vartheta_{i2}$ for at least one i. In this case, a trustor of type i does not maximize her payoff, because the trustee will play D_2 if $\vartheta_{i1} > \theta > \vartheta_{i2}$. The trustee will receive $R_2 + \theta$, while the trustor receives S_1 and in all the following games the trustor plays D_1 and receives a payoff P_1. However, if the trustor had chosen $\vartheta_{i1} = \vartheta_{i2}$, she would have received P_1 in the game mentioned and R_1 or P_1 in all the following games, which is more than she receives now. Thus, if $\vartheta_{i1} \neq \vartheta_{i2}$, the trustor of type i increases her payoff by moving her threshold toward the threshold of the trustee. Therefore, if $\vartheta_{i1} \neq \vartheta_{i2}$ the trustor does not use a best reply. This shows that the trustor's threshold and trustee's threshold with respect to this trustor are the same in equilibrium.

Proof of Theorem 2 (page 54) First, we prove the equilibrium condition of assertion *i)*. The game analyzed here is a repeated game with infinite horizon and exponential discounting. Hence, we can use a well-known result of dynamic programming theory, namely, Bellman's optimality principle (Bellman 1954; Kreps 1990b). Optimality on the equilibrium path is guaranteed if deviations from the prescribed path in any decision node do not increase the payoff. Therefore, it has to be proven that if an actor makes a one-step deviation from the equilibrium path, the actor's payoff will not increase. Without loss of generality, consider deviations at time $t = 0$. The involved trustor is a trustor of type i. First, consider $\theta > \vartheta_i$. Then, both actors defect, and, therefore, no one has an incentive to deviate. Now, consider $\theta \leq \vartheta_i$. Again the trustor has no incentive to deviate, because on the equilibrium path she receives her maximal payoff R_1. The trustee should play D_2 if he can obtain a short-term gain that is higher than the long-term loss due to the punishment by the trustors. Thus, it has to be proven that the restriction in Theorem 2 is exactly the condition that the long-term loss for the trustee will be larger than the short-term gain, if he plays D_2 and $\theta \leq \vartheta_i$. In other words, we will show that the trustee also does not have an incentive to deviate from the equilibrium path.

Let $EU_2(C_2, \theta; \boldsymbol{\vartheta})$ be the expected payoff for the trustee if both actors follow the trigger strategy. In equilibrium, no trustor is ever informed about deviations from the equilibrium path by the trustee. To calculate the expected payoff of the trustee, a matrix is used that contains the probabilities about which trustor will be involved in subsequent games. The information exchange opportunities are not relevant in this case. The probability matrix equals

$$\mathbf{\Pi} = \begin{pmatrix} 1-\delta_1+\delta_1\pi_1 & \delta_1\pi_2 & \cdots & \delta_1\pi_n \\ \delta_2\pi_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta_{n-1}\pi_n \\ \delta_n\pi_1 & \cdots & \delta_n\pi_{n-1} & 1-\delta_n+\delta_n\pi_n \end{pmatrix}, \tag{18}$$

where the diagonal elements are the probabilities that the same trustor continues a series of transactions or stops and is chosen again as the new trustor. The other elements are the probabilities that one trustor stops and another trustor starts a series of transactions.

The trustee's payoff with a trustor of any type j he encounters in the future equals P_2 with probability $Pr(\theta > \vartheta_j) = 1 - F(\vartheta_j)$ and R_2 with probability $F(\vartheta_j)$. The trustee's expected payoff for the equilibrium path, if he starts transactions with a trustor of type i equals

$$EU_2(C_2, \theta; \boldsymbol{\vartheta}) = R_2 + \sum_{t=1}^{\infty} w^t \mathbf{e}_i' \mathbf{\Pi}^t \boldsymbol{\mu} = R_2 + \mathbf{e}_i'(\tilde{\mathbf{\Pi}}_w - \mathbf{I})\boldsymbol{\mu}, \tag{19}$$

where $\boldsymbol{\mu}$ is a n-vector with $\mu_i = F(\vartheta_i)R_2 + (1 - F(\vartheta_i))P_2$.

Now the trustee deviates from the equilibrium path and abuses trust (D_2). The necessary condition for subgame-perfect equilibrium is that $EU_2(C_2, \theta; \boldsymbol{\vartheta}) \geq EU_2(D_2, \theta; \boldsymbol{\vartheta})$ for all $\theta \leq \vartheta_i$. The involved trustor of course knows that trust has been abused. The probabilities whether or not the following trustors will have the information about the abuse of trust are given in matrix

$$\mathbf{Q} = \begin{array}{c} \\ S_{inf} \\ S_{noinf} \end{array} \begin{array}{c} \begin{array}{cc} S_{inf} & S_{noinf} \end{array} \\ \left(\begin{array}{c|c} \mathbf{T} & \mathbf{\Pi} - \mathbf{T} \\ \hline \mathbf{0} & \mathbf{\Pi} \end{array}\right) \end{array}, \tag{20}$$

where $\mathbf{T}$ denotes the probabilities that a trustor who has information about abused trust communicates this to the trustor in the following game (see equation (1)). $\mathbf{\Pi} - \mathbf{T}$ are the probabilities that a trustor with information about abused trust does not communicate this to the next trustor. $\mathbf{\Pi}$ are the probabilities for which trustor will be involved in a following game after the information about abused trust is lost.[16]

The payoff for the trustee is $R_2 + \theta$ in the game in which he abuses trust; the payoff will be P_2 as long as the trustee encounters trustors who know about this deceit. As soon as information is not transmitted to the following trustor, the payoffs are as if no trustor was ever deceived. This is the consequence of the fact that information is lost among the trustors as soon as a new trustor who starts a series of transactions does not obtain information from the foregoing trustor, and the trustee knows that the trustors do not know anymore about his abuse of trust.

---

[16] The stochastic matrix $\mathbf{Q}$ is a transition matrix of a Markov chain in which the states are defined by which trustor has transactions with the trustee and whether this trustor has information about trust abused by the trustee or not.

Thus,

$$\begin{aligned} EU_2(D_2,\theta;\boldsymbol{\vartheta}) &= R_2+\theta+\sum_{t=1}^{\infty} w^t \left(\mathbf{e}_i' \quad \mathbf{0}'\right) \mathbf{Q}^t \begin{pmatrix} P_2\mathbf{1} \\ \boldsymbol{\mu} \end{pmatrix} \\ &= R_2+\theta+\left(\mathbf{e}_i' \quad \mathbf{0}'\right)(\tilde{\mathbf{Q}}_w-\mathbf{I}) \begin{pmatrix} P_2\mathbf{1} \\ \boldsymbol{\mu} \end{pmatrix}. \end{aligned} \tag{21}$$

By using $\mathbf{T}\tilde{\mathbf{T}}_w = \tilde{\mathbf{T}}_w - \mathbf{I}$ and, $\boldsymbol{\Pi}\tilde{\boldsymbol{\Pi}}_w = \tilde{\boldsymbol{\Pi}}_w - \mathbf{I}$, it follows by straightforward computation that

$$\begin{aligned} \tilde{\mathbf{Q}}_w-\mathbf{I} &= \begin{pmatrix} \mathbf{I}-w\mathbf{T} & -w(\boldsymbol{\Pi}-\mathbf{T}) \\ \mathbf{0} & \mathbf{I}-w\boldsymbol{\Pi} \end{pmatrix}^{-1} - \mathbf{I} \\ &= \begin{pmatrix} \tilde{\mathbf{T}}_w & w\tilde{\mathbf{T}}_w(\boldsymbol{\Pi}-\mathbf{T})\tilde{\boldsymbol{\Pi}}_w \\ \mathbf{0} & \tilde{\boldsymbol{\Pi}}_w \end{pmatrix} - \mathbf{I} \\ &= \begin{pmatrix} \tilde{\mathbf{T}}_w & \tilde{\boldsymbol{\Pi}}_w-\tilde{\mathbf{T}}_w \\ \mathbf{0} & \mathbf{I}+\frac{w}{1-w}\boldsymbol{\Pi} \end{pmatrix} - \mathbf{I} \\ &= \begin{pmatrix} \tilde{\mathbf{T}}_w-\mathbf{I} & \tilde{\boldsymbol{\Pi}}_w-\tilde{\mathbf{T}}_w \\ \mathbf{0} & \frac{w}{1-w}\boldsymbol{\Pi} \end{pmatrix}. \end{aligned} \tag{22}$$

By substituting (22) in (21):

$$\begin{aligned} EU_2(D_2,\theta;\boldsymbol{\vartheta}) &= R_2+\theta+\mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I})P_2\mathbf{1}+\mathbf{e}_i'(\tilde{\boldsymbol{\Pi}}_w-\tilde{\mathbf{T}}_w)\boldsymbol{\mu} \\ &= R_2+\theta+\mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I})(P_2\mathbf{1}-\boldsymbol{\mu})+\mathbf{e}_i'(\tilde{\boldsymbol{\Pi}}_w-\mathbf{I})\boldsymbol{\mu}. \end{aligned} \tag{23}$$

Hence, $EU_2(C_2,\theta;\boldsymbol{\vartheta}) \geq EU_2(D_2,\theta;\boldsymbol{\vartheta})$ is equivalent to

$$\begin{aligned} \theta &\leq \mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I})(\boldsymbol{\mu}-P_2\mathbf{1}) \\ &= \mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I}) \begin{pmatrix} (R_2-P_2)F(\vartheta_1) \\ \vdots \\ (R_2-P_2)F(\vartheta_n) \end{pmatrix} \\ &= (R_2-P_2)\mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I})F(\boldsymbol{\vartheta}). \end{aligned} \tag{24}$$

In equilibrium, (24) should hold for all $\theta \leq \vartheta_i$: $EU_2(C_2,\theta;\boldsymbol{\vartheta}) \geq EU_2(D_2,\theta;\boldsymbol{\vartheta})$; clearly $\theta = (R_2-P_2)\mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I})F(\boldsymbol{\vartheta})$ is the most restrictive θ for which the trustee has no incentive to abuse trust placed by a trustor of type i. Consequently, the threshold chosen by a trustor of type i should be smaller or equal to this value, i.e.,

$$\vartheta_i \leq (R_2-P_2)\mathbf{e}_i'(\tilde{\mathbf{T}}_w-\mathbf{I})F(\boldsymbol{\vartheta}) \tag{25}$$

to ensure that the trustee never abuses trust. Because this holds for all i, the trigger strategies are in equilibrium.

Remark. There is a small problem with 'subgame perfectness' in this game. As long as trustors communicate information when a new trustor enters the game, subgame perfectness is well defined. In that case, the trigger equilibrium is subgame perfect, because the actors act according the equilibrium in the constituent game after leaving the equilibrium path, and this is also an equilibrium of the repeated game. Note that to ensure subgame perfectness here, it was necessary to assume that the trustee will continue to abuse trust after abusing trust once as long as the trustors know about the abuse of trust, and, thus, that the trustee knows whether trustors communicate information about the behavior of the trustee. If a trustor enters the game and does not receive information from her predecessor, this trustor does not know anything about what happened in the game before. She even does not know how many rounds are already played. Thus, strictly speaking, the node in which this trustor starts, and all nodes thereafter, cannot be considered as subgames because there is a large set nodes that the game could have reached and that are all in one information set for the new trustor. On the other hand, the trustor cannot distinguish her situation from the first round in the game and, therefore, from the viewpoint of the trustor the whole information set can be 'collapsed' to one node. The only problem is that the trustee knows what happen in previous rounds. But, because the trustee does not use the experiences he had in previous rounds, we consider the round that starts with a trustor who is not informed about the behavior of the trustee in earlier rounds as the start of a new game. For this new game, the trigger equilibria are subgame-perfect equilibria. This finishes the proof of the equilibrium condition.[17]

Assertion *ii)* states that there exists a unique Pareto optimal equilibrium in trigger strategies (assertion *ii)*). First, we show that if inequality (25) is strict for a certain threshold, there exists a Pareto dominant equilibrium in which equality holds. The second step to prove assertion *ii)* consists of a construction that shows that two Pareto not-comparable equilibria thresholds are always Pareto dominated by another equilibrium threshold. Suppose there exists a subgame-perfect equilibrium with $\vartheta_i < (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}F(\vartheta_j)$ for at least one i. Then ϑ_i can be increased until $\hat{\vartheta}_i = (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}F(\vartheta_j)$ because the right side is bounded above by $\frac{(R_2-P_2)w}{1-w}$ and monotonous increasing in ϑ_i, while ϑ_i is not bounded above. The other inequalities for $j \neq i$ still hold because the right sides increase with ϑ_i. In the next step, it has to be checked whether $\vartheta_{i+1} = (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{i+1j}F(\vartheta_j)$, otherwise ϑ_{i+1} can be increased until $\hat{\vartheta}_{i+1} > \vartheta_{i+1}$. Continue this procedure until ϑ_n and start again from ϑ_1, etc. This gives an increasing sequence of vectors $\boldsymbol{\vartheta}$associated with equilibria in trigger strategies, which is bounded because $\vartheta_i \leq \frac{(R_2-P_2)w}{1-w}$. By the convergence theorem that bounded increasing sequences converge (Akkermans/Van Lint 1970: 217), the sequence converges to a limit $\boldsymbol{\vartheta}$ where $\vartheta_i = (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}F(\vartheta_j)$ for all i and the equilibrium associated with the limit is a Pareto improvement of all equilibria found before.

[17] Note that the discussion about *subgame perfectness* of the trigger equilibria is a discussion about 'subgames' and not about 'perfectness'. A reader who does not want to consider a round starting with an uninformed trustor as the start of a new game should introduce beliefs for the new trustor about the situation in which she enters the game. The only way to introduce beliefs that are consistent with the strategies is that the game has followed an equilibrium path. This implies that the trustor is convinced that trust is never abused and, therefore, she will act as if she starts a new game.

Now it has to be shown that there exists a *unique* $\boldsymbol{\vartheta}^*$ that Pareto dominates all other equilibria in trigger strategies. Suppose that there are two Pareto dominant equilibria $\boldsymbol{\vartheta}$and $\hat{\boldsymbol{\vartheta}}$. Then there exists, possibly after relabeling, an l such that

$$\vartheta_i \geq \hat{\vartheta}_i \text{ for } 1 \leq i \leq \text{ and } \vartheta_i < \hat{\vartheta}_i \text{ for } l < i \leq n \tag{26}$$

Define $G_i(\boldsymbol{\vartheta}) = \vartheta_i - (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}F(\vartheta_j)$. Because F is a strictly increasing function, $R_2 - P_2 > 0$, and all matrix elements of $(\tilde{\mathbf{T}}_w - \mathbf{I})$ are positive, $\partial G_i/\partial\vartheta_j < 0$ if $i \neq j$. Then,

$$G_i(\vartheta_1, \ldots, \vartheta_l, \hat{\vartheta}_{l+1}, \ldots \hat{\vartheta}_n) < G_i(\vartheta_1, \ldots, \vartheta_n) = 0, \quad 1 \leq i \leq l; \tag{27}$$

$$G_i(\vartheta_1, \ldots, \vartheta_l, \hat{\vartheta}_{l+1}, \ldots \hat{\vartheta}_n) \leq G_i(\hat{\vartheta}_1, \ldots, \hat{\vartheta}_n) = 0, \quad l < i \leq n. \tag{28}$$

Therefore, $(\vartheta_1, \ldots, \vartheta_l, \hat{\vartheta}_{l+1}, \ldots \hat{\vartheta}_n)$ is a subgame-perfect equilibrium. This equilibrium Pareto dominates the equilibria considered before. Thus, from two Pareto non-comparable equilibria, it is possible to construct a subgame-perfect equilibrium that is a Pareto improvement of the two equilibria found before, contradicting that $\boldsymbol{\vartheta}$and $\hat{\boldsymbol{\vartheta}}$ were payoff dominant equilibria. Thus, it is impossible that two different payoff dominant equilibria exist, which proves uniqueness.

Proof of Theorem 3 (page 57) The iterative algorithmic argument in the proof of Theorem 2 implies that if for a certain change in the parameters a ϑ_i can be increased, it is possible to construct a new subgame-perfect equilibrium that is a Pareto improvement of the equilibria found before. Consequently, if one of the right sides of the equations in (25) increases, it is possible to increase the corresponding ϑ_i. This implies that there exists a Pareto superior equilibrium. Therefore, the comparative statics of ϑ_i can be derived from the following equation directly by studying whether the right side of the equation increases or decreases in a certain parameter, although ϑ_i is only implicitly given in the equations.

$$\vartheta_i^* = (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}F(\vartheta_j^*). \tag{29}$$

Define $H_i = (R_2 - P_2)\sum_{j=1}^{n}(\tilde{\mathbf{T}}_w - \mathbf{I})_{ij}F(\vartheta_j)$. To prove *i)*, note that $F(\theta) \geq 0$ and all matrix elements of $\tilde{\mathbf{T}}_w - \mathbf{I}$ are positive, so $\frac{\partial H_i}{\partial(R_2-P_2)} > 0$, which implies that the right side of (29) increases in $R_2 - P_2$. As stated before this is sufficient to prove assertion *i)*. Assertion *ii)* follows because $\frac{\partial\tilde{\mathbf{T}}_w}{\partial w} = -\tilde{\mathbf{T}}_w\frac{\partial(\mathbf{I}-w\mathbf{T})}{\partial w}\tilde{\mathbf{T}}_w > 0$ element-wise and, therefore, $\frac{\partial H_i}{\partial w} > 0$. To prove *iii)*, note that if $(\tilde{\mathbf{T}}_w)_{ij} > 0$ then $\frac{\partial\sum_{l=1}^{n}(\tilde{\mathbf{T}}_w)_{il}}{\partial\delta_j} > 0$; and, therefore, $\frac{\partial H_i}{\partial\delta_j} > 0$. This is exactly the case if a path exists from a trustor of type i to a trustor of type j. The argument for *iv)* is similar to the argument for *iii)*. Again, if a path exists from a trustor of type i to a trustor of type j, it holds that $(\tilde{\mathbf{T}}_w)_{ij} > 0$, which implies that $\frac{\sum_{l=1}^{n}(\tilde{\mathbf{T}}_w)_{il}}{\alpha_{jn}} > 0$; and, therefore, $\frac{\partial H_i}{\partial\alpha_{jn}} > 0$. To prove *v)*, define F_1 and F_2 to be two probability distributions for which $F_1 > F_2$ in the sense of stochastic ordering. Note that $R_2 - P_2 > 0$ and all matrix elements are positive. Consequently, H_i increases for all i changing from

F_1 to F_2 and, Moreover, the proportion of games in which trust will be placed increases because $F_1(\vartheta^*_{1,i}) < F_2(\vartheta^*_{1,i}) < F_2(\vartheta^*_{2,i})$ for all i, where $\vartheta^*_{j,i}$ is the solution for a trustor i and F_j.

Proof of Theorem 4 (page 59) First, we prove that $\sum_{j=1}^2(\tilde{\mathbf{T}}-\mathbf{I})_{1j} > \sum_{j=1}^2(\tilde{\mathbf{T}}-\mathbf{I})_{2j}$ implies that $\vartheta_1 > \vartheta_2$. To prove this assertion, assume that under the given condition $\vartheta_1 \leq \vartheta_2$. Then the following inequalities hold.

$$\begin{aligned} \vartheta_2 &= (R_2 - P_2)\sum_{j=1}^2(\tilde{\mathbf{T}}_w - \mathbf{I})_{2j}F(\vartheta_j) \\ &\leq (R_2 - P_2)F(\vartheta_2)\sum_{j=1}^2(\tilde{\mathbf{T}}_w - \mathbf{I})_{2j} \\ &< (R_2 - P_2)F(\vartheta_2)\sum_{j=1}^2(\tilde{\mathbf{T}}_w - \mathbf{I})_{1j}. \end{aligned} \tag{30}$$

The last strict inequality follows directly from $\gamma_1 < \gamma_2$. Therefore, $(\vartheta_2, \vartheta_2)$ is a feasible solution of the inequalities of Theorem 2 and from Theorem 2 it is known that there exists a subgame-perfect equilibrium that is a Pareto improvement of the equilibrium $(\vartheta_2, \vartheta_2)$, which is already a Pareto improvement of $(\vartheta_1, \vartheta_2)$ and it still can be improved because (30) is a strict inequality. This is in contradiction with the fact that ϑ was the Pareto optimal solution. As a consequence, $\vartheta_1 > \vartheta_2$.

Now, to prove *i*), define in the transition matrix $\mathbf{T}$ $T_i = 1 - \delta + \delta D_{\text{out}}(i)$. Then,

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_1 - T_{11} \\ T_2 - T_{22} & T_{22} \end{pmatrix}. \tag{31}$$

Assuming that we have a subgame-perfect equilibrium for given T_{11} and T_{22}, we know that $F(\vartheta_1) > F(\vartheta_2)$, because by calculating the matrix $(\tilde{\mathbf{T}} - \mathbf{I})$ it can be seen immediately that $\sum_{j=1}^2(\tilde{\mathbf{T}}-\mathbf{I})_{1j} > \sum_{j=1}^2(\tilde{\mathbf{T}}-\mathbf{I})_{2j}$. It follows from straightforward calculations that $\frac{H_i}{T_{11}} > 0$ for $i = 1, 2$ and $\frac{H_i}{T_{22}} < 0$ for $i = 1, 2$. That means that we can find a subgame-perfect equilibrium that is a Pareto improvement of the initial equilibrium by making T_{11} as large as possible and T_{22} as small as possible, subject to $T_{ii} \leq 1 - \delta + \delta\pi_i$ and $T_{ij} \leq \delta\pi_j, i \neq j$. The solution of the constrained optimization process is provided in *i*).

To prove *ii*), define in the transition matrix $\mathbf{T}$ $T_i = 1 - \delta + \delta D_{\text{out}}(i)$. Thus,

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_1 - T_{11} \\ T_1 - T_{11} & T_2 - T_1 + T_{11} \end{pmatrix}. \tag{32}$$

Assuming that we have a subgame-perfect equilibrium for given T_{11}, we know again that $F(\vartheta_1) > F(\vartheta_2)$. It follows from straightforward calculations that $\frac{H_1}{T_{11}} > 0$, but $\frac{H_2}{T_{11}} < 0$. That means that the two types of trustors have conflicting interests. Using that trustors of type 2 cannot force trustors of type 1 to give up contacts in their own group for contacts with trustors of type 2, we obtain the equilibrium as given in the theorem.

Proof of Theorem 5 (page 60) We prove assertion *i*) of the theorem mainly with arguments from Theorem 4. First, we split the trustors in two groups, namely trustors of type 1 and all the others. For this division a matrix $\mathbf{T}$ can be defined, which has all the properties needed to apply Theorem 4. This means that for optimization we have to maximize the first column, which can be done by taking all the original elements (before the division of the trustors in two groups) separately optimal. Second, we keep the first column fixed and do the same by splitting the whole population of trustors in two groups: the trustors of type 1 and 2 and all the others. Again we have to maximize the first column, which means maximizing the elements of the first two columns in the original matrix. The first column was already fixed. Now the elements of second column have to be taken maximal.

The argument for assertion *ii*) is similar to that for part *i*). The difference is that the trustor with higher numbers depend on types of trustors with lower indices about how much contacts they will have with them.

Proof of Theorem 6 (page 61) Denote $\boldsymbol{\vartheta}^*(\mathbf{A}) = \boldsymbol{\vartheta}^*$ and $\boldsymbol{\vartheta}^*(\mathbf{A}_0) = \boldsymbol{\vartheta}_0^* = (\vartheta_0^*, \ldots, \vartheta_0^*)$ by symmetry of the game in $\mathbf{A}_0$. Furthermore, we define $F(\boldsymbol{\vartheta}^*) = F(\vartheta_0^*)\mathbf{1}$ and the scalar $\mu = (R_2 - P_2)f(\vartheta_0^*)$. Using the implicit function theorem, the first order approximation of (3) is

$$\begin{aligned}
\vartheta^* &= (R_2 - P_2)\Big((\mathbf{I} - w\mathbf{T})^{-1} - \mathbf{I}\Big)F(\boldsymbol{\vartheta}^*) \\
&= (R_2 - P_2)\Big(\tilde{\mathbf{T}}_0 - \mathbf{I} + w\tilde{\mathbf{T}}_0(\mathbf{T} - \mathbf{T}_0)\tilde{\mathbf{T}}_0\Big)\Big(F(\vartheta_0^*)\mathbf{1} + f(\vartheta_0^*)(\boldsymbol{\vartheta}^* - \boldsymbol{\vartheta}_0^*)\Big) + \epsilon \\
&= \vartheta_0^* + \mu(\tilde{\mathbf{T}}_0 - \mathbf{I})(\boldsymbol{\vartheta}^* - \boldsymbol{\vartheta}_0^*) + w(R_2 - P_2)F(\vartheta_0^*)\tilde{\mathbf{T}}_0(\mathbf{T} - \mathbf{T}_0)\tilde{\mathbf{T}}_0\mathbf{1} + \epsilon,
\end{aligned} \tag{33}$$

where $\epsilon = O(\mathbf{A} - \mathbf{A}_0)$ is 'small' if the difference between $\mathbf{A}$ and $\mathbf{A}_0$ is 'small'. By straightforward multiplication, it can be verified that if $x \neq \frac{1}{n}$

$$(\mathbf{I} - x\mathbf{J})^{-1} = \mathbf{I} + \frac{x}{1 - nx}\mathbf{J}, \tag{34}$$

Therefore,

$$\begin{aligned}
\tilde{\mathbf{T}}_0 &= (\mathbf{I} - w\mathbf{T}_0)^{-1} = \tfrac{1}{1-w\eta_1}\Big(\mathbf{I} - \tfrac{w\eta_2}{1-w\eta_1)}\mathbf{J}\Big)^{-1} \\
&= \tfrac{1}{1-w\eta_1}\Big(\mathbf{I} + \tfrac{w\eta_2}{1-w\eta_1-nw\eta_2)}\mathbf{J}\Big) \\
&= \tfrac{1}{1-\zeta_1}\big(\mathbf{I} + \tfrac{w\eta_2}{1-\zeta_2}\mathbf{J}\big).
\end{aligned} \tag{35}$$

where $\zeta_1 = w\eta_1$ and $\zeta_2 = w\eta_1 - nw\eta_2$. Note that $\zeta_1 + nw\eta_2 = \zeta_2$. Substituting (35) and $\big(\mathbf{I} + \frac{w\eta_2}{1-\zeta_2}\mathbf{J}\big)\,\mathbf{1} = \frac{1-\zeta_1}{1-\zeta_2}\mathbf{1}$ in (33) results in

$$\begin{aligned}
\boldsymbol{\vartheta}^* &= \boldsymbol{\vartheta}_0^* + \mu(\tilde{\mathbf{T}}_0 - \mathbf{I})(\boldsymbol{\vartheta}^* - \boldsymbol{\vartheta}_0^*) \\
&\quad + \tfrac{(R_2-P_2)wF(\vartheta_0^*)}{(1-\zeta_1)(1-\zeta_2)}\Big(\mathbf{I} + \tfrac{w\eta_2}{(1-\zeta_2)}\mathbf{J}\Big)(\mathbf{T} - \mathbf{T}_0)\mathbf{1} + \epsilon.
\end{aligned} \tag{36}$$

Using that $\boldsymbol{\vartheta}_0^* = \mu(\tilde{\mathbf{T}}_0 - \mathbf{I})\boldsymbol{\vartheta}_0^*$ and under the assumption that the inverse of $\mathbf{I} - \mu(\tilde{\mathbf{T}}_0 - \mathbf{I})$ exists, it holds that

$$\boldsymbol{\vartheta}^* = \tfrac{(R_2-P_2)wF(\vartheta_0^*)}{(1-\zeta_1)(1-\zeta_2)}\big(\mathbf{I} - \mu(\tilde{\mathbf{T}}_0 - \mathbf{I})\big)^{-1}\Big(\mathbf{I} + \tfrac{w\eta_2}{(1-\zeta_2)}\mathbf{J}\Big)(\mathbf{T} - \mathbf{T}_0)\mathbf{1} + \epsilon. \tag{37}$$

Using (34) and (35) and more 'tedious but straightforward calculation', it follows that

$$\begin{aligned}
\left(\mathbf{I}-\mu(\tilde{\mathbf{T}}_0-\mathbf{I})\right)^{-1} &= \left(\mathbf{I}-\mu\left(\tfrac{1}{1-\zeta_1}\left(\mathbf{I}+\tfrac{w\eta_2}{1-\zeta_2}\mathbf{J}\right)-\mathbf{I}\right)\right)^{-1} \\
&= \tfrac{1-\zeta_1}{1-\zeta_1(1+\mu)}\left(\mathbf{I}-\tfrac{\mu w\eta_2}{(1-\zeta_1(1+\mu))(1-\zeta_2)}\mathbf{J}\right)^{-1} \\
&= \tfrac{1-\zeta_1}{1-\zeta_1(1+\mu)}\left(\mathbf{I}+\tfrac{\mu w\eta_2}{(1-\zeta_1))(1-\zeta_2(1+\mu))}\mathbf{J}\right).
\end{aligned} \tag{38}$$

Because this inverse matrix exists, the existence of the $\left(\frac{G}{\vartheta*}\right)^{-1}$ is guaranteed, which ensures that we indeed could use the implicit function theorem. Furthermore, it follows from the definitions of $\mathbf{T}$ and $\mathbf{T}_0$ that

$$\begin{aligned}
(\mathbf{T}-\mathbf{T}_0)\mathbf{1} &= \left(\left((1-\delta)\mathbf{I}+\tfrac{\delta}{n}\mathbf{A}\right)-\left((1-\delta)\mathbf{I}+\tfrac{\delta}{n}\mathbf{A}_0\right)\right)\mathbf{1} \\
&= \tfrac{\delta}{n}(\mathbf{A}-\mathbf{A}_0)\mathbf{1} \\
&= \delta\Big(\mathbf{D}_{\mathrm{out}}(\mathbf{A})-\mathbf{D}_{\mathrm{out}}(\mathbf{A}_0)\Big).
\end{aligned} \tag{39}$$

By substituting (38) and (39) in (37), the expression for ρ_1 as defined in the theorem follows directly:

$$\rho_1 = \frac{\delta w(R_2-P_2)F(\vartheta_0^*)}{n(1-\zeta_1)(1-\zeta_2(1+\mu))}. \tag{40}$$

Using the same equations and some more calculus leads to

$$\rho_2 = \frac{(1+\mu)w\eta_2}{1-\zeta_2(1+\mu)}. \tag{41}$$

The expression in terms of the outdegrees and density follow immediately from (39) and the fact that $n\Delta = \sum_{i=1}^{n} D_{\mathrm{out}}(i)$.

According to the definition of the transition matrix, $0 < \zeta_1 < \zeta_2 < 1$. Because all other parts of ρ_1 and ρ_2 are clearly positive, the signs of ρ_1 and ρ_2 depend on the magnitude of $f(\vartheta_0^*)(R_2-P_2)$. The assumption that the distribution F is concave and (6) ensures that

$$\mu = f(\vartheta_0^*)(R_2-P_2) < \frac{F(\vartheta_0^*)(R_2-P_2)}{\vartheta_0^*} = \frac{1-\zeta_2}{\zeta_2}. \tag{42}$$

This implies that $0 < \zeta_2(1+\mu) < 1$, which is sufficient to assure that ρ_1 and ρ_2 are positive.

Bibliography

Akkermans, S. T. M./J. H. van Lint (1970), *Algebra en Analyse*, Groningen

Axelrod, R. (1984), *The Evolution of Cooperation*, New York

— (1986), An Evolutionary Approach to Norms, in: *American Political Science Review 80*, 1095–1111

— (1997), *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton

Bellman, R. E. (1954), Some Problems in the Theory of Dynamic Programming, in: *Econometrica 22*, 37–48

Berman, A./R. J. Plemmons (1979), *Nonnegative Matrices in the Mathematical Sciences*, New York

Binmore, K. (1998), *Game Theory and the Social Contract. Volume 2: Just Playing*, Cambridge-MA

Blumberg, B. F. (1997), *Das Management von Technologiekooperationen: Partnersuche und Verhandlungen mit dem Partner aus empirisch-theoretischer Perspektive*, Amsterdam

Buskens, V. (1998), The Social Structure of Trust, in: *Social Networks 20*, 265–289

— (1999), *Social Networks and Trust*, Amsterdam

— /K. Yamaguchi (1999), A New Model for Information Diffusion in Heterogeneous Social Networks, in: M. Becker/M. Sobel (eds.), *Sociological Methodology*, Oxford, 281–325

Coleman, J. S. (1990), *Foundations of Social Theory*, Cambridge-MA

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

Friedman, J. W. (1971), A Non-Cooperative Equilibrium for Supergames, in: *Review of Economic Studies 38*, 1–12

Fudenberg, D./J. Tirole (1991), *Game Theory*, Cambridge-MA

Granovetter, M. S. (1985), Economic Action and Social Structure: The Problem of Embeddedness, in: *American Journal of Sociology 91*, 481–510

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

Harsanyi, J. C./R. Selten (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge-MA

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems: An Introductory Analysis w with Applications to Biology, Control, and Artificial Intelligence*, Ann Arbor

— (1998), *Emergence: From Choas to Order*, New York

Klein, D. B. (Ed.) (1997), *Reputation: Studies in the Voluntary Elicitation of Good Conduct*, Ann Arbor

Kreps, D. M. (1990a), Corporate Culture and Economic Theory, in: J. Alt/K. Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge, 90–143

— (1990b), *A Course in Microeconomic Theory*, New York

— /R. Wilson (1982), Reputation and Imperfect Information, in: *Journal of Economic Theory 27*, 253–279

Kumbasar, E./A. K. Romney/W. H. Batchelder (1994), Systematic Biases in Social Perception, in: *American Journal of Sociology 100*, 477–505

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

Macaulay, S. (1963), Non-Contractual Relations in Business: A Preliminary Study, in: *American Sociological Review 28*, 55–67

Milgrom, P./J. Roberts (1992), *Economics, Organization and Management*, Englewood Cliffs

Raub, W. (1993), Symbiotic Arrangements: A Sociological Perspective, in: *Journal of Institutional and Theoretical Economics 149*, 716–724

— /J. Weesie (1990), Reputation and Efficiency in Social Interactions: An Example of Network Effects, in: *American Journal of Sociology 96*, 626–654

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

Snijders, C. (1996), *Trust and Commitments*, Amsterdam

Vanberg, V./R. D. Congleton (1992), Rationality, Morality, and Exit, in: *American Political Science Review 86*, 418–431

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

— /V. Buskens/W. Raub (1998), The Management of Trust Relations via Institutional and Structural Embeddedness, in: P. Doreian/T. Fararo (eds.), *The Problem of Solidarity: Theories and Models*, Amsterdam, 113–138

Wilson, R. (1985), Reputations in Games and Markets, in A. E. Roth (ed.), *Reputations in Games and Markets*, Cambridge, 267–300