Analyse & Kritik 01/2014 (© Lucius & Lucius, Stuttgart) S. 135–152

Werner Raub, Vincenz Frey and Vincent Buskens

Strategic Network Formation, Games on Networks, and Trust*

Abstract: This paper brings two major research lines in current sociology together. Research on social networks has long focused primarily on network effects but meanwhile also addresses the emergence and dynamics of networks. Research on trust in social and economic relations shows that networks have effects on trust. Using game theory, we provide a simple model that allows for an integrated and simultaneous analysis of network effects on trust and for the endogenous emergence of the network. The model also allows for characterizing the value of the network. We use standard assumptions on full strategic rationality. Testable implications of the model as well as model extensions are sketched.

1. Introduction

Theoretical and empirical research has established that social networks have important effects for micro-level individual behavior as well as macro-level social phenomena. This includes—but is not limited to—individual search behavior on the labor market and labor market outcomes (Granovetter 1973; 1974), individual adoption and macro-level diffusion of innovations (Coleman et al. 1966), the spread of diseases (Morris et al. 1995), social inequality (Lin 2001), and also trust in social and economic exchange (Coleman 1990). Game theory is a useful tool for modeling network effects. Research on *games on networks* (Goyal 2007, chap. 3; Jackson 2008, chap. 9) assumes a network and its characteristics as given and exogenous and analyzes network effects on individual behavior and the macro-consequences of such behavior. Raub and Weesie (1990) is likely to be the first game-theoretic model of network effects for a social dilemma, namely,

* Comments of and discussions with Rense Corten and other colleagues of our Utrecht group *Cooperation in Social and Economic Relations* as well as suggestions of an anonymous reviewer are gratefully acknowledged. The paper is part of the project *Trust in Dynamic Networks*, funded by the Netherlands Organization for Scientific Research (NWO, Graduate Program Grant 2008/2009 for the ICS). Additional support for Raub was provided by NWO (PIONIER-program *The Management of Matches*; grants S 96-168 and PGS 50-370) and for Buskens by Utrecht University (High Potential-program *Dynamics of Cooperation, Networks, and Institutions*). Raub acknowledges the hospitality of Nuffield College, University of Oxford. The paper is part of a series of related studies (Frey et al. 2013; Raub et al. 2012 and 2013) and some overlap cannot be avoided. See *section 4* for details on how the studies are related.

the Prisoner's Dilemma. Buskens (2002) provides models of network effects for trust problems. Buskens and Raub (2013) survey the theoretical and empirical literature on conditions under which networks facilitate cooperation in social dilemmas.

Often, actors can affect their position in a network and the network structure, at least to some degree, by forming, maintaining, or severing relations with others. For example, actors can often choose with whom to exchange goods or information and with whom to collaborate. Thus, networks and their characteristics will often not be exogenous. If actors have opportunities to choose their relations and if networks have important consequences for actors, there is an incentive for 'networking'. Namely, actors will try to form relationships with an eye on optimizing their individual benefits from the network: they will tend to strategically invest in forming and maintaining relations that are beneficial and end relations that are not (see, e.g., Flap 2004). Thus, the emergence and dynamics of networks becomes an issue.

Much of the literature on social networks has focused on effects of social networks. Systematic research on the emergence and dynamics of networks is more recent and has to cope with the problem that the emergence and dynamics of networks is inherently—and even more so than network effects—due to interdependent behavior of actors, thus complicating theoretical and empirical analysis (Flap/Völker 2013; Snijders 2013). However, since the mid-1990's, progress has been made in studying the emergence and dynamics of networks, too. The game-theoretic literature on *strategic network formation* has rapidly developed in economics (see Dutta/Jackson 2003 for a collection of important and pioneering work). In strategic network formation models, links between actors are endogenous. These models employ assumptions about the benefits that actors derive from being in specific network positions and allow for an analysis of how incentive-guided and goal-directed actors form, maintain, or sever links with others and for an analysis of the macro-level properties of the emerging networks. Meanwhile, a broad literature on games on networks as well as strategic network formation has become available (see Jackson/Zenou 2013 for a comprehensive collection of contributions and textbooks such as Goyal 2007; Vega-Redondo 2007; Jackson 2008). Also, research in these fields has become very interdisciplinary with contributions from economics and sociology but also from mathematics, physics, and biology (see, e.g., Newman et al. 2006 for a collection of important work and Newman 2010 for a textbook).

If networks have important effects for actors and if actors can influence these effects by forming, maintaining, or severing links with others, an obvious next step is the *simultaneous analysis of both phenomena—network formation and network effects—in an integrated model.* One then aims at a model that allows for deriving implications for actors' strategic network formation *as well as* implications for network effects on behavior. Flap (2004) has outlined a research program for an integrated analysis of both network formation and network effects, using the notion of "investments in and returns on social capital". Models that actually implement such a program are referred to as models for the *co-evolution of networks and behavior.* Such models are still very scarce. Examples

in sociology and economics include work on cooperation in dynamic networks by Eguiluz et al. (2005), Pujol et al. (2005), and Vega-Redondo (2006; see Corten 2009, chap. 5.1 for a survey).

In this paper we provide a simple model of network effects on behavior in trust problems and of investments of the actors involved in those trust problems in forming the network. We model trust problems using the standard Trust Game, a paradigmatic example of a social dilemma as well as a stylized representation of typical problems that emerge in social and economic exchange (see Buskens/Raub 2013 for review). Our model allows for specifying conditions such that the network fosters trustfulness and trustworthiness and such that actors invest in forming the network in the first place. Also, the model allows to specify the value of the network for the actors. We assume full game-theoretic rationality with respect to actors' behavior, both in the Trust Game and with respect to investments in forming the network. The assumption of full strategic rationality is uncommon in models of strategic network formation. To keep the analysis tractable, models usually assume 'myopic best-reply behavior'. This means that actors, when deciding about forming, maintaining, and severing links, do not take into account the implications of their decisions for future behavior of other actors and the repercussions of that future behavior for themselves. Full strategic rationality requires that such long-term effects are taken into account. In our model, due to its simplicity, the analysis is tractable under the assumption of full strategic rationality. We thus likewise contribute to more theoretical pluralism in research on strategic network formation.

We first introduce our game-theoretic model. Subsequently, we derive implications for network effects and for network formation. We conclude with a summary of our main results, a sketch of testable implications for experimental research, and extensions of the model.

2. The Model[1]

2.1 Embedded Trust Games

In our model, Trust Games are embedded in a game Γ. We wish to consider the simplest conceivable setting for network effects and network formation related to Trust Games and thus assume a triad with three actors, one trustee and two trustors A and B. Each trustor is involved in standard Trust Games with the trustee. The Trust Game (e.g., Dasgupta 1988; Kreps 1990; see also Coleman 1990, chap. 5) is depicted in *figure 1*.

In the Trust Game, actor 1 is a trustor and actor 2 is the trustee. The game starts with a move of the trustor. She[2] can choose between placing or not placing trust. If trust is not placed, the interaction ends and the trustor receives payoff P_i ($i = A, B$), while the trustee receives payoff P. If trust is placed, the trustee

[1] We focus on core features of the game-theoretic model. See a textbook such as Rasmusen 1994 for game-theoretic concepts and assumptions that are employed in our analysis.

[2] To facilitate readability, we use female pronouns for the trustors and male pronouns for the trustee.

chooses between honoring and abusing trust. If he honors trust, the trustor's payoff is $R_i > P_i$ and the trustee receives payoff $R > P$. If trust is abused, the payoff for the trustor is $S_i < P_i$, while the trustee receives $T > R$.

Assume that the Trust Game is played as an 'isolated encounter' in the sense that an actor's behavior cannot have repercussions for future interactions (the actors play a 'one-shot game'). Under standard game-theoretic assumptions, the payoffs in the game represent utilities for the actors, there is common knowledge meaning that all actors know that all actors know all elements of the game, and the actors are rational in the sense that they are able and willing to optimize their own utility. Under these assumptions, if the trustor would place trust, the best reply for the trustee would be to abuse trust, since $T > R$. Because the trustor can anticipate this, she is better off not placing trust than placing trust, since $S_i < P_i$. Hence, not placing trust, while placed trust would be abused is the unique subgame perfect equilibrium of the game.[3] Both actors would be better off if trust would be placed and honored, since $R_i > P_i$ and $R > P$. The Trust Game thus models a social dilemma in the sense that individually rational equilibrium behavior is associated with a Pareto-suboptimal outcome.

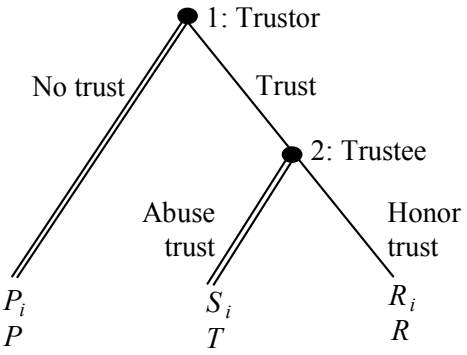

Figure 1: The Trust Game ($S_i < P_i < R_i, P < R < T$); double lines indicate behavior in the unique subgame perfect equilibrium

Standard Trust Games are embedded in the repeated game Γ as follows. The repeated game is played by two trustors and one trustee in rounds $t = 1, 2, \ldots$. Γ likewise has an initial round 0 but we first sketch the structure of Γ for rounds 1, 2,.... In each of those rounds, the trustee plays first of all a Trust Game with one of the trustors and subsequently a Trust Game with the other trustor. It does not matter whether the two trustors play their games in the same sequence in each round. To keep things simple, we assume that each actor's payoff function is the same in each Trust Game. We do not need to assume that the Trust

[3] Subgame perfection is the basic refinement of the Nash equilibrium concept and a common conceptualization of individually rational behavior in situations with strategic interdependence. Throughout, we consider subgame perfect equilibria and refer to them for brevity as 'equilibria'.

Games are symmetric in the sense that $R_i = R$ and $P_i = P$. We also do not need to assume that the two trustors have the same payoff function for their Trust Games beyond $R_i > P_i > S_i$ for each of the trustors.

After each round $t = 1, 2, \ldots$ of Γ, the next round $t + 1$ is played with a constant probability w $(0 < w < 1)$, while Γ stops after each round with probability $1 - w$. Thus, actors play indefinitely often repeated Trust Games and we can apply standard theory for indefinitely often repeated games (e.g., Friedman 1986). Note that our assumptions imply that both trustors play the same number of Trust Games with the trustee.

By way of example, rounds $1, 2, \ldots$ of Γ could be interpreted as business periods (for example, market days) in which two buyers in the role of the trustor purchase goods from a seller in the role of the trustee under conditions of asymmetric information. More precisely, in each period, each buyer decides whether or not to purchase a good from the seller and, if the buyers decide to buy, the seller decides on selling good quality or selling bad quality for the price of good quality. With probability $1 - w$, each period is the final one in the sense that the seller stops business due to some exogenous contingency beyond his control (for example, a new competitor enters the market who offers a superior product). In Axelrod's (1984) formulation, w represents the "shadow of the future": with increasing w, actors' long-term incentives increase.

2.2 Investments in Network Formation

Rounds $1, 2, \ldots$ of Γ are preceded by round 0. In that round, at costs, actors can form links between them that allow for information exchange. Thus, in round 0 actors can invest in network formation. An example would be that market participants set up a consumer organization or a buyer association keeping track of transactions. Links between buyers in the sense of information exchange about the behavior of the seller are then due to the distribution of information on the behavior of market participants by the organization. After round 0 has been played and given that the actors have formed an information network, taking into account what the consequences will be for the subsequent Trust Games, actors play a game on a network and we can study the effects of the network for behavior in those Trust Games.

Without investments in round 0, each trustor is only informed on the history of her own games. More precisely, consider the Trust Game of trustor i in round $t \geq 1$ of Γ. Trustor i is then informed that no investments have been made in round 0 and is also informed about what happened in all Trust Games in which i has been involved in rounds $1, \ldots, t-1$. However, i has no information about what happened in any earlier Trust Game of the trustee with the other trustor j $(j = A, B;\ i \neq j)$.

With investments in round 0, trustors are not only informed about the history of their own Trust Games but also about the history of the Trust Games of the other trustor. More precisely, consider again the Trust Game of trustor i in round $t \geq 1$ of Γ. Trustor i is then informed that investments have been made in round 0. Furthermore, i is informed about what happened in all Trust Games

in which i has been involved in rounds 1,..., $t-1$ as well as what happened in all earlier Trust Games of trustor j in rounds 1,..., $t-1$ and also about the Trust Game of j in round t if j has played the first Trust Game in round t.

For the trustee we assume that he is informed in round $t \geq 1$ about the history of his Trust Games with both trustors as well as about what has happened in round 0. Thus, the network would provide extra information exclusively for the trustors but would not provide extra information for the trustee.

For analytical tractability, we assume that all information—if available—is correct. We thus exclude 'noise' in the sense that, for example, an actor erroneously believes that investments have been made in round 0 or erroneously believes that trust has been abused in an earlier Trust Game.

The *total costs of forming the information network* in round 0 are assumed to be $\tau > 0$. Different scenarios for sharing these costs can be conceived as different kinds of institutions: they are rules of the game in which the actors are involved (North 1990, chap. 1). Seen from this perspective, we model effects of institutions for network formation as well as how institutions and networks interact in affecting cooperation in social dilemmas. The conditions derived below will illustrate that it might depend on the parameters of the game and the costs of investing in information exchange, whether actors are willing to invest in network formation under one or the other institution. Thus, our model allows for inferences on institutional design.

We consider two simple institutional rules for cost sharing. Under the *trustor-institution*, only the trustors can invest. Thus, in round 0, the trustors decide simultaneously and independently about their individual investment. Each trustor can either invest $\frac{\tau}{2}$ or decide not to invest. If each actor invests $\frac{\tau}{2}$, the information network is formed. Conversely, if at least one trustor decides not to invest, the information network is not formed. In that case, a trustor who had been willing to invest does not lose her own investment. This corresponds to common assumptions in models of network formation, namely, two-sided link formation (a link is only formed if both actors wish to be linked) with shared costs of links (Jackson 2008, chap. 6). Our second and alternative institutional rule, the *trustee-institution*, stipulates that only the trustee can form the information network by investing τ.

2.3 Further Assumptions on Γ

We assume that round 1 is always played after round 0. Consequently, we assume that each actor's (expected) payoff for Γ equals the sum of the realized costs in round 0 and the exponentially discounted payoffs in rounds 1, 2,.... This implies, for example, that trustor i's payoff under the trustor-institution is

$$U_i = -\frac{\tau}{2} + R_i + wR_i + w^2R_i + ... + w^{t-1}R_i + ... = -\frac{\tau}{2} + \frac{R_i}{1-w}$$

if both trustors have contributed $\frac{\tau}{2}$ to the information network and if trust is placed and honored in all of i's Trust Games in rounds 1, 2,.... Similarly, the

trustee's payoff would be $U = \frac{2R}{1-w}$ if trust would always be placed and honored also in all Trust Games with the other trustor.

Under the trustee institution,

$$U = -\tau + 2P + w2P + w^2 2P + ... + w^{t-1}2P + ... = -\tau + \frac{2P}{1-w}$$

would be the trustee's payoff after having invested τ in round 0, while trust is never placed by any of the trustors in rounds 1, 2,.... In this case, $U_i = \frac{P_i}{1-w}$ would be the payoff for each trustor.[4]

We further assume that all actors are informed on the structure of Γ, that they know from each other that they have this information, etc. The structure of Γ is thus 'common knowledge' (Rasmusen 1994, 44). Finally, we assume that Γ is played as a noncooperative game in the sense that actors cannot incur binding agreements or binding unilateral commitments that are not explicitly modeled as moves in the structure of the game.[5]

3. Analysis of the Model

We assume rational behavior of actors and are thus interested in specifying conditions for subgame perfect equilibria of Γ such that trust will be placed and honored throughout the Trust Games in all rounds 1, 2,.... We refer to such an equilibrium as a *trust equilibrium*. This approach is based on the commonly used assumption that such an equilibrium can be considered as the 'solution' because each actor maximizes his or her own payoff, given the equilibrium strategies of all other actors, *and* because a trust equilibrium is associated with higher payoffs for each actor than the situation where trust is not placed (see Buskens/Raub 2013). Given our focus on network formation, we are specifically interested in equilibria such that trust is placed and honored after formation of the network in round 0, whereas no trust is placed if the network is not formed. To specify conditions for such equilibria, we first consider the subgame Γ^- that is played after the network has not been formed in round 0. Subsequently, we analyze the subgame Γ^+ that is played after the network has been formed. Finally, we analyze conditions such that investments in the information network in round 0 are implied by equilibrium behavior.

3.1 Trust Based on Conditional Strategies

In Γ^- as well as Γ^+, placing and honoring trust can be the result of equilibrium behavior based on *conditional strategies*. When playing a Trust Game with the

[4] Note that we interpret payoffs as cardinal utilities. Note, too, that the model includes discounting of future payoffs due to the probability that the game might end and that we neglect negative time preferences (see Rasmusen 1994, 108–110). It would be no problem to include negative time preferences and results would remain robust.

[5] We assume a noncooperative game precisely because we wish to specify conditions such that rational actors will place and honor trust 'endogenously' and without external enforcement, based exclusively on the embeddedness of the Trust Game in a sequence of interactions and the information exchange network between the actors.

trustee in round t, trustor i can condition her behavior on information about the trustee's behavior in previous Trust Games. If i has information that the trustee honored trust in earlier games, i can reward this by placing trust. Conversely, if i has information that the trustee defected in earlier games, i can punish this by not placing trust. Thus, with $T > R$, the trustee has a short-term incentive to abuse trust in each Trust Game. However, he also has to take into account that abusing trust now may imply long-term costs in future games since no trust may be placed in at least some of those future games so that the trustee would then obtain only $P < R$ in those future games. Anticipating conditional strategies of the trustors, the trustee thus has to balance short-term incentives (T–R) and long-term incentives (R–P). Repeated Trust Games and, moreover, Trust Games embedded in a network through which trustors exchange information can thus have trust equilibria based on conditional strategies. Placing and honoring trust can then be a result of individually rational behavior of actors who take long-term effects of their present behavior into account.

What is the conditional strategy for trustors that is associated with the most attractive reward for the trustee if he honors trust and the most severe punishment for abusing trust? Such a strategy is commonly known as a 'trigger strategy' (e.g., Friedman 1986). This is the strategy such that a trustor places trust if she has no information that the trustee ever abused trust and such that she never places trust again in all future games with the trustee as soon as she receives information that the trustee has abused trust.[6] Obviously, if both trustors use trigger strategies and the trustee always honors trust, trust will be placed and honored in all Trust Games in all rounds of Γ. If long-term incentives can at all be large enough to induce a trustee to honor trust, these incentives are certainly large enough if the trustors use trigger strategies. Conversely, if the trustors anticipate that the trustee will honor trust due to his long-term incentives, then placing trust becomes attractive for the trustors, since $R_i > P_i$. It follows that an indefinitely often repeated Trust Game has a trust equilibrium if and only if there is an equilibrium that comprises trigger strategies of the trustors and honoring trust by the trustee (see Friedman 1986, chap. 3 for details). We thus derive conditions for such equilibria in Γ^- and Γ^+.

Note that the assumption underlying this approach need not be that trustors do indeed use trigger strategies. Trust equilibria do require the use of conditional strategies by the trustors but these conditional strategies may comprise less severe punishments than implied by a trigger strategy. For example, a trustor may be willing to return to placing trust if the 'punishment period' for the trustee has been 'long enough'. Nevertheless, the existence of a trigger strategy equilibrium is a necessary condition for the existence of trust equilibria based on conditional strategies that involve less extreme sanctioning and thus smaller long-term incentives for honoring trust. Hence, following a common approach in

[6] A trigger strategy requires furthermore that a trustor will place trust in her first Trust Game if she has by then no information about behavior in games of the trustee with the other trustor and that she never places trust again as soon as she or the other trustor has not placed trust in an earlier Trust Game (see the Appendix on why trigger strategies are defined in this specific way).

empirical applications, we assume that placing and honoring trust becomes more likely when the conditions for a trigger strategy equilibrium become less restrictive (see Buskens/Raub 2013 for a more detailed discussion of this approach in empirical applications).

3.2 Network Effects: Analysis of Γ^-

In a subgame Γ^- that is played if the information network is not formed, each trustor is only informed on her own previous games with the trustee but has no information on what happens in the trustee's Trust Games with the other trustor. Our first proposition provides the condition for a trust equilibrium in this subgame.

Proposition 1—Trust without a network: Γ^- has a trust equilibrium if and only if,

$$w \geq TEMP^- := \frac{T-R}{T-P}.$$

Proofs for our propositions are sketched in the Appendix.

Note that $TEMP^-$ is a measure for the trustee's incentives to abuse trust. It is easily seen that $0 < TEMP^- < 1$. Proposition 1 reflects the well-known result (e.g., Kreps 1990) that placing and honoring trust in an indefinitely often repeated Trust Game constitutes equilibrium behavior if and only if the incentives to abuse trust are compensated by a sufficiently large probability w that the game continues. The proposition implies that placing and honoring trust is facilitated and becomes more likely—in the sense that the condition for a trust equilibrium becomes less restrictive—if the short-term incentive $T - R$ to abuse trust decreases, if the trustee's costs $T - P$ of no longer being trusted by a trustor increase, and if the continuation probability w increases. Note that the proposition implies that the existence of a trust equilibrium depends exclusively on the incentives of the trustee and not at all on the trustors' incentives. We will return to this issue in our concluding discussion.

3.3 Network Effects: Analysis of Γ^+

Subgame Γ^+ is played after investments in forming the information network have been pledged and the network has been formed in round 0. Therefore, in Γ^+, each trustor is not only informed on the history of her own interactions with the trustee but also on the history of the other trustor's previous games. Proposition 2 provides the condition for a trust equilibrium if the information network is available.

Proposition 2—Trust in a network: Γ^+ has a trust equilibrium if and only if

$$w \geq TEMP^+ := \frac{T-R}{(T-P)+(R-P)}.$$

In Proposition 2, $TEMP^+$ is a measure for the trustee's incentives to abuse trust, with $0 < TEMP^+ < 1$. Thus, Γ^+ has a trust equilibrium for a large

enough continuation probability w and for a small enough incentive $TEMP^+$ to abuse trust. The comparative static results for the condition in Proposition 1 also hold for Proposition 2. Also, $TEMP^+ < TEMP^-$, i.e., the trustee's incentive to abuse trust is smaller in Γ^+ than in Γ^-, since the trustee's abuse of trust in a Trust Game with trustor i can now be punished in future games not only by i herself but also by the other trustor. Therefore, the condition for a trust equilibrium is less restrictive in Γ^+ than in Γ^-. This is in line with earlier research on network effects in Trust Games and related social dilemmas (e.g., Buskens 2002; Raub/Weesie 1990).

3.4 Network Effects: The Value of the Network

Since the condition for a trust equilibrium is less restrictive in Γ^+, it follows that there are parameter configurations such that Γ^+ has a trust equilibrium while Γ^- has no such equilibrium. This is the case if and only if $TEMP^+ \leq w < TEMP^-$. Under this condition, the network is valuable for the actors. Moreover, we are now able to specify an upper bound[7] for the *value of the network*. To do so, we compare payoffs in a subgame Γ^- without a network of information links and in the subgame Γ^+ with such a network under the assumption that a trust equilibrium, if it exists, will be played in the subgame Γ^+, while trust will never be placed in the subgame Γ^- if Γ^- has no trust equilibrium.

Proposition 3—Value of the network: Assume that $TEMP^+ \leq w < TEMP^-$ so that a trust equilibrium exists only in Γ^+. The upper bound on the value of social capital is then equal to $\frac{2(R-P)}{1-w}$ for the trustee and $\frac{R_i-P_i}{1-w}$ for each trustor i.

The upper bound on the value of the network thus increases if the costs $R - P$ and, respectively, $R_i - P_i$ of not placing trust compared to honored trust increase and if the continuation probability w increases, as long as the costs $R - P$ and $R_i - P_i$ and the continuation probability w are small enough so that Γ^- has no trust equilibrium. Obviously, $\frac{2(R-P)}{1-w}$ and $\frac{R_i-P_i}{1-w}$ can also be interpreted as upper bounds on the costs of investments in network formation that a rational trustee and rational trustors would be willing to incur.

3.5 Network Formation

Having analyzed the effects of the network, we now turn to network formation. We do so by specifying conditions for equilibria of Γ that imply investments in network formation in round 0 and placing and honoring trust in all Trust Games in subsequent rounds 1, 2,....

[7] See the Appendix for why we can only specify an upper bound on rather than the exact value of the network.

Proposition 4—Network formation: Γ has equilibria such that the information network is formed in round 0 and trust is placed and honored in all Trust Games in all rounds 1, 2,... if $TEMP^+ \leq w < TEMP^-$ and
(4.1) $\frac{\tau}{2} \leq \frac{R_i - P_i}{1-w}$ for each trustor i under the trustor-institution;
(4.2) $\tau \leq \frac{2(R-P)}{1-w}$ under the trustee-institution.

Given $w < TEMP^-$, there is no trust equilibrium in Γ^-, while not placing trust in all Trust Games in all rounds 1, 2,... of is equilibrium behavior. In this equilibrium of Γ^-, the trustors realize a payoff of $\frac{P_i}{1-w}$, while the trustee's payoff is $\frac{2P}{1-w}$. Conversely, given $TEMP^+ \leq w$, there is a trust equilibrium in Γ^+. Thus, after investments in network formation, the actors' payoffs associated with the trust equilibrium in Γ^+ are $\frac{R_i}{1-w}$ for each trustor and $\frac{2R}{1-w}$ for the trustee. The costs of investments in social capital are thus small enough under conditions (4.1) and (4.2).

Our proposition on network formation shows that *rational actors will invest in forming the information network if their trust problems are neither too small nor too large.* Small trust problems in the sense of $w \geq TEMP^-$ can be solved without investments in network formation (see Proposition 1). Large trust problems in the sense of $TEMP^+ > w$ cannot be solved even when relying on network effects (see Proposition 2). Trust problems in the interval $TEMP^+ \leq w < TEMP^-$ can only be solved through network formation and rational actors incur the required investments if the costs are small enough and do not exceed the thresholds specified in conditions (4.1) and, respectively, (4.2). With respect to network formation, this is the core implication of our model.

It is useful to note that it depends on the parameters of the game whether an institution for cost sharing contributes to inducing network formation. The trustor-institution has the advantage that the costs of network formation can be divided among two trustors, while the trustee institution requires that the trustee covers the full costs. However, the benefits must be large enough for each trustor to be willing to pay these costs. If the returns on investments in network formation are divided unevenly, it might be that one trustor is willing to bear considerable costs of network formation but the other trustor is not.

4. Discussion

Our game-theoretic model provides conditions such that a network of information exchange relations induces placing and honoring trust. Simultaneously, we have endogenized the network and have specified conditions such that actors incur costs to form the network in the first place. Thus, we have integrated an analysis of games on networks that focuses on network effects with an analysis of strategic network formation that focuses on how networks emerge as a result of rational behavior. Moreover, we have specified the value of the network.

Our proposition on network formation shows that incentives for investments in the network are not restricted to the trustors who would suffer from opportunistic behavior of the trustee. Rather, the trustee likewise has incentives

to invest in network formation among trustors, although the trustee is himself not at all vulnerable to opportunistic behavior of trustors. In the Trust Game, the trustor's equilibrium strategy not to place trust provides protection against abuse of trust but the trustor cannot increase own payoffs through exploiting the trustee. Under the trustee-institution, *only* the trustee can invest in network formation and a rational trustee will do so when condition (4.2) is fulfilled. The trustee then sets up an information network for the trustors that enables them to update each other on the behavior of the trustee. Thus, the trustee 'binds himself'. Through setting up the information network for the trustors, it becomes less attractive for the trustee to abuse trust because the future punishment of abusing trust increases. This, however, can induce trustors to place trust in the first place, thus allowing gains for both trustors and trustees—trust is placed and honored—compared to the payoffs they obtain when no trust is placed. Investments of the trustee in network formation can thus be seen as a 'commitment' that allows for placing and honoring trust (Raub 2004).

Consider empirically testable implications of our model for lab experiments that allow for carefully implementing our model assumptions and for manipulating model parameters in experimental conditions (see Buskens/Raub 2013 for a survey of the meanwhile extensive literature on experimental tests of predictions from repeated game models, including games that involve networks of actors). Lab experiments require assumptions on how monetary incentives for subjects are related to the parameters of our model. We thus need assumptions on the subjects' utility functions or have to establish empirical evidence on relevant properties of their utility functions. Furthermore, we need to assume, as discussed in *section 3*, that the equilibria on which our propositions focus are 'solutions' in the sense that rational actors tend to implement these equilibria. More specifically, we need the common assumption that the likelihood of a certain equilibrium behavior increases when the conditions for the equilibrium become less restrictive. Our predictions then follow from our analysis of the conditions in our propositions. The predictions are on effects of changes in the parameters of the Trust Game, the continuation probability w, and the total costs τ of investments in network formation. Predictions refer to effects on investments in network formation, on behavior in the Trust Games, and on the relation between investment behavior and behavior in the Trust Games.

We can distinguish three scenarios. The first scenario covers $w < TEMP^+ < TEMP^-$ so that *trust problems are large*. We predict a small likelihood of investments in network formation, while the likelihood of investments is only weakly associated with the costs τ. We also predict a small likelihood of placing trust and a small likelihood of honoring trust when trust has been placed. Also, the effects of investments in network formation on behavior in the Trust Games would be expected to be small.

In the second scenario, *trust problems are of intermediate size* such that placing and honoring trust presupposes that the information network is formed, i.e., $TEMP^+ < w < TEMP^-$. We predict sizeable effects of τ on investment behavior as well as sizeable effects of investments on subsequent behavior in Trust Games. More specifically, we predict that the likelihood of investments

decreases with increasing costs τ. Likewise, we predict that investments in network formation have a positive effect on the likelihood of placing and honoring trust.

The third scenario has *small trust problems*, i.e., $TEMP^+ < TEMP^- < w$. Just like for the first scenario, we predict a small likelihood of investments in network formation, with a weak association between the costs τ and the likelihood of investments and small effects of investments on behavior in Trust Games. Other than for scenario 1, we now of course predict a large likelihood of placing and honoring trust.

Our model can be extended in various directions. In this paper, we have considered the Trust Game and the simplest case of a network for the Trust Game, namely the case of three actors, with one trustee and two trustors. In a companion paper (Raub et al. 2012), we analyze the more general case with one trustee and $n \geq 2$ trustors. Such a model allows for deriving additional implications on size effects, i.e., effects of an increasing number of trustors. A further extension would be to consider network effects and network formation for other social dilemmas than the Trust Game. In Raub et al. (2013), we develop a model for $n \geq 3$ actors that covers a class of paradigmatic examples of social dilemmas, including not only the Trust Game but also, among others, the Investment Game (Berg et al. 1995), the Prisoner's Dilemma, and a 2-actor version of the Public Goods Game (e.g., Gächter/Thöni 2011). In addition to allowing for analyzing size effects, that model shows that the major implications derived in the present paper are quite robust in the sense that they hold also for other social dilemmas than the Trust Game.

Finally, a core feature of the model analyzed here is that the Trust Game is assumed to be a game with complete information and is repeated indefinitely often. Together, these two assumptions greatly simplify the analysis. A considerably more complex model results when we assume incomplete information (in the technical sense of game theory, see Rasmusen 1994, 47) and finitely repeated Trust Games. In a finitely repeated Trust Game with incomplete information, the trustor cannot observe all characteristics of the trustee. For example, with some probability the trustee has no incentive to abuse trust (since $T \leq R$) or has no opportunity to do so. The trustor knows the probability but cannot directly observe whether or not the trustee has an opportunity and an incentive to abuse trust. While the well-known backwards induction argument establishes that placing and honoring trust cannot be result of equilibrium behavior in a finitely repeated Trust Game with complete information, it is also well-known that—due to reputation effects—quite some placing and honoring trust can be equilibrium behavior in a finitely repeated Trust Game with incomplete information (see Buskens/Raub 2013 for further discussion and references).

In Frey et al. (2013) we study a model with one trustee and two trustors, playing finitely repeated Trust Games with incomplete information and an opportunity for the trustors to invest in network formation in an initial round 0. The model is far more complex than the one in this paper but yields similar implications, thus again indicating that implications derived in the present paper for an extremely simple model are quite robust when assumptions of the

model are varied. In particular, the model for finitely repeated Trust Games again implies investments in forming the information network if trust problems are neither too small nor too large. In addition, though, the Frey et al. model has an additional implication that enriches the theory of network effects and network formation related to trust problems.

Note that in the model presented in this paper, the conditions for trust equilibria depend exclusively on the incentives for the trustees and not at all on the incentives for the trustors. This seems problematic. Experimental research indicates, for example, that trustor behavior in the Trust Game depends also on the trustor's own incentives. In particular, the likelihood of placing trust increases when $P_i - S_i$ decreases or $R_i - S_i$ increases (e.g., Snijders 1996). A reason for this may be that subjects in the trustor role anticipate that trustee behavior may not only be motivated by the trustee's own monetary outcomes but also by monetary outcomes of the trustor. In a Trust Game with incomplete information it is possible to account for such a possibility with respect to trustee incentives and for the fact that the trustor cannot directly observe those incentives. Then, the information network no longer serves exclusively as a means that allows for sanctioning trustee behavior in a game with trustor i also through future behavior of trustor j. Rather, the information network serves another purpose, too. Namely, via information on how the trustee has behaved in previous games, including games with the other trustor, a trustor can now also learn about unobservable characteristics of the trustee such as whether or not (and if so, how) the trustee's behavior is also affected by the trustors' payoffs. One can then show that the existence of an equilibrium with investments in the information exchange network, which induces placing and honoring trust for possibly many rounds of the finitely repeated game, also depends on the incentives for the trustors.

Our contribution employed formal and more precisely game-theoretic modeling of network formation and effects, including the assumption of rational behavior. Such models have assets in the sense of clear specification of model assumptions. Also, as we have shown, such models can yield implications that include testable hypotheses. Such models, though, also have problematic features. For example, they often include very strong rationality assumptions, including assumptions on Bayesian updating of beliefs, strong assumptions that actors can foresee very complex consequences of their decisions, and strong assumptions on what actors know about the network structure and their position in the network. A common approach in the literature for coping with these issues is to drop or at least to strongly relax rationality assumptions, for example, by adopting a pure learning model or by assuming myopic best-response behavior with respect to network formation rather than farsighted rationality. This easily leads to adopting different assumptions on behavioral *regularities* for different contexts which could be seen as problematic from a methodological perspective. Moreover, it is less than obvious, to say the least, that radically assuming away strategic rationality does lead to models that are better in line with observable behavior (see, e.g., Callander/Plott 2005; Pantz 2006; Berninghaus et al. 2012; Corten/Buskens 2010). Another approach—less frequently chosen but the one

chosen here—is to radically simplify the network by considering a triad and modeling network formation as a once and for all decision at the beginning of the game. Of course, a drawback of this approach is that one abstracts from many core features of network structure and dynamics. Still, the strength of this approach and its contribution to theoretical pluralism is that it allows for an integrated and simultaneous analysis of network formation and effects under standard rationality assumptions. It would be interesting to also explore a third and in a sense 'intermediate' approach that allows for more complexity with respect to network structure and dynamics than ours while assuming 'some' but less than perfect strategic rationality. Such an approach is in its infancy but see Berninghaus et al. (2012) and Morbitzer et al. (2013) for some work in this direction (see Buskens et al. 2014 for a more detailed discussion of these issues and further references).

Appendix

Proof of Proposition 1. Proposition 1 is a direct implication of the fundamental theorem on trigger strategy equilibria in indefinitely often repeated games (e.g., Friedman 1986, 88–89).

Proof of Proposition 2. This proposition is a simple extension of the fundamental theorem on trigger strategy equilibria in indefinitely often repeated games. We sketch core elements of the proof and refer to Friedman (1986) for further details.

More specifically, we show that the condition in Proposition 2 is necessary and sufficient to ensure that the trustee maximizes his payoff in Γ^+ by always honoring trust, given that both trustors use the trigger strategy. We show this indirectly and thus assume that the trustee has a better response than always honoring trust. In that case, there must be some round t such that the trustee abuses trust for the first time against one of the trustors (otherwise, his strategy could not yield a higher payoff than always honoring trust). Immediately afterwards, both trustors will start to never place trust again. Thus, we must have

$$2R+w2R+...+w^{t-2}2R+w^{t-1}((i-1)R+T+(2-i)P)+w^t2P+w^{t+1}2P+... >$$

$$2R+w2R+...+w^{t-2}2R+w^{t-1}((i-1)R+R+(2-i)R)+w^t2R+w^{t+1}2R+... = \frac{2R}{1-w}$$

with $i = 1, 2$. This is equivalent with $(i-1)R+T+(2-i)P+w2P+w^2 2P+... > \frac{2R}{1-w}$. Thus, if it is profitable for the trustee to abuse trust in some round t, it would be likewise profitable to abuse trust already in round 1. The left-hand side of the equation for round 1 is larger for $i = 2$. Hence, payoff maximization by the trustee requires that he abuses trust in the second Trust Game in round 1. This implies

$$R + T + \frac{w2P}{1-w} > R + R + \frac{w2R}{1-w} = \frac{2R}{1-w}$$

which is equivalent with $w < \frac{T-R}{(T-P)+(R-P)}$ and thus contradicts the condition in Proposition 2. This completes our sketch of the proof that the trustee maximizes his payoff by always honoring trust if the trustors use trigger strategies under the condition of Proposition 2.

To see that the equilibrium is also subgame perfect, it suffices to note that a combination of trigger strategies implies that the trustors never place trust for the rest of Γ^+ as soon as trust has been abused or has not been placed once. However, never placing and always abusing trust always constitutes an equilibrium in Γ^+ as well as in all subgames of Γ^+. Hence, the equilibrium is also subgame perfect. This completes the sketch of the proof.

Proposition 3 follows directly from the equilibrium payoffs in $\Gamma^+$and Γ^- under the assumption that $TEMP^+ \leq w < TEMP^-$ so that a trust equilibrium exists only in Γ^+ and that trust will never be placed in Γ^-. Namely, the payoffs associated with the trigger strategy equilibrium in Γ^+ are $\frac{2R}{1-w}$ for the trustee and $\frac{R_i}{1-w}$ for the trustors, while $\frac{2P}{1-w}$ and $\frac{P_i}{1-w}$ are the payoffs in Γ^-. The upper bound on the value of the network is equal to the difference between these payoffs. Note that never placing trust need not be the only equilibrium outcome of Γ^- when $w < TEMP^-$. There can be equilibria of Γ^- such that the trustee sometimes abuses trust, while the trustors always place trust, with equilibrium payoffs larger than $\frac{2P}{1-w}$ and, respectively, $\frac{P_i}{1-w}$. The value of the network would then be smaller. Hence, Proposition 3 indeed only specifies an upper bound.

Proposition 4 can be derived as follows. The subgames Γ^+ and Γ^- always have equilibria such that trust is never placed and would always be abused in all Trust Games in all rounds 1, 2,... of Γ^+ and Γ^-. Moreover, under the condition $TEMP^+ \leq w < TEMP^-$ there is a trust equilibrium in Γ^+ but not in Γ^-. A trust equilibrium is associated with higher payoffs for each actor than payoffs that are obtained when trust is never placed in all Trust Games in all rounds 1, 2,.... Under the conditions of Proposition 4, we thus obtain the following equilibrium for Γ. First, each actor who can invest in network formation does indeed invest in round 0. Second, both trustors never place trust and the trustee would abuse trust unconditionally in each subgame Γ^-. Third, the trustors play trigger strategies and the trustee always honors trust in subgame Γ^+.

The question emerges why the conditions in Proposition 4 are sufficient but not necessary for the existence of an equilibrium such that actors invest in network formation and subsequently always place and honor trust. Assume that $TEMP^+ < TEMP^- \leq w$. Then, a trust equilibrium also exists for subgame Γ^-. Thus, investments in network formation are not necessary in round 0 for ensuring that a trust equilibrium exists in the subgame after round 0. Nevertheless, trustors could never place trust in Γ^- and play trigger strategies in Γ^+. Investments in social capital are then still consistent with equilibrium behavior under the conditions of Proposition 4.

Bibliography

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

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

Berninghaus, S. K./K.-M. Ehrhart/M. Ott (2012), Forward-looking Behavior in Hawk-Dove Games in Endogenous Networks: Experimental Evidence, in: *Games and Economic Behavior* 75, 35–52

Buskens, V. (2002), *Social Networks and Trust*, Boston

—/R. Corten/W. Raub (2014), Social Networks, in: Braun, N./N. J. Saam (eds.), *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*, Wiesbaden

—/W. Raub (2013), Rational Choice Research on Social Dilemmas, in: Wittek, R./T. A. B. Snijders/V. Nee (eds.), *Handbook of Rational Choice Social Research*, Stanford, 113–150

Callander, S./C. R. Plott (2005), Principles of Network Development and Evolution: An Experimental Study, in: *Journal of Public Economics* 89, 1469–1495

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

—/E. Katz/H. Menzel (1966), *Medical Innovation: A Diffusion Study*, Indianapolis

Corten, R. (2009), *Co-evolution of Social Networks and Behavior in Social Dilemmas: Theoretical and Empirical Perspectives*, PhD thesis, Utrecht University

—/V. Buskens (2010), Co-evolution of Conventions and Networks: An Experimental Study, in: *Social Networks* 32, 4–15

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

Dutta, B./M. O. Jackson (2003) (eds.), *Networks and Groups. Models of Strategic Formation*, Berlin

Eguiluz, V. M./M. G. Zimmermann/C. Cela-Conde/M. San Miguel (2005), Cooperation and Emergence of Role Differentiation in the Dynamics of Social Networks, in: *American Journal of Sociology* 110, 977–1008

Flap, H. (2004), Creation and Returns of Social Capital, in: Flap, H./B. Völker (eds.), *Creation and Returns of Social Capital*, London, 3–23

—/B. Völker (2013), Social Capital, in: Wittek, R./T. A. B. Snijders/V. Nee (eds.), *Handbook of Rational Choice Social Research*, Stanford, 220–251

Frey, V./V. Buskens/W. Raub (2013), Embedding Trust: A Game-theoretic Model for Investments in and Returns on Network Embeddedness, *Mimeo*, Utrecht

Friedman, J. W. (1986), *Game Theory with Applications to Economics*, New York

Gächter, S./C. Thöni (2011), Micromotives, Microstructure, and Macrobehavior, in: *Journal of Mathematical Sociology* 35, 26–65

Goyal, S. (2007), *Connections. An Introduction to the Economics of Networks*, Princeton

Granovetter, M. (1973), The Strength of Weak Ties, *American Journal of Sociology* 78, 1360–1380

— (1974), *Getting a Job. A Study of Contacts and Careers*, Cambridge/MA

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

Jackson, M. O. (2008), *Social and Economic Networks*, Princeton

—/Y. Zenou (2013) (eds.), *Economic Analyses of Social Networks*, 2 volumes, London

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

Lin, N. (2001), *Social Capital: A Theory of Social Structure and Action*, New York

Morbitzer, D./V. Buskens/S. Rosenkranz/W. Raub (2013), How Farsightedness Affects Network Formation, in this issue of *Analyse & Kritik*

Morris, M./J. Zavisca/L. Dean (1995), Social and Sexual Networks: Their Role in the Spread of HIV/AIDS among Young Gay Men, in: *AIDS Education and Prevention* 7, 24–35

Newman, M. (2010), *Networks: An Introduction*, Oxford

—/A.-L. Barabási/D. J. Watts (2006) (eds.), *The Structure and Dynamics of Networks*, Princeton

North, D. C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge

Pantz, K. (2006), *The Strategic Formation of Social Networks: Experimental Evidence*, Aachen

Pujol, J. M., A. Flache/J. Delgado/R. Sanguessa (2005), How Can Social Networks Ever Become Complex? Modelling the Emergence of Complex Networks from Local Social Exchanges, in: *Journal of Artificial Societies and Social Simulation* 8, URL:http://jasss.soc.surrey.ac.uk/8/4/12.html

Rapoport, A. (1974), Prisoner's Dilemma—Recollections and Observations, in: Rapoport, A. (ed.), *Game Theory as a Theory of Conflict Resolution*, Dordrecht, 17–34

Rasmusen, E. (1994), *Games and Information: An Introduction to Game Theory*. 2nd edition, Oxford

Raub, W. (2004), Hostage Posting as a Mechanism of Trust: Binding, Compensation, and Signaling, in: *Rationality and Society* 16, 319–365

—/V. Buskens/V. Frey (2012), Vertrouwen als opbrengst van investeringen in sociaal kapitaal: Een eenvoudig theoretisch model [Trust as a Return on Investments in Social Capital: A Simple Theoretical Model], in: Buskens, V./I. Maas (eds.), *Samenwerking in sociale dilemma's: Voorbeelden van Nederlands onderzoek* [*Cooperation in Social Dilemmas: Examples of Research in the Netherlands*] (special issue of *Mens en Maatschappij*), Amsterdam, 17–44

—/—/— (2013), The Rationality of Social Structure: Cooperation in Social Dilemmas through Investments in and Returns on Social Capital, in: *Social Networks* 35, 720–732

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

Snijders, C. (1996), *Trust and Commitments*, PhD thesis, Utrecht University

Snijders, T. A. B. (2013), Network Dynamics, in: Wittek, R./T. A. B. Snijders/V. Nee (eds.), *Handbook of Rational Choice Social Research*, Stanford, 252–279

Vega-Redondo, F. (2006), Building up Social Capital in a Changing World, in: *Journal of Economic Dynamics and Control* 30, 2305–2338

— (2007), *Complex Social Networks*, Cambridge