Network Formation and Trust

Vincenz Frey

Manuscript committee:	Prof. dr. A. Diekmann
	Prof. dr. P.G.M. van der Heijden
	Prof. dr. A. Schram
	Prof. dr. C. Snijders
	Prof. dr. T. Voss

Frey, V.

Network Formation and Trust

ISBN 978-90-393-6509-0

Printed by Ridderprint BV, Ridderkerk.

This book was composed and typeset using LaTeX by Vincenz Frey.

Network Formation and Trust

Netwerkvorming en vertrouwen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus,
prof. dr. G. J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 1 april 2016 des middags te 12.45 uur

door

Vincenz Cajetan Frey

geboren op 3 april 1983
te Ehrendingen, Zwitserland

Promotoren:	Prof. dr. ir. V. Buskens
	Prof. dr. W. Raub
Copromotor:	Dr. R. Corten

The research presented in this book was funded by the Netherlands Organization for Scientific Research (NWO) Graduate Training Program Grant (2008/2009) awarded to the research school Interuniversity Center for Social Science Theory and Methodology (ICS).

Contents

List of tables

List of figures

Chapter 1

Introduction and synthesis

Imagine that Alice just moved to take on a new job. Two of the many items on her to-do list are "hire a cleaning person" and "buy a laser-pointer online." In these affairs, Alice faces trust problems. A cleaning person might, once alone in the house, steal personal belongings and disappear. The owner of an online shop may never send out the laser-pointer or send a product that does not match the description.

Such trust problems abound in social and economic exchanges that would potentially benefit both parties. Formal institutions such as laws and contracts do often not provide sufficient assurance (Macaulay, 1963; Weber, 1976 [1921]). The costs of legally pursuing an online seller who did not deliver may be prohibitively high and even if a court finds a cleaning person guilty of stealing, the victim may not be fully compensated for the loss. Hence, trust is often needed for an exchange to take place. It is in this sense that Arrow (1974, p. 23) referred to trust as an "important lubricant of the social system" and that others argued that overcoming trust problems is a key to social order and prosperous societies (Deutsch, 1958; Putnam, 1993; Zak & Knack, 2001).

The embeddedness of exchange situations in social structures through which information on past behavior of trustees disseminates can warrant trust (Buskens & Raub, 2013; MacLeod, 2007). If Alice's new neighbors praise their cleaning person, Alice may be trusting that this cleaning person will not steal. Similarly, positive ratings that customers of an online shop have left on a reputation system may give Alice the trust necessary to order a laser-pointer and pay it upfront. Thus, social structures for the dissemination of information can make exchanges possible that would not be possible in the absence of such structures.

In this dissertation, we argue that if trust problems can be overcome using social structures for sharing information, such as word-of-mouth networks or websites that aggregate and disseminate consumer experiences, actors may purposively establish such social structures with the aim to benefit from trust and trustworthiness. Alice

may invite her new neighbors for dinner not just out of courtesy but also with the intention to ask them whether they know a reliable cleaning person. The owner of an online shop may invest in a reputation system, expecting that this will bolster the trust of potential customers. We pursue this idea in Chapters 2 to 5, where we investigate, in an integrated framework, investments in establishing social structures for information sharing and effects of social structures for information sharing on trust and trustworthiness.

We, furthermore, argue that trust problems and the sharing of information can have repercussions for the structure of networks for exchanges in trust situations. We argue that the fear of trust abuse can lead large numbers of people to exchange with one or a few others. Suppose that Alice finds two sellers who offer laser-pointers online. One has received many good ratings. The other has not yet received any ratings. It seems natural that Alice buys from the seller with an established reputation (cf. Kollock, 1994, p. 318). The next customer will probably likewise choose that seller, and so on.... The established seller gets more and more business while the other seller is not given a chance to show whether he is trustworthy, even though he may be no less trustworthy. We thus theorize that the sharing of information on past performance not only promotes trust and trustworthiness but may also give rise to unfounded inequality in exchange volumes among actors who could abuse trust. We elaborate on this in Chapter 6.

In this dissertation, we thus study how social structures for the dissemination of information affect behavior in trust problems and, simultaneously, how social structures emerge in the presence of trust problems. The studies presented in this book contribute to the investigation of conditions under which actors overcome trust problems. They, hence, contribute also to the broader literature that studies mechanisms enabling actors to cooperate together despite incentives to take advantage of one another (Barrera, 2014; Dawes, 1980; Diekmann & Lindenberg, 2015; Kollock, 1998; Pennisi, 2005; Raub et al., 2015; Voss, 1985). By studying the emergence of information sharing networks and exchange networks in the presence of trust problems, this dissertation, furthermore, contributes to the literature on the formation of social networks (Corten, 2014; Flap, 2004; Goyal, 2007; Jackson, 2008; Lin et al., 2001).

This chapter gives a general introduction to the studies presented in the subsequent chapters, sketches what we find in these studies, and points out issues for future research. The remainder of this chapter is structured as follows. In Section 1.1, we define what we mean by a "trust situation." In Section 1.2, we discuss reputation mechanisms at work in socially embedded settings as a remedy to trust problems. The research questions are formulated in Section 1.3. In Section 1.4, we sketch our approach to these questions and summarize the findings. Section 1.5 indicates directions

for future research and Section 1.6 contains a few organizational remarks.

1.1 Trust situations as social dilemmas

In this dissertation, *trust* refers to the trust in a social or economic exchange of one party—the trustor—that a specific other party—the trustee—will not act in a way that benefits the trustee and imposes costs or harm on the trustor (cf. Bacharach & Gambetta, 2001; Hardin, 2002, Chap. 2; Rousseau et al., 1998; Snijders, 1996; Yamagishi & Yamagishi, 1994; for discussions of different concepts of trust, see Hardin, 2002, Chap. 3, Mistzal, 1996, and Rotter, 1971). We do not focus on trust in the *abilities* of a trustee ("confidence"; Barber, 1983; Yamagishi & Yamagishi, 1994, pp. 131–132). Instead, the focus is on trust problems due to *incentives* for the trustee to take advantage of the trustor. If we say "Alice does not trust the cleaning person who placed an advertisement in the local grocery store," we mean that she suspects that this cleaning person would behave opportunistically, *not* that she believes that this cleaning person is incompetent.

Coleman (1990, pp. 97–99) identifies four features that characterize a trust situation in social or economic exchange:

- If the trustor places trust, the trustee can choose to honor or abuse trust while the trustee would not have these options otherwise.
- The trustor regrets the placement of trust if trust gets abused.
- Placing trust is a voluntary action of the trustor and there are no formal safeguards that provide assurance for the trustor.
- There is a time-lag: When deciding whether to place trust, the trustor cannot know whether the trustee will honor or abuse trust.

We add two features to this definition of a trust situation. Both are related to incentives and are standard in the literature on trust problems (see Rousseau et al., 1998):

- It is likely that abusing trust leaves the trustee better off than honoring trust (at least in the short run).
- The trustor and the trustee are both better off if trust is placed and honored than if the trustor does not place trust.

A suboptimal outcome may obtain in such a trust situation if actors rationally pursue their interest and expect that others do the same. Given the assumption of

individual rationality, a trustee abuses trust if this leaves him better off than honoring trust. Also, a trustor only places trust if the trustor's expected outcome of placing trust is better than the trustor's expected outcome of not placing trust (compare Williamson's, 1993, definition of "calculative trust"). If a rational trustor cannot expect good behavior of the trustee with a large enough probability, the trustor will thus withhold trust while trust being placed and honored would leave both actors better off. In this case, individually rational behavior leads to a Pareto-suboptimal outcome; the situation represents a social dilemma (Dawes, 1980; Kollock, 1998; Raub et al., 2015).

Research on the so-called Trust Game shows that—in line with the rationality assumption—people often withhold trust if in the role of the trustor and abuse trust if in the role of the trustee, although placed and honored trust is also observed regularly (see Camerer, 2003, Chap. 7, and Johnson & Mislin, 2011, for reviews). The Trust Game (Dasgupta, 1988; Kreps, 1990a) is a stylized model for the study of trust situations. We introduce this game in later chapters.

Research also shows that the tendencies for withholding and abusing trust depend on characteristics of the trust situation. For example, Snijders & Keren (2001) find that there is less trust if the trustor loses a lot if trust is abused compared to the no trust situation. They also observe less trust when trustees can earn a large extra benefit by abusing trust. This suggests that trust situations may differ with respect to the "size of the trust problem." In later chapters we will often use the notion of the "size of the trust problem" and make it more precise.

1.2 Trust, networks, and network formation

The finding that people often withhold trust or abuse trust in stylized trust situations pertains to "isolated exchanges"—exchanges between trustors and trustees who are anonymous strangers that have no common past or future and are not connected through third parties. Often, however, trust situations occur in social contexts in which people meet each other repeatedly or in which people exchange information about one another. Such "embeddedness" (Granovetter, 1985) enables a reputation mechanism that can warrant trust (Granovetter, 1985; Klein, 1997).

Consider how trust can build if a trustor and trustee deal with one another over time. If a trustor and trustee interact repeatedly—so-called *dyadic embeddedness*—the trustor can learn about the trustee's trustworthiness (Buskens & Raub, 2002). A natural way to behave for the trustor is then to place trust again if the trustee has shown to be reliable. This creates an incentive for the trustee for good behavior. It creates a "shadow of the future" (Axelrod, 1984) which, in turn, allows the trustor to

initially place trust with realistic expectations of good behavior. Experimental studies with Trust Games confirm that subjects tend to be more trusting and trustworthy if they interact together repeatedly than if they play only a single Trust Game (Camerer, 2003, Chaps. 2 and 8; Schneider & Weber, 2013).

This reputation mechanism can be amplified when there is *network embeddedness*—when trustors share their opinions about trustees, for example, in word-of-mouth networks or on online feedback platforms (see Buskens & Raub, 2002, 2013, for the theoretical and empirical literature as well as for the distinction between dyadic and network embeddedness). With network embeddedness, a trustor can learn from the experiences of other trustors and a trustee has to fear that many trustors avoid him if he abuses the trust of a single trustor. The reputation mechanism can then warrant trust even in the absence of personal experiences and when the potential for future exchanges between the same two parties does not suffice to make a trustee withstand the temptation of abuse.[1]

That trust problems can be overcome if there is embeddedness suggests that people may actively seek embeddedness (cf. Flap, 2004). To benefit from dyadic embeddedness, people can exchange with others with whom they have good experiences and, thereby, form long-term exchange relations. Such "commitment formation" in the presence of trust problems has been documented in several case studies and experiments (see Cook et al., 2004, for a review). Kollock (1994), for example, proposed that trust problems explain that, in Southeast Asia, raw rubber is often traded in committed relationships between particular plantation owners and particular brokers while rice is traded in open markets between strangers. It appears that the quality of raw rubber is difficult to assess (which leads to a trust problem) while the quality of rice is easy to judge. Kollock supports his hypothesis on the importance of long-term exchange relations to overcome trust problems by demonstrating that such relations are more likely to emerge in an experimental condition featuring a trust problem than in a control condition (see also Brown et al., 2004; Kirman, 2001; Yamagishi et al., 1998).

To benefit from network embeddedness, people can choose to exchange with others who are embedded in their communication networks. There are indictions that criminal organizations use the strategy of doing business in existing family and ethnic networks to handle the risks associated with illegal behaviors (Gambetta, 1993). Ordinary consumers leverage their networks, too. DiMaggio & Louch (1998) show that they buy certain goods and services more often from friends of friends than one would expect in anonymous markets, particularly if trust problems are large.

[1] Reputation mechanisms are not the only remedy to overcome trust problems. Riegelsberger et al. (2005) provide an overview of the mechanisms that have been identified to mitigate trust problems.

Another strategy for actors to benefit form network embeddedness is to establish relations or institutions for information sharing to "embed" previously unembedded exchange partners. Guseva & Rona-Tas (2001) report that American as well as Russian banks use this strategy to reduce trust problems in the credit card business. In the United States, groups of money lenders run credit bureaus to collect and share information on the credit histories of borrowers (see also Jappelli & Pagano, 2002). Russian banks did not achieve such collaboration, which hampered the growth of the credit market. To assure the creditworthiness of at least a small select group of cardholders, Russian bankers use their existing social ties and build new ties.

Online reputation systems are another example for institutions for information sharing that were established to resolve trust problems (Dellarocas, 2003; Diekmann et al., 2014; Kollock, 1999; Resnick et al., 2000, 2006). Online reputation systems facilitate trust in peer-to-peer market places, such as ebay.com, and also substitute or complement more traditional institutions for the sharing of information, such as guide books with recommendations for hotels and bed & breakfasts or magazines of consumer associations. While online reputation systems are a contemporary example, historical reputation systems that were established to resolve trust problems are discussed in Greif (1989), Diekmann et al. (2014, pp. 65–66), Klein (1997) and Milgrom et al. (1990).

1.3 Research questions

In Chapters 2 to 5, we study "*investments in and returns on network embeddedness*"—that is, the establishment of relations or institutions for information sharing with the aim to embed previously unembedded exchange partners and the effects of such relations or institutions on trust and trustworthiness. In the literature on social capital, it is not a new idea that actors may purposively establish beneficial social relations or institutions (see Flap, 2004; Lin et al., 2001). Still, explicit theoretical models for such "investments in social capital" are scarce and no theoretical models are available for the understanding of investments in establishing network embeddedness as a means to foster trust and trustworthiness (cf. Raub & Buskens, 2012). In addition, while there are indications that actors may purposively establish social structures for information sharing to overcome trust problems (Guseva & Rona-Tas, 2001; Klein, 1997), it is not known under what circumstances this is particularly likely. We aim to fill these gaps in the literature. The overarching question that we address theoretically as well as empirically in Chapters 2 to 5 is: *Under what circumstances are actors particularly likely to establish network embeddedness to overcome trust problems?*

Purposive actors should be especially inclined to exert effort to establish embed-

dedness when embeddedness promotes trust and trustworthiness particularly strongly. Therefore, the above question leads us to also pose the question: *Under what circumstances does network embeddedness promote trust and trustworthiness particularly strongly?* Addressing this question contributes to the literature on the effects of network embeddedness. This literature provides indications that the degree to which networks help overcome trust problems varies across contexts, but it does not provide a systematic analyses of the context dependence of network effects (see Portes & Sensenbrenner, 1993; Mizruchi et al., 2006; Simpson & McGrimmon, 2008).

While we endogenize information sharing and assume relations for social or economic exchanges in trust situations as given in Chapters 2 to 5, we assume information sharing as given and endogenize exchange relations in Chapter 6. As discussed in Section 1.2, it has been shown that trust problems affect the formation of exchange networks: Trust problems lead actors to form dyadic, long-term exchange relations (Brown et al., 2004; Kirman, 2001; Kollock, 1994; Cook et al., 2004; Yamagishi et al., 1998). This research has focused on contexts where there is no network embeddedness and it is, hence, not known what exchange structures emerge in embedded trust situations where trustors can learn about trustees from the experiences of other trustors. To address this gap in the literature, we formulate the question: *How does information sharing in trust problems affect exchange networks?*

1.4 Approaches and findings

1.4.1 Investments in and returns on network embeddedness

The aim of Chapters 2 to 5 is to establish regularities at the macro level. Under what societal, macro conditions is it particularly likely that trust problems lead to the establishment of social structures for information sharing between trustors and under what macro conditions does network embeddedness promote the rate of mutually beneficial exchanges particularly strongly? We address these questions in the framework of methodological individualism (Udehn, 2002; Weber, 1976 [1921]). Figure 1.1 sketches our "micro-macro approach" in a stylized scheme that was developed by Coleman (1986, pp. 1320ff; 1990, pp. 5-10) and is discussed extensively in Raub et al. (2011). In Figure 1.1, we use two Coleman schemes in a sequence: The first (Boxes A to D) represents investments in establishing network embeddedness and the second (Boxes D to G) represents effects of network embeddedness on trust and trustworthiness. In terms of Figure 1.1, our aim is to identify regularities at the macro level represented in Arrows 4 and 8 using explanatory arguments that run via Arrows 1, 2, and 3, and 5, 6, and 7, respectively.

Box A in Figure 1.1 represents propositions describing the initial macro-level circumstances; it maintains that there is a trust problem of a certain size. Box A could also describe circumstances such as the lack of a legal system that guarantees full compensation of the trustor in the case of abuse, the average extra profit that trustees can earn by abusing trust, or the likelihood that a trustee can earn an extra profit by abusing trust. Box A furthermore maintains that there is no network embeddedness but that establishing it is possible. Box A could also specify whether the macro-level situation is such that only trustors or only trustees or either type of actor can establish network embeddedness. Arrow 1 represents "bridge assumptions" (Wippler & Lindenberg, 1987) that describe how these macro-level conditions shape the situation of individual actors at the micro level. Box B describes the micro-level situation of individual actors: The behavioral alternatives of establishing or not establishing network embeddedness and the incentives for choosing one of the alternatives. Arrow 2 represents assumptions on individual behavior. In our case, this is the assumption of rational behavior, made precise as game-theoretic equilibrium behavior. Box C represents micro-level outcomes: The choices of actors whether to invest in establishing embeddedness. Arrow 3 represents "transformation rules" (Wippler & Lindenberg, 1987) that specify how individual choices aggregate to a new macro-level condition. For example, it is here specified whether one actor alone can establish network embeddedness or whether this requires the joint effort of several actors. Box D describes the resulting macro-level situation, which may differ from the initial situation (Box A) in that there may now be network embeddedness.

The macro-level situation represented in Box D shapes the incentives for placing and honoring trust (Box E), thus providing the independent variables for decisions in trust situations (Arrow 6). Behavioral outcomes in trust situations represented in Box F aggregate to the macro-level situation represented in Box G. This situation can be characterized by a lower or higher rate of mutually beneficial exchanges.

Chapters 2 to 4 provide theoretical studies in the spirit of this micro-macro framework. In the game-theoretic models analyzed in these chapters, actors have the option to pledge a costly investment to establish a network for information sharing before they interact in Trust Games. In Chapter 2 we model how network embeddedness affects behavior ("Boxes D to G") in the framework of indefinitely repeated games.[2] The model presented in Chapter 2 allows for an analysis of investments in and returns on network embeddedness in large populations. This model generalizes also to interactions in other social dilemma situations than trust situations, such as situations that are represented in the Prisoner's Dilemma Game (Rapoport & Chammah, 1965)

[2]See Mailath & Samuelson (2006) for a handbook on game-theoretic concepts for the study of reputation effects. Further introduction of these concepts will be given in the respective chapters.

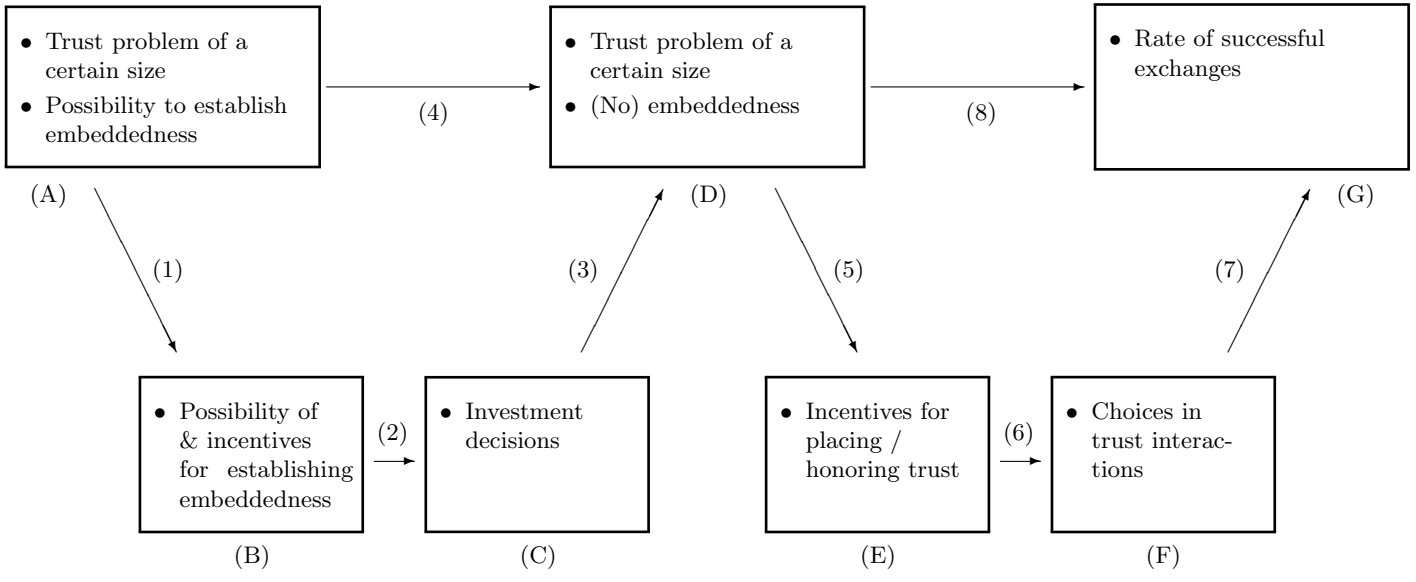


Figure 1.1: The micro-macro approach to the study of investments in and returns on network embeddedness.

or the Investment Game (Berg et al., 1995).

The model analyzed in Chapters 3 and 4 is restricted to the study of interactions in trust situations. It is also much simpler than the model presented in Chapter 2 in that it focuses on a scenario in which two trustors interact repeatedly with a single trustee. On the other hand, the model is more complex because it accounts for incomplete information—the trustee may be of a type who has no incentive or possibility to abuse trust but the trustors cannot directly observe whether the trustee is of this trustworthy type.[3] Technically speaking, in Chapters 3 and 4, we study a finitely repeated Trust Game with incomplete information. Accounting for incomplete information allows modeling embeddedness effects more comprehensively. In a model with complete information, where trustors know the behavioral alternatives and incentives of trustees (as in Chapter 2), network embeddedness affects behavior exclusively because it makes it possible that a single abuse of trust is sanctioned by several trustors. A model with incomplete information allows modeling also that network embeddedness can promote trust and trustworthiness because it enables a trustor to learn about the trustworthiness of a trustee from the experiences of other trustors (see Buskens & Raub, 2002). Furthermore, in a model with complete information, an investment in establishing embeddedness has exclusively the purpose of enabling reputation effects. In a model in which the trustors are unsure about the trustee's type, incomplete information, it can be shown that a trustee's investment in establishing embeddedness can also serve as a credible, costly signal that the trustee is of the trustworthy type. We devote two chapters to the analysis of the model with incomplete information: We study the scenario in which the trustors can establish embeddedness in Chapter 3 and the scenario in which the trustee can estabish embeddedness in Chapter 4.

Many results of the two different models are similar. Thus, the use of different modeling approaches enabled us to check the robustness of predictions to specific model assumptions. Also, however, each model provides some predictions that do not follow from the other model.

A key prediction of our theoretical models is that network embeddedness promotes trust and trustworthiness most strongly if the trust problem is of intermediate size, neither very small nor very large. That is, the magnitude of the embeddedness effect is expected to have an inverse U-shape over the size of the trust problem. This prediction concerns Arrow 8 in Figure 1.1. Furthermore, our theoretical analyses then also suggest that the likelihood of investments in establishing network embeddedness follows an inverse U-shape in the size of the trust problem (this prediction concerns Arrow 4 in Figure 1.1). Both the model for indefinitely repeated games and the model

[3]See Rasmusen (1994, p. 47) for a definition of incomplete information in the technical sense of game theory.

for finitely repeated games with incomplete information imply these effects of the size of the trust problem. The two models differ from each other in that the trustees' incentives are the major driving force in the model for indefinitely repeated games, while the trustors' incentives are the major driving force in the model for finitely repeated games with incomplete information.

To provide some intuition for the predictions on how network effects and investments in network embeddedness depend on the size of the trust problem, suppose that Alice's to-do list contains, next to "hire a cleaning person", two further trust problems, namely, the items "find a carpenter to install a roof window" and "have a landscaper take care of the garden." If a carpenter tries to save time and money, water may leek into the house in stormy weather conditions; the owner of a landscaping business may send incompetent employees who do the garden more harm than good. Alice is not overly attached to the plants in her new garden. She entrusts the job to the local landscaper, thinking that it is not worth her while to try to find out about the experiences of others. Hiring a cleaning person is a more sensitive issue for Alice. She believes that the odds are high that a self-proclaimed cleaning person is a thief in disguise. She would not hire a cleaning person that has not been recommended to her. She, therefore, asks around in her neighborhood and at her workplace for recommendations. Finally, Alice is quite suspicious of carpenters and knows that a leaking roof window can cause a lot of trouble. She writes "find a carpenter to install a roof window" again on her next to-do list and then again on the next one .... Alice knows that she would not trust a carpenter even if others tell good stories and, therefore, does not to try to find out about experiences of others. Eventually, she ditches the idea of having her home office flooded with sunlight and, with some regret, buys a large lamp instead.

Chapter 5 reports a laboratory experiment conducted to test predictions derived in Chapters 2 to 4. The discussion of reputation systems in Section 1.2 suggests that institutions for the dissemination of information were created purposively to promote trust and trustworthiness. However, there is so far no clear empirical evidence for this. That many exchanges take place in settings where there are such institutions could reflect that trust flourishes in these settings. It could also be that trust problems occur frequently in situations in which information sharing networks were established for other reasons (not with the intention to foster trust). The laboratory allows investigating whether the presence of a trust problem indeed leads actors to establish network embeddedness. In addition, the laboratory allows some control over theoretically relevant parameters that affect the size of the trust problem, such as the likelihood that a trustee can benefit from abusing trust or the payoffs in the trust interactions, which are difficult to measure outside the laboratory.

The experiment implemented the theoretical models discussed in Chapters 3 and 4: Two subjects in the role of trustors interacted in Trust Games with a subject in the role of a trustee. The size of the trust problem was manipulated via the probability that the trustee has a monetary short-term incentive to abuse trust. Embeddedness was exogenous or could be established at costs before the trust interactions by the trustors or the trustee. The results show that, if given the possibility, a considerable portion of trustors and trustees pledges an investment to establish embeddedness. The results also confirm that embeddedness promotes trust and trustworthiness and, furthermore, indicate that the effect of embeddedness is stronger if embeddedness is endogenous rather than exogenous. However, we find little support for the inverted U-shape hypotheses: There was no systematic variation in investments in embeddedness or in effects of embeddedness related to the size of the trust problem.

1.4.2 Reputation cascades

In Chapter 6, we assume information sharing between trustors as exogenous and study the endogenous formation of exchange relations between trustors and trustees. We assume that trustors can choose their exchange partners. The question that we address is: How does information sharing in trust problems affect exchange networks?

Figure 1.2 sketches our micro-macro approach to this question using the Coleman scheme. Box A represents the initial macro-level situation. It states that there is a trust problem and that trustors are free to choose their exchange partners. Box A, furthermore, states that trustors share information on the behavior of trustees in disjoint groups and how large these groups are. Arrow 1 represents how the macro-level conditions translate into the situation of individual actors. Box B describes the situation of an individual trustor: The information the trustor has about the trustees, the trustor's options of withholding trust or placing trust in a particular trustee and the trustor's incentives for choosing among these options. Box B also describes the situation of a trustee who gets trusted: The trustee's options of and incentives for honoring or abusing trust. Arrow 2 represents individual decision making, which we assume to be rational, and Box C represents behavioral outcomes: Whether a trustor chose to place trust and if so, in which trustee and whether this trustee honored trust. Arrow 3 represents how the choices at the micro level give rise to a new macro-level situation. This situation (Box D) can be characterized by the rate of successful exchanges and the inequality among trustees in exchange volume.

The game-theoretic model discussed in Chapter 6 suggests that high inequality in exchange volumes among trustees may emerge. What drives this is that, to minimize the risk of abuse, rational trustors avoid dealing with trustees who lack a transaction

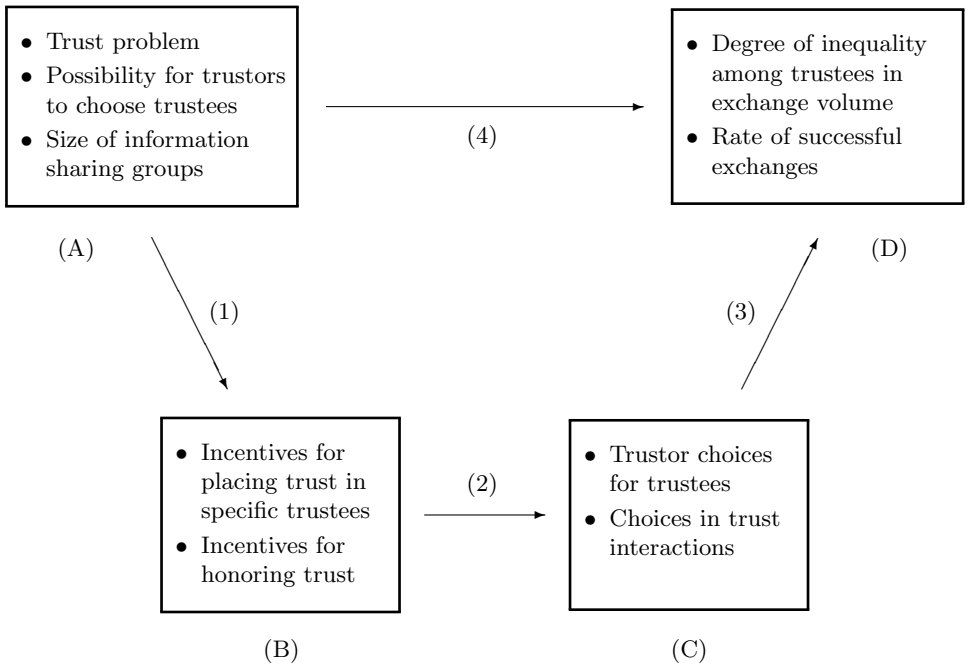


Figure 1.2: The micro-macro approach to the study of reputation cascades.

history in favor of trustees of good repute. After the first time trust is honored in some group of trustors who share information, *all* trustors of that group will in subsequent exchanges place trust in the trustee who was lucky to be the first to honor trust. This leads to a lack of opportunity for potentially trustworthy newcomers to get traction and perpetuates arbitrary initial inequities among trustees. Our analysis thus suggests that trust problems and the sharing of information give rise to a form of cumulative advantage (Gould, 2002; Merton, 1968; Salganik et al., 2006; Van de Rijt et al., 2014) resulting in arbitrary inequality in transaction volume among trustees. The theory implies that the inequality among trustees is larger if the trustors share information in larger communities. This means that higher levels of information pooling have not only a greater potential to resolve trust problems but can also lead to high inequality as an unintended consequence.

The results of a laboratory experiment support the theory. If trustors share information in larger groups, the rate of successful exchanges is higher but there is also more inequality among trustees in exchange volumes. There is clearly some arbitrariness in the inequality: Groups of trustors were often locked in on a trustworthy

trustee without having information on the trustworthiness of any of the other trustees. Furthermore, different (disjoint) groups of trustors were often locked in on different trustees. Our experiment also offers evidence that these cascades were driven by a fear of abuse of trust and not, for example, by a general tendency to imitate the choices of others. In a control condition in which abusing trust was costly for trustees—i.e., in a condition in which there was no trust problem—trustors did not preferentially choose trustees with an established reputation over unknown trustees.

1.5 Issues for future research

1.5.1 Investments in and returns on network embeddedness

Our theoretical models for the simultaneous study of investments in and returns on network embeddedness comprise several assumptions that could be relaxed in future theoretical research. We, first, discuss assumptions concerning the nature of information exchange and, then, assumptions on the structure and dynamics of information sharing networks. Finally, we point out directions for future empirical research.

Our models assume that passing on information is free of costs once a platform or network for information sharing is set up. However, in some settings, sharing information requires effort. For example, it costs Alice some time to post a rating on her experience with an online seller. She might be willing to incur this cost out of gratitude and because she feels an obligation to reciprocate the favor (Diekmann et al., 2014) or because of an urge to exact revenge (see Heyes & Kapur, 2012, p. 814, and the references therein). The empirical studies of Abraham et al. (2014), Diekmann et al. (2014), and Gërxhani et al. (2013) indicate that a small cost of sharing information does not need to undermine the trust and trustworthiness promoting effects of the possibility for information sharing (see Rockenbach & Sadrieh, 2012, for related results). Still, it seems plausible that the effects of network embeddedness are smaller if sharing information is more expensive. Future studies could extend the models presented in this book by including costs of sharing information. Gazzale (2009) provides a theoretical study of reputation effects with costly information sharing that could prove useful for work in this direction. Instead of modeling costs directly, one could also engage alternative assumptions on when trustors share information. Such an assumption could be that trustors voice their opinion exclusively when angered by an abuse of trust. We believe that the model presented in Chapter 2 would under this assumption lead to the same results as reported here. In this model, trust is placed in every single interaction whenever there is an equilibrium involving any trust. Hence,

if a trustor does not voice, other trustors can infer that trust was honored. However, in the equilibria of the models in Chapters 3 and 4, trustors sometimes randomize between placing and withhold trust. If they only share information on trust abuse, some information is lost and embeddedness effects would probably be smaller.

We also assumed that information on a trustee's past choice is always accurate. However, a timely sent out item that was ordered online can arrive late and, conversely, a roof window does not necessarily leak if it was installed sloppily. Furthermore, misunderstandings can occur when a rumor travels through a network. Random contingencies and noise in the transmission of information probably reduce the trust and trustworthiness promoting effects of network embeddedness. But does random inaccuracy in information also change under what circumstances the effects of network embeddedness are largest? To investigate such questions and decrease the level of abstraction (Lindenberg, 1992), the presented models could be enriched by concepts developed in the literature on "optimal punishments" (see Porter, 1983) and reputation building in noisy environments (see Aperjis et al., 2014; Fudenberg & Maskin, 1990).

Actors can also manipulate information purposively. There exist underground markets where online traders can buy good ratings (Xu et al., 2015) and trustors can be competitors who try to trick one another by spreading wrong information. If trustors expect that information may be faked, they should at best attach less weight to information received from third parties than to own experience. This would presumably also diminish the effects of network embeddedness as well as the incentives to establish it. The models presented here could be applied to a broader class of social settings if they were extended to account for the possibility of information manipulation. In addition, studying what can prevent the manipulation of information is of value in itself (see Resnick et al., 2006) and could also shed light on the range of applicability of the studies presented here.

We, furthermore, kept the complexity with respect to the structure of information sharing networks to a minimum, considering only "empty networks" (no information sharing between any pair of trustors) and "complete networks" (immediate information sharing between *all* pairs of trustors). For some applications, such as an online trader's consideration whether to invest in setting up an online reputation system, these seem to be the only relevant alternatives. In other contexts, such as the sharing of information about a carpenter in a town, it is more plausible that some pairs of trustors share information while other pairs of trustors are not connected, or only indirectly. Modeling such structured information sharing networks is a logical step to proceed from the studies presented here. Buskens (1998) uses computer simulations to investigate the effects of static, structured network on trust and trustworthiness. Sim-

ilar simulation approaches could allow deriving predictions on structural properties of word-of-mouth networks that are formed endogenously to mitigate trust problems.[4]

We also assumed that information sharing networks are established once and then remain unchanged over time. These networks might, however, evolve over time. For example, once Alice is convinced of the trustworthiness of her cleaning person, she might not maintain contact with her informants. The small but growing literature on the co-evolution of networks and behavior provides frameworks for modeling concurrent dynamics of networks and behavior in interactions such as in Trust Games (Corten, 2014; Perc & Szolnoki, 2010; Skyrms & Pemantle, 2000). This literature typically uses computational methods and assumes backward-looking actors. Our game-theoretic studies complement this literature by providing an investigation of network formation and network effects under the assumption of fully rational actors.

Future empirical studies could extend on the experiment reported in Chapter 5 along the lines sketched for further theoretical research. Another logical step ahead is to study investments in and returns on network embeddedness outside the laboratory (inspiration for such research could come from the empirical studies on returns on embeddedness reviewed in Buskens & Raub, 2013). Here, we confine ourselves to pointing out directions for future experimental research that follow more directly from the results of our experiment.

In the experiment reported in Chapter 5, subjects had no especially high tendency to establish embeddedness in the conditions in which the observed returns on embeddedness (the increase in earnings in Trust Games due to embeddedness) were largest. It remains, therefore, somewhat unclear whether they established embeddedness in anticipation of the returns on embeddedness. Put more sharply, our experiment offers no solid evidence that the observed investments in establishing network embeddedness were pledged *because* network embeddedness can resolve trust problems. We cannot strictly rule out the possibility that subjects paid for establishing information exchange, for example, out of curiosity or to reduce boredom. In hindsight, it would have been wise to include a control condition in which there is no trust problem (similar as in the experiment on reputation cascades reported in Chapter 6). By including such a control condition, future experiments could obtain more compelling evidence for the hypothesis that trust problems induce actors to establish social structures for information sharing.

Our findings offer little evidence for the prediction that the degree to which embeddedness promotes trust and trustworthiness varies in an inverted U-shape over the

[4]To keep models tractable, complexity should be added step by step. Still, it is conceivable that interesting insights could be obtained from allowing for noise in information transmission and structured networks simultaneously. If information is less reliable, actors might attempt to tap several sources of information and, therefore, form denser networks.

size of the trust problem. We attempted to manipulate the size of the trust problem by manipulating the probability that a trustee has no monetary incentive to abuse trust. However, this manipulation worked imperfectly; subjects did not perceive the trust problem as distinctly larger in one condition than in another condition. Therefore, we should not yet dismiss the hypothesis that embeddedness effects may be small if the trust problem is very small or very large. Further tests of this hypothesis could rely on manipulations of interaction durations or payoffs in the Trust Games.

1.5.2 Reputation cascades

The study presented in Chapter 6 is the first to demonstrate the endogenous emergence of inequality in exchange volumes in situations of trust with information sharing. More research is needed to find out under what circumstances reputation cascades occur and what their consequences are. We first point out some aspects in which future models and empirical tests of reputation cascades could be made more realistic and, thereby, shed light on the conditions under which reputation cascades can be expected.

First, if trustors cannot always choose their partners and face search costs, cascading may not occur. Someone with a broken car may have no choice but to go to the next best car mechanic and someone who heard from one person that a specific car mechanic is trustworthy may choose not to search further to find the car mechanic that most people have good experiences with. Empirical research outside the laboratory will be needed to investigate to what extent these issues undermine the potential for cascading in specific domains. It seems plausible that these forces countering cascades are less strong in the digital economy where search costs are low.

Second, our theory and experimental test ignored that a trustee who lacks a reputation may attempt to attract trustors by offering a lower price, at least in economic exchange. However, we conjecture that a trustee with an established reputation can lower his price such that trustors still prefer his offer over taking the risk of trying an untested trustee.[5] Tentative support for this conjecture and inspiration for experimental tests in this direction can be found in the studies of Kollock (1994) and Yamagishi et al. (1998). They show that trustors are willing to forgo potentially more lucrative offers of untested parties in favor of the relative safety of exchanging in ongoing relationships with proven partners.

Third, we assumed that exchanges occur strictly sequentially and that information is available without any delay. However, before one house-owner detects flaws

[5] In a different context, the literature on "limit pricing" shows that the pricing strategy of a monopolist can prevent market entry (see Milgrom & Roberts, 1982).

in a carpenter's work, another house-owner may have to contract a carpenter. Simultaneously occurring exchanges (Huck et al., 2012), time lags in discovering fraud, and delays in the transmission of information (Manapat & Rand, 2012) might lead a group of trustors who share information to lock in on more than one single trustee.

Fourth, we conjecture that tendencies for choosing popular trustees may be even stronger than in our experiment in settings where information on past dealing is not always accurate, e.g., because it is possible for a trustee to fake a few good ratings. The literature on reputation building under imperfect monitoring (Cripps et al., 2004) provides insights that could prove useful for developing theoretical models that account for the possibility of inaccurate information.

In addition to providing more realistic models and tests of reputation cascades, future research should investigate their consequences. Our experiment indicates that trustors tend to avoid unequal outcomes if they do not have to fear abuse of trust. This suggests that the inequality emerging in situations where trust is an issue is an undesired consequence of the reputation mechanism. The emergent inequality could be undesired even if arbitrary inequality among equally trustworthy trustees is not judged as socially undesirable per se. The reputational advantage gives a market leader a monopoly position that can potentially be abused e.g., by asking high premiums (although, the argument on price competition above suggests that the threat of entrants may discipline a market leader). Inefficiencies could also occur when trustors differ with respect to their preferences for different trustees, e.g., due to variation in geographical proximity between trustor-trustee pairs. Trustors might then ignore their personal preferences in their attempts to avoid the risk of abuse.

Finally, our theory suggests that some fragmentation in reputation systems may lead to high efficiency in trust interactions but prevent excess inequality. The experimental results provide a vague indication of this. Experiments with larger groups—potentially run online—are needed to investigate whether reputation systems can have an ideal size that simultaneously prevents opportunistic tendencies as well as the emergence of overly large inequalities.

1.6 Organizational remarks

Chapters 2 through 6 are written as stand-alone articles, published in or submitted to scientific journals. Each of these chapters is self-contained. This entails that there is some overlap between these chapters as well as between these chapters and this general introduction and synthesis. Cross-references between these chapters are included as references to the published papers and working papers. Appendices to Chapters 2 through 6 are collected at the end of the book and contain additional theo-

retical analyses, proofs for propositions, and additional information on the laboratory experiments and data analyses.

Chapter 2

The rationality of social structure: Cooperation in social dilemmas through investments in and returns on social capital[1]

Abstract: We develop a game-theoretic model of how social structure in the sense of a network of relations between actors helps to mitigate social dilemmas. We simultaneously endogenize the network by modeling actors' incentives to establish the network. Since the network of relations helps to mitigate social dilemmas, it constitutes social capital. We thus analyze investments in and returns on social capital in social dilemmas and characterize the value of social capital. Our model covers a class of social dilemma games including the Trust Game, the Investment Game, the Prisoner's Dilemma, the two-actor Public Goods Game, and others.

[1] A slightly different version of this study is published as Raub, W., Buskens, V., & Frey, V. (2013). The rationality of social structure: Cooperation in social dilemmas through investments in and returns on social capital. *Social Networks, 35*(4), 720–732. Raub wrote main parts of the manuscript and initiated theory development; Buskens and Frey wrote substantial parts of the manuscript and contributed to the theory development. We thank Rense Corten, Arnout van de Rijt, and an anonymous reviewer for useful suggestions.

2.1 Introduction

We model how social structure in the sense of a network of relations between actors helps to mitigate social dilemmas. We simultaneously endogenize the network by modeling actors' incentives to establish the network in the first place. Since the network of relations helps to mitigate social dilemmas, it constitutes social capital: The network helps to achieve ends—cooperation in social dilemmas—that could not be achieved without the network (e.g., Coleman, 1988, 1990, Chap. 12). Our model highlights the rationality of social structure in three respects: First, we derive conditions for individually rational investments in social capital through incurring costs that allow for establishing the network. We conceive of individual rationality as incentive-guided and goal-directed behavior, made precise as equilibrium behavior in noncooperative games. Second, cooperation in a social dilemma is collectively rational in the sense of Pareto-optimality (see Rapoport's, 1974, distinction between individual and collective rationality). Third, the network of relations allows for individually rational cooperation in social dilemmas. Thus, cooperation in social dilemmas can be a return on investments in social capital.

Social dilemmas are situations such that actors face individual incentives to defect by behaving opportunistically, thus impairing their partners. If actors follow these incentives, however, they are worse off than had they cooperated. Hence, social dilemmas represent situations in which individual rationality is at odds with collective rationality (e.g., Buskens & Raub, 2013; Kollock, 1998). While "social dilemma" is a label commonly used in sociology and social psychology, "problem of collective action" and "tragedy of the commons" are labels often used in political science, and "public goods problem" is used in economics (see Ledyard, 1995, p. 122).

Social dilemmas are related to Parsons's (1937) problem of order. In Hobbes's (1991 [1651], Chap. 13) "naturall condition of mankind," actors are interdependent in a world of scarcity, while binding and externally enforced contracts are unfeasible. Actors may thus end up in the "warre of every man against every man." In that situation, the life of man is "solitary, poore, nasty, brutish, and short" and everybody is worse off compared to a peaceful situation. This is a social dilemma among many actors. Parsons posed the challenge to specify conditions such that individually rational actors solve the problem of order, in his view "the most fundamental empirical difficulty of utilitarian thought" (1937, p. 91). Coleman (1964, pp. 166–167) realized the challenge in his early contribution to rational choice sociology and formulated it even more radically:

> "Hobbes took as problematic what most contemporary sociologists take as given: that a society can exist at all, despite the fact that individuals are

> born into it wholly self-concerned, and in fact remain largely self-concerned throughout their existence. Instead, sociologists have characteristically taken as their starting point a social system in which norms exist, and individuals are largely governed by those norms. Such a strategy views norms as the governors of social behavior, and thus neatly bypasses the difficult problem that Hobbes posed [...] I will proceed in precisely the opposite fashion [...] I will make an opposite error, but one which may prove more fruitful [...] I will start with an image of man as wholly free: unsocialized, entirely self-interested, not constrained by norms of a system, but only rationally calculating to further his own self interest."

While it is part of the sociological folklore that Parsons' challenge focuses on how rational choice social research can cope with empirically observed cooperation in social dilemmas, Durkheim's similar argument in his analysis of the division of labor in society (1973 [1893], book I, Chap. 7) is less known. Durkheim recognizes that economic exchange often deviates from what is assumed in neo-classical models of spot exchange on perfect markets. He argued that the governance of exchange exclusively via bilateral contracts requires that the present and future rights and obligations of the partners involved in the transaction are specified explicitly for all circumstances and contingencies that might arise during and after the transaction. Anticipating much of the modern economic and game-theoretic literature on incomplete and implicit contracts, Durkheim pointed out that such purely contractual governance of economic exchange is problematic: Typically, many unforeseen or unforeseeable contingencies could or actually do arise during or after a transaction. To negotiate a contract explicitly covering all these contingencies is costly or even unfeasible. Likewise, renegotiations in the case that contingencies arise are costly (for similar arguments on the limits of contractual governance, see Weber, 1976 [1921], p. 409 in his sociology of law). Such renegotiations may offer incentives for opportunistic behavior since an unexpected contingency will often strengthen the bargaining position of one partner while weakening the position of the other. Hence, according to Durkheim, mutually beneficial economic exchange involves a social dilemma and the problem emerges of how to explain mutually beneficial exchange between individually rational actors.

Our analysis applies to a class of games modeling social dilemmas between two actors. Throughout, we use social and economic exchange to exemplify. Exchange may involve different kinds of social dilemmas, with different games lending themselves as formal models. Consider exchange problems that result from one-sided incentives for opportunistic behavior. In social exchange (Blau, 1964), Ego helps Alter today, trusting that Alter will help Ego tomorrow. If Alter indeed provides help tomorrow, both Ego and Alter are better off than without helping each other. However, Alter faces an

incentive to benefit from Ego's help today without providing help himself tomorrow. Anticipating this, Ego might not provide help in the first place. In economic exchange between a buyer and a seller (e.g., Dasgupta, 1988), the buyer may be insufficiently informed on the quality of a good and thus she has to trust the seller that he will sell a good product for a reasonable price.[2] The seller can honor trust by indeed selling a good product for a reasonable price. Buyer and seller are then both better off compared to the situation without a transaction. Trust thus increases efficiency in economic exchange (Arrow, 1974). However, the seller could also abuse trust by selling a bad product for the price of a good one, thus securing an extra profit. The buyer is then worse off than had she decided not to buy. In both examples, only one actor has incentives for opportunistic behavior: Alter in the case of social exchange and the seller in the case of economic exchange. The other actor—Ego and, respectively, the buyer—can foresee this and may thus avoid entering the exchange. The Trust Game (e.g., Kreps, 1990a), to be introduced below, is a standard model for such exchange problems and is included in the class of games to which our analysis applies.[3]

As a variant that comes close to Durkheim's scenario, consider one-sided incentive problems in exchange that are more complex in the sense that buyer and supplier do not make binary choices but have a larger set of feasible actions. For example, the buyer does not choose between "buying" and "not buying." Rather, she chooses how much time and effort the exchange partners allocate to writing an externally enforceable contract that reduces the seller's opportunities for exploiting the buyer but likewise reduces the gains from trade. Conversely, the seller chooses the degree to which he behaves opportunistically by not sharing these gains. If the buyer anticipates "much" opportunism of the seller, she may prefer an extensive but costly contract that reduces the sellers opportunities for exploiting the buyer. Both actors, however, would be better off without costly contracting and with larger and shared gains from trade. Below, we will see that the Investment Game (Berg et al., 1995) models such features. Our analysis applies to this game, too.

Incentive problems in exchange are often two-sided. For example, the seller has an incentive to sell a bad product for the price of a good one, while the buyer has an incentive to delay payment. Also, both actors may choose the degree to which they cooperate or, respectively, defect. The Prisoner's Dilemma can be used as a model for two-sided incentive problems in exchange with binary choices for buyer and

[2] To facilitate identifying the two actors, we use female pronouns for the buyer and male pronouns for the seller.

[3] In principle, the incentive problems for buyer and seller could be reversed so that only the buyer has an incentive for opportunistic behavior such as delaying payment. Again, the Trust Game could be used as a formal model.

seller (Hardin, 1982), while the two-actor Public Goods Game is a model for two-sided incentive problems and actors being able to choose their degree of cooperation or defection. Both games are included in our analysis.

Until now, we have implicitly assumed that the actors are involved in an "isolated encounter," similar to Granovetter's (1985) atomized exchange on perfect markets. First, there are no previous or future interactions between buyer and seller that could affect their behavior in the focal exchange. For example, an actor cannot sanction the partner's behavior in the focal exchange through own behavior in future interactions with the partner. Second, there are no previous or future interactions with third parties that could affect behavior in the focal exchange. For example, the buyer has no links with other clients of her seller with whom she can exchange information about the seller's behavior. Thus, the seller's behavior in the focal exchange cannot have repercussions for the behavior of other clients in the future. Often, however, exchange is embedded in the sense that buyer and seller interact repeatedly and that they are involved in a network of relations with third parties. For example, the buyer may have repeated opportunities to purchase goods from the seller. Then, the amount of costly contractual safeguards the buyer requires for future exchanges may depend on the seller's behavior in the focal exchange and this may induce the seller to abstain from opportunism by sharing the gains from trade. Or, buyer and seller are part of a network in the sense of a consumer organization or a buyer association that keeps track of transactions and distributes information on behavior among buyers and sellers, thus establishing an information network. Then, the behavior of another buyer in a future exchange with the seller may depend on the seller's behavior in the focal exchange and this may again induce the seller to abstain from opportunism.

Assume that buyers and sellers are part of a network that allows for information exchange. Such a network then indeed constitutes social capital for buyers and sellers if the network helps to achieve ends, namely, gains from trade that could not be achieved without the network. Put differently, there are returns on social capital in the sense that the gains from trade make buyers and sellers better off compared to the situation without exchange or compared to the situation where costly contractual safeguards diminish the gains from trade. Research on "games on networks" (Goyal, 2007, Chap. 3; Jackson, 2008, Chap. 9) provides models of such effects of social capital. In this research, the network is assumed as given and exogenous and network effects on individual behavior are analyzed. Raub & Weesie (1990) analyze what is likely to be the first game-theoretic model of network effects for a social dilemma, namely, the Prisoner's Dilemma. Buskens (2002) provides models of network effects for trust problems. Buskens & Raub (2013) survey the theoretical and empirical literature on conditions under which networks facilitate cooperation in social dilemmas.

Given that there are returns on social capital residing in networks, it follows that actors have incentives to consider their networks not as given but to actively establish, maintain, or sever links with other actors with an eye on returns that can be expected. Buyers can search for other buyers who transact with the same seller and can thus establish or enlarge a network for information exchange about sellers. Research on such investments in social capital drops the assumption of an exogenously given network. Rather, links between actors are modeled as being endogenous. The game-theoretic literature on "strategic network formation" has rapidly developed in recent years (see Dutta & Jackson, 2003; Jackson & Zenou, 2013 for collections of important contributions; for textbooks: Goyal, 2007; Jackson, 2008; Vega-Redondo, 2007; see also Snijders, 2013). Buskens & Van de Rijt (2008) show how this approach can be used in sociology. These models allow for an analysis of how incentive-guided and goal-directed actors establish, maintain, or sever links with others and for an analysis of the macro-level properties of the emerging networks.

If there are returns on social capital because it mitigates social dilemmas and if actors can invest in social capital through establishing, maintaining, or severing links with others, an obvious next step is the simultaneous analysis of both phenomena—investments in and returns on social capital—in an integrated model. The aim is a model that allows for deriving implications for actors' strategic network formation as well as implications for network effects on behavior in social dilemmas. Flap (2004) provides a clear outline of a research program for an integrated analysis of investments in and returns on social capital. However, models that actually implement such a program are very scarce. Examples of such models in sociology and economics include work on cooperation in dynamic networks by Eguíluz et al. (2005), Pujol et al. (2005), and Vega-Redondo (2006; see Corten, 2014, Chap. 3.1 for a more detailed survey). Some research in informatics (e.g., Carbone et al., 2003) employs similar terminology but considers exclusively applications to computer networks and online interactions, using rather different kinds of models.

In this chapter, we provide a game-theoretic model for the simultaneous analysis of network formation, i.e., investments in social capital, and effects of networks on behavior in social dilemmas, i.e., returns on social capital. We also characterize the value of social capital. Our model and its implications are robust in the sense that they cover a class of social dilemmas between two actors, including paradigmatic examples of such dilemmas such as the Trust Game, the Investment Game, the Prisoner's Dilemma, the Public Goods Game, and others. Also, our model assumes full strategic rationality with respect to behavior in social dilemmas as well as with respect to network formation. Note that the assumption of full strategic rationality is rather uncommon in models of strategic network formation. To keep the analysis tractable,

most of these models assume "myopic best-reply behavior:" When deciding about establishing, maintaining, and severing links, actors 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, the analysis is tractable under the assumption of full strategic rationality. We thus contribute to theoretical pluralism in research on strategic network formation.

By focusing on individually rational investments in networks and on network effects on individually rational behavior, we use the theoretical strategy that Coleman (1988, p. S97) described as the attempt "to import the economist's principle of rational action for use in the analysis of social systems proper, including but not limited to economic systems, and to do so without discarding social organization in the process." This is similar to Granovetter's (1985) program of incorporating the effects of "embeddedness"—his label for a network of relations—in the analysis of economic action (see Coleman, 1988, p. S97; Raub & Weesie, 1990, pp. 627–629 for a more comprehensive discussion). We likewise follow Coleman (1988, S105ff; see also Coleman, 1990, Chaps. 11 and 12) in focusing on a specific feature of social structure, namely, network closure. In terms of economic exchange, closure implies that if a supplier is related through transactions with two buyers, then the buyers are likewise related, for example, through information exchange. Coleman and Granovetter highlighted the effects of such close-knit structures for the existence of effective norms as well as for trust and trustworthiness. Our model likewise shows that closure, compared to isolated encounters and atomized exchange, facilitates cooperation in exchange and, more generally, cooperation in social dilemmas. While Coleman as well as Granovetter assumed such a structure as given and exogenous, we extend the analysis by modeling not only effects but also the emergence of closure as a result of rational action.

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 suggestions for future research.

2.2 The model

2.2.1 Social dilemmas

We consider social dilemma games G that satisfy the following properties.[4] First, G is a game with two actors, one actor in role 1 and the other in role 2. We use $i, j = 1, 2$ $(i \neq j)$ to denote actors as well as roles. We assume that G does not involve random moves of a pseudo-actor Nature (see Rasmusen, 1994, p. 10). Second, G has a unique subgame perfect equilibrium in pure strategies which we denote as $D = (D_1, D_2)$. This means that each actor's strategy D_i maximizes i's payoff U_i against the other actor's strategy D_j for each situation that may emerge in G (see Rasmusen, 1994, p. 94 for an explicit definition of a subgame perfect equilibrium). Subgame perfection is the basic refinement of the Nash equilibrium concept and a common conceptualization of individually rational behavior in situations with strategic interdependence. We consider subgame perfect equilibria and we refer to them for brevity as "equilibria" in case this does not cause any confusion. We denote actor i's payoff associated with D as $U_i(D) = P_i$. Third, G has a combination $C = (C_1, C_2)$ of pure strategies that is Pareto-optimal and is a strict Pareto-improvement compared to D.[5] Hence, $U_i(C) = R_i > P_i$ for each actor i. Using common terminology, we refer to strategy C_i as "cooperation." Since D is the unique subgame perfect equilibrium in pure strategies, it follows that C cannot be a subgame perfect equilibrium and we assume, third, that C is likewise not a Nash equilibrium that is not subgame perfect. That is, C_i is not a best-reply strategy against C_j for at least one actor i. Rather, for at least one actor i there is a pure strategy $B_i \neq C_i$ that maximizes i's payoff against C_j. Let T_i denote player i's best-reply payoff against C_j. We thus have $U_i(B_i, C_j) = T_i > R_i$ for at least one actor i. Note that often, and also always in our examples below, $B_i = D_i$ for actors with $T_i > R_i$. It is useful to assume, fifth, that the strategies D_i are minimax strategies. Since D is an equilibrium, the strategies D_i are likewise maximin strategies. In less technical terms, D_i maximizes i's payoff under the assumption that j tries to make i's payoff as low as possible (maximin) and D_i is i's strategy to keep j's payoff as low as possible (minimax; see Rasmusen, 1994, pp. 126–127 for explicit definitions). "Defection" will refer to any deviation from C_i.

[4] We focus on core features of the game-theoretic model. Likewise, we offer intuitive sketches of game-theoretic terminology rather than explicit definitions. See a textbook such as Rasmusen (1994) for a more detailed overview of game-theoretic concepts and assumptions that are employed in our analysis.

[5] By assuming that C as well as D are combinations of pure (rather than mixed) strategies in G and by avoiding random moves of Nature in G, we ensure that the crucial payoffs in G used in the proofs are certain and do not depend on randomization. This hardly jeopardizes the generality of the formulation of G and does not substantially affect the conclusions. However, it helps to avoid unnecessary complications (see Appendix A).

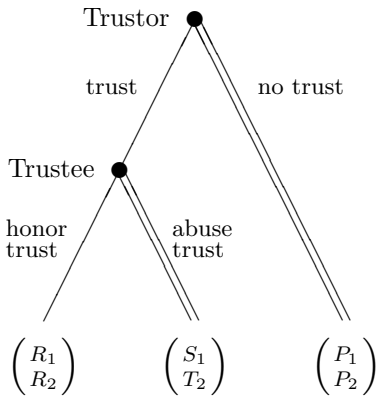


Figure 2.1: The Trust Game ($S_1 < P_1 < R_1$, $P_2 < R_2 < T_2$); double lines indicate behavior in the unique subgame perfect equilibrium.

In general, actor i can deviate in more ways from C_i than exclusively by playing B_i or D_i. Similarly, actors may achieve a strict Pareto-improvement compared to D in more ways than exclusively by playing the strategy combination C. We reserve the term "cooperation," however, to refer to play of the Pareto-optimal strategy combination C. This finalizes the characterization of the "stage game" that we later embed in a repeated game.

Canonical examples of social dilemma games with two actors satisfy these properties. Figure 2.1 depicts the standard Trust Game (e.g., Dasgupta, 1988; Kreps, 1990a; see also Coleman, 1990, Chap. 5) that models one-sided incentive problems in economic and social exchange. In this game, actor 1 is the trustor and actor 2 is the trustee, with D_1 = no trust, D_2 = abuse trust, while C_1 = place trust and C_2 = honor trust. Note that only actor 2 has an incentive for defection in the sense that $U_2(C_1, D_2) = T_2 > U_2(C_1, C_2) = R_2$. The trustor's best-reply strategy against cooperation of the trustee (C_2 = honor trust) would be to cooperate herself by playing C_1 = place trust. The trustor's equilibrium strategy D_1 = no trust implies protection against the trustee's opportunism rather than an attempt to increase the trustor's payoff by exploiting cooperation of the trustee.

Compared to the Trust Game, the Investment Game Berg et al. (1995) is a more complex model of a trust problem and is likewise an example of a social dilemma game G. Again, actor 1 is the trustor and actor 2 is the trustee. While the actors make binary choices in the Trust Game, the trustor can choose the degree to which she trusts the trustee in the Investment Game and the trustee can choose the degree to which he honors trust. More precisely, each actor has an endowment E. The trustor

chooses an amount e of her endowment to send to the trustee ($0 \leq e \leq E$). In our example of economic exchange, sending a larger e would mean that the buyer requires less extensive and thus less costly contractual safeguards. This "investment" e of the buyer is then multiplied by $m > 1$ and the trustee receives me. The parameter m can be seen as indicating the size of the gains from trade. Subsequently, the trustee chooses an amount g he returns to the trustor, with $0 \leq g \leq E + me$. In terms of our example, the seller decides on how to share the gains from trade. The game ends with the trustor receiving $U_1 = E - e + g$ and the trustee receiving $U_2 = E + me - g$. While e indicates how much the trustor trusts the trustee, g indicates how trustworthy the trustee is. It is easily seen that the Investment Game has a unique subgame perfect equilibrium such that the trustee would never return anything, while the trustor would send nothing. Thus, D_1 is the strategy to choose $e = 0$, while D_2 is the strategy to choose $g = 0$ for all e. Hence, for the Investment Game, $U_i(D) = P_i = E$ for both actors.

In the Investment Game, both actors are better off than in the subgame perfect equilibrium and the outcome is Pareto-optimal if the trustor sends everything and the trustee chooses g^* so that both actors end up with equal payoffs. These are the cooperative strategies C_1 and C_2, respectively. If the trustor indeed chooses C_1, i.e., $e = E$, then C_2 implies that the trustee returns $g^* = E(m + 1)/2$.[6] Thus, $U_i(C) = R_i = E(m + 1)/2 > P_i = E$ for both actors. Like in the Trust Game, only actor 2 has an incentive for defection in the Investment Game. The trustee's best-reply against C_1 is to return nothing so that $U_2(C_1, D_2) = T_2 = E + mE > U_2(C_1, C_2) = R_2$, while the trustor's best-reply strategy against C_2 is to cooperate herself by sending the complete endowment E. The trustor's equilibrium strategy D_1 to send nothing again implies protection against the trustee's opportunism rather than an attempt to increase the trustor's payoff by exploiting cooperation of the trustee.

Another paradigmatic example of a social dilemma with two actors that satisfies our properties is the Prisoner's Dilemma as depicted in Table 2.1. In the Prisoner's Dilemma, actor 1 represents the row player, while actor 2 represents the column player. In this game, both actors have an incentive to exploit cooperation of the other actor, since $T_i > R_i$ for both actors. Note, too, that our notation for payoffs in social dilemma games G is derived from common notation for payoffs in the Prisoner's Dilemma, with T_i "temptation," R_i "reward," P_i "punishment," and S_i "sucker's payoff" (e.g., Axelrod, 1984).

[6] Note that g^* could be defined differently, since the trustee could divide the sum of what he receives from the trustor (multiplied by m) plus his own endowment in many ways that imply Pareto-optimality as well as a strict Pareto-improvement compared to the outcome if the trustor sends nothing. Our characterization of g^* is an intuitively plausible one, based on distributive justice. The implications of our model are valid for other characterizations of g^*, too. Furthermore,

Table 2.1: The Prisoner's Dilemma ($S_i < P_i < R_i < T_i$); the shaded cell indicates the unique equilibrium.

		Actor 2	
		Cooperation (C_2)	Defection (D_2)
Actor 1	Cooperation (C_1)	R_1, R_2	S_1, T_2
	Defection (D_1)	T_1, S_2	P_1, P_2

Still another example of a social dilemma game G satisfying our properties is the two-actor version of the Public Goods Game, a variant of the Prisoner's Dilemma in the sense that both actors can choose the degree to which they cooperate or defect. Like in the Investment Game, each actor has an endowment E. They simultaneously choose a contribution e_i to the public good ($0 \leq e_i \leq E$). The joint contribution $e = e_1 + e_2$ is then multiplied by m ($1 < m < 2$) and the gains from trade are distributed equally so that $U_i(e_1, e_2) = U_i(e) = E - e_i + me/2$.[7] Contributing nothing maximizes each actor's payoff, irrespective of the behavior of the other actor. The game thus has a unique subgame perfect equilibrium such that both actors contribute nothing. Hence, $D_i = 0$ and $U_i(D) = P_i = E$ for both actors. Both actors are better off than in the equilibrium and the outcome is Pareto-optimal in the Public Goods Game if both actors contribute their complete endowment, choosing $C_i = E$. Thus, for the Public Goods Game, $U_i(C) = R_i = mE > P_i = E$ for each actor. Both actors have an incentive for defection in the Public Goods Game in the sense that their best-reply strategy against C_j is to contribute nothing so that $U_i(D_i, C_j) = T_i = E + mE/2 = E(1 + m/2) > U_i(C) = R_i = mE$.

These examples show that our properties for a social dilemma game G are satisfied by standard models of such games. Indeed, one can easily show that variants of our examples likewise fulfill these properties. Such variants include the sequential Prisoner's Dilemma and the sequential Public Goods Game (actor j moves after actor i, knowing what i has chosen). Still another example is Rosenthal's (1981) Centipede Game, of which the Trust Game is a special case.

---

it is not necessary to assume that trustor and trustee have the same endowment.

[7] Similar to the Investment Game, the gains from trade could be distributed in other ways that imply Pareto-optimality as well as a strict Pareto-improvement compared to the equilibrium outcome. Note, too, that we could again assume different endowments for the actors.

2.2.2 Embedding the social dilemma game G in a network of actors and a repeated game Γ

We now embed G in a network of actors and in a repeated game Γ. The repeated game is played by $N \geq 3$ actors in rounds $t = 1, 2, \ldots$ and likewise has an initial round 0 (note that we continue—with mild abuse of notation—to use i and j for indicating actors as well as roles in G). We first sketch the structure of the game for rounds $1, 2, \ldots$. In each of those rounds, pairs of actors play G. Each actor always plays in the same role, either role 1 or role 2. We assume that N_1 actors play G in role 1, while N_2 actors play in role 2, with $N_1 + N_2 = N$. In each round $1, 2, \ldots$ of Γ, each actor in role 1 plays G once with each actor in role 2. Thus, each actor in role 1 plays N_2 games in each round $1, 2, \ldots$, while each actor in role 2 plays N_1 games, with $N_1 N_2$ being the total number of games G played per round. The $N_1 N_2$ games per round are played one after the other. To avoid trivialities, we assume that $N_j \geq 2$ if actors in role i have an incentive to defect ($T_i > R_i$).[8] To keep things simple, we assume homogeneity in the sense that actors in the same role have the same payoff function.[9] Actors in different roles may have different payoff functions, i.e., we allow for $T_i \neq T_j$, $R_i \neq R_j$, and $P_i \neq P_j$.

Actors need not play with their partners in the same sequence in each round. However, we do assume that each actor i in the role with the highest value for $(T_i - R_i)/(T_i - P_i)$ plays his games with the actors in the other role in subsequent games within each round. Thus, when playing his first game with an actor in role j in a given round, i plays the games with all other $N_j - 1$ actors j immediately afterwards in that round. We also assume that actors know this rule. We do not need further assumptions on how games are ordered within a round.[10] 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 of these rounds with probability $1 - w$. Thus, actors play indefinitely often repeated social dilemma games G with each other and we can apply standard theory for indefinitely often repeated games (e.g., Friedman, 1986).

Using the Trust Game as an example, rounds $1, 2, \ldots$ of Γ could be interpreted as business periods (for example, market days) in which $N_1 \geq 2$ buyers in the role of the trustor purchase goods from one or more sellers in the role of the trustee. More precisely, in each period, each buyer decides whether or not to purchase a good from

[8] This implies, for example, that there are at least two trustors if G is the Trust Game or the Investment Game as well as that $N_i \geq 2$ for $i = 1, 2$ and $N \geq 4$ if G is the Prisoner's Dilemma or the Public Goods Game.

[9] Strictly speaking, we only need to assume that actors who have an incentive to defect in the sense that $T_i > R_i$ have the same payoff function.

[10] See Appendix A for a more detailed discussion of our assumptions on the ordering of games within rounds.

each of the sellers. With probability $1 - w$, each period is the final one in the sense that all sellers stop business due to some exogenous contingency beyond their own control (for example, new competitors enter the market who offer superior products). In Axelrod's (1984) apt formulation, w represents the "shadow of the future:" With increasing w, actors' long-term incentives increase.

2.2.3 Investments in social capital

Rounds $1, 2, \ldots$ of Γ are preceded by round 0. In that round, at costs, actors can establish links between them that allow for information exchange. Thus, in round 0 actors can invest in social capital. Assuming that the underlying social dilemma is a trust problem between buyers and sellers, 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 sellers are then due to the distribution of information on the behavior of market participants by the organization.

Without investments in round 0, actors are only informed on the history of their own games. More precisely, consider game G that actor i is playing with j in round $t \geq 1$ of Γ. Actor i is then informed that no investments have been made in round 0 and is also informed about what happened in all games G in which i has been involved in rounds $1, \ldots, t-1$ as well as what happened in all games G in round t in which i played with other actors $j_1, j_2, \ldots, \neq j$ before playing with j. However, i has no information about what happened in any earlier game G in which i has not been involved. In particular, i has no information about j's behavior in previous games G in rounds $1, \ldots, t$ with other actors $i_1, i_2, \ldots \neq i$.[11] Thus, in the case of a trust problem, a buyer is informed about her own behavior and the behavior of all sellers in previous interactions with her, but she is not informed about the behavior of sellers in transactions with other buyers.

With investments in round 0, actors are not only informed about the history of their own games but also about the history of all other games. More precisely, consider game G that actor i is playing with j in round $t \geq 1$ of Γ. Actor i is then informed that investments have been made in round 0. Furthermore, i is informed about what happened in all games G in which i has been involved in rounds $1, \ldots, t-1$ as well as what happened in all games G in round t in which i played with other actors $j_1, j_2, \ldots \neq j$. In addition, i also has information about what happened in all earlier

[11] Technically speaking, we refer to information on behavior in the sense of information on moves in previous games that is (or is not) available for actors. Since the strategies for G on which we focus are all pure strategies, moves are not the result of randomization and therefore it is no problem for our analysis that actors cannot observe how others randomize but can only observe the move that is actually chosen.

games G in which i has not been involved. In particular, i has information about j's behavior in previous games G in rounds $1, \ldots, t$ with other actors $i_1, i_2, \ldots \neq i$. Thus, in the case of a trust problem, a buyer is informed about her own behavior and that of all sellers with whom she transacts in all previous trust problems, as well as about the behavior of sellers in transactions with other buyers and the behavior of those other buyers.

Note that a complete network of information links is established among the N actors through investments in round 0 in the sense that each actor is informed not only on the behavior in the actor's own games but also about behavior in all other previous games between all other actors. In principle, we could consider a scenario with less information links and with less information available after investments in round 0. Our results in the next section depend, strictly speaking, only on the assumption that after investments in round 0 the behavior of an actor i who has incentives to defect ($T_i > R_i$) in a game G with actor j becomes known also to all other partners $j_1, j_2, \ldots \neq j$ of i.

For analytical tractability, we assume that all information—if available—is also 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 a partner has defected in an earlier game G.

It is a noteworthy feature of our model that "social capital" is clearly defined. It refers, first, to the links between actors that allow for exchange of information. Second, these links increase actors' sanction potential because they can condition their behavior in games with a partner not only on the previous behavior of the partner in their own games but also on previous behavior of the partner in games with other actors, thus providing more incentives for actors to take into account the long-term effects of their present behavior for future interactions. In other words, the sanction potential of other actors with whom actor i is connected is now also valuable for i.

The total costs of establishing the information network in round 0 and thus the total costs of investments in social capital are assumed to be $c > 0$. 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 social capital in 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 N_i-institution, only actors in role i can invest. Thus, in round 0, actors in role i decide simultaneously and independently about their individual investment. Each actor i can either invest c/N_i or decide not to invest. If each actor i invests c/N_i, the information network is established. Conversely, if at least one actor i decides not to invest, the information network is not established. In that case, actors i who had been willing to invest do not lose their own investment c/N_i. This corresponds to common assumptions in models of network formation, namely, two-sided link formation (a link is only established if both actors wish to be linked) with shared costs of links (Jackson, 2008, Chap. 6).

Our second and alternative institutional rule, the N-institution, is a variant of the first one and stipulates that actors in both roles can invest. Thus, in round 0, actors decide simultaneously and independently about their individual investment. Each actor can either invest c/N or decide not to invest. Again, if each actor invests c/N, the information network is established. Conversely, if at least one actor decides not to invest, the information network is not established and actors who had been willing to invest do not lose their own investment c/N.

2.2.4 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 $1, 2, \ldots$. This implies, for example, that actor i's payoff is

$$\begin{aligned} U_i &= -\frac{c}{N} + N_j R_i + w N_j R_i + w^2 N_j R_i + \ldots + w^{t-1} N_j R_i + \ldots \\ &= -\frac{c}{N} + \frac{N_j R_i}{1-w} \end{aligned} \tag{2.1}$$

if each actor has contributed c/N to the information network under the N-institution and if i as well as all partners j of i cooperate in all games G in rounds $1, 2, \ldots$. Similarly,

$$\begin{aligned} U_i &= N_j P_i + w N_j P_i + w^2 N_j P_i + \ldots + w^{t-1} N_j P_i + \ldots \\ &= \frac{N_j P_i}{1-w} \end{aligned} \tag{2.2}$$

is actor i's payoff if at least one actor who could have contributed to the costs of establishing the information network has refused to do so and if i as well as all

partners j of i played D in all games G in rounds $1, 2, \ldots$.[12]

We further assume that all actors are informed on the structure of Γ (and thus also on the structure of G), that they know from each other that they have this information, etc. The structure of Γ is thus "common knowledge" (Rasmusen, 1994, p. 44). Finally, we assume that Γ (and thus also G) 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.[13]

2.3 Analysis of the model

Since we assume rational behavior of actors and are interested in specifying conditions for cooperation in social dilemmas, we derive conditions for subgame perfect equilibria of Γ such that actors cooperate throughout all rounds $1, 2, \ldots$ in all games G. We refer to such an equilibrium as a "cooperation equilibrium." This approach is based on the commonly used assumption for the analysis of a repeated game that a cooperation equilibrium can be considered as the "solution" of the game because each actor maximizes his or her own payoff, given the equilibrium strategies of all other actors and because a cooperation equilibrium is associated with higher payoffs for each actor than the situation where D is played throughout all games in all rounds. Given our focus on investments in and returns on social capital, we are specifically interested in equilibria such that actors cooperate after the establishment of the network in round 0 while D is played if the network is not established. To specify conditions for such equilibria, we first consider the subgame Γ^- that is played after the network has not been established in round 0. Subsequently, we analyze the subgame Γ^+ that is played after the network has been established.[14] Finally, we analyze conditions such that investments in social capital in round 0 are implied by equilibrium behavior.[15]

[12] 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, pp. 108–110). It would be no problem to include negative time preferences and results would remain robust.

[13] We assume a noncooperative game precisely because we wish to specify conditions such that rational actors will cooperate "endogenously" and without external enforcement in social dilemmas, based exclusively on the embeddedness of the dilemma in a sequence of interactions and the information exchange network between the actors.

[14] Note that, strictly speaking, Γ has $2N_i - 1$ or $N_2 - 1$ subgames Γ^-, depending on the institution for investments, while it has only one subgame Γ^+.

[15] Two earlier papers analyze the model for the special case of the standard Trust Game and for scenarios with a triad comprising one trustee and two trustors (Raub et al., 2014) as well as one trustee and $n \geq 2$ trustors (Raub et al., 2012).

2.3.1 Cooperation based on conditional strategies

In Γ^- as well as Γ^+, cooperation can be the result of equilibrium behavior based on conditional strategies. When playing G with actor i in round t, actor j can condition behavior on information about i's behavior in previous games G, in previous games with j or, in subgame Γ^+, also on previous games with other actors $j_1, j_2, \ldots \neq j$. If j has information that i cooperated in earlier games G, j can reward this by own cooperation. Conversely, if j has information that i defected in earlier games, j can punish this by playing D_j. Thus, with $T_i > R_i$, i has a short-term incentive to defect in each game G but i also has to take into account that defection in a focal game G may imply long-term costs in future games since j or other partners may play D_j in those future games so that i can obtain only $P_i < R_i$ in future games. Anticipating conditional strategies of partners j, actor i thus has to balance short-term incentives $(T_i - R_i)$ and long-term incentives $(R_i - P_i)$. Repeated social dilemmas and, moreover, social dilemmas that are embedded in a network through which actors exchange information can thus have cooperation equilibria based on conditional strategies. Cooperation can then be a result of individually rational behavior of actors who are "enlightened" in the sense that they take long-term effects into account.[16]

Assume that actor i indeed has an incentive to defect $(T_i > R_i)$. We can then specify the conditional strategy for his partners j that is associated with the most attractive reward for i's cooperation and the most severe punishment for i's defection. Such a strategy is commonly known as a "trigger strategy" (e.g., Friedman, 1986). This is the strategy such that j cooperates in games G with i if j has no information that i ever defected and such that j plays D_j in all future games with i as soon as j receives information that i has defected.[17] Obviously, if all actors use trigger strategies, they will cooperate in all games G in all rounds of Γ. Given our assumptions on G, when actors play D they use maximin as well as minimax strategies. This implies that the indefinitely often repeated game G has a cooperation equilibrium if and only if there is an equilibrium that comprises trigger strategies (see Friedman, 1986, Chap. 3 for details). We thus derive conditions for trigger strategy equilibria in Γ^- and Γ^+. Note that the assumption underlying this approach need not be that actors do indeed

[16] Coleman (1964, p. 180) has characterized "enlightened self-interest" when conceiving of "socialization" in a way that contrasts with the common view of "internalization of values," namely, socialization as allowing an actor "to see the long-term consequences to oneself of particular strategies of action, thus becoming more completely a rational, calculating man."

[17] A trigger strategy requires furthermore that j will cooperate in j's first game G with i if j has by then no information about behavior in games of i with other partners $j_1, j_2, \ldots \neq j$ and that j stops to cooperate in games with i as soon as j has defected him- or herself in an earlier game with i or some other actor $i_1, i_2, \ldots \neq i$. Also, j stops to cooperate in games with i as soon as j receives information that any other actor $j_1, j_2, \ldots \neq j$ or $i_1, i_2, \ldots \neq i$ has defected in an earlier game (see Appendix A on why trigger strategies are defined in this specific way).

use trigger strategies. Cooperation equilibria do require the use of a conditional strategy by actor j who interacts with a partner i who has incentives to defect ($T_i > R_i$) but the conditional strategy of j may comprise less severe punishments than implied by a trigger strategy. For example, j may be willing to return to cooperation if i's "punishment period" has been "long enough." Nevertheless, the existence of a trigger strategy equilibrium is a necessary condition for the existence of cooperation equilibria based on conditional strategies that involve less extreme sanctioning. Hence, following a common approach in empirical applications, we assume that cooperation 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 of game-theoretic models of cooperation in social dilemmas).

2.3.2 Analysis of Γ^-

Our first proposition provides the condition for a cooperation equilibrium in a subgame Γ^-, i.e., for a cooperation equilibrium without investments in social capital. When playing games G in Γ^-, each actor i is only informed on own previous games with the partner j but has no information on how j behaved in other previous games with partners $i_1, i_2, \ldots \neq i$.

Proposition 2.1 (Cooperation without a network)**.** Γ^- *has a cooperation equilibrium if and only if* $w \geq w^- := \max(w_1^-, w_2^-)$*, with* $w_i^- := (T_i - R_i)/(T_i - P_i)$ *for* $i = 1, 2$*.*

Appendix A provides sketches of proofs for our propositions.

Note that w_i^- is a measure for the actors' incentives to defect. It is easily seen that $0 \leq w_i^- < 1$ for $i = 1, 2$ and hence also $0 < w^- < 1$. Proposition 2.1 reflects the well-known result that cooperation in an indefinitely often repeated social dilemma game G constitutes equilibrium behavior if and only if the incentives to defect are compensated by a sufficiently large probability w that the game continues. The proposition implies that cooperation is facilitated—in the sense that the condition for a cooperation equilibrium becomes less restrictive—if the short-term incentive $T_i - R_i$ to play B_i decreases, if there are increasing costs $T_i - P_i$ of a conflict in the sense that both actors simultaneously try to exploit the partner's cooperation by playing B_i, if the costs $R_i - P_i$ of playing D compared to mutual cooperation increase, and if the continuation probability w increases. Note that there always exists sufficiently large w as well as sufficiently small $T_i - R_i$ so that the condition in Proposition 2.1 is fulfilled. Note, too, that $w_i^- = 0$ for an actor i without incentives to defect, i.e., $T_i = R_i$. Thus, for games like the Trust Game or the Invest Game, the proposition

implies that the existence of a cooperation equilibrium depends exclusively on the incentives of the trustees and not at all on the trustors' incentives.

2.3.3 Analysis of Γ^+

In order to assess the returns on social capital, we now derive the condition for a cooperation equilibrium in subgame Γ^+. Thus, we consider cooperation equilibria after investments in social capital have been pledged and the network of information links has been established in round 0. Therefore, in Γ^+, each actor i is not only informed on the history of own interactions with j but also on the history of j's interactions in previous games with partners $i_1, i_2, \ldots \neq i$. Proposition 2.2 provides the condition for a cooperation equilibrium if the network of information links is available.

Proposition 2.2 (Cooperation in a network)**.** *Γ^+ has a cooperation equilibrium if and only if $w \geq w^+ := \max(w_1^+, w_2^+)$, with $w_i^+ := (T_i - R_i)/(T_i - P_i + (N_j - 1)(R_i - P_i))$ for $i = 1, 2$.*

In Proposition 2.2, w_i^+ is a measure for the actors' incentives to defect. Again, $w_i^+ = 0$ for an actor i without incentives to defect, i.e., $T_i = R_i$. The comparative static results for the condition in Proposition 2.1 also hold for Proposition 2.2. Moreover, for actors with an incentive to defect, balancing of short-term and long-term incentives now also depends on N_j, the number of partners of i and long-term incentives for defection decrease with an increasing number of partners. This follows from the fact that i's defection can be punished in Γ^+ in future games not only by the partner j against whom i defected but also by other actors with whom i interacts.

It is useful to summarize properties of w_i^+ and w^+ in a separate proposition.

Proposition 2.3 (Properties of w_i^+ and w^+)**.** *The following properties hold for w_i^+ and w^+:*

(1) *$0 \leq w_i^+ < 1$ for $i = 1, 2$ and $0 < w^+ < 1$;*

(2) *for actors in role i with $T_i > R_i$: $\frac{\partial w_i^+}{\partial N_j} < 0$ and $w_i^+ \to 0$ for $N_j \to \infty$;*

(3) *if $T_i > R_i$ for role $i \neq j$, then w^+ weakly decreases in N_j and $w^+ \to 0$ for $N_j \to \infty$;*

(4) *$w^+ < w^-$ and for actors in role i with $T_i > R_i$: $w_i^+ < w_i^-$ for $N_j \geq 2$.*

As (1) of Proposition 2.3 shows, Γ^+ has a cooperation equilibrium for a large enough continuation probability w and for a small enough short-term incentive $T_i - R_i$

to play B_i so that this short-term incentive is sufficiently compensated by the long-term incentives for cooperation in order to avoid future punishment by the partners. Conversely, for a sufficiently small continuation probability w and for a sufficiently large short-term incentive $T_i - R_i$ to play B_i, the network of information links does not suffice to ensure the existence of a cooperation equilibrium. It can be seen in (2) of Proposition 2.3 that for actors in role i with an incentive to defect, w_i^+ decreases when the number of i's partners increases. Thus, $w \geq w_i^+$ for a sufficiently large number N_j of partners of such actors. Similarly, (3) of Proposition 2.3 shows that w^+ becomes smaller and converges to 0 when the number of partners N_j increases for actors in role i with an incentive to defect. Finally, (4) shows that the condition for cooperation equilibria is less restrictive in Γ^+ than in Γ^-.

2.3.4 Returns on social capital: Comparing Γ^- and Γ^+

Since the condition for cooperation equilibria is less restrictive in Γ^+, it follows that there are parameter configurations such that Γ^+ has a cooperation equilibrium while Γ^- has no such equilibrium. This is the case if and only if $w^+ \leq w < w^-$. Under this condition, there are returns on social capital. Moreover, we are now able to specify an upper bound for the value of social capital. 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 cooperation equilibrium, if it exists, will be played in the subgame Γ^+, while D will be played throughout in the subgame Γ^- if Γ^- has no cooperation equilibrium.

Proposition 2.4 (Value of social capital)**.** *Assume that $w^+ \leq w < w^-$ so that cooperation equilibria exist only in Γ^+. The upper bound on the value of social capital is then equal to*

$$\frac{N_j(R_i - P_i)}{1 - w} \text{ for } i = 1, 2.$$

The upper bound on the value of social capital for actors in role i thus increases if the number N_j of partners increases, if the costs $R_i - P_i$ of playing D compared to mutual cooperation increase, and if the continuation probability w increases, as long as the costs $R_i - P_i$ and the continuation probability w are small enough so that Γ^- has no cooperation equilibrium. Obviously, $N_j(R_i - P_i)/(1 - w)$ can also be interpreted as the upper bound for the costs of investments in social capital that a rational actor in role i would be willing to incur.

2.3.5 Investments in social capital

Having analyzed the returns on social capital, we now turn to investments. We do so by specifying conditions for equilibria of Γ that imply investments in round 0 and cooperation in subsequent rounds $1, 2, \ldots$. The conditions are very similar for the N_i-institution, with only actors in role i being able to invest, and the N-institution that allows for investments of all actors.

Proposition 2.5 (Investments in social capital under the N_i-institution)**.** *Γ has equilibria such that all actors in role i (either $i = 1$ or $i = 2$) invest in social capital in round 0 and all actors i and j subsequently cooperate in all games G in all rounds $1, 2, \ldots$ if*

(1) $w^+ \leq w < w^-$ *and*

(2) $\frac{c}{N_i} \leq \frac{N_j(R_i - P_i)}{1-w}$ *for actors in role i (either $i = 1$ or $i = 2$).*

Proposition 2.6 (Investments in social capital under the N-institution)**.** *Γ has equilibria such that all actors $i = 1, 2$ invest in social capital in round 0 and all actors i subsequently cooperate in all games G in all rounds $1, 2, \ldots$ if*

(1) $w^+ \leq w < w^-$ *and*

(2) $\frac{c}{N} \leq \frac{N_j(R_i - P_i)}{1-w}$ *for actors in role $i = 1, 2$.*

Under conditions (1) of Propositions 2.5 and 2.6 there are no cooperation equilibria in Γ^- (compare Proposition 2.1 for the condition for cooperation equilibria in Γ^-), while playing D in all games G in all rounds $1, 2, \ldots$ of Γ^- is equilibrium behavior. In this equilibrium of Γ^-, an actor in role i realizes a payoff of $N_j Pi/(1-w)$. Conversely, there are cooperation equilibria in Γ^+ under conditions (1) of Propositions 2.5 and 2.6, as can be seen by comparing these conditions with the condition for cooperation equilibria in Proposition 2.2. Thus, after investments in social capital, actor i's payoff associated with a cooperation equilibrium in Γ^+ is $N_j R_i/(1-w)$. Also, the costs of investments in social capital are small enough under conditions (2) of Propositions 2.5 and 2.6, as can be seen by comparing these conditions with Proposition 2.4.

Note that the conditions in Proposition 2.5 and Proposition 2.6 are fulfilled for sufficiently large N_j. This could be interpreted as implying that rational actors will invest in social capital when they are involved in repeated social dilemmas if the number of actors involved is large enough. We return to this issue in the concluding discussion.

It is useful to note that Proposition 2.5 and Proposition 2.6 show that incentives for investments in social capital are not restricted to those actors who would suffer

from opportunistic behavior of their partners. Rather, actors who themselves have incentives to defect likewise have incentives to invest in social capital, even if these actors are not themselves vulnerable to opportunistic behavior of their partners. To see this, consider the Trust Game. In this game, only the trustee has incentives for defection. 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. Proposition 2.5 and Proposition 2.6 show that not only trustors have an incentive to invest in social capital. Trustees have such investment incentives likewise. For example, consider the N_2-institution when G is the Trust Game. Then, *only* the trustees can invest in social capital and rational trustees will do so when conditions (1) and (2) of Proposition 2.5 are fulfilled. In the extreme, there could be only one trustee who interacts in each round of Γ with a sizeable number of trustors and sets up an information network for the trustors that enables them to update each other on the behavior of the trustee. Thus, trustees have an incentive "to bind themselves." Through setting up the information network for the trustors, it becomes less attractive for trustees to behave opportunistically because the future punishment of opportunistic behavior increases. This, however, can induce trustors to cooperate in the first place, thus allowing gains for both trustors and trustees through mutual cooperation—trust is placed and honored—compared to the payoffs they obtain when no trust is placed. A similar analysis applies to the Investment Game. Thus, investments of the trustees in social capital can be seen as a "commitment" that stabilizes cooperation in a social dilemma (Raub, 2004).

Finally, Propositions 2.5 and 2.6 reveal that rational actors will invest in social capital if their cooperation problems are neither too small nor too large. Small cooperation problems in the sense of $w \geq w^-$ can be solved without investments in social capital (see also Proposition 2.1). Large cooperation problems in the sense of $w < w^+$ cannot be solved even with investments in social capital (see also Proposition 2.2). Cooperation problems in the interval $w^+ \leq w < w^-$ can only be solved through investments in social capital and rational actors incur such investments if the costs are small enough and do not exceed the thresholds specified in conditions (2) of Propositions 2.5 and 2.6.

It depends on the parameters of the game whether an institution for cost sharing contributes to inducing investments in social capital. Under the N-institution, costs of social capital can be divided among more actors. However, the benefits must be large enough for all actors to be willing to pay these costs. If returns on social capital are divided very unevenly, it might be that one group is willing to bear considerable costs for social capital but the other group is not. In that case, social capital might only be generated under the N_i-institution where i indicates the actors in the role

that has the higher returns on social capital.

2.4 Conclusions and discussion

We have developed a game-theoretic model that allows for deriving conditions such that a network of information exchange relations between actors induces cooperation in social dilemmas and, simultaneously, allows for endogenizing the network by specifying conditions such that actors incur costs to establish the network in the first place. In other words, we have specified conditions such that the network helps to achieve ends that could not be achieved otherwise, namely, cooperation in social dilemmas that is beneficial for the actors. This means that the network constitutes social capital (Coleman, 1988). Thus, we have modeled investments in and returns on social capital, applying Coleman's and in fact Granovetter's (1985) strategy to combine rational choice assumptions with assumptions on social organizations and social networks. More precisely, we have extended Coleman's and Granovetter's approach by providing an integrated analysis of "games on networks" that focuses on network effects with an analysis of "strategic network formation" that focuses on how networks emerge and develop as a result of incentive-guided and goal-directed behavior.

Our model covers a class of social dilemma games between two actors that includes paradigmatic examples such as the Trust Game, the Prisoner's Dilemma, the Investment Game, and the two-actor Public Goods Game. In our model, "social capital" has a precise meaning, referring to links between actors that allow for exchange of information and thus increase actors' sanction potential because they can condition their behavior vis-à-vis a partner not only on the previous behavior of the partner in their own interactions but also on the partner's behavior with third parties. Moreover, an upper bound on the value of social capital can be specified. We have derived conditions such that rational actors invest in social capital and such that these investments yield returns in the sense that actors subsequently cooperate.

This study aimed at developing and analyzing a theoretical model. What are empirically testable implications? It seems useful to first of all consider 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). Obviously, 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 2.3, 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 T_i, R_i, and P_i of game G, the continuation probability w, the number of partners N_j, and the total costs c of investments in social capital on investments in social capital, on behavior in social dilemmas, and on the relation between investments in social capital and behavior in social dilemmas. Predictions on effects of changes in T_i, R_i, P_i, w, and N_j on behavior in social dilemmas are similar to predictions that follow from earlier models (see Buskens & Raub, 2013). The other predictions are new and highlight the added value of our model.

For experimental tests, we can roughly distinguish between three scenarios. The first scenario covers $w < w^+ < w^-$. Here, cooperation problems are large. We predict a small likelihood of investments in social capital, while the likelihood of investments is only weakly associated with the costs c. We also predict a small likelihood of cooperation as well as small effects of investments in social capital on behavior in game G. We finally predict a small likelihood of cooperation in game G.

The second scenario covers $w^+ \leq w < w^-$. Cooperation problems are now less severe but rational cooperation presupposes that the information network is established. We predict sizeable effects of c on investments in social capital as well as sizeable effects of such investments on subsequent behavior in game G. More specifically, we predict that the likelihood of investments in social capital decreases with increasing costs c. Likewise, we predict that investments in social capital have a positive effect on the likelihood of cooperation.

Finally, $w^+ < w^- < w$ is our third scenario. Cooperation problems are now small. Just like for the first scenario, we predict a small likelihood of investments in social capital, with a weak association between the costs c and the likelihood of investments and small effects of investments in social capital on behavior in game G. Other than for scenario 1, we now of course predict a large likelihood of cooperation in game G.

Note, too, that our model implies other stark predictions of the following kind. Namely, assume a system with $N_1 \geq 2$ actors in role 1 and one actor in role 2 and assume that the information network has been established (i.e., the actors play subgame Γ^+). Also, assume that $T_2 > R_2$ for the actor in role 2 (for actors in role 1, we may have $T_1 > R_1$ or $T_1 = R_1$). Contrast such a system with a scenario with only one actor in role 1 who plays N_1 games G with the actor in role 2 in each round of Γ^+. It follows from our model that behavior in G should not differ significantly between the two settings, strictly speaking not even for sizeable N_1 (see Bolton & Ockenfels,

2009; Camerer & Weigelt, 1988, for earlier work in this direction).

We have seen that the conditions for investments in social capital as specified in Proposition 2.5 and Proposition 2.6 are fulfilled for sufficiently large N_j. This could be interpreted as implying that rational actors who are involved in repeated social dilemmas will invest in social capital if the number of actors involved is large enough. However, this interpretation is problematic. Namely, one would expect that reliable distribution of information, which we have assumed, often becomes more problematic when the size of the network increases. Then, however, it becomes plausible to assume that the costs c of establishing the network increase with increasing network size. Since the number of links in a network of N actors is equal to $N(N-1)/2$, one could even argue that increasing network size has a larger impact on the costs of distributing information than on the returns on social capital so that there would be an optimal size of the network (e.g., Bowles & Gintis, 2004). Of course, if information is distributed via a website, costs of information distribution may hardly depend on the size of the network. This shows that the applicability of our model to larger networks will depend on details such as the technology of information distribution.[18]

Consider other restrictive assumptions of our model that could be replaced in future research. First, consider applications of our model to social dilemma games with one-sided incentives for defection such as the Trust Game and the Investment Game. In such applications, the conditions for cooperation equilibria depend exclusively on the incentives for the trustees and not at all on the incentives for the trustors. This seems problematic. Experimental research strongly suggests, 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 $P_1 - S_1$ decreases or $R_1 - S_1$ 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 trustees' own monetary outcomes but, at least for some trustees, also by monetary outcomes of the trustor.

This relates to another simplifying assumption in our model, namely, that trustees always have short-term incentives to abuse trust and that trustors have complete information on the trustees' incentives. An implication of this assumption is that the information network serves exclusively as a means that allows for sanctioning actors'

[18] We owe an interesting suggestion concerning the effects of information that is available online to Arnout van de Rijt and an anonymous reviewer. For various kinds of trust problems and similar social dilemmas, the internet meanwhile provides easily accessible reputation information. An example is reputation information relevant for economic transactions. For other kinds of trust problems and social dilemmas—co-authorship relations as well organized crime may be examples—such online reputation information is less feasible or not practical. Our analysis might suggest that actors are more inclined to invest in offline ties that provide information for the latter type of relations and that the willingness to invest in offline ties that provide information for the former type of relations has declined since the advent of the internet.

behavior in a focal game G also through the behavior of third parties in future games G. Thus, our model neglects that the information network may serve another purpose, too. Namely, via information on how the partner has behaved in previous games with third partners, an actor may also learn about unobservable characteristics of the partner such as whether or not (and if so, how) a trustee's behavior is also affected by trustors' payoffs.

Finitely repeated social dilemmas with incomplete information (in the technical sense of game theory, see Rasmusen, 1994, p. 47) can be used to address these issues. In a finitely repeated game G with incomplete information one assumes that actors do not know their partner's incentives for sure. For example, for a finitely repeated Trust Game, one assumes that with some—possibly small—probability the trustee has no incentive to abuse trust (since $T_2 \leq R_2$ for the trustee) or he has no opportunity to do so. The trustor knows the probability but she cannot directly observe whether or not the trustee has an opportunity and an incentive to abuse trust. The analysis of finitely repeated games with incomplete information is rather complex. However, one can show, for example, that trustor behavior does also depend on the trustor's own incentives. Frey et al. (2015b) analyze finitely repeated Trust Games between two trustors and one trustee with an initial round 0 in which actors can establish, at costs, an information link between the two trustors. On the one hand, this is a much simpler version of the model presented here because $N_1 = 2$ and $N_2 = 1$ and results only apply to the Trust Game rather than also to other social dilemmas. On the other hand, it is a more complex version of the model presented here because it accounts for incomplete information. The Frey et al. model yields various implications that are similar to those of our model but adds additional implications about the effects of incentives of the trustors.

We have analyzed equilibria such that either all actors cooperate throughout or all actors play D throughout. Also, in our model actors can only establish a complete network that links all actors to all actors. An interesting extension would be a model that allows for incomplete networks and for equilibria with cooperation between some pairs of actors, while other pairs of actors play D. One way to account for such a situation would be to assume that who invests in social capital and the total amount of investments have implications for the set of actors connected through information links (another way to tackle this issue is briefly addressed in Remark A.10 in Appendix A). Note, however, that such extensions of the model are far from trivial. For example, our analysis uses the assumption that information on the behavior of an actor who has incentives to defect in an interaction with some partner, if such information diffuses at all, becomes immediately known to all other partners of the actor. A model allowing for incomplete networks would require more complex assumptions

on information diffusion, while the implications of such more complex assumptions are not easy to deduce (see Buskens, 2002, Chaps. 3 and 4 for work in this direction). Hence, the value of social capital is difficult to establish. Obviously, analyzing the incentives to invest in social capital becomes much more complex, too, when the implications of these investments for information diffusion are less straightforward.

We have used simple assumptions on information and on incentives for providing information. First, we have assumed that information, if available, is reliable. We have thus excluded "noise," unintentional miscommunication, and also strategic distortion of information (see Buskens & Raub, 2013, for further discussion). However, if actors in the same role, say, trustors, are not only interacting with the same partners, say, trustees, but are also each other's competitors, they may have incentives for distorting information and may take into account that information they receive could be distorted. Modeling such circumstances is complex but we would roughly expect that they attenuate the effects of the availability of an information network.

Finally, another problem related to providing information and investments in social capital is that social capital is a collective good (Coleman, 1990, Chap. 12). If an information network is established, it serves also those actors who did not contribute to the costs of providing the network. In principle, this induces incentives for free riding and the risk of suboptimal investments in social capital. We have used assumptions on institutional rules for sharing the costs of investments that mitigate opportunities and incentives for free riding and thus facilitate investments in social capital. Specifically, the information network is established only if each actor who can invest, does indeed invest and actors do not lose their own contribution if an actor who can invest has refused to do so. Thus, if the costs of establishing the information network are low enough in the sense of conditions (2) of Propositions 2.5 and 2.6, the best what every actor who can invest can do, is to indeed invest. This property of our institutional rules for cost sharing facilitates that investing is consistent with equilibrium behavior and with collective rationality. Future research on investments in social capital should analyze implications of other institutional rules for investments. Also, obviously, one could extend our model by also endogenizing the institutional rules for investments. This could be done by considering still another round of Γ that is played before round 0. In that round, actors could establish the institution, again at costs. Such an extended game Γ can then be analyzed using the same techniques that we have employed for endogenizing the network.

The limitations of our model and our agenda for future research should not distract from the model's main strength. It provides an integrated and simultaneous analysis of strategic network formation and of network effects on cooperation: Investments in and returns on social capital in mitigating social dilemmas.

Chapter 3

Embedding trust: A game-theoretic model for trustors' investments in and returns on network embeddedness[1]

Abstract: Social relations through which information disseminates promote efficiency in social and economic interactions that are characterized by problems of trust. This provides incentives for rational actors to invest in their relations. In this chapter, we study a game-theoretic model in which two trustors interact repeatedly with the same trustee and decide, at the beginning of the game, whether to invest in establishing an information exchange relation between one another. We show that the costs the trustors are willing to bear for establishing the relation vary in a non-monotonic way with the severity of the trust problem. The willingness to invest in establishing the information exchange relation is high particularly for trust problems that are neither too small nor too severe.

[1] A slightly different version of this study is published as Frey, V., Buskens, V., & Raub, W. (2015). Embedding trust: A game theoretic model for investments in and returns on network embeddedness. *Journal of Mathematical Sociology, 39*(1), 39–72. Frey wrote the main part of the manuscript and developed the main theoretical argumentation. Buskens and Raub contributed to theory development. We thank Ozan Aksoy, Rense Corten, André Grow, Andreas Flache, Thomas Gautschi, and Manuel Muñoz-Herrera, participants of the June 2011 International Conference on Social Dilemmas and the July 2011 Game Theory and Society Conference for helpful comments.

3.1 Introduction

The idea that social relations are a resource for actors to achieve various goals is widely accepted. It is due to this idea that social relations are often referred to as "social capital" (Lin, 2002). People receive information about job openings through their weak ties (Granovetter, 1973; Ioannides & Loury, 2004). Firms profit from close and committed relations to their buyers and suppliers (Kirman, 2001; Uzzi, 1996), and information exchange relations with third parties mitigate problems of trust (Buskens & Raub, 2002; Coleman, 1990; DiMaggio & Louch, 1998). Often, however, social relations are not simply an exogenously given constraint. If they have instrumental value, actors have incentives to actively establish and maintain relations with an eye on returns that can be expected (Flap, 2004; Lin, 2002, Chap. 8). People have incentives to maintain weak ties in order to gain access to valuable information. Firms have incentives to engage in committed buyer-seller relations. And in situations in which we need to trust others, we have incentives to establish information exchange relations.

In this chapter, we devise and analyze a game-theoretic model for the simultaneous study of investments in and returns on social relations. How information exchange relations with third parties facilitate trust will be our focus. We model how social structure in the sense of relations through which information about reputations can spread mitigates the social dilemma that is inherent to situations characterized by trust problems. We simultaneously endogenize the social structure by modeling actors' incentives to establish information exchange relations.

We want to be specific about what we mean with a "situation characterized by trust problems" and about the benefits that social relations have in this context. In a trust situation, a trustor first decides whether or not to place trust in a trustee. If the trustor places trust, the trustee can choose between honoring and abusing trust. The trustor regrets having placed trust if the trustee abuses trust but benefits if the trustee honors trust. The trustee likewise benefits from honored trust compared to no trust being placed but it is likely that he could earn an extra profit by abusing trust. Therefore, if the interaction is happening in isolation (i.e., if the trustor has no information about the trustee's behavior in past interactions and behavior in the current interaction will not affect future interactions), the trustee is expected to take this extra profit if he has the possibility. Anticipating this, the trustor is expected not to place trust. Because both actors would be better off if trust was placed and honored, such a trust situation represents a social dilemma. The Trust Game (Dasgupta, 1988; Kreps, 1990a), which we will introduce in Section 3.2, provides a formal model for trust situations.

It is well established theoretically and as an empirical finding that the cooperative outcome with trust being placed and honored can be reached if the interaction is embedded in a long-term relation and, especially, if it is embedded in a network through which information about behavior disseminates (Buskens et al., 2010; Coleman, 1990; DiMaggio & Louch, 1998; Huck et al., 2010; Raub & Weesie, 1990; for a survey see Buskens & Raub, 2013). Buskens & Raub (2002) distinguish two mechanisms through which "embeddedness" (Granovetter, 1985) can promote trust and trustworthiness, namely, learning and control. In a long-term relation, a trustor can *learn* from her experiences about the behavior of the trustee. In addition, the trustor has the possibility to sanction an abuse of trust by not placing trust again in future interactions. This gives the trustor some *control* over the trustee: It creates a "shadow of the future" (Axelrod, 1984) that can deter untrustworthy behavior and, therefore, make trust warranted. Embeddedness in a network through which information spreads amplifies these effects: It allows a trustor to learn also from the experiences of other trustors and a trustee may be sanctioned for an abuse of trust also by other trustors who receive information about his behavior. Therefore, embeddedness in an information network can mitigate the social dilemma, making high levels of trust and trustworthiness possible that could not be reached without the network.

In this chapter, we move beyond the analysis of the returns on information exchange relations. We treat such relations as endogenous and assume that actors establish them with an eye on the returns that can be expected. We thus model the co-evolution of relations for information exchange and behavior in trust situations. The question that we pose is: Under what circumstances are trustors who interact with the same trustee most likely to invest in an information exchange relation between one another in order to reap the benefits of trust and trustworthiness? To derive theoretical answers to this question, we focus on the smallest possible scenario, namely, a triad. The game-theoretic model that we devise assumes two trustors who both interact a finite number of times with the same trustee and who are uncertain about whether the trustee has an incentive to abuse trust. Before interacting with the trustee, the two trustors can decide whether or not to establish a relation between one another at costs. If they establish the relation, they subsequently communicate about the behavior of the trustee after every interaction. Consequently, each trustor can learn about the trustee also from the other trustor's past interactions and each trustor benefits also from the other trustor's opportunities to sanction the trustee.

In our analysis, we first establish the returns on the information exchange relation. To this end, we identify the sequential equilibrium of the interactions between the trustors and the trustee after the trustors have or have not established the information exchange relation. We, thus, model the effects of network embeddedness using the

theory of reputation building in sequential equilibria that was pioneered by Kreps et al. (1982) and Kreps & Wilson (1982a). This approach assumes fully rational actors and accounts for effects of learning as well as effects of control. Related models for reputation building and embeddedness effects often consider only control effects (Eguíluz et al., 2005; Raub & Weesie, 1990; Vega-Redondo, 2006) or assume boundedly rational, backward-looking actors and consider only learning effects (Macy & Flache, 2002; Nowak & Sigmund, 2005; Roca et al., 2012). After having established the returns on the information exchange relation, we identify under what conditions the trustors will establish this relation. We assume fully rational behavior also at this stage of the game, while models for the co-evolution of networks and behavior typically assume that actors choose to create, maintain, or sever a link based on some simple backward-looking criterion (Pujol et al., 2005; Skyrms & Pemantle, 2000). Finally, by means of a comparative statics analysis, we show that the maximum cost that the trustors are willing to bear for establishing the information exchange relation, i.e., the trustors' willingness to invest, varies in a non-monotonic way with the size of the trust problem. The trustors' willingness to invest first increases as the trust problem becomes more severe and then decreases again as trust gets ever more problematic. This suggests that the formation of information exchange relations as a means to support trust and trustworthiness is most likely in trust problems that are neither too small nor too severe.

The chapter is organized as follows. In Section 3.2, we present the model in detail. In Section 3.3, we analyze the model and present our results. Finally, in Section 3.4, we conclude and point out directions for future research. An appendix provides additional results and the proofs.

3.2 The model

Before we introduce our model, we want to introduce its main building block, namely, the Trust Game (TG, Dasgupta, 1988).[2] The TG has two players—a trustor (*she*) and a trustee (*he*)—and it starts with the trustor's decision whether or not to place trust. In the case of no trust, the TG ends and trustor and trustee receive the payoffs P_1 and P_2, respectively. In the case of trust, the trustee decides whether to honor or abuse trust. Honored trust leaves both actors better off than no trust, earning them $R_1 > P_1$ and $R_2 > P_2$, respectively. If the trustee abuses trust, the trustor earns $S_1 < P_1$ and, hence, regrets having placed trust. Below we use the TG in different contexts. It depends on the context in which the TG is used whether it is assumed

[2]Throughout this chapter, we use standard game theory terminology and assumptions. See, e.g., Fudenberg & Tirole (2000) for a textbook.

that the trustee could earn a higher payoff than R_2 by abusing trust and whether the trustor is informed about the incentives of the trustee.

3.2.1 One-shot Trust Games with complete and incomplete information

In the standard Trust Game with complete information (Dasgupta, 1988; Kreps, 1990a), it is assumed that the trustee could earn $T_2 > R_2$ by abusing trust and that the trustor knows that the trustee has an incentive to abuse trust. It is easily seen that the standard Trust Game has a unique subgame perfect equilibrium such that the trustee would abuse trust and the trustor does not place trust. As trust being placed and honored would leave both actors better off, the standard Trust Game represents a social dilemma.

It is known from experiments, however, that a considerable portion of trustees actually honor trust also if the standard Trust Game or a similar game is played only once (see e.g., Camerer, 2003, Chap. 2 for an overview). Moreover, if the trustor was certain about the behavior of the trustee, the notion of trust would be superfluous. In fact, one can argue that the trust problem arises from the trustor's uncertainty about the behavior of the trustee, which will be determined by the trustee's preferences and constraints. The trustee might, for example, honor trust because he derives more utility from honoring trust than from abusing trust due to internalized norms and values that trigger internal sanctions when he abuses trust.[3]

The possibility that the trustee may have no incentive to abuse trust and the uncertainty on the side of the trustor is accounted for in the Trust Game with incomplete information (Camerer & Weigelt, 1988; Dasgupta, 1988). The Trust Game with incomplete information starts—as shown in Figure 3.1—with a random move of Nature that determines the trustee's incentives (his *type*). After this move, the trustor and the trustee play a TG together in which the trustor does not know whether the trustee does have an incentive to abuse trust. We interpret the actors' payoffs as utilities and model the trustee's type via his payoffs. With probability π the trustee's payoff from abusing trust is $T_2 - \theta < R_2$ and with probability $1 - \pi$ the trustee's payoff from abusing trust is $T_2 > R_2$. That is, with probability π the trustee has no incentive to abuse trust and with probability $1 - \pi$ he does (just as in the standard Trust Game) have an incentive to abuse trust. The trustee knows his incentives, whereas the trustor cannot directly observe the outcome of the move of Nature and is only informed on the probability π. So, in Figure 3.1, the trustor does not know whether she has to move at the left or the right node. This is indicated by the dashed

[3] Alternatively, the trustee might have the desire but not the opportunity to abuse trust.

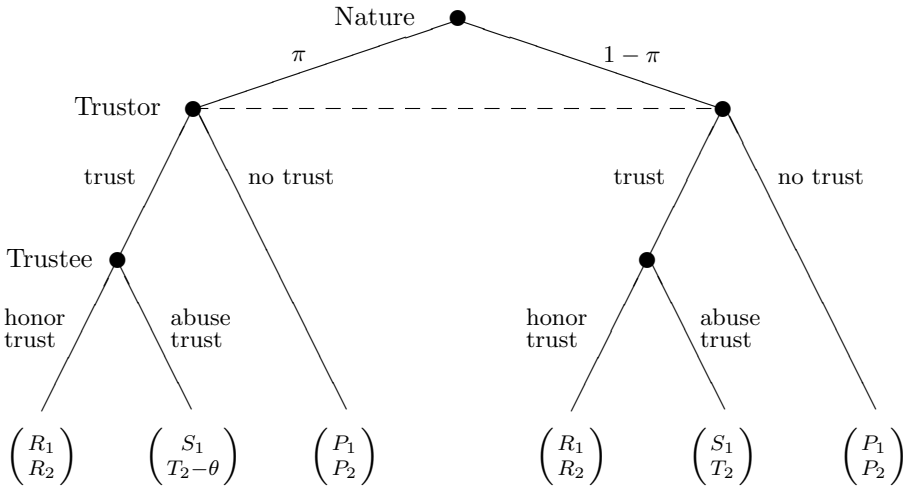


Figure 3.1: The Trust Game with incomplete information. $S_1 < P_1 < R_1$, $P_2 < R_2 < T_2$, and $T_2 - \theta < R_2$.

line that includes these nodes in one information set. The equilibrium solution is straightforward to identify. A trustee with payoffs $T_2 - \theta < R_2$ (a *friendly trustee*) would always honor trust; a trustee with payoffs $T_2 > R_2$ (an *opportunistic trustee*) would always abuse trust. Hence, the trustor's expected payoff from placing trust is $\pi R_1 + (1 - \pi)S_1$ while her payoff from not placing trust is P_1. The trustor's unique equilibrium strategy is thus not to place trust if $\pi < (P_1 - S_1)/(R_1 - S_1)$ and to place trust if $\pi > (P_1 - S_1)/(R_1 - S_1)$. If $\pi = (P_1 - S_1)/(R_1 - S_1)$, the game has multiple equilibria because both strategies of the trustor yield the same expected payoff. The quantity $(P_1 - S_1)/(R_1 - S_1)$ can be interpreted as a measure of the risk a trustor incurs when placing trust (see Buskens, 2002; Snijders, 1996). For later use in our analysis, we define $RISK := (P_1 - S_1)/(R_1 - S_1)$.

3.2.2 A game with investments in network embeddedness

We thus far assumed a one-shot interaction between a trustor and a trustee. In many contexts, however, trust interactions are embedded in long-term relations and/or in a network of relations through which reputations can spread. In the game Γ that we study, we therefore assume that the same trustor and trustee interact together some finite number ($N \geq 1$) of times. Moreover, we assume that the trustee interacts also with a second trustor and we allow the two trustors to invest in network

embeddedness—a relation between one another through which they can exchange information about the behavior of the trustee.

The structure of Γ***:*** Γ starts in period 0.1 with a random move of Nature "choosing" a trustee of the friendly type or of the opportunistic type with probabilities π and $1-\pi$, respectively. The probability π is common knowledge. In period 0.2, the two trustors, who are not informed about the outcome of the move of Nature, can decide whether or not to establish—at costs—the information exchange relation between one another for the rest of the game. Then, one TG (the *stage game*) is played in each of the remaining periods $1, 2, ..., 2N$. Each trustor i, $i = 1, 2$, plays in half of these periods and the trustee plays in every period. Every odd period starts with a move of Nature that determines with equal and independent probability whether trustor 1 plays a TG with the trustee in that period while trustor 2 plays a TG with the trustee in the subsequent even period or vice versa. In which sequence the two trustors interact with the trustee in an odd and the subsequent even period is made common knowledge before the TG of the odd period is played.[4]

The investment decision that the trustors take in period 0.2 is specified as follows. Each trustor chooses independently to propose to invest or not to propose to invest. If both trustors propose to invest, the information exchange relation (henceforth, often simply referred to as *relation*) gets established and each trustor carries half of the total investment cost $c > 0$ that is necessary to establish the relation. If only one trustor proposes to invest, the relation does not get established (as if no trustor proposed to invest) but also the trustor who proposed to invest incurs no cost. We thus assume that a trustor cannot freeride on the investment of the other trustor. This investment rule corresponds to a prevalent assumption in the literature on network formation, namely, two-sided link formation with shared costs of creating a link (Jackson, 2008, Chap. 6).

Whether or not the relation got established has the following consequences for the information available to each trustor in periods 1 to $2N$. If the trustors established the relation in period 0.2, they subsequently exchange information about the outcomes of their interactions with the trustee directly after every TG. So, when making her choice in a given TG, a trustor knows the outcomes (the realized moves) of all TGs that have been played prior to that TG. On the other hand, if the trustors have *not* established the relation, the trustors never exchange information and, hence, the trustor at play in a given TG knows the outcomes of her own previous TGs but does not know the

[4]It could alternatively be assumed that trustor 1 always interacts with the trustee in the odd periods while trustor 2 always plays in the even periods. The analysis of this alternative scenario yields very similar but somewhat more complicated results.

outcomes of the previously played TGs that the other trustor participated in. Note that we assume that information is always truthful.

The outcome of the trustors' investment decision is made common knowledge before the trustors interact with the trustee. Hence, when choosing whether to honor or abuse trust in a given TG, the trustee knows whether only the trustor with whom he is playing the current TG will be informed on his choice or whether the other trustor will be informed, too. The trustors know that the trustee knows this and so forth. Note further that the trustee always knows the outcomes of all past TGs irrespectively of whether the relation between the trustors got established.

Further assumptions on Γ ***and an illustrating example:*** We assume that the structure of Γ is common knowledge and that Γ is played as a non-cooperative game. We also assume that the two trustors' stage-game payoffs are identical and we continue denoting them with R_1, P_1, and S_1. The trustee's stage-game payoffs (indexed with 2) may differ from the trustors' payoffs in the sense that, for example, $P_1 \neq P_2$. An actor's total payoff in Γ is the sum of the undiscounted payoffs that the actor received in the TGs, minus the cost of an investment in period 0.2. For instance, consider the situation that $N = 3$ and both trustors propose to invest in period 0.2. Subsequently, trustor i places trust in her first two TGs but not in her last TG and the trustee honors her trust in the first two TGs. Trustor i's total payoff is then

$$U^{\Gamma}_{Trustor\ i} = R_1 + R_1 + P_1 - \frac{c}{2}. \tag{3.1}$$

Alternatively, if trustor i is the only one who proposed to invest or if none of the trustors proposed to invest and, subsequently, trustor i never places trust, her total payoff is

$$U^{\Gamma}_{Trustor\ i} = P_1 + P_1 + P_1. \tag{3.2}$$

3.3 Analysis of the model

In our analysis of Γ, we assume rational behavior in the trust interactions as well as in the investment decision. Moreover, we assume rational beliefs in the sense that the trustors update their beliefs about the trustee's type following Bayes' rule. We analyze under what conditions Γ has an equilibrium such that the trustors invest in the establishment of the information exchange relation. To identify under what conditions Γ has such an "investment equilibrium" we, first, establish what the trustors can expect to happen in their TGs after they have or have not established the relation. In Subsection 3.3.1, we sketch the concept of reputation building in a sequential equilibrium and introduce the necessary notation for the formal specification of a sequential

equilibrium. In Subsection 3.3.2 and 3.3.3, we specify the sequential equilibrium for the scenario that the relation has or has not been established. Comparing the payoffs the trustors can expect in these two scenarios for periods 1 to $2N$, we, then, in Subsection 3.3.4, identify the expected return on investment in the information exchange relation and specify under what conditions Γ has an investment equilibrium. Finally, in order to derive testable predictions, we analyze in Subsection 3.3.5 how changes in the parameters of the game affect the return on investment and, hence, the maximum cost of investment for which Γ has an investment equilibrium.

3.3.1 Trust and trustworthiness as a result of conditional behavior and reputation building

If the trustors knew with certainty that the trustee is of the opportunistic type, backward induction would predict that they never place trust, irrespectively of whether the relation has been established. With incomplete information about the trustee's incentives, however, trust being placed and honored during all but the last few periods can be an equilibrium outcome. This has been established by Camerer and Weigelt (1988; see also Bower et al., 1997; Buskens, 2003) who apply the analysis of reputation building in sequential equilibrium pioneered by Kreps et al. (1982) and Kreps & Wilson (1982a) to finitely repeated Trust Games. Informally, a combination of beliefs and strategies constitutes a sequential equilibrium (Kreps & Wilson, 1982b) if the beliefs are justified by the strategies following Bayesian updating and the strategies are best replies against the others' strategies given the beliefs.

To illustrate why trust being placed and honored in all but the last few periods can be a sequential equilibrium, let us assume that a trustor places trust in an early period of the repeated game. In this case, trust may be honored for two different reasons. First, the trustee may have no incentive at all to abuse trust because he is of the friendly type. Second, the trustee may have a short-term incentive to abuse trust but follow an incentive for reputation building. Specifically, an opportunistic trustee may honor trust because he knows that if a trustor gets the information that he ever abused trust, the trustor can infer that he must be of the opportunistic type and, therefore, decide never to place trust again in future periods. On the other hand, if the trustee does honor trust, the trustor, while remaining uncertain about the trustee's type, might become more confident that he is of the friendly type and place trust again in the future. A trustor can anticipate on such reputation building by an opportunistic trustee. She may, therefore, be inclined to indeed place trust in an early period of the game even if the probability that the trustee is of the friendly type is small. As the end of the game approaches, however, an opportunistic trustee's

incentive to maintain a reputation for being trustworthy decreases and he might, therefore, abuse trust, leading to no trust being placed anymore by a trustor who is informed about the abuse of trust. Conversely, anticipating this, a trustor might choose not to place trust anymore even if she has no information that the trustee has abused trust previously. Thus, the sequential equilibrium with trust being placed and honored in all but the last few periods results from a subtle interplay of the trustors who try to learn about and to control the trustee, taking the trustee's incentives for reputation building into account, and a trustee who balances the long-term effects of his reputation and the short-term incentives for abusing trust, taking into account that the trustors anticipate on this balancing.

In the following two subsections we establish the sequential equilibrium of the periods 1 to $2N$ of Γ after the relation has or has not been established. In this, we build closely on the analyses of finitely repeated TGs with one, two, or more trustors by Anderhub et al. (2002), Bower et al. (1997), Buskens (2003), and Camerer & Weigelt (1988). We refer the reader to these studies as well as to Fudenberg & Tirole (2000, Chap. 8) for a detailed derivation of the sequential equilibrium.

Before proceeding, we need to introduce some more notation. First, we refer to the continuation of the game after the relation has or has not been established in period 0.2 as (*continuation game*) Γ^+ and (*continuation game*) Γ^-, respectively. To describe the actors' strategies, we let t_n^i denote the probability that trustor i at play in period n places trust in that period and we let h_n denote the probability that a trustee of the opportunistic type honors trust in that period. It is clear that because a friendly trustee has no short-term incentive to abuse trust, he will always honor trust with probability 1. To describe the trustors' beliefs, we let π_n^i stand for trustor i's belief at the start of period n that the trustee is of the friendly type. At the beginning of period 1 this belief equals the prior probability ($\pi_1^i = \pi$, for $i = 1, 2$). At the end of every period, each trustor updates her belief following Bayes' rule. Note that π_n^i also is the trustee's reputation; it indicates what type he is thought to be. Finally, similar to $RISK := (P_1 - S_1)/(R_1 - S_1)$, we define $TEMP := (T_2 - R_2)/(T_2 - P_2)$ as a second measure pertaining to the payoffs of the stage game. While *RISK* measures the risk a trustor incurs when placing trust (Section 3.2.1), *TEMP* measures an opportunistic trustee's temptation to abuse trust (see Buskens, 2002; Snijders, 1996).

3.3.2 Trust and trustworthiness without network embeddedness

Our first proposition specifies the unique sequential equilibrium of Γ^-, i.e., in periods 1 to $2N$ of Γ after the trustors have not established the information exchange relation.[5] In Γ^-, each trustor is only informed on the outcomes of her own past TGs but not on the outcomes of the TGs that the other trustor played with the trustee.

Proposition 3.1. *The beliefs and strategies specified below constitute the unique sequential equilibrium of* Γ^-.

- *Belief of trustor i in period n that the trustee is of the friendly type:*
  - *If, in period $n-1$, trustor i did not place trust or was not at play, then* $\pi_n^i = \pi_{n-1}^i$.
  - *If, in period $n-1$, trustor i placed trust and trust was honored, then* $\pi_n^i = \max(RISK^{\lceil \frac{2N-n+1}{2} \rceil}, \pi_{n-1}^i)$.
  - *If, in period $n-1$, trustor i placed trust and trust was abused, then* $\pi_n^i = 0$.
- *Probability that (if at play) trustor i places trust in period n:*
  - *If* $\pi_n^i > RISK^{\lceil \frac{2N-n+1}{2} \rceil}$, *then* $t_n^i = 1$.
  - *If* $\pi_n^i = RISK^{\lceil \frac{2N-n+1}{2} \rceil}$, *then* $t_n^i = TEMP$.
  - *If* $\pi_n^i < RISK^{\lceil \frac{2N-n+1}{2} \rceil}$, *then* $t_n^i = 0$.
- *Probability that an opportunistic trustee honors trust of trustor i at play in period n:*
  - *If* $\pi_n^i \geq RISK^{\lfloor \frac{2N-n}{2} \rfloor}$, *then* $h_n = 1$.
  - *If* $\pi_n^i < RISK^{\lfloor \frac{2N-n}{2} \rfloor}$, *then* $h_n = \frac{\pi_n^i}{1-\pi_n^i}\left(\frac{1}{RISK^{\lfloor \frac{2N-n}{2} \rfloor}} - 1\right)$.

Appendix B provides the proofs for our propositions.

We describe the course of behavior and beliefs in the equilibrium defined in Proposition 3.1 focusing on the interactions between some trustor i and the trustee. In Γ^-, the sequential equilibrium of the interactions between some trustor i and the trustee is identical to the sequential equilibrium of the finitely repeated TG with incomplete information and only one trustor (see Anderhub et al., 2002, Bower et al., 1997,

[5] In the game Γ, the set of sequential equilibria coincides with the set of perfect Bayesian equilibria. That is, a combination of beliefs and strategies that is a sequential equilibrium of Γ^+ or Γ^- is also a perfect Bayesian equilibrium of the respective continuation game (see Fudenberg & Tirole, 2000, Theorem 8.2).

or Camerer & Weigelt, 1988). What happens in the interactions between the focal trustor and the trustee is independent of what happens in the interactions between the other trustor and the trustee.

The equilibrium of the interactions between each trustor i and the trustee can be described as evolving over three phases. Initially, trustor i places trust and the trustee honors trust. In this first phase, trustor i does not change her belief, knowing that either type of trustee would always honor trust. As the end of the game comes closer, the second phase starts. In the first TG of the second phase, trustor i still places trust with probability 1 while the opportunistic trustee begins to randomize (because trustor i would afterwards not place trust anymore without being convinced that the probability that she is playing with a friendly trustee exceeds the prior probability π). Then, both actors randomize and trustor i becomes more and more confident that the trustee is of the friendly type until the first instance that she does not place trust or that the trustee abuses her trust.[6] Thereafter, the third phase starts: Trustor i does not place trust anymore.

We let τ denote how many TGs need to be left to play between trustor i and the trustee for trust still being placed and honored with certainty. If, for example, the opportunistic trustee's randomization (the second phase) starts in the next to last TG of trustor i, $\tau = 2$. It follows from Proposition 3.1 that the integer τ is such that π lies in the interval $[RISK^{\tau}, RISK^{\tau-1})$. This implies that

$$\tau = \left\lceil \frac{\log \pi}{\log RISK} \right\rceil. \tag{3.3}$$

It can be seen from Equation (3.3) that τ increases stepwise in $RISK$ and decreases stepwise in π. That is, the second phase tends to start earlier if the risk associated with placing trust is higher and if the probability of a friendly trustee is smaller.

Note furthermore that τ is independent of N. Hence, if τ is larger or N is smaller, the phase in which trust is placed and honored with certainty (the first phase) is shorter. Even more, the equilibrium evolves over three phases as described above only if $\tau < N$. If $\tau = N$, the opportunistic trustee randomizes already in the first TG. If $\tau > N$, trustor i never places trust because, given the parameters, the game is too short for the opportunistic trustee to start building a reputation.

Proposition 3.2 specifies a trustor's expected payoff associated with the unique sequential equilibrium of Γ^{-}.

[6] While the trustee becomes more likely to abuse trust as the end of the game approaches, the trustors randomize with a constant probability.

Proposition 3.2. *The expected payoff for a trustor in* Γ^- *is*

$$U_1^{\Gamma^-} = \begin{cases} (N-\tau)R_1 + (S_1 + \pi\frac{R_1-S_1}{RISK^{\tau-1}}) + (\tau-1)P_1 & \text{if } \tau \leq N \\ NP_1 & \text{if } \tau > N \end{cases} \tag{3.4}$$

Previous studies on reputation building in finitely repeated games provide no explicit account of expected payoffs. In our analysis, however, knowing the expected payoff of a trustor in Γ^- (and in Γ^+) is crucial and we, therefore, want to briefly sketch the intuition behind Proposition 3.2. If $\tau > N$, a trustor never places trust and receives a payoff of NP_1. If $\tau \leq N$, each trustor receives R_1 in her first $N-\tau$ TGs. Then, in the TG in which an opportunistic trustee starts to randomize while the trustor still places trust with probability 1, a trustor's expected payoff is $R_1\big(\pi + (1-\pi)h_{N-\tau+1}\big) + S_1(1-\pi)(1-h_{N-\tau+1})$. This reduces to $S_1 + \pi\frac{R_1-S_1}{RISK^{\tau-1}}$, which we, henceforth, sometimes denote by X_1 and which must be marginally smaller than R_1 and, in equilibrium, must be at least as large as P_1 (i.e., $P_1 \leq X_1 < R_1$). Finally, a trustor's expected total payoff for her last $\tau-1$ TGs is $(\tau-1)P_1$, which follows from the fact that in these TGs a trustor is (at best) indifferent between placing and withholding trust.

To summarize, the course of behavior in the N interactions between each trustor i and the trustee in Γ^- depends essentially on π and *RISK* that together determine τ. N and τ determine whether trust is possible at all, and if so, in how many of her TGs trustor i will benefit from trust being placed and honored with certainty. With every unit increase in τ, the number of TGs of trustor i in which her trust is placed and honored with certainty decreases by 1 and, accordingly, trustor i's expected payoff decreases by $R_1 - P_1$.[7] The analysis furthermore reveals how much a trustor suffers from the trust problem. Compared to an ideal world in which a trustor would earn NR_1, trustor i's loss due to the trust problem is $(\tau-1)(R_1-P_1) + R_1 - X_1$, where $P_1 \leq X_1 < R_1$. We use this as a measure for the size of the trust problem. Specifically, we define the (approximate) size of the trust problem as $\tau(R_1-P_1)$, i.e., as the number of TGs a trustor does not benefit from trust being placed and honored with certainty multiplied by the value that trust being placed and honored has for a trustor. We, thus, say that the trust problem is larger (more severe) if $R_1 - P_1$ is larger or if τ is larger (because *RISK* is larger or π smaller).

[7] Note that π and *RISK* also determine the trustor's expected payoff for the period in which the opportunistic trustee starts to randomize (X_1). The stage-game payoffs of the trustee only determine the randomization probability of the trustors but do not affect their expected payoffs.

3.3.3 Trust and trustworthiness with network embeddedness

An opportunistic trustee has a stronger incentive to build and maintain a reputation for being trustworthy in Γ^+ than in Γ^-. In Γ^+, each trustor receives information not only on the outcomes of her own TGs but also on the outcomes of the TGs that the other trustor plays with the trustee. This is common knowledge and the trustee, hence, knows that his choice in a given TG might affect not only the future choices of the trustor with whom he plays that TG but also the future choices of the other trustor. He knows, for example, that if he abuses trustor i's trust, also the other trustor will from then on know that she must be playing with an opportunistic trustee and will not place trust anymore. Hence, the long-term consequences that an opportunistic trustee has to consider when making his choice in a given TG in Γ^+ are the same as if he played *all* remaining TGs with the trustor with whom he plays that TG. The long-term costs of an abuse of trust are, thus, larger in Γ^+ than in Γ^-, whereas, obviously, the short-term incentive to abuse trust is the same in both continuation games. Our third proposition specifies the unique sequential equilibrium that results in the interactions between the trustors and the trustee in Γ^+.

Proposition 3.3. *The beliefs and strategies specified below constitute the unique sequential equilibrium of* Γ^+.

- *Belief of trustor i in period n that the trustee is of the friendly type:*
  - *If, in period $n-1$, trust was not placed, then* $\pi_n^i = \pi_{n-1}^i$.
  - *If, in period $n-1$, trust was placed and honored, then* $\pi_n^i = \max(RISK^{2N-n+1}, \pi_{n-1}^i)$.
  - *If, in period $n-1$, trust was placed and abused, then* $\pi_n^i = 0$.
- *Probability that (if at play) trustor i places trust in period n:*
  - *If* $\pi_n^i > RISK^{2N-n+1}$, *then* $t_n^i = 1$.
  - *If* $\pi_n^i = RISK^{2N-n+1}$, *then* $t_n^i = TEMP$.
  - *If* $\pi_n^i < RISK^{2N-n+1}$, *then* $t_n^i = 0$.
- *Probability that an opportunistic trustee honors trust of the trustor i at play in period n:*
  - *If* $\pi_n^i \geq RISK^{2N-n}$, *then* $h_n = 1$.
  - *If* $\pi_n^i < RISK^{2N-n}$, *then* $h_n = \frac{\pi_n^i}{1-\pi_n^i}\left(\frac{1}{RISK^{2N-n}} - 1\right)$.

In the sequential equilibrium of Γ^+ specified in Proposition 3.3, the trustee and the trustor at play in a given TG behave as if they played all $2N$ TGs together, i.e., as if there was only one trustor playing $2N$ TGs with the trustee. The sequential equilibrium of Γ^+ evolves over the same three phases as the sequential equilibrium of the interactions between each trustor i and the trustee in Γ^-. First, both trustors place trust and the trustee honors trust with probability 1 until and including the TG after which there are *in total* τ TGs left to be played, i.e., until and including period $2N - \tau$, where τ is determined by *RISK* and π as specified in Equation (3.3). Then, the randomization begins and after the first instance that one of the trustors did not place trust or that the trustee abused trust, both trustors do not place trust anymore. These three phases obtain if $\tau < 2N$. If $\tau = 2N$, the trustee randomizes already in the first period; if $\tau > 2N$, the trustors never place trust.

Proposition 3.4 specifies a trustor's expected payoff associated with this equilibrium of the continuation game Γ^+.

Proposition 3.4. *In Γ^+, the expected payoff for a trustor is*

$$U_1^{\Gamma^+} = \begin{cases} \frac{(2N-\tau)R_1+(S_1+\pi\frac{R_1-S_1}{RISK^{\tau-1}})+(\tau-1)P_1}{2} & \text{if } \tau \leq 2N \\ NP_1 & \text{if } \tau > 2N \ . \end{cases} \tag{3.5}$$

To understand how a trustor's expected payoff for Γ^+ is calculated, realize that in the case that $\tau \leq 2N$, each trustor plays expectedly in half of the periods 1 to $2N - \tau$ in which the trustor at play earns R_1 as well as in half of the $\tau - 1$ periods for which the expected payoff of the trustor at play is P_1 and each trustor has a 50% chance of playing in the period in which the trustee's randomization starts.

We have now established what the trustors can expect to happen in equilibrium in their interactions with the trustee after they have or have not established the information exchange relation. In the next section, we establish the expected return on investment and the condition for the existence of an equilibrium such that the trustors establish the relation.

3.3.4 Returns on and investments in network embeddedness

When choosing whether or not to propose to invest in the information exchange relation in period 0.2, a rational trustor will weigh the cost of investment against the expected return on investment. The expected return on investment—the value that the relation has for a trustor—derives from the difference in a trustor's expected payoffs in Γ^+ and Γ^- and can be calculated as $r_1 = U_1^{\Gamma^+} - U_1^{\Gamma^-}$, where r_1 denotes a

trustor's expected return on investment.[8] r_1 can also be interpreted straightforwardly as a trustor's "willingness to invest," i.e., the maximum cost of investment a rational trustor is willing to incur in order to establish the relation. Proposition 3.5 specifies r_1 and establishes that Γ always has an investment equilibrium if the cost of investment per trustor is smaller than r_1.

Proposition 3.5. *In Γ, an equilibrium such that both trustors propose to invest (an investment equilibrium) exists if and only if for each trustor the cost of investment ($c/2$) does not exceed the expected return on investment (r_1), i.e., iff $c/2 \leq r_1$, where r_1 falls in the following intervals:*

$$\begin{array}{lrcccl} \textit{if } \tau \leq N, & \frac{\tau-1}{2}(R_1-P_1) & < & r_1 & \leq & \frac{\tau}{2}(R_1-P_1) \\ \textit{if } N < \tau \leq 2N, & \frac{2N-\tau}{2}(R_1-P_1) & \leq & r_1 & < & \frac{2N-(\tau-1)}{2}(R_1-P_1) \\ \textit{if } \tau > 2N, & & & r_1 & = & 0. \end{array}$$

Proposition 3.5 distinguishes three scenarios. If $\tau \leq N$, there is an equilibrium phase in which trust is placed and honored with certainty in Γ^+ as well as Γ^- but this phase is longer in Γ^+ because an opportunistic trustee remains trustworthy in Γ^+ until he has only half as many TGs left with each trustor compared to the situation in Γ^-.[9] If $N < \tau \leq 2N$, trust is placed with certainty in at least the first TG in Γ^+ but the trustors never place trust in Γ^- because, given π and *RISK*, the game is too short for the trustee to start building a reputation if the information exchange relation has not been established. Finally, if π or N is very small or *RISK* very large such that $\tau > 2N$, trust is not even possible in Γ^+.

In the third scenario with trust being not even possible in Γ^+, there is no return on embeddedness and a trustor is not willing to incur any cost $c/2 > 0$ for establishing the relation. For the other two scenarios, Proposition 3.5 specifies intervals for r_1. To understand the specification of the intervals, note that r_1, roughly, equals the number of TGs in which a trustor would profit from trust being placed and honored with certainty in Γ^+ but not in Γ^- multiplied with the benefit of honored trust compared to no trust $(R_1 - P_1)$. Thus, if $\tau \leq N$, $r_1 \approx \frac{\tau}{2}(R_1 - P_1)$ because a trustor expectedly benefits from trust being placed and honored with certainty until she has $\tau/2$ TGs left in Γ^+ but only until she has τ TGs left in Γ^-. If $N < \tau \leq 2N$, implying that trust is only possible in Γ^+, the number of *additional* TGs in which a trustor would profit from trust being placed and honored with certainty if the relation gets established

[8] The specification of Γ implies that, as $U_1^{\Gamma^+}$ and $U_1^{\Gamma^-}$, r_1 is identical for the two trustors.

[9] In other words, the "endgame" of τ TGs in which trust and trustworthiness are not certain anymore occurs for each trustor separately and in its full length in Γ^-. In Γ^+, however, the endgame occurs only once and each trustor plays in half of the τ TGs of the endgame.

simply equals the number of TGs in which she would expectedly do so in Γ^+ (namely, $(2N-\tau)/2$) and, thus, $r_1 \approx \frac{2N-\tau}{2}(R_1 - P_1)$.

This calculation of r_1 yields the precise r_1 if the payoff that a trustor can expect for the interaction in which an opportunistic trustee begins to randomize (X_1) equals P_1. If $X_1 \neq P_1$, i.e., if a trustor's expected payoff for the TG in which randomization starts is not the same as her expected payoff for the subsequent TGs, r_1 is somewhat larger or smaller. Specifically, the intervals specified in Proposition 3.5 show that, for a given N and τ, r_1 can be up to $\epsilon < (R_1 - P_1)/2$ smaller (larger) if $\tau \leq N$ ($N < \tau \leq 2N$). We provide the exact formulas for r_1 in the proof of Proposition 3.5. Note, however, that the specification in Proposition 3.5 and the approximation of r_1 are sufficient to derive the main comparative statics.

Proposition 3.5 states that $c/2 \leq r_1$ is a necessary and sufficient condition for the existence of an equilibrium such that the trustors establish the relation. Such an equilibrium is never unique, however. That both trustors do not propose to invest is always part of an equilibrium. Because the relation gets only established if *both* trustors propose to invest, each trustor is indifferent between proposing to invest and not proposing to invest given the other trustor does not propose to invest. We note, however, that if $c/2 < r_1$, the investment equilibrium risk-dominates and payoff-dominates the "no investment equilibrium." If $c/2 < r_1$, a trustor will (in expectations) never lose by proposing to invest (because she only incurs the cost if the relation does get established), while she (in expectations) would gain if the other trustor proposed to invest, too.

3.3.5 Comparative statics

In this section, we investigate the comparative statics of r_1 in order to derive testable predictions. What we present can be interpreted interchangeably as the comparative statics of (i) the value of embeddedness, (ii) the potential return on investment, or (iii) the maximum cost of investment per trustor for which Γ has an investment equilibrium. Because $r_1 = U_1^{\Gamma^+} - U_1^{\Gamma^-}$, it is clear that the parameters that determine $U_1^{\Gamma^-}$ as well as $U_1^{\Gamma^+}$, namely π, S_1, P_1, R_1, and N, also fully determine r_1. In the following, we treat the *ceteris paribus* effect of a change in each of these parameters in a separate subsection. We formulate our results for changes in π, S_1, P_1, and R_1 such that they state how r_1 changes as the parameter under study changes in the direction that tends to lead to an increase in τ (an earlier start of the randomization). We thus, for example, establish how r_1 changes if π *decreases*. To focus on parameter changes that tend to lead to an earlier start of the randomization phase and, therefore, to deviate from the standard practice to focus on effects of increases in *all* parameters

allows presenting the results more efficiently.

Our approximation of r_1 suggests that a change in some parameter may affect r_1 by (i) affecting in how many additional TGs a trustor would benefit from trust being placed and honored with certainty if the relation gets established and/or (ii) by affecting the value of honored trust compared to no trust ($R_1 - P_1$). The latter will be the case only if R_1 or P_1 changes. The former may occur if N changes or if a change in π, S_1, P_1, or R_1 triggers a change in τ, i.e., in how early randomization starts. There is a third and somewhat more subtle way in which a parameter change can affect r_1. Namely, r_1 may change due to a change in π, S_1, P_1, or R_1 that does *not* trigger a change in τ but "only" leads to a change in the payoff a trustor can expect for the TG in which randomization starts (X_1). Our analyses show that r_1 is affected in the same direction by a change in π, S_1, P_1, or R_1 irrespectively of whether this change triggers a change in τ or only affects X_1. In order to provide some intuition for our results without being overwhelmed by details, we focus our explanations on parameter changes that affect τ and do not consider the effects of parameter changes that affect X_1 but not τ.

Changes in π: Proposition 3.6 establishes how r_1 changes as the probability that the trustee is of the friendly type (π) decreases.

Proposition 3.6. *Given the specification of* Γ *and the definitions of* τ *and* r_1*, it holds that:*

- *If* $\tau \leq N$*,* r_1 *increases as* π *decreases.*
- *If* $N < \tau \leq 2N$*,* r_1 *decreases as* π *decreases.*

Proposition 3.6 states that as π decreases the value of the relation first increases and then decreases again. To understand this result, recall that if π is smaller, τ tends to be larger, i.e., randomization tends to start earlier. Recall further that if trust is also possible if the relation has not been established, a trustor would benefit in Γ^+ from trust being placed and honored with certainty until she has $\tau/2$ TGs left, whereas she would do so in Γ^- only until she has τ TGs left. Hence, if π decreases such that τ increases and π is large enough also after the decrease such that, given *RISK* and N, trust is possible also in Γ^-, the number of additional TGs in which a trustor could benefit from honored trust with certainty ($\tau/2$) increases and, consequently, r_1 increases. The relation becomes more valuable because avoiding half of the phase in which trust and trustworthiness are not certain anymore is more valuable the longer this phase is. The effect of a decrease in π on r_1 is opposite if trust is not possible in Γ^-, both before and after the decrease in π, but possible in Γ^+, at least before

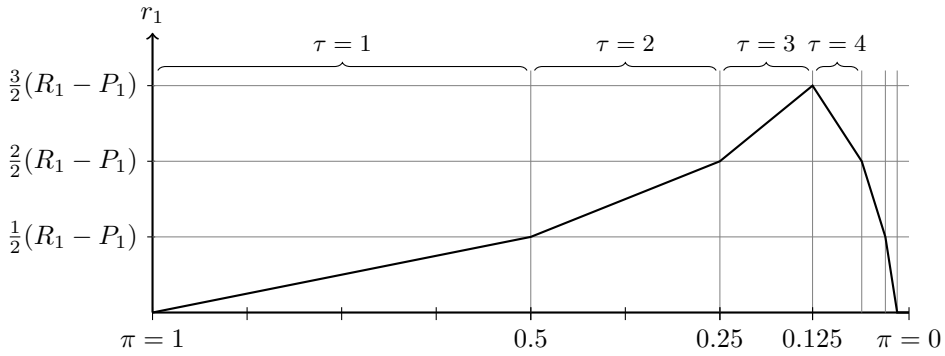


Figure 3.2: The effect of changes in π on r_1 in an example with $RISK = 0.5$ and $N = 3$.

the decrease in π. In this case, a decrease in π that leads to an earlier start of the randomization (an increase in τ) leads to a decrease in the number of TGs in which a trustor could benefit from trust being placed and honored with certainty in Γ^+ while leaving $U_1^{\Gamma^-}$ unchanged at NP_1. Consequently, it leads to a decrease in r_1. Note further that if π is so small that, given $RISK$ and N, trust is not even possible in Γ^+, $r_1 = 0$ irrespectively the precise π.

Figure 3.2 visualizes how r_1 depends on π in an example with $N = 3$ and $RISK = 0.5$. It provides a more detailed picture than Proposition 3.6, illustrating also the additional results established in Lemma B.1 in Appendix B. Figure 3.2 quantifies r_1 in terms of $R_1 - P_1$ and shows that r_1 is largest (namely, $r_1 = N(R_1 - P_1)/2$) if trust is placed and honored with certainty in half of the $2N$ TGs in Γ^+ while in Γ^- the trustee's randomization starts in the first TG of each trustor and π is just large enough such that the trustee does not want to start randomizing a TG earlier (in Figure 3.2 at the border of the intervals $\tau = 3$ and $\tau = 4$). The figure further shows that the increase towards the maximum as well as the decrease thereafter is monotonic and stepwise linear.[10]

Changes in S_1***:*** Proposition 3.7 establishes how r_1 changes as the payoff a trustor receives if the trustee abuses trust (S_1) decreases. If S_1 is smaller, $RISK$ is larger and, consequently, τ tends to be larger, i.e., the randomization tends to start earlier.

[10]It can be seen from Figure 3.2 that the increase towards the maximum and the decrease thereafter is not linear even though r_1 depends linearly on π for changes in π that do not affect τ because the range of π for which τ is constant is smaller, the smaller π.

Proposition 3.7. *Given the specification of* Γ *and the definitions of* τ *and* r_1*, it holds that:*

- *If* $\tau \leq N$*,* r_1 *increases as* S_1 *decreases.*
- *If* $N < \tau \leq 2N$*,* r_1 *decreases as* S_1 *decreases.*

Proposition 3.7 shows that the potential return on investment increases as S_1 decreases as long as trust remains possible also in Γ^-. As S_1 decreases further and trust is only possible in Γ^+, the potential return on investment decreases. It is also clear that as S_1 gets so small that trust is not even possible in Γ^+, $r_1 = 0$. Thus, r_1 changes in the same manner if S_1 decreases as it changes if π decreases. This is not surprising. Neither a decrease in S_1 nor a decrease in π affects the value of honored trust compared to no trust $(R_1 - P_1)$ while both lead to an earlier start of randomization and, hence, affect in the same way the expected number of additional TGs in which a trustor can benefit from trust being placed and honored with certainty if the relation gets established. We provide additional details on how r_1 depends on S_1 in Appendix B and we note that a figure could be drawn for the dependence of r_1 on S_1 that is similar to Figure 3.2.

Changes in P_1***:*** Proposition 3.8 establishes how r_1 is affected by an increase in the payoff a trustor receives if she does not place trust (P_1). If P_1 is larger, the trustors are more reluctant to place trust and τ tends to be larger (the randomization tends to start earlier).

Proposition 3.8. *Given the specification of* Γ *and the definitions of* τ *and* r_1*, it holds that:*

- *If* $\tau \leq N$*,* r_1 *increases as* P_1 *increases.*
- *If* $N < \tau \leq 2N$*,* r_1 *decreases as* P_1 *increases.*

Proposition 3.8 shows the maximum cost for which Γ has an investment equilibrium changes in the same direction due to an increase in P_1 as due to a decrease in π or S_1. The mechanics associated with a change in P_1 are more complicated, however. An increase in P_1 may lead to an earlier start of the randomization and, hence, affect the number of additional TGs in which a trustor would benefit from trust being placed and honored with certainty in Γ^+. At the same time, an increase in P_1 also reduces the value of honored trust compared to no trust $(R_1 - P_1)$. If trust is not possible in Γ^- while (at least before the increase in P_1) trust is possible in Γ^+, these two effects have the same direction: both contribute to a decrease in r_1. However, if trust is also possible in Γ^- before and after the increase in P_1, the two effects are

opposed to one another. To see that in this case r_1 increases if P_1 increases requires going into the details of how a change in P_1 affects X_1 (which is partly via the effect of a change in P_1 on the probability that an opportunistic trustee honors trust in the interaction in which he starts to randomize).

Changes in R_1***:*** Proposition 3.9 establishes how r_1 changes as the payoff a trustor gets if she places trust and the trustee honors her trust (R_1) decreases. If R_1 is smaller, τ tends to be larger, i.e., randomization tends to start earlier.

Proposition 3.9. *Given the specification of* Γ *and the definitions of* τ *and* r_1*, it holds that if* $\tau \leq 2N$*,* r_1 *decreases as* R_1 *decreases.*

Similar to a change in P_1, a decrease in R_1 affects r_1 through a decrease in $(R_1 - P_1)$ and, potentially, through an increase in τ as well as a decrease in the probability that an opportunistic trustee honors trust in the interaction in which he starts to randomize. Proposition 3.9 shows that, as long as trust is possible in at least the first TG of Γ^+, the total of these effects is such that r_1 decreases as R_1 decreases.

Changes in N***:*** Finally, Proposition 3.10 specifies how a change in the number of repetitions (N) affects r_1.

Proposition 3.10. *Given the specification of* Γ *and the definitions of* τ *and* r_1*, it holds that:*

- *If* $N + 1 \leq \tau \leq 2(N + 1)$*,* r_1 *increases as* N *increases.*
- *If* $\tau \leq N$ *or* $\tau > 2(N + 1)$*,* r_1 *does not change as* N *increases.*

Proposition 3.10 establishes that r_1 increases due to an increase in N if trust is possible in Γ^+ but not in Γ^- before and after the increase in N, whereas r_1 does not change if trust is also not possible in Γ^+ after the increase in N or if trust is possible also in Γ^- already before the increase. In the former case, an increase in N means adding a period in which each trustor would earn P_1 in Γ^- but R_1 in Γ^+. In the latter case, a trustor's expected payoff for Γ^+ and Γ^- both change in the same way and, consequently, r_1 does not change. In the proof, we also quantify how r_1 changes due to changes in N.

Summarizing: Propositions 3.6 and 3.7 show that as π or S_1 decreases, the return on the information exchange relation first increases and then decreases again. These effects can be summarized in relation to the size of the trust problem that we defined in Section 3.3.2 as $\tau(R_1 - P_1)$. Recall that if π or S_1 decreases, randomization tends to start earlier (i.e., τ tends to increase) and, hence, the trust problem becomes

larger. Thus, Propositions 3.6 and 3.7 show that the maximum cost of investment per trustor for which Γ has an investment equilibrium varies in a non-monotonic way if the size of the trust problem increases due to a decrease in π or in S_1. The trustors' willingness to invest first increases as the trust problem gets more severe but after some point (if the trust problem becomes so severe that trust is not possible without the information exchange relation) the trustors' willingness to invest decreases again as the trust problem gets even more severe. The trustors' willingness to invest changes in the same non-monotonic manner if P_1 increases, whereas it always decreases if R_1 decreases (Propositions 3.8 and 3.9). Changes in P_1 and R_1 cannot be related straightforwardly to changes in the size of the trust problem. Changes in P_1 and R_1 affect τ and $R_1 - P_1$ simultaneously but not in the same direction.

Finally, let us note that if N becomes large, the range in which the trustors' willingness to invest decreases if π or S_1 decreases or P_1 increases becomes small. In the example shown in Figure 3.2 with $RISK = 0.5$ and $N = 3$, the trustors' willingness to invest (r_1) increases if the proportion of friendly trustees (π) decreases from close to 100% to 12.5% (0.5^3) and it decreases if π decreases from 12.5% to 1.6% (0.5^6). For $N = 4$, r_1 would decrease only if π decreases from 6.3% (0.5^4) to 0.4% (0.5^8). More generally, for large N, r_1 will increase for most of the parameter space if π or S_1 decrease or if P_1 increases. Only in a small part, the effects in the opposite direction are expected. This can be interpreted as follows: If the game is repeated often enough, the trust problem is unlikely to be too severe for an investment in network embeddedness to pay off.[11]

3.4 Conclusions and discussion

We devised and analyzed a model for the simultaneous investigation of investments in and returns on information exchange relations in the context of trust problems. We modeled trust problems using the Trust Game and assumed that two trustors interact a finite number of times with the same trustee. The trustors do not know whether the trustee maximizes his payoff in the one-shot Trust Game by abusing trust but they do know the probability of interacting with such an "opportunistic trustee." We specified the conditions for the existence of an equilibrium such that the trustors establish an information exchange relation between one another in order to benefit from an extended phase of trust and trustworthiness. The major results of the analyses can be summarized in the prediction that the maximum cost that the trustors are willing to incur for establishing the information exchange relation (the trustors' willingness to invest) varies in a non-monotonic way in the size of the

[11] We owe this remark to an anonymous reviewer of the *Journal of Mathematical Sociology*.

trust problem. The trustors' willingness to invest is largest if the trust problem is neither too small nor too severe. This suggests that the formation of information exchange relations as a means to support trust and trustworthiness is most likely in trust problems of intermediate severity.

This new prediction suggests that transitivity in networks—the proportion of closed triads—might be larger in contexts in which actors interact in situations that feature substantial but also not too extreme trust problems. In addition, our model provides one possible explanation for homophily—the tendency of people in similar situations or with similar interests to link to one another. Studies provide evidence that people form long-term relations and choose to transact within such relations in order to mitigate trust problems and that they do this particularly if the trust problem is neither too small nor too severe (DiMaggio & Louch, 1998; Kollock, 1994; Simpson & McGrimmon, 2008; Yamagishi et al., 1998). However, it remains to be investigated empirically whether people establish information exchange relations in order to reap the benefits of trust and trustworthiness and whether they tend to do this especially if the trust problem is of intermediate severity.

Before we point out directions for future theoretical research, we want to briefly discuss our related work. In Raub et al. (2013), we study a similar game but assume complete information and indefinite repetition; the game ends with some positive probability after each of the periods in which all trustors interact with the trustee about whom they know that he has a short-term incentive to abuse trust. In this model, a network for information exchange does not give the trustors additional opportunities to learn about the trustee (they anyway know that he has a short-term incentive to abuse trust). However, the network does give the trustors more control over the trustee. It makes it possible that the trustee gets sanctioned for an abuse of trust by not being trusted again not only by the focal trustor but also by other trustors. Therefore, also in a situation with complete information, a network for information exchange can make trust and trustworthiness possible in situations in which it would not be possible without the network (cf. Raub & Weesie, 1990). Other than in the game studied in this chapter, the equilibria of this alternative model depend crucially on the trustee's incentives (rather than the incentives of the trustors). This alternative model, which also covers scenarios with more than two trustors as well as social dilemmas other than the Trust Game, allows deriving a number of additional results but the main conclusion is likewise that the formation of an information exchange network between the trustors is most likely if the trust problem is neither too small nor too severe.

We believe that the game introduced in this chapter offers a promising framework for addressing further questions on the formation of social relations as a means to

mitigate trust problems. Moreover, we conjecture that our main prediction is invariant to some alternative specifications of the game. First, we assumed two-sided link formation with shared costs of creating a link. This "investment rule" does not reflect that a trustor may have an incentive to freeride on the other trustor's effort to establish the relation (cf. Coleman, 1990, Chap. 12). Our main results would also hold, however, if the investment rule was such that the trustors share the cost of establishing the relation (c) if they both propose to invest and that a trustor pays the total cost if she is the only one who proposes to invest. One can check that also in this scenario, in which freeriding is possible, there exists an equilibrium such that the relation gets established only if the return on embeddedness for each trustor is at least as large as half of the total cost of establishing the relation (if $r_1 \geq c/2$).[12] The investment rule that we assumed furthermore neglects that one may regret having exerted an effort if the relation does not get established because one's effort is not reciprocated. To account for this, it could be assumed that a trustor loses her investment ($c/2$) if she is the only one who proposes to invest. It can be checked that also with this investment rule, there is an equilibrium such that the trustors establish the relation only if $r_1 \geq c/2$. Thus, our main results appear to be rather robust to the exact specification of the investment rule.

Second, our model could be adapted for the study of the formation of a complete network for information exchange between $k \geq 2$ trustors who interact with the trustee. Suppose that in total kN Trust Games are played such that every subsequent k periods, each trustor plays once with the trustee and that the order in which the trustors interact with the trustee within some k periods is determined randomly (and announced publicly) at the beginning of these periods. If there is no network for information exchange, each trustor earns the same expected payoff as in the corresponding scenario of the presented model ($U_1^{\Gamma^-}$). On the other hand, if there is a network for information exchange, each trustor can avoid $(k-1)/k$ of the (τ) interactions in which trust and trustworthiness are not certain anymore and earn an expected payoff of $((kN-\tau)R_1 + X_1 + (\tau-1)P_1))/k$, which is equal to $U_1^{\Gamma^+}$ if $k = 2$ and where $P_1 \leq X_1 < R_1$. Our approach to model returns on and investments in information exchange relations (network embeddedness) may also be adapted to model returns on and investments in long-term relations (dyadic embeddedness). Specifically, one could calculate the benefit a trustor derives from interacting repeatedly with the same trustee instead of interacting with different trustees as her expected payoff of playing a finitely repeated game of N periods minus her expected payoff of

[12] Given this alternative investment rule, the relation will get established in equilibrium by the investment of one trustor if $r_1 > c$, although this does create coordination problems similar to a Chicken Game. If $c \geq r_1 \geq c/2$, it will be an equilibrium that the trustors establish the relation jointly and if $r_1 < c/2$, there cannot be an equilibrium such that the relation gets established.

playing N one-shot games.

The presented game could furthermore be used to model the formation of interaction relations instead of information exchange relations. Suppose two trustors have a priori an information exchange relation between each other. If they interact with different trustees, each gets the payoff $U_1^{\Gamma^-}$, whereas each receives the payoff $U_1^{\Gamma^+}$ if they both interact with the same trustee. This suggest that two trustors who share an information exchange relation may be willing to "pay a premium" to a trustee who has the capacities to enter an interaction relation with both of them, particularly if the trust problem is neither negligible nor extremely severe.

Finally, our model could be adapted for the study of investments in information exchange by the trustee. It might, at first sight, seem counterintuitive that a trustee with an incentive to abuse trust in the one-shot game could want to make such an investment. After all, information exchange between the trustors restricts his possibilities to abuse trust. Yet, information exchange between the trustors leads to an extended phase of trust and trustworthiness precisely because it restricts the trustee's opportunities to abuse trust and this also benefits the trustee. Therefore, a trustee can invest in information exchange as an act of incurring a credible commitment (Raub, 2004; Schelling, 1960), namely, he can commit to remaining trustworthy for a longer phase. The analysis of a game with investments in information by the trustee will be more complicated, however, because a trustee's investment decision might signal whether he has a short-term incentive to abuse trust.

The major strength of our study is that we devised a model for an integrated and simultaneous analysis of the formation of social relations and the effects of such relations on behavior in trust problems. We derived a new prediction from this model, namely, that two trustors who interact with the same trustee are most likely to establish an information exchange relation between one another in order to reap the benefits of trust and trustworthiness if the trust problem is neither too small nor too severe.

Chapter 4

Embedding trust: Trustees' investments in network embeddedness as credible commitments and signals of trustworthiness[1]

Abstract: This chapter presents a game-theoretic model for the understanding of trust due to a trustee's investment in establishing network embeddedness—a network or platform that allows for the exchange of information among trustors. The trustee's investment can promote trust for two reasons. First, the investment of the trustee can serve as a self-binding commitment that allows for trust and trustworthiness because it creates a context in which an abuse of trust would get punished more fiercely. Second, the trustee's investment can promote trust by serving as a credible signal of intrinsic trustworthiness. The analysis of the model establishes under what circumstances a trustee is likely to establish network embeddedness to promote trust by either of these two mechanisms.

[1]This chapter is single-authored. A slightly different version of this study is to be resubmitted as a "revise and resubmit" to an international journal. I thank Vincent Buskens and Werner Raub for comments and suggestions at various stages of the research process and Jacob Dijkstra and Victor Stoica for comments on an earlier version of the manuscript.

4.1 Introduction

Consider the following example of a trust problem involving a trustee's investment in establishing network embeddedness. You find a marvelous antique bowl in an online shop. If you buy it, the seller has to decide whether to ship the advertised bowl or, instead, to abuse your trust by shipping an inferior bowl or not shipping anything at all. You would like to buy the advertised bowl but you incur a risk if placing trust and buying because it is likely that the seller can earn an extra profit by abusing your trust. On the seller's website, you see that he pays TrustorNet for its services; TrustorNet provides a platform on which the seller's customers can leave feedback on their transactions. The fact that the seller makes this expense gives you the feeling that he is trustworthy. The feedback he received on TrustorNet confirms your impression. You furthermore reason that the seller should take into account that you will leave bad feedback if he abuses your trust. You buy the bowl and place it a couple of days later happily in your living room.

The example indicates that an investment of a trustee (the seller) in establishing network embeddedness—a network or platform for the exchange of information between several trustors (the customers)—can mitigate the trust problem in two ways. First, the investment of the trustee in establishing network embeddedness can serve as a self-binding commitment (cf., Raub, 2004; Schelling, 1960; Snijders & Buskens, 2001; Williamson, 1983). If the trustors can exchange information, the trustee can get punished more fiercely for an abuse of trust—for example, by the withdrawal from further interactions not only by the focal trustor but also by other trustors (cf. Buskens & Raub, 2002). That is, network embeddedness increases the long-term costs of an abuse of trust (the "shadow of the future," Axelrod, 1984). Network embeddedness, therefore, promotes trustworthiness and—as trustworthiness begets trust—it also fosters trust. This is also expected and observed if network embeddedness is given exogenously (Bohnet & Huck, 2004; Bohnet et al., 2005; Bolton et al., 2004; Bolton & Ockenfels, 2009; Buskens et al., 2010; Cook & Hardin, 2001; DiMaggio & Louch, 1998; Huck et al., 2010; for a survey see Buskens & Raub, 2013), as, for example, if customers and sellers are embedded in a local community. If network embeddedness is not given, a trustee can have opportunities to invest in establishing it. Even a trustee who could earn an extra profit by abusing trust can then benefit from creating a context in which an abuse of trust will get sanctioned harshly. The trustee can thereby bind himself credibly to being more trustworthy and this may induce a trustor to place trust in the first place.

Second, a trustee's investment in establishing network embeddedness can resolve the trust problem by serving as a costly and honest signal of *intrinsic* trustworthiness.

Empirical evidence suggests that some trustees would honor trust even in the absence of contextual factors that deter untrustworthy behavior (see Camerer, 2003, Chap. 2, Fehr & Fischbacher, 2003, and Johnson & Mislin, 2011, for overviews of experimental results; see Hardin, 2002, Chap. 2 and Riegelsberger et al., 2005, for the distinction between intrinsic and contextual properties that induce trustworthiness). A trustor would benefit from knowing whether she is facing such an intrinsically trustworthy trustee and the trustee, too, would gain from being identified by the trustor as trustworthy.[2] However, *every* trustee has an interest in being trusted and this makes it difficult for the trustee to convince the trustor of his trustworthiness (Granovetter, 2002). Signaling theory (e.g., Gambetta, 2009; Spence, 1973; Zahavi, 1975; Bliege Bird & Smith, 2005) suggests that taking some costly, observable action can allow an actor to credibly communicate his unobservable characteristics, such as trustworthiness (Bacharach & Gambetta, 2001; Paik & Woodley, 2012; Patel, 2012; Przepiorka & Diekmann, 2013; Raub, 2004; Schroeder & Rojas, 2002). We show that a trustee's costly investment in establishing network embeddedness can serve this purpose—it can be a credible signal of intrinsic trustworthiness. Specifically, observing whether the trustee does or does not pay the cost of establishing network embeddedness enables a trustor to discriminate reliably between the trustworthy type and the mimic if the benefit from being held for intrinsically trustworthy and, hence, being trusted does compensate the trustworthy type for the cost of investment, whereas being trusted does not compensate an untrustworthy trustee for the cost of investment.

Under what circumstances is a trustee likely to make a costly investment in establishing network embeddedness to bind himself to being trustworthy? Under what circumstances is a trustee likely to establish network embeddedness to signal his intrinsic trustworthiness? How does the effect of a trustee's investment on behavior in trust interactions depend on the circumstances? This study develops theoretical answers to these questions.

To address these questions in a parsimonious way, we study a setting with only three actors and use a finitely repeated Trust Game with incomplete information (see James, 2002) as the building block of our model. In the presented game, two trustors interact with the same trustee in finitely repeated Trust Games. The trustors do not know whether the trustee is of the *friendly type* (the intrinsically trustworthy type) or the opportunistic type (the type of trustee who can earn an extra benefit by abusing trust). The trustee knows his own type and, at the beginning of the game, has the possibility to invest in establishing network embeddedness—he can make a costly investment to set up a relation for information exchange *between the*

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

two trustors. Our analysis identifies under what conditions this game has equilibria in which the trustee invests in establishing network embeddedness and in which the investment leads to more trust and trustworthiness because it serves as a self-binding commitment or because it signals that the trustee is of the friendly type. To be specific, our propositions specify what the range of the cost of establishing network embeddedness is for which the game has such equilibria and how this cost range depends on payoff relations in the Trust Game, the number of interactions between each trustor and the trustee, the probability of interacting with a friendly trustee, and the specific way the friendly trustee is modeled.

We assume that a trustee of the friendly type has non-material motivations that offset the material incentives to abuse trust.[3] A trustee could, in this sense, be intrinsically trustworthy for various reasons, for example, due to inequity aversion (Fehr & Schmidt, 1999), guilt aversion (Battigalli & Dufwenberg, 2007), altruism (Becker, 1976), or warm-glow (Andreoni, 1989). We analyze two scenarios that subsume these more specific non-material motivations. In the *game with guilt-avoiding trustees* a friendly trustee is assumed to be intrinsically trustworthy because he would suffer from an *internal sanction* if he abused trust. In the *game with reward-seeking trustees* a friendly trustee has never an incentive to abuse trust because he feels an *internal reward* if he honors trust.

The results suggest that a trustee is particularly likely to invest in establishing network embeddedness to bind himself if (i) there will be quite some trust if the trustee established network embeddedness while the situation is so severe that the trustors would not dare to place trust if they cannot exchange information and (ii) a trustee can gain a lot from being trusted. This holds for the game with guilt-avoiding trustees as well as the game with reward-seeking trustees. However, the results concerning investments in establishing network embeddedness as signals are quite different for the two versions of the game. In the game with guilt-avoiding trustees, a trustee's investment cannot serve as a signal of intrinsic trustworthiness. If a friendly trustee is trustworthy because he would feel guilty if he betrayed trust, he never benefits more from being trusted than a trustee who does not feel guilt (an opportunistic trustee). In the game with guilt-avoiding trustees it is, therefore, not possible that the friendly trustee is willing to invest some high cost to convince the trustors of his trustworthiness and that this cost is so high that it is not worthwhile

[3]We assume intrinsic trustworthiness to be a stable trait rather than a state. We assume that a trustee is either of the friendly type or of the opportunistic type; we do not take into account that non-material motives may depend on situational factors that frame the specific decision situation (e.g., Andreoni, 1995; Lindenberg & Steg, 2007). Note, furthermore, that our model is in principle not restricted to the scenario that friendly trustees are trustworthy due to non-material motivations. A friendly trustee could also be trustworthy because he has no material incentive to abuse trust, e.g., because he simply does not have an attractive option to abuse trust.

for the opportunistic trustee to mimic the friendly type by pledging the investment, too. The situation is reversed in the game with reward-seeking trustees. The friendly trustee always benefits more from being trusted than the opportunistic trustee because he derives an internal reward when honoring trust that is larger than the extra benefit that the opportunistic trustee earns when abusing trust. Therefore, a friendly trustee who gets such an internal reward when honoring trust may pledge the investment to induce the trustors to place trust even if the material benefit of inducing the trustors to trust does not fully compensate the trustee for the investment.

Our model is related to so-called "co-evolution games" (Vega-Redondo, 2007, Chap. 6) in that we treat network embeddedness as endogenous and at the same time focus on behavior in some underlying game (in our case the Trust Game). Theoretical studies on the co-evolution of networks and behavior typically assume boundedly rational actors that make choices according to some backward-looking criterion. For example, Fosco & Mengel (2011) assume that actors choose to interact with those whose partners earned high payoffs in the past. We follow a different approach and assume full strategic rationality with respect to investments in network relations as well as behavior in the Trust Games. We assume that a trustee weighs the costs and benefits of an investment in network embeddedness rationally. This is a contribution to the literature on the co-evolution of networks and behavior on its own. Frey et al. (2015b), henceforth FBR, and Raub et al. (2013) take a similar approach. The paper by FBR is a companion paper to the current study and investigates, in essentially the same game, investments in network embeddedness by the trustors rather than the trustee. Raub et al. study likewise a similar game but assume indefinite repetition and complete information; the game ends with some positive probability after each of the periods in which all trustors interact with the trustee and the trustors know that the trustee has a short-term incentive to abuse trust. As in the current study, the assumption of full strategic rationality allows Raub et al. to model and interpret investments in network embeddedness by trustees as self-binding commitments. Their model does, however, not feature the signaling of intrinsic trustworthiness. While one prerequisite for such signaling is given—actors are assumed to be strategically rational—a more fundamental prerequisite is missing from the Raub et al. model: Actors do not have any unobservable properties. To our knowledge, the current study is the first to provide an analysis of investments in establishing network embeddedness as signals of intrinsic trustworthiness. Thereby, the study provides novel insights into mechanisms that may drive the formation as well as the effects of networks.

The next section introduces the game. Subsequently, we first present a generic analysis of behavior in the Trust Games, then the analysis of investments in establishing network embeddedness as commitments, and then the analysis of investments as

signals of trustworthiness. The last section draws conclusions and indicates directions for future research. The proofs for most of the propositions are in Appendix C.

4.2 The game

This section describes the game Γ and the difference between its two versions—the game with *g*uilt-*a*voiding trustees (Γ^{ga}) and the game with *r*eward-*s*eeking trustees (Γ^{rs}). Throughout, we use the superscripts ga and rs to let notation refer specifically to either of these versions. We omit these superscripts if we refer to both versions simultaneously.

4.2.1 Actors and moves

The game Γ has three actors—two trustors and one trustee—and proceeds as follows: First, the trustee's type is determined and the trustee learns his type (period 0.1); then, the trustee decides whether to invest in establishing network embeddedness (period 0.2); finally, the trustors interact with the trustee in a finite series of binary Trust Games (periods 1 to $2N$). Specifically, in period 0.1, a random move of a pseudo-actor Nature determines the type of the trustee. With probability π, the trustee is of the *friendly type* (the intrinsically trustworthy type) and with probability $1-\pi$, he is of the *opportunistic type*, where $0 < \pi < 1$. The trustee's type is modeled via the payoffs in the Trust Game as described below. While the probability π is common knowledge and the trustee knows his own type, the trustors are not informed about the type of the trustee. In period 0.2, the trustee decides whether or not to invest the cost c to establish network embeddedness—a relation for information exchange *between the two trustors*. Then, one binary Trust Game (TG) is played in each of the periods 1 to $2N$. In every TG, the trustor at play, first, decides whether or not to place trust. If the trustor does not place trust, the TG ends. If the trustor places trust, the trustee chooses whether to honor or abuse trust and the TG ends. Every odd period starts with a random move of Nature that determines with equal and independent probability whether trustor 1 or trustor 2 plays a TG with the trustee in that period while the other trustor plays a TG with the trustee in the subsequent even period. In which sequence the two trustors interact with the trustee in an odd and the subsequent even period is publicly announced before the TG of the odd period is played.[4]

[4]This assumption on the order of play is the same as in the model of FBR. The results presented in this chapter would also obtain if one simply assumed that trustor 1 and trustor 2 take turns in interacting with the trustee (see the remark in Appendix C).

4.2.2 Payoffs

The payoffs in any TG are as follows. If trustor i who is at play does not place trust, trustor i and the trustee earn the payoffs P_1 and P_2, respectively, in that TG. If i places trust, she gets $R_1 > P_1$ if the trustee honors trust and $S_1 < P_1$ if the trustee abuses trust. The trustee gets $R_2 > P_2$ if he honors trust and $T_2 > R_2$ if he abuses trust. These are the material payoffs and we assume that the trustors and a trustee of the opportunistic type are exclusively concerned with own material payoffs. An opportunistic trustee would thus maximize his utility in *the focal* TG by abusing trust. A friendly trustee, however, has non-material motivations such that if trustor i places trust, he maximizes his utility by honoring trust. We model this in two ways.

- In Γ^{ga} (the game with guilt-avoiding trustees), a friendly trustee has no incentive to abuse i's trust because if he abuses trust, he suffers from an *internal sanction* θ that offsets the material incentive to abuse trust ($T_2 - \theta < R_2$).
- In Γ^{rs} (the game with reward-seeking trustees), a friendly trustee could abuse i's trust without feeling remorse but he has no incentive to do this because he receives an *internal reward* υ if he honors trust such that $T_2 < R_2 + \upsilon$.

An actor's total payoff for Γ (i.e., Γ^{ga} as well as Γ^{rs}) is the sum of the undiscounted payoffs that the actor earns in the TGs in the periods 1 to $2N$, minus—in case of the trustee—the cost of an investment in period 0.2 if such an investment has been pledged.

4.2.3 Information sets

The information a trustor has when the TGs are played depends on the trustee's investment decision. We refer to the continuation of Γ after the trustee did *not* establish the relation for information exchange between the trustors in period 0.2 as (*continuation game*) Γ^- and to the continuation of the game after the trustee did establish that relation as (*continuation game*) Γ^+. In Γ^-, each trustor, when making a choice in a TG, is informed about the choices that were made in her own past TGs but not about the choices that were made in the other trustor's past TGs. On the other hand, in Γ^+, the trustors exchange information directly at the end of every TG and do this always truthfully. Hence, in Γ^+ each trustor is always informed about the history of all past TGs. Which continuation game is played, i.e., whether the trustee has established network embeddedness, is common knowledge.

4.3 Analysis of the game

We solve the games Γ^{ga} and Γ^{rs} for sequential equilibria (Kreps & Wilson, 1982b; see Rasmusen, 1994, Chap. 6 for a textbook). The structure of this section reflects that one analyzes a game backwards to identify a sequential equilibrium. We first specify the sequential equilibria of the TGs after the trustee did or did not establish network embeddedness in a generic manner. Specifically, we specify the equilibria of the continuation games Γ^{-} and Γ^{+} without fixing the beliefs that the trustors hold about the trustee's type at the beginning of these continuation games. We also specify the expected payoffs of a trustee associated with the sequential equilibria of Γ^{-} and Γ^{+}. The beliefs of the trustors that enter Γ^{-} and Γ^{+} affect how much trust and trustworthiness there will be and they depend on the choices of friendly and opportunistic trustees to invest in establishing network embeddedness in period 0.2. We analyze these investment choices in the second subsection. In this analysis we further refine the sequential equilibrium concept. It is well-known that signaling games can have sequential equilibria that require implausible out-of-equilibrium beliefs. We will employ a simple refinement to discard such equilibria. Namely, we will assume that the trustors do not change their belief about the trustee's type upon observing that the trustee deviated from a conjectured equilibrium in period 0.2. This refinement is referred to as *passive conjecture* (Rasmusen, 1994; Rubinstein, 1985).[5] Throughout, we refer to sequential equilibria and sequential equilibria involving the passive conjecture for brevity as "equilibria" in case this does not cause any confusion.

4.3.1 Behavior and payoffs in the Trust Games

Our analysis of the continuation games Γ^{-} and Γ^{+} builds on the analyses of FBR, Buskens (2003), and Camerer & Weigelt (1988). Camerer & Weigelt (see also Anderhub et al., 2002, and Bower et al., 1997) apply the analysis of reputation building in sequential equilibria to finitely repeated Trust Games with incomplete information. Roughly, they show that in the sequential equilibrium also a trustee with an incentive to abuse trust in the one-shot Trust Game (an opportunistic trustee) will typically honor trust for many periods. The trustee balances the short-term incentive to abuse trust and the long-term benefit of having a reputation for being trustworthy and chooses to mimic a trustee who is intrinsically trustworthy and who would never abuse trust. However, as the end of the game comes close, the long-term benefit of a

[5] Further analyses show that very similar results as those reported in this chapter are obtained under the assumption of two other prominent equilibrium refinements for signaling games—the criterion of universal divinity (Banks & Sobel, 1987) and the concept of undefeated equilibria (Mailath et al., 1993).

good reputation decreases and the trustee (if he is of the opportunistic type) begins to abuse trust with positive probability. The trustor anticipates this and likewise begins to randomize. The trustor becomes more and more confident that the trustee is of the trustworthy type with every additional time that the trustee honors trust but the trustor will not place trust anymore after a single abuse of trust or after she did not place trust and the trustee had thus no possibility to further convince the trustor of his trustworthiness.

Buskens (2003) uses the sequential equilibrium model to study network effects on trust and trustworthiness in a scenario with two trustors and one trustee. He shows, among other things, that compared to the situation without information exchange, the trustee remains trustworthy with each trustor during more TGs if each trustor passes on information about the trustee's behavior to the other trustor with sufficiently high probability. This reflects that information exchange between the trustors gives the trustee a stronger incentive to honor trust because a single abuse of trust will lead *both* trustors to not place trust anymore. FBR simplify the game studied by Buskens by assuming that if information is exchanged at all, it is exchanged with probability 1 after every TG. And they extend the game by endogenizing the opportunity for information exchange: Before the TGs are played, the trustors can choose whether to pay some cost to establish information exchange. The analysis of FBR establishes under what conditions the trustors are most likely to pledge this costly investment.

We can lend directly from FBR for the analysis of the continuation of the game Γ after the trustee did or did not establish information exchange between the trustors. The game Γ^{ga} with guilt-avoiding trustees differs from the game studied by FBR only in that the trustee, rather than the trustors, can establish network embeddedness in period 0.2. A crucial difference in Γ^{ga} (as well as in Γ^{rs}) is that rational trustors may be able to draw inferences about the trustee's type (change their belief) already upon observing the trustee's investment decision while in the game of FBR, the trustors cannot learn anything about the type of the trustee before the TGs are played (because the trustee does not move in period 0.2). If the trustors change their belief about the type of the trustee in period 0.2, this can largely affect the course of play in the TGs. Still, given the beliefs that the trustors hold after period 0.2 and that enter the continuation games Γ^{-} and Γ^{+}—which we will denote by π^{-} and π^{+}, respectively—the same sequential equilibria apply to Γ^{ga-} and Γ^{ga+} as to the respective continuation games studied by FBR. These equilibria are specified formally in Propositions 1 and 3 of FBR, which we, henceforth, refer to as P_FBR 1 and P_FBR 3, respectively.

The equilibria specified in P_FBR 1 and P_FBR 3 apply also to the continuation games Γ^{rs-} and Γ^{rs+} of the game with reward-seeking trustees. In the TGs, the

friendly trustee honors trust whenever trust gets placed. Clearly, the *reason* for which a friendly trustee behaves like this does not affect the equilibrium behavior of the opportunistic trustee or the trustors (cf. Kreps et al., 1982). Hence, in this section we will hardly have to distinguish between Γ^{ga} and Γ^{rs} because, given the beliefs π^- and π^+, the equilibrium course of play in the TGs is the same in the two versions of Γ. We here sketch this course of play and specify the associated expected payoffs of a trustee. To this end, we define two measures pertaining to the TG, namely, $RISK := (P_1 - S_1)/(R_1 - S_1)$ and $TEMP := (T_2 - R_2)/(T_2 - P_2)$. *RISK* measures the risk a trustor incurs when placing trust; *TEMP* measures an opportunistic trustee's temptation to abuse trust (Snijders, 1996).

Behavior and payoffs in Γ^-***:*** Assume first that the trustee did not establish network embeddedness. In Γ^- (that is, in Γ^{ga-} as well as in Γ^{rs-}), the equilibrium in the interactions between each trustor i and the trustee is independent of what happens in the interactions between the other trustor and the trustee; the equilibrium is the same as if there was only trustor i, but no second trustor, who plays a finitely repeated TG with the trustee. This equilibrium evolves typically over three phases. First, trustor i places trust and the trustee honors trust until trustor i and the trustee have τ^- TGs left to play together, where τ^- is the smallest integer for which it holds that $\pi^- \geq RISK^{\tau^-}$, i.e.,

$$\tau^- := \left\lceil \frac{\log \pi^-}{\log RISK} \right\rceil \text{ for } 0 < \pi^- < 1. \tag{4.1}$$

Then, in the next TG of trustor i, the trustee—if he is of the opportunistic type—starts to randomize. Still one TG later, trustor i begins to randomize too, placing trust with probability *TEMP*. In this second phase (the randomization phase), trustor i and the trustee randomize until trustor i does not place trust or the trustee abuses i's trust. After the first instance that trustor i did not place trust or that the trustee abused i's trust, they enter the third and final phase: trustor i does not place trust anymore. Note, furthermore, that if trustor i still places trust in her last TG, an opportunistic trustee abuses trust in that TG.

It bears emphasizing that the equilibrium does not evolve over all these three phases if the situation is too severe. Eq. (4.1) shows that τ^- tends to be larger (the randomization phase tends to start earlier) if π^- is smaller or *RISK* is larger, whereas how many interactions prior to the end of the game the randomization starts is independent of N. If π^- is so small or *RISK* so large that $\tau^- = N$, the opportunistic trustee's randomization begins already in the first TG. Moreover, if $\tau^- > N$, trustor i never places trust because, given the parameters, there are not enough interactions

between trustor i and the trustee for the trustee to start building a reputation.

There will also be no randomization phase if $\pi^- = 0$ or $\pi^- = 1$. In these cases, it is as if trustor i had complete information and τ^- is not specified by eq. (4.1). It follows from backward induction that if trustor i is totally convinced that the trustee is of the opportunistic type ($\pi^- = 0$), she should never place trust, irrespectively of how long the game is. With some abuse of notation we, therefore, say that $\tau^- = \infty$ if $\pi^- = 0$. On the other hand, if $\pi^- = 1$, trustor i would even place trust in the equilibrium of the one-shot TG. In the repeated game, trustor i will then place trust in every TG and if the trustee is indeed of the friendly type, the trustee honors trust in every TG. If—contrary to i's belief—the trustee is of the opportunistic type, he honors trust in every TG except for the last TG, in which he abuses trust. This is the same course of play as if $RISK < \pi^- < 1$ and thus, by Equation (4.1), $\tau^- = 1$. We, therefore, say that $\tau^- = 1$ if $\pi^- = 1$. Proposition 4.1 specifies the expected payoffs of friendly and opportunistic trustees in Γ^{ga-} and Γ^{rs-}.

Proposition 4.1. *The expected payoffs of a friendly trustee in* Γ^{ga-} ($U_F^{\Gamma^{ga-}}$) *and* Γ^{rs-} ($U_F^{\Gamma^{rs-}}$) *are:*

$$U_F^{\Gamma^{ga-}} = \begin{cases} 2\Big((N-\tau^-)R_2 + \tau^- P_2 + (T_2 - P_2)(1 - TEMP^{\tau^-})\Big) & \text{if } \tau^- \leq N \\ 2NP_2 & \text{if } \tau^- > N \end{cases}$$

$$U_F^{\Gamma^{rs-}} = \begin{cases} 2\Big((N-\tau^-)(R_2+\upsilon) + \tau^- P_2 + (T_2 - P_2)\frac{R_2+\upsilon-P_2}{R_2-P_2}(1 - TEMP^{\tau^-})\Big) & \text{if } \tau^- \leq N \\ 2NP_2 & \text{if } \tau^- > N \end{cases}$$

The expected payoffs of an opportunistic trustee in Γ^{ga-} ($U_O^{\Gamma^{ga-}}$) *and* Γ^{rs-} ($U_O^{\Gamma^{rs-}}$) *are:*

$$U_O^{\Gamma^{ga-}} = U_O^{\Gamma^{rs-}} = \begin{cases} 2\Big((N-\tau^-)R_2 + T_2 + (\tau^- - 1)P_2\Big) & \text{if } \tau^- \leq N \\ 2NP_2 & \text{if } \tau^- > N \end{cases}$$

Proof. The equilibrium described above and specified formally in P_FBR 1 implies Proposition 4.1. For the case that $\tau^- \leq N$, a trustee's expected payoff for the interactions with each trustor i in Γ^- can be thought of as consisting of two main components. The first component pertains to the phase at the beginning of the game in which trust is placed and honored with certainty. In the game with guilt-avoiding trustees Γ^{ga} this component is $(N-\tau^-)R_2$ for either type of trustee; in the game with reward-seeking trustees Γ^{rs} this component is likewise $(N-\tau^-)R_2$ for an opportunistic trustee while it is $(N-\tau^-)(R_2+\upsilon)$ for a friendly trustee. The second component pertains to the τ^- last TGs that the trustee plays with trustor i (the second and third phase of the equilibrium). Assume, first, that the trustee is of the friendly type.

In this case, trust may break down from the second of these τ^- last TGs on because trustor i then begins to randomize, placing trust with probability *TEMP* as long as she placed trust before. A friendly trustee's expected payoff for the τ^- last TGs played with trustor i thus equals

$$\sum_{i=0}^{\tau-1}\Big(TEMP^i \cdot R_2 + (1 - TEMP^i)P_2\Big) \tag{4.2}$$

in Γ^{ga} and

$$\sum_{i=0}^{\tau-1}\Big(TEMP^i \cdot (R_2 + \upsilon) + (1 - TEMP^i)P_2\Big) \tag{4.3}$$

in Γ^{rs}. Using $\sum_{i=0}^{n} x^i = \frac{1-x^{n+1}}{1-x}$, Equations (4.2) and (4.3) can be rearranged to $\tau^- P_2 + (T_2 - P_2)(1 - TEMP^{\tau^-})$ and $\tau^- P_2 + (T_2 - P_2)\frac{R_2+\upsilon-P_2}{R_2-P_2}(1 - TEMP^{\tau^-})$, respectively. Now assume that the trustee is of the opportunistic type. If trustor i places trust in one of her τ^- last TGs, an opportunistic trustee is indifferent between, on the one hand, honoring trust and maybe being trusted again by trustor i and, on the other hand, abusing trust and certainly never being trusted again by trustor i (this indifference holds if trustworthiness in the current TG may lead to trust in a subsequent TG whereas an opportunistic trustee would always abuse trust in the last TG that he plays with trustor i). Hence, an opportunistic trustee's expected payoff for his τ^- last TGs with trustor i equals what he receives if he abuses trust in the first of these TGs and, consequently, is not trusted again in the subsequent $\tau^- - 1$ TGs, i.e., $T_2 + (\tau^- - 1)P_2$. □

Behavior and payoffs in Γ^+***:*** Now assume that the trustee established network embeddedness. In Γ^+, with information exchange between the trustors, the choice of the trustee in a given TG can affect the future choices of *both* trustors (rather than only the future choices of the trustor at play in the focal TG). This increases the long-term costs of an abuse of trust. In fact, in Γ^+, an opportunistic trustee's incentive to honor trustor i's trust in a given period n is the same as if he played *all* the remaining $2N - n$ TGs with that trustor. Therefore, the sequential equilibrium of Γ^+ is similar to that of Γ^- in the sense that the course of play that applies to the interactions between *one* trustor i and the trustee in Γ^- applies to the interactions between *both* trustors and the trustee in Γ^+. In Γ^+, both trustors place trust and the trustee honors trust until there are *in total* τ^+ TGs left, where τ^+ is calculated as τ^- but with π^+ instead of π^- (thus, if $\pi^+ = \pi^-$, $\tau^+ = \tau^-$). Then, the randomization begins and after the first instance that the trustee abuses the trust of one trustor or that one trustor does not place trust, both trustors never place trust again. Analogous

to the situation in Γ^-, the trustors never place trust in Γ^+ if $\tau^+ > 2N$ and we also assume that $\tau^+ = \infty$ if $\pi^+ = 0$ and $\tau^+ = 1$ if $\pi^+ = 1$. Proposition 4.2 specifies the expected payoffs of a trustee in Γ^{ga+} and Γ^{rs+}.

Proposition 4.2. *The expected payoffs of a friendly trustee in* Γ^{ga+} $(U_F^{\Gamma^{ga+}})$ *and* Γ^{rs+} $(U_F^{\Gamma^{rs+}})$ *are:*

$$U_F^{\Gamma^{ga+}} = \begin{cases} (2N-\tau^+)R_2 + \tau^+P_2 + (T_2-P_2)(1-TEMP^{\tau^+}) & \text{if } \tau^+ \leq 2N \\ 2NP_2 & \text{if } \tau^+ > 2N \end{cases}$$

$$U_F^{\Gamma^{rs+}} = \begin{cases} (2N-\tau^+)(R_2+\upsilon) + \tau^+P_2 + (T_2-P_2)\frac{R_2+\upsilon-P_2}{R_2-P_2}(1-TEMP^{\tau^+}) & \text{if } \tau^+ \leq 2N \\ 2NP_2 & \text{if } \tau^+ > 2N \end{cases}$$

The expected payoffs of an opportunistic trustee in Γ^{ga+} $(U_O^{\Gamma^{ga+}})$ *and* Γ^{rs+} $(U_O^{\Gamma^{rs+}})$ *are:*

$$U_O^{\Gamma^{ga+}} = U_O^{\Gamma^{rs+}} = \begin{cases} (2N-\tau^+)R_2 + T_2 + (\tau^+-1)P_2 & \text{if } \tau^+ \leq 2N \\ 2NP_2 & \text{if } \tau^+ > 2N \end{cases}$$

Proof. Proposition 4.2 is implied by the sequential equilibrium described above and specified in P_FBR 2 in a similar manner as Proposition 4.1 is implied by the sequential equilibrium of the interactions between the trustee and one trustor i in Γ^-. The only difference is that in Γ^+ it is as if the trustee played with just one trustor and played $2N$ TGs with that trustor. □

Comparison of the equilibria of Γ^- ***and*** Γ^+***:*** Before we turn to the analysis of investments in network embeddedness, we briefly compare the equilibria of the continuation games Γ^- and Γ^+. To this end, assume that the beliefs π^- and π^+ that enter these continuation games are identical and thus $\tau^- = \tau^+$. Note that $\pi^- = \pi^+$ and, hence, $\tau^- = \tau^+$ would hold if network embeddedness was exogenously given or not given. It also holds, as we shall see, if friendly and opportunistic trustees establish network embeddedness with equal probability. The comparison of the equilibria shows that if $\tau^- = \tau^+$, trust gets placed and honored in more TGs in Γ^+ than in Γ^-. The opportunistic trustee remains trustworthy with each trustor i until he has only half as many TGs left with that trustor if there is network embeddedness than if the trustors cannot exchange information. Specifically, trust gets placed and honored with certainty in the TGs of each trustor i expectedly until trustor i has $\tau^+/2$ TGs left in Γ^+ but only until trustor i has τ^- TGs left in Γ^- (where $\tau^+ = \tau^-$). This extension of the phase of trust and trustworthiness occurs if the trustors place trust in some TGs also if an abuse of trust can only be punished by the focal trustor (i.e., if $N \geq \tau^-$). A more pronounced effect may obtain if the trustors would never place trust if there is no network embeddedness ($N < \tau^-$). If $N < \tau^- = \tau^+ \leq 2N$, there will

be some phase of trust and trustworthiness if the trustors can exchange information while the trustors will never place trust otherwise.

4.3.2 Investments in network embeddedness

We now turn to the analysis of investments in network embeddedness in period 0.2. We let ρ_F and ρ_O, respectively, denote the probabilities with which a friendly trustee and an opportunistic trustee invest in establishing network embeddedness. To identify whether some combination of investment probabilities ρ_F and ρ_O can be part of a sequential equilibrium, one can proceed over two steps. First, one derives from that combination of ρ_F and ρ_O the beliefs π^- and π^+. Second, one checks whether trustees of the friendly type as well as trustees of the opportunistic type would, *given* these beliefs π^- and π^+, indeed maximize their expected payoffs by investing with the initially assumed probabilities ρ_F and ρ_O. If this is the case, there is a sequential equilibrium involving the assumed combination of investment probabilities ρ_F and ρ_O. Otherwise, the assumed combination of ρ_F and ρ_O cannot be part of a sequential equilibrium.

The beliefs π^- and π^+ that rational trustors will hold at the beginning of the continuation games Γ^- and Γ^+ follow from ρ_F and ρ_O by Bayes' rule. Specifically, in a sequential equilibrium it must hold, by Bayes' rule, that $\pi^- = \frac{(1-\rho_F)\pi}{(1-\rho_F)\pi+(1-\rho_O)(1-\pi)}$ and $\pi^+ = \frac{\rho_F\pi}{\rho_F\pi+\rho_O(1-\pi)}$. To specify with what probability a trustee should establish network embeddedness *given* the beliefs π^- and π^+, we use U^{Γ^-} and U^{Γ^+} without the subscripts F or O to refer to a trustee's expected payoffs for Γ^- and Γ^+, respectively, without explicit reference to a specific type of trustee (so U^{Γ^-} means "$U_i^{\Gamma^-}$, where $i = F, O$"). In a sequential equilibrium a trustee must invest with probability 1 if, given the beliefs π^- and π^+, it holds for his type that $c < U^{\Gamma^+} - U^{\Gamma^-}$. Conversely, a trustee must not invest if given π^- and π^+, $c > U^{\Gamma^+} - U^{\Gamma^-}$. If it holds for a trustee that $c = U^{\Gamma^+} - U^{\Gamma^-}$, he can randomize between investing and not investing. However, this equality can hold only for very specific parameter constellations because U^{Γ^+} and U^{Γ^-} do not depend in a smooth manner on ρ_F and ρ_O. That is, randomizing between investing and not investing can only be optimal for a trustee under very specific parameter constellations. We restrict the analysis to generic games Γ (i.e., we assume that the payoffs at the end of different branches of the game tree are not identical) and hence exclude that, in equilibrium, a trustee randomizes in period 0.2.

There is a complication. Namely, Bayes' rule does not always allow the trustors to draw inferences about the trustee's type in period 0.2. For example, if in a conjectured equilibrium either type of trustee invests with probability 1 (if $\rho_F = \rho_O = 1$), Bayes' rule does not allow the trustors to update their belief about the trustee's type if

the trustee deviates by not investing. If $\rho_F = \rho_O = 1$, it should never occur that the trustee does not invest. The belief π^- is then not specified (Bayes' rule yields a division by 0) and the trustors are in principle free to hold any out-of-equilibrium belief $0 \leq \pi^- \leq 1$ (the same holds for π^+ in a conjectured equilibrium in which $\rho_F = \rho_O = 0$). They could, for example, assume that only an opportunistic trustee would ever deviate from an equilibrium in which $\rho_F = \rho_O = 1$ by not investing ($\pi^- = 0$). While there is no obvious reason for such pessimism, it will artificially induce the trustees to be willing to establish network embeddedness even if the cost of investment c is "very large." As mentioned, we employ the passive conjecture refinement—the assumption that the trustors do not change their beliefs upon observing a deviation from a conjectured equilibrium (here $\pi^- = \pi$)—to avoid constructing equilibria that require out-of-equilibrium beliefs that are difficult to justify.[6]

Given that trustees do not randomize in period 0.2 but invest either with probability 0 or 1, there are just four combinations of ρ_F and ρ_O that can be part of an equilibrium of Γ: Two potential "pooling equilibria" in which the trustee invests with probability 1 (0) regardless of his type and two potential "separating equilibria" in which one type of trustee invests while the other type of trustee does not invest. In the following, we first establish under what circumstances Γ^{ga} and Γ^{rs} have pooling equilibria in which $\rho_F = \rho_O = 1$, i.e., equilibria in which the trustee invests in establishing network embeddedness regardless of his type. In such equilibria, the investment serves as a commitment. Then, we investigate the conditions for the existence of separating equilibria in which the trustee establishes network embeddedness only if he is of the friendly type. In such equilibria (in which $\rho_F = 1$ and $\rho_O = 0$) the investment is a signal of intrinsic trustworthiness.

Investments in network embeddedness as commitments: In an equilibrium in which $\rho_F = \rho_O = 1$, a trustee's investment in establishing network embeddedness serves as a self-binding commitment because it makes it possible that a potential abuse of trust gets punished by both trustors. The investment does, however, not signal intrinsic trustworthiness. If the trustee invests regardless of his type, the investment does not convey any information about his intrinsic trustworthiness and, hence, the trustors do not change their beliefs upon observing the investment. That is, in equilibria in which $\rho_F = \rho_O = 1$, π^+ equals the prior probability π. Furthermore, given the assumption of a passive conjecture, the out-of-equilibrium belief π^- likewise equals π.

[6]Note that we assume the passive conjecture only for period 0.2. In the sequential equilibrium of the TGs, it is assumed that the trustors are totally convinced that the trustee is of the opportunistic type if there is an unexpected abuse of trust during the phase of the equilibrium in which trust should be placed and honored with certainty (cf. Kreps & Wilson, 1982a). This reflects that an opportunistic trustee could benefit from such an abuse of trust if the trustors "are forgiving" whereas a friendly trustee would regret the abuse of trust whatever the trustors' reaction.

The trustors are not more or less optimistic about the trustee's intrinsic trustworthiness than they were initially after observing that the trustee indeed invested as well as after observing that the trustee deviated form the conjectured equilibrium by not investing. Formally, in equilibria in which $\rho_F = \rho_O = 1$, $\pi^- = \pi^+ = \pi$ and, hence, $\tau^- = \tau^+$. Proposition 4.3 establishes what the maximum cost of investment c is for which Γ has such equilibria. We use τ without a superscript $-$ or $+$ to simplify the notation in propositions and discussions concerned with such equilibria. Specifically, we use τ without a superscript to denote simultaneously how many TGs need to be left *between trustor i and the trustee* before an opportunistic trustee would start to randomize in the interactions with trustor i in Γ^- (i.e., τ^-) as well as how many TGs need to be left *in total* before the trustee's randomization starts in Γ^+ (i.e., τ^+).

Proposition 4.3. *The game Γ^{ga} has a sequential equilibrium in which $\rho_F = \rho_O = 1$ if and only if $c \leq \bar{c}^{ga}$, where*

$$\bar{c}^{ga} = \begin{cases} \tau(R_2 - P_2) - (T_2 - P_2) & \text{if } \tau \leq N \\ (2N - \tau)(R_2 - P_2) + (T_2 - P_2)(1 - TEMP^{\tau}) & \text{if } N < \tau \leq 2N. \end{cases}$$

The game Γ^{rs} has a sequential equilibrium in which $\rho_F = \rho_O = 1$ if and only if $c \leq \bar{c}^{rs}$, where

$$\bar{c}^{rs} = \begin{cases} \tau(R_2 - P_2) - (T_2 - P_2) & \text{if } \tau \leq N \\ (2N - \tau)(R_2 - P_2) + (T_2 - P_2) & \text{if } N < \tau \leq 2N. \end{cases}$$

The proof of Proposition 4.3 is in Appendix C, together with the proofs of all further propositions of this chapter.

In general, Γ has an equilibrium in which either type of trustee invests if the cost of investment c does not exceed the benefit of the type of trustee who benefits least from investing, i.e., if $c \leq \bar{c} = \min(U_F^{\Gamma^+(\tau)} - U_F^{\Gamma^-(\tau)}, U_O^{\Gamma^+(\tau)} - U_O^{\Gamma^-(\tau)})$, where, for example, $U_F^{\Gamma^+(\tau)}$ denotes the expected payoff of a friendly trustee for Γ^+ given τ^+ as it follows from $\pi^+ = \pi$. If $c > \bar{c}$, at least one type of trustee would be better off if he deviates from the conjectured equilibrium by not investing in establishing network embeddedness. Note that, given τ, the expected payoffs for the TGs of the game with guilt-avoiding trustees and the game with reward-seeking trustees are the same for an opportunistic trustee ($U^{\Gamma^{ga-}(\tau)} = U^{\Gamma^{rs-}(\tau)}$ and $U^{\Gamma^{ga+}(\tau)} = U^{\Gamma^{rs+}(\tau)}$) and rather similar for a friendly trustee. It is, therefore, not surprising that $\bar{c}$ is quite similar for both versions of Γ.

Proposition 4.3 distinguishes two scenarios in the specification of $\bar{c}$. In the scenario with $\tau \leq N$, both trustors place trust in some TGs also if the trustee deviates from the

conjectured equilibrium by not establishing network embeddedness and the trustee gets the opportunity to abuse each trustor's trust. If the trustee does establish network embeddedness, trust will be placed and honored with certainty in the interactions of each trustor i until that trustor has half as many TGs left than if the trustee did not invest (expectedly until trustor i has $\tau/2$ TGs left instead of until i has τ TGs left) but the trustee gets only once the opportunity to abuse trust. In this scenario, $\bar{c}$ is the same in Γ^{ga} and Γ^{rs} and equals an opportunistic trustee's "return on investment," i.e., $U_O^{\Gamma^+(\tau)} - U_O^{\Gamma^-(\tau)}$. It equals how much an opportunistic trustee benefits from trust being placed and honored in τ additional TGs minus what he loses from having the opportunity to abuse trust only once rather than twice. A friendly trustee does expectedly benefit more from establishing network embeddedness than an opportunistic trustee because he has no interest in abusing trust, anyway.

In the second scenario, where $N < \tau \leq 2N$, the trustors never place trust if the trustee deviates by not establishing network embeddedness because, given the parameters, the threat of the future punishment of one trustor alone is not sufficient to induce the trustee to start building a reputation. In this case, the maximum cost of investment $\bar{c}$ for which there exists an equilibrium in which $\rho_F = \rho_O = 1$ reflects a trustee's benefit from being trusted in some TGs in Γ^+ compared to never being trusted in Γ^-. If $N < \tau \leq 2N$, $\bar{c}$ is somewhat smaller in the game Γ^{ga} with guilt-avoiding trustees than in the game Γ^{rs} with reward-seeking trustees; $\bar{c}^{ga}$ is by $(T_2 - P_2)TEMP^{\tau}$ smaller than $\bar{c}^{rs}$. An opportunistic trustee's return on investment is the same in Γ^{ga} and Γ^{rs} but a friendly trustee's return on investment is smaller in Γ^{ga} than in Γ^{rs}. In Γ^{ga}, it is the friendly trustee who benefits less from pledging the investment, because establishing network embeddedness enables an opportunistic trustee not only to benefit from honored trust in some TGs but also gives him the opportunity to earn an extra profit by abusing trust. In Γ^{rs}, however, a friendly trustee benefits always more from inducing the trustors to place trust than an opportunistic trustee (irrespective of whether the latter would honor or abuse trust when trusted) and, hence, the opportunistic trustee's return on the investment determines $\bar{c}^{rs}$.

Our next proposition establishes how changes in the parameters of the game affect $\bar{c}^{ga}$ and $\bar{c}^{rs}$. Almost all parameters affect $\bar{c}^{ga}$ and $\bar{c}^{rs}$ in the same manner (there is only a minor difference in the effect of changes in R_2). We, therefore, use $\bar{c}$ without the superscripts ga or rs to refer to $\bar{c}^{ga}$ and $\bar{c}^{rs}$ simultaneously.

Proposition 4.4. *The maximum cost $\bar{c}$ for which Γ has a sequential equilibrium in which $\rho_O = \rho_F = 1$ depends on the parameters of the game as follows.*

(1) • *If $\tau \leq N$, $\bar{c}$ increases if τ increases due to an increase in P_1 or a decrease in π, S_1, or R_1.*
• *If $N < \tau \leq 2N$, $\bar{c}$ decreases if τ increases due to an increase in P_1 or a decrease in π, S_1, or R_1.*

(2) • *If $N < \tau \leq 2N$, $\bar{c}$ increases in N.*

(3) • *If $\tau \leq 2N$, $\bar{c}^{ga}$ increases in R_2.*
• *If $\tau < 2N$, $\bar{c}^{rs}$ increases in R_2.*
• *If $1 < \tau \leq 2N$, $\bar{c}$ decreases in P_2.*
• *If $\tau \leq N$, $\bar{c}$ decreases in T_2.*
• *If $N < \tau \leq 2N$, $\bar{c}$ increases in T_2.*

(1) of Proposition 4.4 concerns the effects of changes in the probability of interacting with a friendly trustee (π) and the risk a trustor incurs when placing trust ($RISK = (P_1 - S_1)/(R_1 - S_1)$). These parameters affect how much a trustee profits from establishing network embeddedness as they determine (together with N) in how many additional TGs the trustee benefits from being trusted if he establishes network embeddedness. Recall that π and *RISK* together determine how many interactions need to be left before the randomization phase starts; the randomization tends to start earlier (τ tends to be larger) if π is smaller or *RISK* is larger. (1) of Proposition 4.4 shows that $\bar{c}$ first increases if π or *RISK* changes such that the randomization phase starts earlier (τ increases) and then, after some point, $\bar{c}$ decreases again if π or *RISK* further changes such that there need to be more and more TGs left before the randomization starts. $\bar{c}$ begins to decrease in τ when τ gets so large that $\tau > N$. To understand this result, assume first that π and *RISK* are such that $\tau \leq N$. In this case, the trustee's investment leads to an extension of the phase of trust and trustworthiness; it enables the trustee to benefit from being trusted with certainty in $\tau - 1$ additional TGs. Clearly, this is more valuable if τ is larger; thus $\bar{c}$ is larger if τ is larger. But once τ is so large that it exceeds N, ($N < \tau \leq 2N$), $\bar{c}$ becomes smaller if τ further increases because there will then be less trust and trustworthiness in Γ^+ while there was already no trust in Γ^- before the increase in τ.

In how many additional TGs a trustee benefits from being trusted if he establishes network embeddedness also depends on the number of interactions between each trustor and the trustee (N). (2) of Proposition 4.4 establishes that $\bar{c}$ does not change in N if trust is also possible in Γ^- before and after the increase in N. In this case, an increase in N means adding one TG for each trustor in which the trustee

would benefit from trust being placed and honored in Γ^- as well as in Γ^+. However, $\bar{c}$ increases in N if trust is possible in Γ^+ but not in Γ^- before and after the increase in N. In this case, an increase in N means adding one TG for each trustor in which the trustee would not be trusted in Γ^- but benefits from trust being placed and honored in Γ^+.

(3) of Proposition 4.4 concerns the effects of changes in the TG payoffs of the trustee. It shows that if the benefit a trustee derives from honored trust compared to no trust ($R_2 - P_2$) is larger, the trustee profits more from trust being placed and honored in more TGs after establishing network embeddedness. Thus, $\bar{c}$ increases in $R_2 - P_2$. (3), furthermore, establishes that $\bar{c}$ decreases in the payoff an opportunistic trustee earns when abusing trust (T_2) if $\tau \leq N$ but increases in T_2 if $N < \tau \leq 2N$. If $\tau \leq N$, $\bar{c}$ decreases in T_2 because if T_2 is larger, an opportunistic trustee suffers more from having the opportunity to abuse trust only once in Γ^+ rather than twice in Γ^-. On the other hand, if $N < \tau \leq 2N$, an opportunistic trustee has a larger incentive to invest if T_2 is larger because he will be trusted and get an opportunity to abuse trust only after an investment. If $N < \tau \leq 2N$, a friendly trustee's incentive to invest increases in T_2, too. If T_2 is larger, the trustors place trust in the randomization phase with higher probability. Hence, the number of TGs in which a friendly trustee expectedly benefits from trust being placed and honored after having established network embeddedness increases in T_2.

There are two further observation that relate to the TG payoffs of the trustee. First, the described dependence of $\bar{c}$ on T_2 implies that if T_2 is especially large, $\bar{c}$ will be small if $\tau \leq N$ but large if $N < \tau \leq 2N$. Hence, if T_2 is especially large, a small change in π, $RISK$, or N can bring about a large change in $\bar{c}$. Second, if $\tau \leq N$, Γ may have no equilibrium in which $\rho_F = \rho_O = 1$ even for an arbitrarily small cost of investment $c > 0$ and even though trust would be placed an honored in more TGs in Γ^+ than in Γ^-. This occurs if the loss that an opportunistic trustee incurs from having only once the opportunity to abuse trust in Γ^+ rather than twice in Γ^- (namely, $T_2 - R_2$) is large while the benefit an opportunistic trustee derives from an extended phase of trust and trustworthiness ($\tau(R_2 - P_2)$) is small. In that case, an opportunistic trustee is better off in Γ^- than in Γ^+.

Summarizing, Γ can have pooling equilibria in which the trustee establishes network embeddedness, regardless of his type, to bind himself to being more trustworthy. The condition for the existence of such equilibria is almost the same in the game with guilt-avoiding trustees (Γ^{ga}) and in the game with reward-seeking trustees (Γ^{rs}). In general, the maximum cost for which Γ^{ga} and Γ^{rs} have equilibria in which either type of trustee establishes network embeddedness is high if (i) a trustee's gain from honored trust compared to the situation of withheld trust ($R_2 - P_2$) is large and (ii) the

probability of interacting with an intrinsically trustworthy trustee (π) and the risk a trustor incurs in a TG when placing trust ($RISK$) are neither too small nor too large such that compared to the situation without network embeddedness, trust will be placed and honored with certainty in many additional TGs if the trustee establishes network embeddedness. Let us consider three scenarios to illustrate condition (ii). First, if π is large and $RISK$ is small, trust will be placed in almost all TGs even without network embeddedness and, hence, a trustee cannot gain much from establishing network embeddedness. Similarly, second, if π is very small and $RISK$ is very large, trust will be placed and honored in very few TGs even if the trustee establishes network embeddedness and the trustees are, therefore, likewise not willing to pay a large cost for doing this. However, third, the trustee can benefit a lot from establishing network embeddedness and $\bar{c}$ is large if π and $RISK$ are such that trust will be placed and honored with certainty in about half of the TGs if the trustee establishes network embeddedness while if there is no network embeddedness, the trustors are rather indifferent between never placing trust and giving the trustee a chance to start building a reputation (τ is close to N). The analysis shows, furthermore, that $\bar{c}$ will be highest if π and $RISK$ are such that trust will be placed and honored with certainty in almost half of the interactions if the trustee establishes network embeddedness while trust will never be placed if the trustee does not establish network embeddedness (τ is just larger than N) and the extra benefit an opportunistic trustee can earn by abusing trust ($T_2 - R_2$) is large.

Investments in network embeddedness as signals: We now focus on separating equilibria in which the trustee invests in establishing network embeddedness if he is of the friendly type ($\rho_F = 1$) but not if he is of the opportunistic type ($\rho_O = 0$). In such an equilibrium, the investment in establishing network embeddedness is a perfectly discriminating signal of intrinsic trustworthiness. After observing the trustee's investment decision, the trustors know whom they are dealing with ($\rho_F = 1$ and $\rho_O = 0$ together imply $\pi^- = 0$ and $\pi^+ = 1$) and always place trust if the trustee invested but never place trust if the trustee did not invest ($\tau^- = \infty$ and $\tau^+ = 1$).[7] The results concerning such separating equilibria are quite different for the game Γ^{ga} with guilt-avoiding trustees and the game Γ^{rs} with reward-seeking trustees and are, therefore, presented separately. We have the following proposition for the game with guilt-avoiding trustees.

[7] It is obvious that the other candidate for a separating equilibrium, namely the situation that only the opportunistic trustee invests, can never be part of a sequential equilibrium. The opportunistic trustee would take the cost of establishing network embeddedness, thereby reveal himself as being of the opportunistic type, and, consequently, never be trusted. More formally, $\rho_F = 0$ and $\rho_O = 1$ together imply (by Bayes' rule) $\pi^- = 1$ and $\pi^+ = 0$ and, hence, $\tau^- = 1$ and $\tau^+ = \infty$; but given $\tau^- = 1$ and $\tau^+ = \infty$, an opportunistic trustee would better deviate and not invest.

Proposition 4.5. *The game* Γ^{ga} *cannot have a sequential equilibrium in which* $\rho_F = 1$ *and* $\rho_O = 0$.

This result reflects the well-known fact that signaling can only work if the "good type" benefits more from being identified as such than the "bad type" (given that both types face the same cost of emitting the signal). More specifically, an equilibrium in which an investment in network embeddedness serves as a signal of intrinsic trustworthiness requires that a friendly trustee benefits more from always being trusted compared to never being trusted than an opportunistic trustee. However, this is not the case in Γ^{ga}, where a friendly trustee honors trust because he would suffer an internal sanction $\theta > T_2 - R_2$ if he abused trust. If the trustors never place trust, the trustee earns $2NP_2$ regardless of his type. But if the trustors always place trust, the trustee earns more if he is of the opportunistic type than if he is of the friendly type, namely, $(2N-1)R_2 + T_2 > 2NR_2$, because an opportunistic trustee can abuse trust in the last TG without feeling remorse. Hence, in Γ^{ga}, it cannot be that the trustors interpret an investment as a signal of intrinsic trustworthiness and that it is beneficial for the friendly trustee to pledge the investment while, at the same time, the cost of investment is high enough to deter mimicry by the opportunistic trustee.

A remark is in order. The game Γ^{ga} could have a separating equilibrium in which $\rho_F = 1$ and $\rho_O = 0$ if friendly trustees could establish network embeddedness at lower costs than opportunistic trustees. Specifically (as can be inferred from the calculations provided in the proof of Proposition 4.5), Γ^{ga} could have a separating equilibrium in which $\rho_F = 1$ and $\rho_O = 0$ if $c_F + T_2 - R_2 \leq c_O$, where c_F and c_O denote the cost of establishing network embeddedness for friendly and opportunistic trustees, respectively. Assuming that different types face different costs of sending some signal is not uncommon (see, e.g., Gintis et al., 2001; Patel, 2012). In the current context, this assumption could be defended based on arguments along the line that a trustee who has social preferences (a friendly trustee) is more adept at bringing people together than a trustee who is exclusively interested in material outcomes or that a friendly trustee even derives some pleasure from doing this. We focus here, however, on whether signaling trustworthiness is possible if trustworthy and untrustworthy actors differ only in the benefit they derive from how emitting the costly signal (establishing network embeddedness) changes the subsequent course of play. Our next and last proposition establishes that such signaling is possible in Γ^{rs} where a friendly trustee receives an internal reward υ every time that he honors trust.

Proposition 4.6. *The game Γ^{rs} has a sequential equilibrium in which $\rho_F = 1$ and $\rho_O = 0$ (and in which the trustee is always trusted after having established network embeddedness but never trusted otherwise) if and only if $2N(R_2 - P_2) + T_2 - R_2 \leq c \leq 2N(R_2 + \upsilon - P_2)$.*

To understand why signaling intrinsic trustworthiness via establishing network embeddedness is possible in Γ^{rs}, realize that in Γ^{rs} a friendly trustee benefits always more from being trusted in some TG than an opportunistic trustee, irrespectively of whether the latter honors or abuses trust (because $R_2 + \upsilon > T_2$). Hence, in Γ^{rs}, a friendly trustee benefits more from inducing the trustors to trust than an opportunistic trustee. It can, therefore, be worthwhile for a friendly trustee to pay some high cost of investment c while an opportunistic trustee is better off if he does not pledge the investment and is never trusted than if he mimics the friendly trustee by establishing network embeddedness, too.

An equilibrium in which the investment serves as a credible signal of trustworthiness can only exist if the cost of establishing network embeddedness is "rather high." An opportunistic trustee will abstain from mimicking the friendly type only if the cost of investment is so large that he prefers *never being trusted* over pledging the investment and *always being trusted*. Clearly, this can only be the case if the cost of investment is higher than the cost $\bar{c}^{rs}$ for which Γ^{rs} has a pooling equilibrium in which also the opportunistic trustee invests and in which the investment has a less pronounced effect on the behavior in the TGs. The lower bound for c for which there can be an equilibrium in which $\rho_F = 1$ and $\rho_O = 0$ is high in particular if N is large, $R_2 - P_2$ is large, and / or T_2 is large. On the other hand, the cost c must be small enough such that a friendly trustee can afford the investment. The maximum cost of investment c for which a friendly trustee will signal his trustworthiness by pledging the investment increases in N, $R_2 - P_2$, and υ. The cost range for which Γ^{rs} has a separating equilibrium is larger if the difference in the benefit that friendly and opportunistic trustees derive from always being trusted compared to never being trusted is larger. This is the case if N is larger (since a friendly trustee can then get the internal reward υ more often), if υ is larger, and if an opportunistic trustee's extra benefit from abusing trust ($T_2 - R_2$) is smaller.

Summarizing, in the game Γ, the investment in establishing network embeddedness by the trustee can only signal the trustee's intrinsic trustworthiness if trustworthy trustees benefit more from being trusted than untrustworthy trustees. The game Γ^{ga} with guilt-avoiding trustees can, therefore, not have a separating equilibrium in which the trustee's investment is a credible signal of intrinsic trustworthiness. There would be room for such an equilibrium in Γ^{ga} only if one assumed that intrinsically trustworthy trustees face lower costs of establishing network embeddedness than opportunistic

trustees. However, the game Γ^{rs} with reward-seeking trustees can have an equilibrium in which the trustee establishes network embeddedness only if he is of the friendly type (even if friendly and opportunistic trustees face the same cost of investment). Such an equilibrium can only exist if the cost of investment c is rather high and the condition for the existence of such an equilibrium is less restrictive if the number of interactions between each trustor and the trustee is large, trustworthy trustees derive a large internal reward from honoring trust, and untrustworthy trustees cannot gain too much from abusing trust.

4.4 Conclusions and discussion

In this chapter, a game-theoretic model has been developed for the understanding of trust due to an investment in establishing network embeddedness by the trustee. A costly investment of the trustee in establishing information exchange between several trustors with whom he interacts can promote trust in two ways. First, a trustee's investment can serve as a self-binding commitment because it makes it possible that an abuse of trust gets punished more severely. Second, a trustee's costly investment can serve as a credible signal of intrinsic trustworthiness.

A key result of this study is the prediction that a trustee is particularly likely to establish network embeddedness in order to bind himself if (i) a trustee's gain from honored trust compared to the situation of withheld trust ($R_2 - P_2$) is large and (ii) the probability of interacting with a trustworthy trustee (π) and the risk a trustor incurs when placing trust ($RISK$) are neither too small nor too large such that compared to the situation without network embeddedness, trust will be placed and honored with certainty in many additional TGs if the trustee establishes network embeddedness. To return to the example given in the introduction, this suggests that an antiquities seller who has specialized in bowls may be more likely to pay the provider of an external reputation platform to make it possible for his customers to exchange information than a seller of ordinary vases or a seller of old-timer cars. For a seller of ordinary vases the investment might not pay off because potential customers also dare to trust him in the absence of network embeddedness. A seller of old-timer cars may have only a small incentive to establish network embeddedness because potential customers may not be ready to assume the large risk associated with buying even if there is network embeddedness (and the seller might have to provide a warrenty in addition).

This result is in line with and adds to the results of FBR and Raub et al. (2013; see also Raub et al., 2014). FBR analyze the scenario that the trustors, rather than the trustee, can establish network embeddedness. They report qualitatively the same effects of changes in π and $RISK$ (as well as N) on the maximum cost the trustors

are willing to pay to establish network embeddedness. Raub et al. (2013) study a similar game but assume indefinite repetition and complete information. Other than in the games studied in the current chapter as well as in the game studied by FBR, the course of play in the TGs in the equilibria of the game studied by Raub et al. depends exclusively on the trustee's incentives and not at all on the incentives of the trustors (as is common for indefinitely repeated games with complete information). Still, this alternative model, which also covers scenarios with more than two trustors as well as social dilemmas other than the Trust Game, leads to a similar conclusion: Investments in the establishment of network embeddedness are predicted to be more likely if the incentives to abuse trust are neither too small nor too large. Together, these complementary models thus suggest that investments in establishing network embeddedness are most likely if the trust problem is intermediate—not negligibly small and also not insurmountably large. Future research could test this prediction empirically.

The second key result of this study is that an intrinsically trustworthy trustee may invest in establishing network embeddedness to credibly signal to the trustors that he is of the trustworthy type. Such signaling may occur in particular if the cost of investment is high, the trustors interact with the trustee many times, and the trustworthy trustee derives a large internal reward from honoring trust while untrustworthy trustees cannot earn a too high extra benefit by abusing trust. Signaling intrinsic trustworthiness via establishing network embeddedness is not possible, however, if intrinsically trustworthy trustees are trustworthy because they would suffer from an internal punishment when abusing trust. This reflects that if different types face the same costs of sending some signal, communicating unobservable properties via sending that signaling is only possible if the "good" types profit more from being identified as such than the "bad" types profit from being wrongly held for being of the good type.

It bears emphasis that a trustee's investment in establishing network embeddedness promotes trust for different reasons and to a different extent in a separating equilibrium (in which a trustee of the trustworthy type invests to signal his trustworthiness) than in a pooling equilibrium (in which the trustee invests regardless of his type to bind himself to being (more) trustworthy). In a pooling equilibrium, the investment of the trustee does not convey any information about his type and promotes trust and trustworthiness exclusively via the effects of information exchange between the trustors. In a separating equilibrium, on the other hand, the trustors know that they are dealing with a trustworthy trustee if the trustee pledges the investment and there is, hence, no more need for the disciplining effects of information exchange. The effect of the trustee's investment on behavior in the TGs is more pronounced in the

latter type of equilibrium. In a pooling equilibrium, the investment of the trustee leads to trust being placed and honored in some more TGs. In a separating equilibrium, the investment has a stronger effect: The trustors never place trust if the trustee does not invest while they place trust throughout if the trustee does invest. Hence, the presented results suggest that an investment by the trustee in establishing network embeddedness may have a strong effect on behavior in the trust interactions in particular if the cost of this investment is especially high while the trustors interact with the trustee many times and trustworthy trustees derive a large internal reward from honoring trust. Furthermore, this also suggests that network embeddedness may promote trust more if it has been established by the trustee than if it was given exogenously or established by the trustors.

It should, furthermore, be noted that in a separating equilibrium, the trustee's investment is a "wasteful signal." The trustee could convince the trustors of his trustworthiness by burning resources in any other manner than by creating a context that *would* promote trust and trustworthiness *if* the trustors were uncertain about his type. It is not difficult to see that the presented results concerning separating equilibria can be interpreted more generally as predicting under what circumstances a trustee who interacts in finitely repeated Trust Games with two trustors will send *any* signal of cost c to convey his intrinsic trustworthiness. In this sense, the presented analysis of signaling trustworthiness is complementary to earlier work on signaling trustworthiness in one-shot games Bacharach & Gambetta (2001); Raub (2004) and in infinitely repeated games Przepiorka & Diekmann (2013).

Although there is an abundance of actions a trustee could take to signal intrinsic trustworthiness, we think that establishing network embeddedness may be a particularly attractive option. First, trustors have an interest in acquiring information about a trustee's reputation. It is therefore unlikely, for example, that it escapes the attention of potential customers that the keeper of an online shop pays for the use of some external platform on which his customers can report on their experiences. At least, it seems plausible that it is more likely that the potential customers would overlook, for example, a notice that reports a charitable donation (see Fehrler & Przepiorka, 2013; Milinski et al., 2002, for studies on charitable giving as a signaling device). An investment in establishing network embeddedness might thus be a signal of trustworthiness with high "broadcast efficiency" (Bliege Bird & Smith, 2005; Gintis et al., 2001). Second, it may be uncertain whether trustors interpret some signal as a sign of trustworthiness. If this is the case, the trustee might rather want to establish network embeddedness than to display some other signal because if the signal fails to convey its intended message, he can still benefit from having created a context that promotes trustworthiness and, thus, trust. Finally, trustors might not be totally convinced of

the trustee's trustworthiness also after having observed the signal (for example due to some randomness in the cost of sending the signal that may make it possible for an untrustworthy trustee to afford sending the signal). Information exchange between trustors could then still further increase the level of trust after signaling.

These considerations indicate that it might be interesting for future research to study models in which signaling and network effects complement one another. This could provide further insights into the circumstances under which signaling intrinsic trustworthiness via investments in network embeddedness might be particularly likely. Furthermore, it might provide insight into how the cumulative working of signaling and network effects can render some level of cooperation possible in particularly adverse social dilemma situations in which neither of the effects alone suffices to allow for cooperation.

Chapter 5

Investments in and returns on embeddedness: An experiment with Trust Games[1]

Abstract: Theory implies that trust problems can be overcome if the interacting actors are embedded in social structures through which reputation information disseminates. That "embeddedness" is expected to promote trust and trustworthiness suggests that actors may exert effort to establish embeddedness. Theory, furthermore, suggests stronger effects of embeddedness on trust and trustworthiness and, hence, also stronger incentives to establish embeddedness if the size of the trust problem is intermediate, rather than very small or very large. We tested these predictions in a laboratory experiment in which 342 subjects played repeated Trust Games with exogenous or endogenous embeddedness under varying sizes of the trust problem. The results confirm that embeddedness promotes trust and trustworthiness and show that the effect of embeddedness tends to be stronger if embeddedness is endogenous rather than exogenous. However, we find no systematic variation in investments in embeddedness or effects of embeddedness in the size of the trust problem.

[1]This chapter presents joint work with Vincent Buskens and Rense Corten. This study is currently under review at an international journal. Frey (first author) wrote the manuscript; Frey, Buskens and Corten jointly designed the experiment; Frey programmed and executed the experiment; Frey, Buskens, and Corten analyzed the data. Comments of Werner Raub and participants of the June 2014 International Conference on Experimental Social Science on Social Dilemmas are gratefully acknowledged.

5.1 Introduction

The dissemination of reputation information often promotes trust and trustworthiness in social and economic exchange. Good ratings of past buyers give us the trust necessary to buy goods online (Diekmann et al., 2014; Resnick et al., 2006), bureaus that document credit histories facilitate credit lending (Brown & Zehnder, 2007; Djankov et al., 2007), and we do not hesitate to lend valuable personal belongings to friends, expecting that they do not like if we have to tell common friends of their neglect. The hypothesis that the "embeddedness" (Granovetter, 1985) of exchanges in networks or other social structures for the sharing of reputation information ("network embeddedness") promotes trust and trustworthiness found support in several empirical studies (e.g., Bohnet & Huck, 2004; Bolton et al., 2004; Buskens et al., 2010; DiMaggio & Louch, 1998; Huck et al., 2010; see Buskens & Raub, 2013, for an overview of the theoretical and empirical literature). Trust problems are also mitigated by embeddedness in long-term relations of the same partners ("dyadic embeddedness"). However, in the current study we focus on network embeddedness and, henceforth, often refer to "network embeddedness" simply as "embeddedness" (see Buskens & Raub, 2002, for the distinction).

This chapter presents an experimental study that expands the literature on embeddedness and trust in two respects. First, if embeddedness can make mutually beneficial exchanges possible that would not be possible without embeddedness, actors may actively seek embeddedness (cf. Flap, 2004). The few studies that investigate this idea indicate that people choose exchange partners such as to benefit from embeddedness. Brown et al. (2004), Kirman (2001), Kollock (1994), and Simpson & McGrimmon (2008) show that people respond to trust problems by forming long-term exchange relations to profit from dyadic embeddedness. DiMaggio & Louch (1998) find that people tend to turn to sellers who are embedded in their network when purchasing goods of which they cannot readily assess the quality.

Choosing to transact with embedded partners is one way to reap the benefits of embeddedness. Alternatively, actors can establish network relations or other social structures for the exchange of reputation information to "embed" previously unembedded transaction partners. In this sense, trust problems create incentives for online traders to pay for the services of external reputation platforms, for banks to invest in setting up information sharing systems, and for people to introduce their friends to each other. Frey (2014), Frey et al. (2015b), and Raub et al. (2013) develop game-theoretic models for the understanding of such costly investments in establishing embeddedness as a means to support trust and trustworthiness. Here, we provide a first empirical investigation of this idea.

Our second contribution is investigating under what conditions network embeddedness promotes trust more or less strongly. As Mizruchi et al. (2006, p. 310; see also Portes & Sensenbrenner, 1993) note, studies on embeddedness "have gone far in demonstrating that networks matter, but they have contained the seeds of something more: That the extent to which networks matter varies across actors and situations." Work in this direction includes Bohnet et al. (2005) who show that the effect of networks for the exchange of reputation information may be stronger if actors who can abuse trust can observe how others in the same position build a reputation for trustworthiness, Huck et al. (2012) who show that competition for interaction partners may amplify the effect of embeddedness, and Simpson & McGrimmon (2008) who find that low-trustors are more sensitive to embeddedness.

In this chapter, we test the hypothesis derived by Frey (2014), Frey et al. (2015b), and Raub et al. (2013) that the degree to which network embeddedness promotes trust and trustworthiness follows an inverted U-shape in the size of the trust problem. The embeddedness effect should be strongest if the trust problem is of intermediate size. If the trust problem is very small—e.g., if you are utterly convinced of the conscientiousness of your friend—there may be a lot of trust also without embeddedness and, hence, the effects of embeddedness are small. If the trust problem is very large—e.g., if a client of a bank asks for a credit to do business in an industry that is known for notoriously unreliable entrepreneurs—there will be hardly any trust even if there is embeddedness and, hence, the effects of embeddedness are likewise small. We, furthermore, test a second and related inverted U-shape hypothesis. As the theoretically expected returns on network embeddedness (the increase in earnings in trust interactions due to embeddedness) follow an inverted U-shape in the size of the trust problem, Frey (2014), Frey et al. (2015b), and Raub et al. (2013) hypothesize that also the inclination of actors to exert effort to establish network embeddedness is higher if the size of the trust problem is intermediate rather than very small or very large.

The results of our experiment, furthermore, shed light on whether there are differences in the effects of endogenous and exogenous network embeddedness. Studies indicate that the degree to which specific institutions mitigate social dilemmas can depend on whether they were chosen endogenously or imposed exogenously (e.g., Gürerk et al., 2014; Sutter et al., 2010). Schneider & Weber (2013), for example, show that a longer interaction duration (dyadic embeddedness) promotes cooperation more if chosen endogenously rather than imposed exogenously. It is conceivable that this is also the case for network embeddedness. Endogenous network embeddedness might promote trust especially strongly because a trustee's investment in establishing embeddedness could serve as a costly signal of trustworthiness and because trustors who

are more sensitive to information from third parties could be particularly inclined to establish embeddedness.

The remainder of the chapter is organized as follows. In Section 5.2.1, we describe the strategic setting in which we study investments in and returns on embeddedness. That is, in Section 5.2.1, we describe the "game" in which the participants of our experiment interacted. To enhance the understanding of this game, we explain in this section also how it was implemented in the experiment. In Section 5.2.2, we provide information on the design of the experiment. In Section 5.3, we sketch the game-theoretic analysis of the specific games played in the experiment and state the hypotheses. We present the results in Section 5.4 and conclude and point out directions for future research in Section 5.5.

5.2 The strategic setting and its implementation in the experiment

5.2.1 The game

The Trust Game (TG): We use the binary Trust Game with incomplete information (TG; Dasgupta, 1988; Kreps, 1990b) as an experimental paradigm. In this game, the trustor, first, decides to place trust in the trustee or withhold trust. If the trustor withholds trust, the TG ends and the trustor and the trustee earn $P = 30$ "points" each. If the trustor places trust, the trustee can honor or abuse trust. If the trustee honors trust, the trustor and the trustee both receive $R = 50$ points. If the trustee abuses trust, the trustor receives $S = 0$ points and the trustee receives either $T = 100$ points or $T - \theta = 0$ points, depending on his type. This TG is illustrated in Figure 5.2 that is discussed further below.

The TG with incomplete information thus assumes two types of trustees: Trustees who have an incentive to abuse trust and trustees who have no such incentive. The trustee types could differ in non-monetary motives, such as altruistic preferences (Manapat et al., 2013), and θ could represent an "internal sanction" an altruistic trustee suffers when abusing trust. Non-monetary motives are difficult to manipulate experimentally and we, therefore, manipulate the type of the trustee via the points a trustee gets when abusing trust, as it was done, for example, in the experiments of Camerer & Weigelt (1988) and Neral & Ochs (1992). The trustee knows whether he is of the type who can get 100 points by abusing trust (the *"opportunistic type"*) or the type who earns 0 points when abusing trust (the *"friendly type"*).[2] However, the

[2]It is possible that in our experiment some trustees refrain from abusing trust for non-monetary motives. We do not take this into account in the theoretical analysis presented in Section 5.3 but

trustor does not directly observe the trustee's type and only knows the probability π with which the trustee is of the friendly type.

Under suitable assumptions, the TG represents a social dilemma (Dawes, 1980; Kollock, 1998). In a subgame perfect equilibrium the trustee would abuse trust if he is of the opportunistic type and honor trust if he is of the friendly trustee. Given any of the probabilities of a friendly trustee (π) implemented in our experiment, the trustor's equilibrium strategy is then to withhold trust because the trustor's payoff from withholding trust, P, is larger than the trustor's expected payoff from placing trust, $\pi R + (1-\pi)S$. Individual rationality thus leads to trust not being placed while both actors would be better off if trust was placed and honored.

The condition for an equilibrium involving trust, $\pi \geq (P-S)/(R-S)$, is more restrictive if π is smaller or $(P-S)/(R-S)$ is larger. We, therefore, interpret π and $(P-S)/(R-S)$ as measures of the size of the trust problem and say that the trust problem is larger if π is smaller or $(P-S)/(R-S)$ is larger. In the experiment, we vary π to manipulate the size of the trust problem.[3]

The Repeated Triad Trust Game (RTTG): To study investments in and effects of network embeddedness, we make the TG part of a repeated game played by two trustors and one trustee. In this "Repeated Triad Trust Game" (RTTG), trustor 1 and trustor 2 each interact three times in a TG with the trustee. In our experiment, participants played such RTTGs in different conditions regarding network embeddedness—information exchange between the two trustors about the behavior of the trustee—and the probability that the trustee is of the friendly type.

Figure 5.1 illustrates the course of the RTTG. First, in period 0.1, the trustee's type is determined and announced to the trustee. With probability $\pi = 0.05, 0.2$, or 0.4, the trustee is of the *friendly type* and with probability $1-\pi$ the trustee is of the *opportunistic type*. The probability π is common knowledge and the three actors know that the trustee's type does not change over the course of the RTTG. However, the trustors are not informed about the trustee's type.

In the conditions with endogenous embeddedness, period 0.1 is followed by period 0.2 in which the trustors or the trustee can establish embeddedness (information exchange between the trustors) by pledging a costly investment of $c = 40$ points. In the condition ENDO_R, each trustoR chooses in period 0.2 whether to propose to invest. If both trustors propose to invest, information exchange gets established and each trustor carries half of the cost of investment $c = 40$ points. If only one trustor

will consider it when interpreting the experimental results in Section 5.5.

[3]While we use a symmetric TG, the TG can be asymmetric in the sense that, for example, the payoff associated with honored trust, R, is not the same for the trustor and the trustee. In an asymmetric TG, the payoffs in the condition $\pi \geq (P-S)/(R-S)$ refer to the payoffs of the trustor.

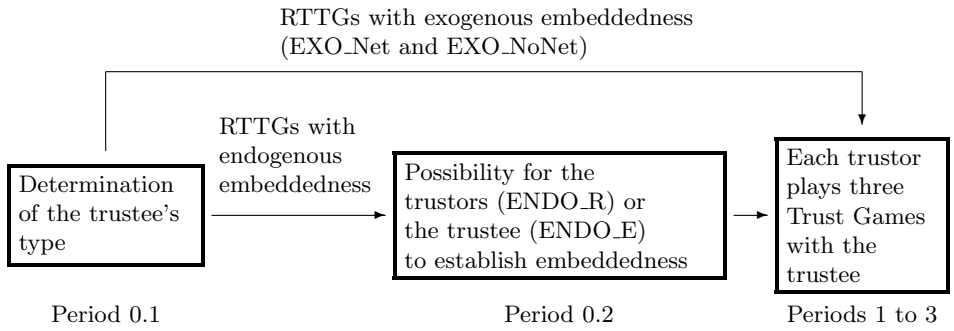


Figure 5.1: Timeline of the Repeated Triad Trust Games (RTTGs).

proposes to invest, embeddedness does not get established and neither trustor incurs a cost. All three actors get informed about whether embeddedness is established (and each actor knows that the others get informed about this, too).[4] In the condition ENDO_E, it is the trusteE who can choose in period 0.2 whether to establish embeddedness at a cost of $c = 40$ points.

In the conditions in which embeddedness is exogenous, EXO_Net and EXO_NoNet, there is no period 0.2. The game proceeds directly from the determination of the trustee's type (period 0.1) to the play of the Trust Games in periods 1 to 3.

In each of the periods 1, 2, 3, each trustor plays one TG with the trustee. In which order the two trustors interact with the trustee within a period is determined randomly in each period. If there is no embeddedness, the trustors receive no information about the choices that are made in the other trustor's TGs. If there is embeddedness, also the trustor who was not involved in some TG gets informed automatically and truthfully about the outcome of that TG directly at the end of that TG.

The RTTG ends after the second TG of period 3. In EXO_Net and EXO_NoNet, the total payoffs for each actor are the sum of points the actor earned in the TGs. In ENDO_R and ENDO_E, it is this sum minus the cost of an investment in period 0.2 (i.e., minus $c/2 = 20$ points per trustor in ENDO_R and minus $c = 40$ for the trustee in ENDO_E).

The computer interface: To strengthen the understanding of the RTTG, we now discuss how it was implemented in the computer interface that was programmed in z-Tree (Fischbacher, 2007). Figure 5.2 illustrates how the TGs were played. Figure 5.2 shows two example screens that a participant in the role of a trustor 1 might see

[4]The actors get informed about the outcome of the two trustors' investment decisions but they do not get explicitly informed about the individual decisions. For example, if a trustor does not propose to invest, that trustor cannot find out whether the other trustor proposed to invest.

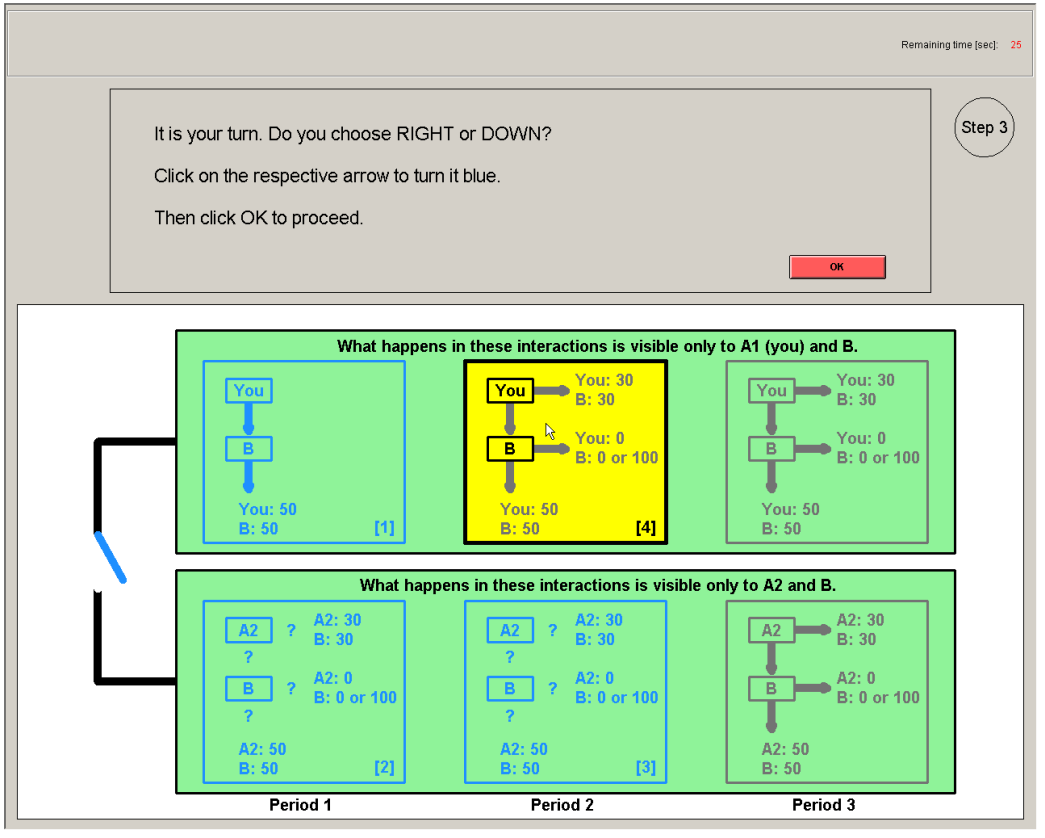
when playing a TG in period 2. Trustor 1 is asked to choose "RIGHT" or "DOWN." To avoid inducing normative associations we used neutral labels rather than explicitly referring to trust. A1, A2, and B represented "trustor 1," "trustor 2," and the "trustee." The ongoing TG is displayed on a yellow background, the past TGs are shown in blue, and numbers in square brackets inform about the order in which the past TGs were played. Furthermore, the trustor viewing the screen ("You") sees what happened in his/her first TG. In the case that there is embeddedness (Figure 5.2b), the trustor sees also what happened in the two TGs the other trustor (A2) has already played. However, if there is no embeddedness (Figure 5.2a), the trustor does not see what choices were made in the past TGs of the other trustor and sees question marks in these TGs instead.

Note two further features of the screens. First, we tried to avoid any confusion about what information the different actors receive. Therefore, it was written above the three TGs of each trustor who receives information about the choices in these TGs. In addition, the switch in the line that connects the boxes of the three TGs of trustor 1 and trustor 2 was "open" if there was no information exchange and "closed" if there was information exchange. Second, in Figure 5.2 it says "B: 0 or 100" next to the arrows that represent the trustee's choice to abuse trust. This reflects that trustor 1 (as well as trustor 2) is uncertain about the trustee's type. This looks different on the screen of the trustee. It either says "You: 100" in each TG or "You: 0" in each TG.

5.2.2 Experimental design and procedures

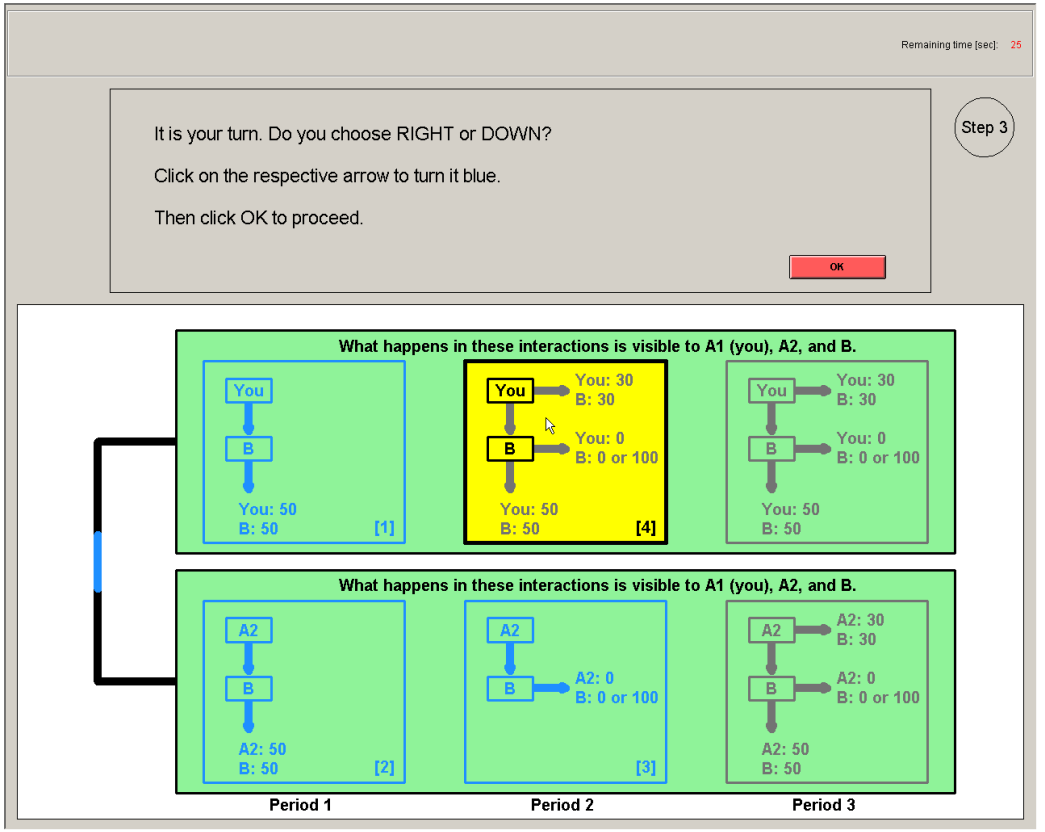
The participants of each of our experimental sessions played twelve RTTGs, one after the other. They, first, played six RTTGs of one type (e.g., EXO_NoNet) and then six RTTGs of a different type (e.g., EXO_Net). A typical session had 18 participants divided into 6 triads. Between each of the twelve RTTGs, the participants were assigned to new roles and triads. For the first RTTG, each subject was assigned the role of trustor 1, trustor 2, or trustee. The roles were then rotated such that those in the role of trustor 1 were in the role of a trustor 2 in the next RTTG, then in the role of a trustee and then again in the role of trustor 1.... After the roles were assigned, the triads were formed randomly. It was unavoidable that participants played in more than one RTTG together. The participants were informed about this but the instructions also stressed that if this occurs, they would not be able to recognize it.

The participants read printed instructions (see Appendix D.2) and took a quiz at the beginning of a session. They read further instructions before the second six RTTGs. The sessions ended with a small questionnaire and the participants privately



(a) Without embeddedness.

Figure 5.2: Example screens from the experiment illustrating the play of the Trust Games and the availability of information on past choices to a trustor 1 ("You").



(b) With embeddedness.

Table 5.1: Number of sessions by combination of the type of RTTG played first and second and the probability of a friendly trustee (π). Participants per session in parentheses.

	π.05	π.2	π.4
EXO_Net, EXO_NoNet	1 (21)	1 (18)	1 (15)
EXO_NoNet, EXO_Net	1 (21)	1 (18)	1 (21)
ENDO_R, ENDO_E	2 (18, 18)	2 (24, 18)	2 (18, 15)
ENDO_E, ENDO_R	2 (18, 21)	2 (21, 15)	2 (21, 21)

receiving cash for the points they earned in the RTTGs (1 Euro for every 150 points).

The probability of a trustee being friendly (π) was varied between sessions. We use the shorthands π.05, π.2, and π.4 to refer to the conditions with $\pi = .05$, $\pi = .2$, and $\pi = .4$, respectively. Some sessions used RTTGs with exogenous embeddedness while other sessions used RTTGs with endogenous embeddedness. For each π condition we conducted one session with EXO_Net in the first six RTTGs and EXO_NoNet in the second six RTTGs as well as one session with the reversed order. For ENDO_R and ENDO_E we conducted two sessions per π condition and order (see Table 5.1). In total, we collected data on 223 and 228 RTTGs EXO_NoNet and EXO_Net, respectively, and on 456 RTTGs played in each of the conditions ENDO_R and ENDO_E.[5]

We conducted the experiment in the ELSE laboratory at Utrecht University in 2013 and recruited the 342 participants (average age = 23, 55.3% females, mostly Bachelor students) online from the subject pool using ORSEE (Greiner, 2004). The sessions lasted between one and a half and two hours and the participants earned on average 16.9 Euros.

5.3 Theory and hypotheses

In this section, we sketch the equilibrium analysis of the specific RTTGs played in our experiment to derive hypotheses on investments in and effects of network embeddedness (see Buskens, 2003, Frey, 2014, and Frey et al., 2015b, for more elaborate theoretical analyses). We first identify the effect of embeddedness on behavior and payoffs in the TGs and how the effect of embeddedness depends on the probability π of a trustee being of the friendly type. Having identified how embeddedness affects earnings in the TGs, we then focus on investments in establishing embeddedness.

[5]Due to technical problems in the session with $\pi = 0.4$ and EXO_NoNet being used in the first six RTTGs, the data from the last of these six RTTGs is not available and we have data on 223 rather than 228 RTTGs EXO_NoNet.

Camerer and Weigelt's (1988) analysis of a finitely repeated Trust Game with incomplete information provides the basis for the analysis of the effects of embeddedness in the TGs. They analyze a game in which just one trustor interacts with the trustee and show that in such a game there is a unique sequential equilibrium that typically evolves as follows.[6] First, trust gets placed and honored. Second, as the end of the game comes close, the trustor and the trustee—if he is of the opportunistic type—begin to randomize between placing and withholding trust and honoring and abusing trust, respectively. Finally, the trustor does not place trust anymore after the first instance of abused or not placed trust.

In this equilibrium, an opportunistic trustee honors trust only because there is a shadow of the future; the trustee knows that he will not be trusted again after an abuse. As the end of the game comes close, however, the shadow of the future shrinks and an opportunistic trustee becomes less inclined to honor trust. That the trustor continues to place trust, at least with positive probability, is only possible because the trustor can learn about the trustees type in the randomization phase. In this phase, gradual reputation building occurs in the sense that after every TG in which the trustee honors trust, the trustor becomes more optimistic that the trustee is of the friendly type (the trustor's belief that the trustee is of the friendly type increases).

If the prior probability of a trustee being of the friendly type (π) is larger, less reputation building is needed and the randomization phase starts later. Specifically, the randomization starts when the number of TGs left reaches some critical value which decreases in π (and also depends on the trustor's payoffs in the TG). If π is very small and the trustor and the trustee interact only a few times together, the trustor will never place trust in the equilibrium. It would be too risky for the trustor to give the trustee an opportunity to start building a reputation.

In the RTTG, where *two* trustors interact with the trustee, the equilibrium course of play in the TGs depends also crucially on whether there is embeddedness (information exchange between the trustors). If there is no embeddedness, the sequential equilibrium for the three TGs of each trustor is the same as if there was no second trustor. However, if there is embeddedness, trustworthiness in one TG may make *both* trustors more optimistic about the trustee's trustworthiness and, conversely, a single abuse of trust reveals the trustee's incentives for opportunistic behavior to *both* trustors. We follow the standard assumption of game-theoretic analysis that actors respond to reliable information from own experiences and others' experiences in the same manner (cf. Bolton et al., 2004). It can then be shown that if there is embed-

[6] Informally, a combination of beliefs and strategies constitutes a sequential equilibrium (Kreps & Wilson, 1982b) if the beliefs are justified by the strategies following Bayesian updating and the strategies are best replies against the others' strategies given the beliefs (see Rasmusen, 1994, Chap. 6, for a textbook).

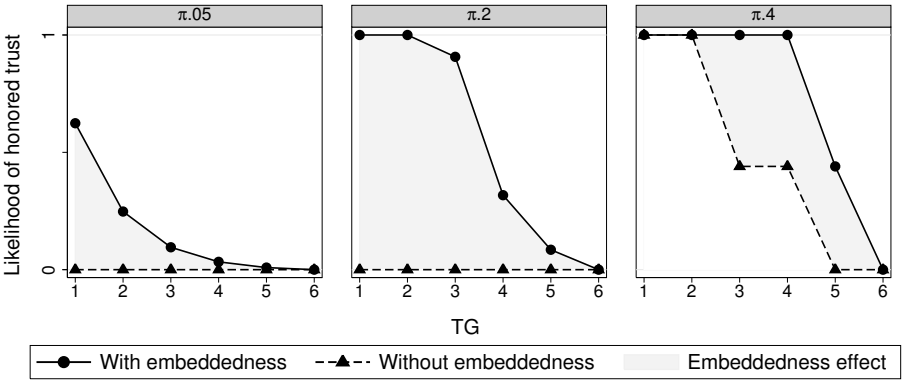
dedness, the equilibrium is the same as if one trustor played all six TGs. If there is embeddedness, the trustee remains trustworthy with as many TGs left *in total* as with each trustor individually if there is no embeddedness. Thus, both trustors can trust until they have each only half as many TGs left if there is embeddedness than if there is no embeddedness.

Figure 5.3a illustrates how embeddedness affects the equilibria of the experimental games. In each panel, the gray area highlights how much higher the expected equilibrium rate of honored trust is if there is embeddedness than if there is no embeddedness. If the probability of a friendly trustee is $\pi = 0.05$, both trustors never place trust in equilibrium if there is no embeddedness. If there is embeddedness, trust is possible in equilibrium but an opportunistic trustee randomizes already from the first TG on. If $\pi = 0.2$, embeddedness makes it possible that trust is placed and honored with certainty in equilibrium in the first two TGs while there is still no trust if there is no embeddedness. If $\pi = 0.4$, the sequential equilibrium is such that the randomization begins only in the fifth TG if there is embeddedness while without embeddedness, trust gets placed and honored in each trustor's first TG and the randomization begins in each trustor's second TG.

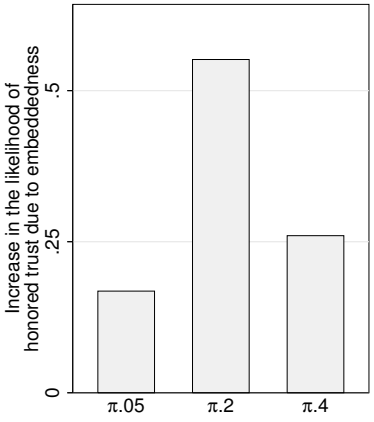
In these equilibria, the effect of embeddedness on the rate of honored trust follows an inverted U-shape in the size of the trust problem (π), as it is illustrated in Figure 5.3b. Embeddedness has the largest effect in the condition with a trust problem of intermediate size ($\pi = 0.2$). If $\pi = 0.05$, the embeddedness effect is small because the trust problem is so large that there is in the equilibrium little honored trust even with embeddedness. If $\pi = 0.4$, the trust problem is so small that there is quite some honored trust also without embeddedness and embeddedness makes the equilibrium phase with honored trust only modestly longer.

We mention that, more generally, as long as the trust problem is so large that trust is not possible without embeddedness, a decrease in the size of the trust problem leads to more trust only in the scenario with embeddedness. Hence, a decrease in the size of the trust problem then leads to an increase in the effect of embeddedness. Thus, between $\pi = .05$ and $\pi = .2$ the embeddedness effect increases in π. On the other hand, if the trust problem is not that large and some trust is also possible without embeddedness, a decrease in the size of the trust problem leads to an increase in trust and trustworthiness in each trustor's TGs to the same extent if there is no embeddedness as it does *in total* if there is embeddedness. Hence, a decrease in the size of the trust problem then leads to a decrease in the embeddedness effect (see Frey et al., 2015b, for details).

Now consider investments in establishing embeddedness. A rational actor will invest only if the costs of investment do not exceed the expected returns on investment.



(a) Development of the likelihood of honored trust with embeddedness and without embeddedness over the six TGs.



(b) Increase in the likelihood of honored trust due to embeddedness averaged over all six TGs.

Figure 5.3: Embeddedness effect on the likelihood of honored trust in the sequential equilibrium in the different conditions with respect to the probability π of a friendly trustee. (Displayed are expected equilibrium rates of honored trust in RTTGs with an opportunistic trustee, to maintain consistency with the empirical analyses. The development of the likelihood of honored trust and the embeddedness effect are not qualitatively different in RTTGs with a friendly trustee).

The expected payoffs that rational trustors and trustees will consider when evaluating the returns on an investment in establishing embeddedness are those associated with the equilibria illustrated in Figure 5.3a. The game-theoretic analysis of the RTTG suggests that the equilibria in the TGs illustrated in Figure 5.3a apply equally to the scenarios with exogenous and endogenous embeddedness. That is, if actors choose not to establish embeddedness, they should behave in the TGs as if there had not been a possibility to establish embeddedness; if they do establish embeddedness, they should play in the TGs as if embeddedness had been imposed exogenously. One might conjecture that a trustee's investment in establishing embeddedness could serve as a costly signal that the trustee is of the friendly type. However, the analysis of Frey (2014) shows that is not the case for the RTTG as implemented in our experiment.[7]

The incentives to establish embeddedness reflect the inverted U-shape in the effect of embeddedness on the rate of honored trust. A trustor earns expectedly 28.2 points more in the TGs if there is embeddedness than if there is no embeddedness if $\pi = 0.2$ while it is expectedly only 1.1 points and 18.3 points more if $\pi = 0.05$ and $\pi = 0.4$, respectively. Hence, in ENDO_R, a rational trustor proposes to invest if $\pi = 0.2$ (as $28.2 > c/2 = 20$) but not if $\pi = 0.05$ or $\pi = 0.4$. Friendly and opportunistic trustees likewise benefit expectedly from paying $c = 40$ points to establish embeddedness if $\pi = 0.2$ (expected earnings are, respectively, 91.8 and 110.0 points higher in the TGs with embeddedness). If $\pi = 0.05$, friendly and opportunistic trustees earn, in the TGs, expectedly 70.0 and 60.7 points more if there is embeddedness. They thus expected to benefit from an investment also if $\pi = 0.05$, but to a lesser extent than if $\pi = 0.2$. However, if $\pi = 0.4$, neither type of trustee can expect to recuperate the investment (friendly trustees' expected earnings are only 5.7 points higher if there is embeddedness; opportunistic trustees are expected to be even by 30 points worse off in the TGs with embeddedness because an opportunistic trustee may be able to abuse trust twice if there is no embeddedness but only once if there is embeddedness and the increase in trust due to embeddedness does not compensate an opportunistic trustee for this).

From this sketch of the equilibrium analysis of the RTTG we infer hypotheses for investments in establishing embeddedness and effects of embeddedness on trust and trustworthiness. Concerning the latter, we formulate hypotheses focusing on the effects of embeddedness averaged over the six TGs of an RTTG. To infer hypotheses for the behavior of the participants of the experiment from the game-theoretic

[7] Frey (2014) shows a trustee's investment in establishing embeddedness cannot credibly signal that the trustee is of the friendly type in the RTTG implemented in our experiment, while such signaling is possible in a slightly different model. In this alternative model, friendly trustees are assumed to have no incentive to abuse trust in the TG because they receive a high payoff when honoring trust (rather than because they receive a low payoff when abusing trust).

analysis, we assume that the only payoffs that the participants are concerned with are the monetary payoffs. In addition, we interpret an equilibrium as a "solution of the game" in the sense that rational participants tend to implement the equilibrium. Finally, we assume that the likelihood of some behavior increases if the conditions for an equilibrium involving that behavior become less restrictive (see Buskens & Raub, 2013). We then have the following hypotheses:

H1: (a) Trustors and (b) trustees are more likely to (propose to) invest in establishing embeddedness in the condition $\pi.2$ than in the conditions $\pi.05$ and $\pi.4$.

H2: (a) Trustors are more inclined to place trust and (b) opportunistic trustees are more inclined to honor trust if there is embeddedness than if there is no embeddedness.

H3: The effect of embeddedness on (a) placing trust and (b) honoring trust is larger in $\pi.2$ than in $\pi.05$ and $\pi.4$.

We will, furthermore, investigate whether there is indeed no difference in the degree to which exogenous and endogenous embeddedness promotes trust and trustworthiness. This null effect is implied by our game-theoretic analysis. Intuition outside the confines of the game-theoretic model as well as previous experiments on endogenous institutions (e.g., Gürerk et al., 2014; Sutter et al., 2010; Schneider & Weber, 2013) suggest that endogenous embeddedness may have particularly strong effects. A trustee's investment in establishing embeddedness might serve as a costly signal of trustworthiness and, hence, promote trust particularly strongly.[8] Embeddedness established by the trustors could have particularly strong effects due to a self-selection: Trustors who are especially sensitive to information from third parties could be particularly inclined to establish embeddedness.

5.4 Results

For the test of our hypotheses we, first, present results on investments in establishing embeddedness (Section 5.4.1) and then results on how embeddedness affected behavior in the TGs (Section 5.4.2). We, furthermore, look at investments in establishing embeddedness in view of the observed monetary returns on embeddedness (Section 5.4.3). Finally, an analysis of how trust and trustworthiness were affected by the π condition provides tentative explanations for anomalies in the effects of embeddedness (Section 5.4.4).

[8] Compare footnote 7.

Table 5.2: Decisions of trustors and trustees to (propose to) invest in establishing embeddedness, overall and by π condition. Proportions in parentheses.

	All		π.05		π.2		π.4	
Trustors	0.49	(446/912)	0.46	(137/300)	0.50	(156/312)	0.51	(153/300)
Trustees (all)	0.26	(119/456)	0.31	(46/150)	0.20	(31/156)	0.28	(42/150)
Friendly	0.29	(30/103)	0.17	(1/6)	0.17	(5/30)	0.36	(24/67)
Opportunistic	0.25	(89/353)	0.31	(45/144)	0.21	(26/126)	0.22	(18/83)

5.4.1 Investments in establishing embeddedness

Table 5.2 summarizes the investment decisions. In ENDO_R, trustors took 912 decisions whether to propose to invest in establishing embeddedness, and they proposed to invest in 446 (49%) of these decisions.[9] They proposed to invest in about half of the instances in all π conditions and not, as hypothesis *H1a* predicts, more often in π.2 than in π.05 and π.4. Trustees did also not establish information exchange most often in π.2, contrary to *H1b*. Friendly trustees invested more frequently in the condition π.4 (36%) than in the conditions π.05 and π.2 (17% each). Opportunistic trustees invested more often in π.05 (31%) than in π.2 and π.4 (21% and 22%, respectively).

We tested *H1a* and *H1b* statistically in multi-level logistic regressions of the decisions to (propose to) invest in establishing information exchange on dummies for the conditions π.05 and π.4. These regressions are reported in Table 5.3 and include a random intercept at the level of individual participants to account for the nesting of investment decisions in participants. The results provide no support for hypotheses *H1a* and *H1b*. Neither trustors, friendly trustees, nor opportunistic trustees had a significantly lower tendency to (propose to) invest in the condition π.05 or π.4 than in the condition π.2, the reference category.

Additional multi-level regressions show that neither trustors, friendly trustees, or opportunistic trustees were in one π condition significantly more or less inclined to (propose to) invest compared to the situation in the other two π conditions together ($p > 0.05$). Finally, regressing the investment choices of trustees for each π condition on a dummy for the trustee's type, we find that in the condition π.4, friendly trustees were significantly ($p < 0.05$) more inclined to establish embeddedness than opportunistic trustees.

[9] Information exchange was established (both trustors proposed to invest) in 24% of the 456 observed RTTGs played in the condition ENDO_R.

Table 5.3: Multi-level logistic regressions of the decisions to (propose to) invest in establishing embeddedness. Random intercept at the subject level.

	Investment proposals by trustors		Investments by friendly trustees		Investments by opportunistic trustees	
$\pi.05$	−0.33	(0.44)	−0.07	(1.69)	0.68	(0.40)
$\pi.4$	0.03	(0.44)	1.55	(1.06)	0.05	(0.46)
Constant	0.03	(0.31)	−2.34*	(1.18)	−1.74***	(0.35)
Variance subject level	0.83***	(0.10)	0.58	(0.79)	0.26	(0.30)
Number of decisions	912		103		353	
Number of subjects	229		86		211	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4.2 Effects of embeddedness

How did embeddedness affect behavior in the Trust Games? We address this question for trustors and opportunistic trustees but do not investigate the behavior of friendly trustees. As expected, friendly trustees did almost always honor trust (in 98.9% of the instances). Furthermore, we restrict the focus to behavior in TGs in which the trustor at play has not yet observed an abuse of trust in the focal RTTG (83% of the observed TGs). This allows analyzing embeddedness effects on trustfulness keeping trustee behavior in preceding TGs of the focal RTTG constant without using variables to control for the history of play. In the remaining sample, the average "*trustfulness*" (0, 1) is 0.65 and the average "*trustworthiness*" (0, 1; defined only if trust was placed) of opportunistic trustees is 0.67.[10]

Figure 5.4 gives a descriptive overview of how network embeddedness affected trustfulness and trustworthiness in the different conditions.[11] In line with *H2a* and *H2b*, trustfulness and trustworthiness were consistently higher if there was embeddedness than if there was no embeddedness. The average trustfulness was up to 29%-points higher and average trustworthiness was up to 33%-points higher. The only exception is that in $\pi.4$, trustors were about as trustful if there was exogenous embeddedness as if there was no exogenous embeddedness.

Figure 5.4 offers no systematic indication for an inverted U-shape in the effect of embeddedness on trustfulness and trustworthiness (*H3a* and *H3b*). Embeddedness

[10] For the analysis of trustworthiness, it is conceivable to include the choices trustees made when trusted by a trustor who already observed an abuse of trust. However, these are only 100 choices of trustees (because after observing an abuse of trust, trustors placed trust in only 8.2% of the TGs) and including them does not affect the results qualitatively.

[11] Figures D.1 and D.2 in Appendix D.1 show how the averages of trustfulness and trustworthiness developed over the six TGs of an RTTG in the different conditions.

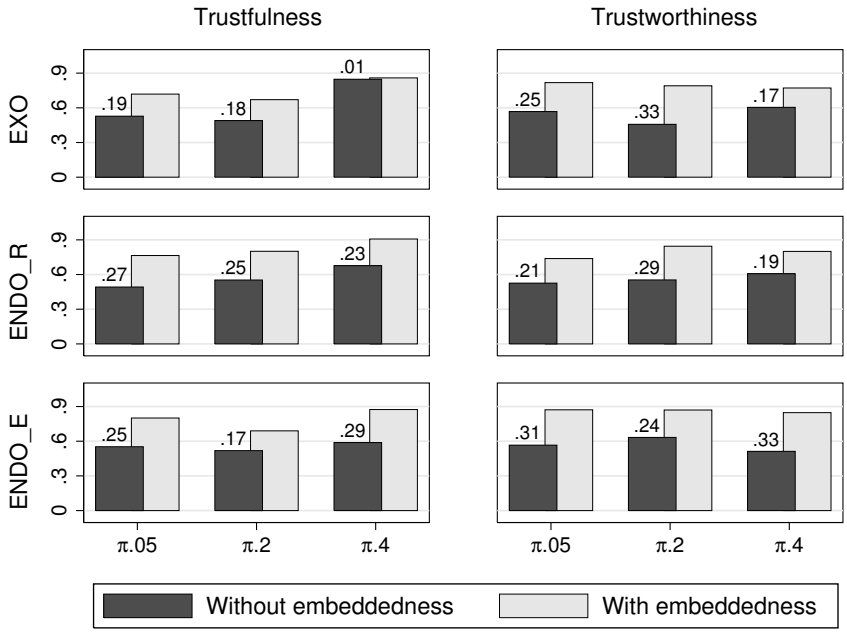


Figure 5.4: Average trustfulness and trustworthiness with and without embeddedness by the likelihood of a friendly trustee (π) in the condition with exogenous embeddedness (EXO) and the conditions with embeddedness choice for the trustors (ENDO_R) or the trustee (ENDO_E). The numbers above the bars report how much higher the rate of trustfulness or trustworthiness was with embeddedness than without embeddedness.

promoted the trustworthiness of opportunistic trustees indeed more in $\pi.2$ than in $\pi.05$ and $\pi.4$ if embeddedness was exogenous or could be established by the trustors. However, the embeddedness effect on trustfulness was in no scenario strongest in $\pi.2$ and in ENDO_E, embeddedness even promoted trustfulness as well as trustworthiness the least in $\pi.2$.

Finally, Figure 5.4 suggests that endogenously established embeddedness may have stronger effects than exogenously imposed embeddedness. The trustors' trustfulness tended to be more strongly affected by endogenously established embeddedness than by exogenously imposed embeddedness. The trustworthiness of opportunistic trustees was about equally affected by embeddedness established by the trustors as by exogenous embeddedness while it was somewhat more strongly affected by embeddedness established by the trustee than by exogenous embeddedness.

To investigate the effects of embeddedness statistically, we use multi-level logistic

regression models of the individual trusting and honoring decisions, controlling for the position of the TG in which a choice was made in an RTTG and accounting for the nesting of individual choices in subjects. Specifically, we model the propensity of trustor i to place trust in TG j of RTTG k as a function of the experimental condition and the position of TG j in RTTG k as follows:

$$\begin{aligned}\text{logit}(\text{trust}_{ijk}) = \beta_0 &+ \beta_1 \text{ENDO\_R} + \beta_2 \text{ENDO\_E} + \beta_3 \text{COND} + \\ &\beta_4(\text{ENDO\_R X COND}) + \beta_5(\text{ENDO\_E X COND}) + \\ &\beta_6 \text{POS} + u_i + \varepsilon_{ijk},\end{aligned} \tag{5.1}$$

where β_3 indicates the vector of coefficients for the CONDition dummies NET, $\pi.05$, and $\pi.4$ and the interactions NET X $\pi.05$ and NET X $\pi.4$. β_4 and β_5 are the coefficient vectors for these CONDition variables interacted with the dummies ENDO_R and ENDO_E, respectively. β_6 indicates the coefficient vector for the variables that control for the POSition in which TG j was played in RTTG k, namely, PERIOD (1, 2, 3), TG2InPeriod (1 for the second TG in a period, 0 for the first TG in a period), and the interactions of these two variables with NET. Finally, u_i is a random intercept for trustor i and ε_{ijk} is a stochastic error for the specific decision of trustor i in TG j of RTTG k. We estimated the same model for trustworthiness as the dependent variable. The results of the two regressions are in Table D.1 in Appendix D.1.

Here we discuss the Average Marginal Effects (AMEs; Long, 1997, Chap. 3) of embeddedness on trustfulness and trustworthiness and the differences in these AMEs between the π conditions and the scenarios EXO, ENDO_R and ENDO_E. These are reported in Table 5.4 and obtained from post-estimations on the regression results. The AMEs of embeddedness inform how many %-points the probability of trustfulness (trustworthiness) is higher in a TG with embeddedness than without embeddedness, averaged over the six TGs played in an RTTG. Differences between the AMEs in Table 5.4 and the numbers reported in Figure 5.4 result from controlling for the position of a TG in an RTTG and accounting for the nesting of decisions in individual participants.

The large and highly significant AMEs reported in the panel at the top of Table 5.4 lend support for the hypothesis that embeddedness fosters trust and trustworthiness (*H2a* and *H2b*). The only exception is that in the condition with $\pi = 0.4$ and exogenous embeddedness there is no significant effect of embeddedness on trustfulness. In the same condition, trustees are predicted to be 18.1%-points more likely to honor trust if there is embeddedness.

The middle panel of Table 5.4 reports the test of the inverted U-shape hypothesis—whether the AMEs of embeddedness were larger in $\pi.2$ than in $\pi.05$ and $\pi.4$ (*H3a* and

Table 5.4: Average Marginal Effects (AMEs) of embeddedness on trustfulness and trustworthiness in the different experimental conditions (top panel), differences in the AMEs of embeddedness between π conditions (middle panel), and differences in the AMEs of endogenous embeddedness and exogenous embeddedness (bottom panel).

		$\pi.05$		$\pi.2$		$\pi.4$	
Trustfulness	Overall	0.25^{***}	(0.02)	0.21^{***}	(0.02)	0.17^{***}	(0.02)
	EXO	0.21^{***}	(0.03)	0.20^{***}	(0.04)	0.01	(0.02)
	ENDO_R	0.26^{***}	(0.05)	0.22^{***}	(0.04)	0.19^{***}	(0.03)
	ENDO_E	0.27^{***}	(0.04)	0.20^{***}	(0.04)	0.28^{***}	(0.03)
Trustworthiness	Overall	0.29^{***}	(0.03)	0.32^{***}	(0.03)	0.26^{***}	(0.04)
	EXO	0.30^{***}	(0.04)	0.40^{***}	(0.05)	0.18^{**}	(0.06)
	ENDO_R	0.21^{***}	(0.06)	0.29^{***}	(0.05)	0.23^{***}	(0.06)
	ENDO_E	0.35^{***}	(0.04)	0.28^{***}	(0.05)	0.38^{***}	(0.06)

		Difference $\pi.2 - \pi.05$		Difference $\pi.2 - \pi.4$	
Trustfulness	Overall	−0.04	(0.03)	0.04	(0.03)
	EXO	−0.02	(0.05)	0.18^{***}	(0.04)
	ENDO_R	−0.03	(0.06)	0.03	(0.05)
	ENDO_E	−0.07	(0.05)	−0.07	(0.05)
Trustworthiness	Overall	0.03	(0.04)	0.06	(0.05)
	EXO	0.10	(0.07)	0.22^{**}	(0.08)
	ENDO_R	0.08	(0.08)	0.07	(0.08)
	ENDO_E	−0.07	(0.06)	−0.10	(0.07)

		Difference EXO − ENDO_R		Difference EXO − ENDO_E	
Trustfulness	Overall	-0.08^{**}	(0.03)	-0.11^{***}	(0.03)
	$\pi.05$	−0.05	(0.06)	−0.06	(0.05)
	$\pi.2$	−0.03	(0.05)	−0.01	(0.06)
	$\pi.4$	-0.18^{***}	(0.04)	-0.27^{***}	(0.04)
Trustworthiness	Overall	0.06	(0.04)	−0.03	(0.04)
	$\pi.05$	0.09	(0.07)	−0.04	(0.06)
	$\pi.2$	0.11	(0.07)	0.12	(0.07)
	$\pi.4$	−0.05	(0.08)	-0.20^{*}	(0.08)

Based on the regressions in Appendix D.1
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

H3b). There is no systematic support for this hypothesis. The effect of embeddedness on trust and trustworthiness was not significantly stronger in $\pi.2$ than in $\pi.05$ in any of the scenarios EXO, ENDO_R and ENDO_E and it was significantly stronger in $\pi.2$ than in $\pi.4$ only in the scenario with exogenous embeddedness.

The bottom part of Table 5.4 reports whether, contrary to what the game-theoretic analysis of the RTTGs suggests but in line with findings of the literature on endogenous and exogenous institutions (Gürerk et al., 2014; Schneider & Weber, 2013; Sutter et al., 2010), there are significant differences in the effects of endogenous and exogenous embeddedness. The trustors' inclination to place trust, aggregated over the π conditions, was significantly more strongly promoted by embeddedness if the trustors or the trustee could establish it than if it was exogenous. This difference was also significant in $\pi.4$ separately, but not in $\pi.05$ and $\pi.2$. The trustworthiness of opportunistic trustees was not significantly differently affected by embeddedness established by the trustors than by exogenous embeddedness. However, in the condition $\pi.4$, trustworthiness was significantly more strongly associated with embeddedness if the trustee could establish embeddedness than if embeddedness was exogenous.

5.4.3 Linking observed investments in and returns on embeddedness

The reason for which we expected an inverted U-shape in investments in establishing embeddedness over the conditions $\pi.05$, $\pi.2$, and $\pi.4$ is that we expected an inverted U-shape in the effects of (endogenous) embeddedness. However, in the experiment, there was no inverted U-shape in the effects of endogenous embeddedness (Section 5.4.2). Given that there was no inverted U-shape in the effects of endogenous embeddedness, it is not surprising that there was no inverted U-shape in investments either. In fact, it could still be that subjects were, as we expect theoretically, most inclined to establish embeddedness in the conditions in which the *observed* monetary returns on embeddedness were largest. In this section, we investigate whether this was the case. We first discuss the *observed* monetary returns on embeddedness and then compare them with the *observed* investments in establishing embeddedness.

Table 5.5 shows how embeddedness affected earnings in the TGs. Table 5.5 reports how many additional points participants earned on average in total in the TGs in RTTGs with embeddedness compared to the situation in RTTGs without embeddedness. Potential costs of establishing embeddedness are *not* subtracted. TG earnings were higher with embeddedness than without embeddedness for trustors and friendly trustees in almost all conditions. Opportunistic trustees often benefited little or even suffered from embeddedness. This was predicted only for the condition $\pi.4$, but also in

Table 5.5: Embeddedness effect on TG earnings for the different actors: difference in the average sum of earnings in the TGs in RTTGs with embeddedness compared to RTTGs without embeddedness. Potential costs of establishing embeddedness are *not* subtracted. Sample: all TGs.

		$\pi.05$	$\pi.2$	$\pi.4$
Trustor	EXO	19.7	18.7	10.4
	ENDO_R	14.3	21.3	26.2
	ENDO_E	31.0	20.0	31.5
Friendly trustee	EXO	72.2	18.2	−7.7
	ENDO_R	23.3	0.1	35.0
	ENDO_E	68.0	18.4	31.1
Opportunistic trustee	EXO	−2.3	−5.3	−20.3
	ENDO_R	11.3	8.0	−17.4
	ENDO_E	−7.2	5.7	20.9

$\pi.05$ and $\pi.2$ there was some trust even without embeddedness, making it possible to abuse the trust of both trustors. Furthermore, Table 5.5 implies that embeddedness increased the sum of the three actors' earnings in the TGs in all conditions.

Establishing embeddedness paid off only in few conditions. That is, the average returns on embeddedness (Table 5.5) exceeded the costs of investment only in few conditions. Trustors could on average recuperate an investment of 20 points in the conditions $\pi.2$ and $\pi.4$. Friendly trustees had a net benefit from establishing embeddedness only in $\pi.05$.[12] For opportunistic trustees the investment of 40 points did not pay off in any of the π conditions.

The data offer no clear indication that actors were most inclined to establish embeddedness in the conditions in which the returns on embeddedness were largest. This can be seen from a comparison of observed investments (Table 5.2) with observed returns (Table 5.5). For trustors, the returns increased in π and so did their tendency to establish embeddedness (although not significantly; Section 5.4.1). Friendly trustees did not only benefited more from embeddedness than opportunistic trustees, they also invest more often in establishing embeddedness than opportunistic trustees. However, there are also several contradictions.

A logistic regression analysis does not reject the null-hypothesis that investments are independent of returns. We regressed investment decisions on the returns reported in Table 5.5. The regression included a dummy for whether the investing actor was

[12]Some differences reported in Table 5.5 for friendly trustees rely on small numbers of observations, especially in the condition with $\pi = 0.05$. In the condition ENDO_E with $\pi = 0.05$, only one friendly trustee was observed in the condition with embeddedness; compare Table 5.2.

trustor or trustee and a random intercept at the level of individual subjects. A 1-point increase in returns increases the odds of an investment by a factor of 1.003, which is not significantly different form 1 ($p = 0.822$). Thus, we do not find that subjects were most inclined to establish embeddedness in the conditions in which the observed monetary returns on investment were largest.

5.4.4 Explaining the lack of an inverted U-Shape in the effect of embeddedness over the π conditions

We conclude this section with an attempt to explain why there was no inverted U-shape in the effect of embeddedness on trust and trustworthiness over the π conditions. We, first, consider the lack of an increase in the embeddedness effect from $\pi.05$ to $\pi.2$. Then, we consider the lack of a decrease in the effect of endogenous embeddedness form $\pi.2$ to $\pi.4$.

Recall, the theoretical basis for the expectation of an increase in the effect of embeddedness from $\pi.05$ to $\pi.2$. Generally, the sequential equilibrium theory suggests that if π increases, there will be more trust and trustworthiness. The sequential equilibrium theory, furthermore, suggests that from $\pi.05$ to $\pi.2$ the level of trust and trustworthiness should increase exclusively in the situation with embeddedness. If there is no embeddedness, trust and trustworthiness should remain absent even in $\pi.2$. This implies that embeddedness should promote trust and trustworthiness more in $\pi.2$ than in $\pi.05$.

In the experiment, however, there was no more trust and trustworthiness in $\pi.2$ than in $\pi.05$, contrary to the theoretical expectation. Specifically, post-estimations of the regressions reported in Appendix D.1 reveal no significantly negative average marginal effect of an RTTG being played in the condition $\pi.05$ rather than in the condition $\pi.2$ on trustfulness and trustworthiness (this result holds for any of the scenarios EXO, ENDO_R and ENDO_E and also separately for RTTGs with embeddedness). This suggests that subjects did not perceive the trust problem as being smaller in $\pi.2$ than in $\pi.05$. Manipulating the size of the trust problem via the probability π worked imperfectly and there was no increase in trust and trustworthiness that could have taken place disproportionately in the condition with embeddedness.

It also bears mentioning that the incidence of honored trust was considerable even without embeddedness already in the condition $\pi.05$, where no trust was expected. This observation can be reconciled with the sequential equilibrium theory when assuming that some trustees are intrinsically trustworthy and that trustors anticipate this (cf. Camerer & Weigelt, 1988). However, it means that none of the experimental conditions represented a trust problem so large that there is no trust without embed-

dedness (which implies that if the rate of honored trust had increased from $\pi.05$ to $\pi.2$, the effect of embeddedness should have decreased from $\pi.05$ to $\pi.2$, as from $\pi.2$ to $\pi.4$).

From $\pi.2$ to $\pi.4$, the increase in trustfulness and trustworthiness should disproportionately take place in the situation without embeddedness and, consequently, the effect of embeddedness should become smaller (see Section 5.3). In the experiment, trust and trustworthiness increased from $\pi.2$ to $\pi.4$, in each of the scenarios EXO, ENDO_R, and ENDO_E. Post-estimations of the regressions in Appendix D.1 reveal for any of the scenarios EXO, ENDO_R and ENDO_E a significantly positive, $p < 0.05$, average marginal effect on trustfulness and trustworthiness for an RTTG being played in $\pi.4$ rather than in $\pi.2$. Hence, that the effect of embeddedness was not diminished in $\pi.4$ in the scenarios ENDO_R and ENDO_E reflects that the increase in trustfulness and trustworthiness from $\pi.2$ to $\pi.4$ was not larger if there was no embeddedness than if there was embeddedness, which challenges the sequential equilibrium theory.

5.5 Conclusions and discussion

In this chapter we have discussed an experiment in which two trustors interacted with one trustee in finitely repeated Trust Games. Embeddedness—a relation between two trustors for the automatic, truthful exchange of information about outcomes in the Trust Games—was exogenous or could be established endogenously at costs by the trustors or the trustee. The game was played under different probabilities π of a trustee being of the "friendly type." This set-up allows investigating whether actors attempt to benefit from the trust and trustworthiness promoting effect of embeddedness by actively establishing it (see Frey, 2014, Frey et al., 2015b, and Raub et al., 2013, for theoretical models), similar to actors who form long-term relations (Brown et al., 2004; Kirman, 2001; Kollock, 1994; Simpson & McGrimmon, 2008) or choose to exchange with embedded partners (DiMaggio & Louch, 1998) to mitigate trust problems. Varying the likelihood of a trustee being of the friendly type enabled us to test two further hypotheses. Namely, that the degree to which embeddedness promotes trust and trustworthiness follows an inverted U-shape in the size of the trust problem and that, hence, the inclination of actors to actively establish embeddedness follows likewise an inverted U-shape in the size of the trust problem (Frey et al., 2015b; Raub et al., 2013). Finally, the experiment allows investigating whether the degree to which embeddedness promotes trust and trustworthiness depends on whether embeddedness is established endogenously or imposed exogenously.

We are the first to observe the endogenous formation of information exchange

relations in the presence of trust problems in a controlled environment. When given the opportunity, a substantial portion of trustors and trustees indeed pledged a costly investment to establish embeddedness. We also observed considerably higher levels of trust and trustworthiness if there was embeddedness than if there was no embeddedness, which adds to previous evidence for the effects of embeddedness (e.g., Bohnet et al., 2005; Bolton et al., 2004; Buskens et al., 2010; Buskens & Raub, 2013; Huck et al., 2010; Van Miltenburg et al., 2012).

Other findings are less in line with the theory. First, the incidence of investments in establishing embeddedness did not have the predicted inverted U-shape in the probability of a trustee being of the friendly type. This is not surprising given that the effect of embeddedness on trust and trustworthiness did not have the expected inverted U-shape either. However, the data also do not indicate that investments in establishing embeddedness were most frequent in the conditions in which the *actual* monetary returns on embeddedness were largest (see Prendergast, 1999, for a similar finding regarding the choice of remuneration schemes by firms). Our results thus suggest that it is probably not realistic to expect that actors correctly anticipate the size of the returns on embeddedness, at least unless the actors are highly incentivized to do so or highly experienced with the interaction situation (see Binmore, 1998, Chap. 0.4.2, Kreps, 1990b, and Camerer, 2003, for general arguments on the role of experience in experimental studies).

The observed investments in establishing embeddedness also have to be interpreted with some caution. We did not observe more investments in the conditions in which the observed returns were highest. It is, therefore, somewhat unclear whether subjects invested in establishing embeddedness in anticipation of the returns on embeddedness. We cannot fully rule out the possibility that actors paid for establishing information exchange, for example, out of curiosity or to reduce boredom. Hence, our experiment offers no solid evidence that the observed investments in establishing network embeddedness were pledged *because* network embeddedness can resolve trust problems. Future studies could address this issue by experimentally contrasting situations that feature a trust problem to situations that do not feature a trust problem.

The results also offer little support for the prediction that the degree to which embeddedness promotes trust and trustworthiness follows an inverted U-shape. These conditions were supposed to present subjects with a trust problem of decreasing size. However, subjects did not perceive the trust problem as smaller in $\pi.2$ than in $\pi.05$ (they were not more trusting). Furthermore, the level of trust was considerable also in the condition $\pi.05$, where very little trust was predicted even if there is embeddedness. Hence, our study offers no test of the prediction that the trust and trustworthiness promoting effects of embeddedness may be diminished due to a very large trust prob-

lem. The results do suggest, however, that the effects of embeddedness may obtain even if the trust problem is "objectively" quite large.

These results furthermore add to the mixed findings regarding comparative-statics predictions of the sequential equilibrium theory (see Anderhub et al., 2002, and the references therein). Brandts & Figueras (2003) find in their experiment the expected effects of changes in the likelihood of a friendly trustee. This might reflect that they had subjects play many more repeated games (see Camerer & Weigelt, 1988; Neral & Ochs, 1992; Van Miltenburg et al., 2012, for results on behavior approaching the sequential equilibrium with experience). However, additional analyses did also not reveal a difference emerging in trust and trustworthiness (or even in the effect of embeddedness) between the conditions $\pi.05$ and $\pi.2$ in games played towards the end of our sessions.

Exogenous embeddedness promoted trust and trustworthiness indeed more in the condition featuring an intermediate trust problem ($\pi.2$) than in the condition representing a small trust problem ($\pi.4$). This was not the case for the effect of endogenous embeddedness, even though the tendencies for placing and honoring trust were also in these conditions higher in $\pi.4$ than in $\pi.2$. This latter finding begs for further theoretical investigation and links to the last set of findings that we want to discuss.

Under some conditions, embeddedness fostered trust and trustworthiness more if chosen endogenously rather than imposed exogenously, contrary to the theoretical expectation of no difference, given the sequential equilibrium. If the trustors could establish embeddedness, embeddedness promoted trust more than if it was exogenous. We conjecture that this might be due to self-selection, namely, that trustors who are particularly sensitive to embeddedness were more inclined to propose to establish embeddedness than trustors who are less sensitive to embeddedness.[13] An alternative explanation is that the additional text describing the investment decision in the printed instructions and the fact the investment decision had to be taken instigated subjects to reflect more on possible effects of embeddedness. However, if that had been the case, the trustworthiness of trustees should likewise have been more dependent on embeddedness if the trustors could establish it than if it was exogenous (trustees read the same instructions, they were informed that the trustors are "now" taking the investment decision, and they have likely already been in the situation of a trustor taking an investment decision).

If the trustee could establish embeddedness in the $\pi.4$ condition, trust as well as trustworthiness was more strongly associated with embeddedness than if embeddedness was exogenous. Self-selection by trustees according to their sensitivity to

[13] Our data are not suited for a thorough investigation of this hypothesis because trustors who did consistently not propose to invest were never observed in the condition with embeddedness.

embeddedness could be an explanation but it would additionally require that trustors are aware of the self-selection. Costly signaling is an alternatively explanation (Frey, 2014). Some trustees with monetary incentives for opportunistic behavior may have been intrinsically trustworthy due to experiencing an internal reward when honoring trust, a "warm-glow" (Andreoni, 1989). These trustees might have been particularly likely to invest in establishing embeddedness and trustors may have interpreted an investment as a signal of intrinsic trustworthiness (see footnote 7). Such signaling might have occurred in the $\pi.4$ condition but not in $\pi.05$ and $\pi.2$ because the expected returns on establishing embeddedness are considerably larger in $\pi.4$ for a trustee who has no intention to abuse trust than for a trustee who intends to abuse trust (see Section 5.3).

Our interpretations of the stronger effects of endogenous embeddedness are speculative and meant to inspire future research. The results certainly do indicate that embeddedness effects identified in survey research (Gulati, 1995; Gulati & Gargiulo, 1999; Robinson & Stuart, 2007) may reflect the cumulative working of several mechanisms while the results of laboratory experiments with exogenous embeddedness (Bohnet & Huck, 2004; Bolton et al., 2004; Buskens et al., 2010; Huck et al., 2010; Van Miltenburg et al., 2012) might underrepresent the total potential of social structures in alleviating trust problems.

That several questions remain open should not distract from what the study contributes to the literature on embeddedness and trust. First, the results consolidate the evidence for the trust and trustworthiness promoting effects of embeddedness. Second, we are the first to observe that actors who face trust problems invest in establishing embeddedness. Third, the results indicate that investments in and effects of embeddedness may obtain even under unfavorable conditions, namely, if the trust problem is, objectively, small or large. Fourth and finally, the results suggest that embeddedness tends to foster trust and trustworthiness more if established endogenously rather than imposed exogenously.

Chapter 6

Reputation cascades[1]

Abstract: Reputation systems are lauded for their effectiveness in fostering trust between strangers. In this chapter, we study a previously overlooked side-effect: The production of reputational differentiation between equally trustworthy individuals. This endogenous inequality is caused by feedback effects in the reputation-building process. "Reputation cascades" driven by trustors choosing to exchange with partners who have shown themselves to be trustworthy can make entry difficult for those who lack a reputation, while allowing established parties to perpetuate their dominance. Results from a laboratory experiment support the prediction that information sharing leads not only to higher levels of honored trust but also to higher inequality among trustees. We conclude that while large reputation systems enabled by modern technology facilitate large volumes of otherwise unviable transactions, they may also set in motion reputational snowballs that generate unfounded inequities.

[1]This chapter presents joint work with Arnout van de Rijt. Frey (first author) and Van de Rijt (second author) jointly developed the theory, designed the experiment, analyzed the data, and wrote the manuscript; Frey developed the formal theory and programmed and executed the experiment. We thank Idil Akin, Ensieh Eftekhari, Urmimala Senn, and Hyang-Gi Song for assistance in conducting the experiment and Vincent Buskens, Rense Corten, Jasson Jones, Wojtek Przepiorka, and Werner Raub for helpful comments.

6.1 Introduction

Trust problems hamper mutually beneficial exchange across a broad swath of social settings. Trust is an issue whenever exchange requires that one party—the "trustor"—first expose herself to the risk of abuse by the other party—the "trustee" (Coleman, 1990, Chap. 5; Dasgupta, 1988; Kreps, 1990a). Abuse of trust can range from refusal to pay, failure to deliver on a paid order or compromising on quality to negligence or theft. While in many cases abuse of trust is punishable under the law, legal recourse may be too costly or fail, and even when a case is won may not fully compensate the trustor's loss (Macaulay, 1963). Similarly, direct forms of punishment without involvement of a judicial system can be associated with prohibitively high costs for the trustor (Nikiforakis, 2008), especially when the target lives far away or is unidentified.

Reputation systems can effectively resolve trust problems when other mechanisms fall short (Buskens & Raub, 2013; Klein, 1997; Kuwabara, 2015; Przepiorka, 2013; Resnick et al., 2000). For motives of altruism or reciprocity, trustors may choose to exert some effort to share their experiences, for example by posting an internet rating (Diekmann et al., 2014; Heyes & Kapur, 2012, p. 814; Wehrli, 2014, Chap. 3). Reputation systems aggregate and collate trustor reports and thereby enable a form of indirect reciprocity—trustors can learn from the experiences of other trustors and selectively exchange with trustees of good repute while avoiding abusive trustees. The risk of developing a bad reputation and the associated loss of future exchange opportunities with potentially many trustors allows trustees to resist the temptation to take advantage, in anticipation of which trustors may place trust with realistic expectations of benevolent trustee behavior. While prominent historical examples of reputation systems exist (Diekmann et al., 2014, pp. 65–66; Klein, 1997), advances in communication technology of the past decades have made it possible to share reputation information on a wide range of social or economic exchanges among trustors dispersed across vast geographical areas. This has enabled enormous volumes of exchange between remote strangers who could not have feasibly transacted otherwise.

Here we study a previously overlooked side-effect of reputation systems: Reputation building exhibits a form of cumulative advantage (Gould, 2002; DiPrete & Eirich, 2006; Manzo & Baldassarri, 2015; Merton, 1968; Salganik et al., 2006; Van de Rijt et al., 2014), resulting in arbitrary inequality in transaction volume among trustees. To minimize the risk of abuse, rational trustors may avoid trustees who lack a transaction history in favor of a trustee of good repute. The unintended consequence is a "reputation cascade" that keeps solidifying reputational advantages of an established party while preventing others from getting a chance to prove their trustworthiness. When reputation systems combine ratings on large numbers of actors, such repu-

tation cascades can lead to the exclusion of many in favor of a fortunate few and, hence, high inequality in exchange volumes. The inequality is "arbitrary" to the extent that the differentiation is baseless, namely, when the excluded trustees are no less trustworthy than the established trustees. In this chapter, we develop a parsimonious game-theoretic model that demonstrates such cascades and the endogenous production of arbitrary inequality in reputation systems, and we report the results of a laboratory experiment in which we tested this model.

The idea that trust problems affect exchange patterns is not new. In the literature on market entry barriers, it has been shown theoretically and empirically that consumer learning can give a pioneering brand, about which consumers have already learned the quality, an enduring advantage over a later entrant, about which consumers would have to invest in additional learning (Bain, 1956; Bagwell, 1990; Bronnenberg et al., 2009; Farrell, 1986; Schmalensee, 1982). A sociological literature shows, mainly in controlled experiments, that people often remain with a partner who honored their trust in the past instead of dealing with unknown partners, even if offers of unknown partners are potentially more profitable (Cook et al., 2004; Kirman, 2001; Kollock, 1994; Yamagishi et al., 1998). In both these literatures, trustors are assumed to learn privately from own experience and, hence, sample trustees independently from one another. Information sharing in reputation systems enables a repeated interdependent sampling (Denrell & Le Mens, 2007) that, we argue, gives rise to cascading and high inequality in exchange volumes among trustees, especially when reputation systems pool large numbers of actors.

In Section 6.2, we introduce a simple model for interactions in trust situations where trustors can select their exchange partners and share information on past dealings. Our model offers a stylized representation of trust interactions in real-world reputation systems. We reduce complexity, for example, by assuming that the transmission of information on past behavior is automatic and free of costs, that such information is (if available) always accurate, and that there is no price competition. Section 6.3 sketches how, in our model, game-theoretic equilibrium behavior implies that information sharing promotes trust and trustworthiness and leads to arbitrary inequality among trustees. Appendix E.2 contains a mathematically more explicit analysis of the model. Section 6.4 motivates and describes the design of the laboratory experiment conducted to test implications of the theory. Section 6.5 presents the results and Section 6.6 contains some concluding remarks.

6.2 A model of trust interactions with trustee choice and information sharing

We study the production of arbitrary inequality in reputation systems formally in an indefinitely repeated game. In our game, $N_1 > 1$ trustors and $N_2 > 1$ trustees interact in consecutive rounds 1, 2, 3, ... The trustors take turns; trustor i plays in rounds i, $i + N_1$, $i + 2N_1$, etc. After every round, the next round gets played with probability $0 < w < 1$ while the game ends with probability $1 - w$.

Every round, the trustor i in turn chooses whether to withhold trust or to select one of the trustees and place trust in that trustee. If trustor i does not place trust in any trustee, i receives payoff P_1, the same payoff that any trustor not in turn receives. Any trustee who is not selected by trustor i receives payoff P_2. If trustor i chooses to place trust in trustee j, then j has the option to honor or abuse trust. If j honors trust, the payoffs for i and j are $R_1 > P_1$ and $R_2 > P_2$, respectively. If j abuses trust, i's and j's payoffs are $S_1 < P_1$ and T_j, respectively.

Trustees differ in how much they earn when abusing trust, i.e., in T_j, but throughout the repeated game, T_j stays the same. T_j is drawn independently for each trustee j before round 1 from a probability distribution with unbounded density $\mathbf{F}$. While $\mathbf{F}$ is common knowledge, the actual manifestation of T_j is private information of trustee j.

Reputation systems are operationalized as subsets of trustors sharing information on outcomes of prior rounds. The N_1 trustors are divided into n equally sized, disjoint information sharing communities, where $1 \leq n \leq N_1$. While trustees are always informed about all past choices, trustors have always only information about the choices made in past rounds in which a trustor of their own information sharing community was at play.

The trustors of different information sharing communities alternate in taking decisions. For example, a trustor from community 1 plays in rounds 1, $n+1$, $2n+1$, ... So each trustor plays in every N_1^{th} round and every n^{th} round is played by a trustor from some given information sharing community.

6.3 Informal analysis of the model

Reputation cascades can be game-theoretically derived as Nash equilibria in our indefinitely repeated game. We provide the formal analysis in Appendix E.2. Key to this analysis is a strategy for the trustors that resonates Kollock's (1994) statement that when faced with "a situation in which one can be taken advantage of, the natural response is to restrict one's transactions to those who have shown themselves to be

trustworthy" (p. 318). We suppose that if a trustor knows one or more trustees who were trustworthy in the past, she places trust in one of them. Otherwise, the trustor tries out an untested trustee if such an untested trustee is available and withholds trust if neither an untested nor a reputable trustee is available (if she knows of a past abuse by every trustee).

This strategy, loosely speaking, adapts the concept of a so-called trigger strategy (Friedman, 1986) to the situation with trustee choice and provides an incentive for trustees to resist the temptation of trust abuse. It can be shown that a trustee j best-responds to this strategy either by always honoring or by always abusing trust when trusted, depending on the size of T_j (Lemma E.1 in Appendix E.2). Whether there exists an equilibrium involving honored trust and the trustors playing the above-described strategy then depends on whether this strategy induces a large enough proportion of trustees to be trustworthy (compare Proposition E.1 in Appendix E.2). It may be that the chance that any one trustee is trustworthy is so low that no equilibrium involving honored trust exists because trust abuse is so likely during the search for a trustworthy trustee that the trustors are then be better off when never placing trust.

We consider the two extreme cases of "no reputation system" ($n = N_1$) and a "full reputation system" ($n = 1$) to illustrate how information sharing among trustors can enable trust and how it leads to the production of inequality among trustees. We assume that the situation is such that in the absence of a reputation system, the potential for future exchange following honored trust is insufficient to produce trust, because it does not induce a large enough proportion of trustees to be trustworthy. If trustors are not informed about what happened in prior rounds involving other trustors (no reputation system), then exclusively trustees with a low T_j will honor trust. For any other trustee the immediate payoff improvement from abusing a given trustor ($T_j - R_2$) is greater than what he could earn if he were chosen again by that trustor on every future occasion ($(R_2 - P_2) \cdot (1 + w^{N_1} + w^{2N_1} \ldots)$). Anticipating this, no trustor ever places trust and a socially inefficient outcome is reached, as the distribution of T_j, $\mathbf{F}$, together with the other parameters implies too remote a chance that any one trustee is trustworthy.

If an encompassing reputation system is present and all trustors are informed about all trustor and trustee actions in prior rounds, trust may become feasible. When the first trustor places trust in any of the trustees, that trustee can anticipate a "reputation cascade" following trustworthy behavior. Trustors will in all future rounds rationally copy the first trustor's choice of trustee to avoid the risk of abuse by an untested trustee. In the presence of a reputation system, the gain from repeatedly honoring trust ($(R_2 - P_2) \cdot (1 + w + w^2 + \ldots)$) may then outweigh the short-term

gain from trust abuse ($T_j - R_2$) for some trustees who in the absence of a reputation system could not be trusted. Given a suitable distribution $\mathbf{F}$, the first trustor can then rationally place trust in any of the trustees, anticipating a low chance of abuse because even a trustee with a relatively high T_j is incentivized to be trustworthy. Hence, with a full reputation system, honored trust can become possible because whichever trustee is given the first opportunity to honor trust can establish a monopoly by honoring trust, expecting trustors in all future rounds to prefer interaction with him over any of the other trustees who lack a reputation, as long as he keeps honoring trust.

Between the extreme cases of every trustor being "an isolate" ($n = N_1$) and all trustors sharing information in one encompassing information sharing community ($n = 1$), our theoretical analysis (Appendix E.2) predicts that a decrease in the fragmentation of reputation systems (decrease in n) leads to a higher potential for trust as well as more inequality among trustees. The trustors' strategy together with trustees falling apart into a dichotomy of trustworthy and unstrustworthy trustees implies a monopoly for one trustworthy trustee within each of the n disjoint information sharing communities. If n is smaller, it is more likely that the prospect of occupying such a monopoly position incentivizes any one trustee to be trustworthy (because current trustworthiness is then rewarded with trust in a larger fraction of the potential future rounds). A small n thus induces a larger proportion of trustees to be trustworthy and can, therefore, make it safe enough for trustors to search for trustworthy trustees in the first place when no equilibrium involving trust would be possible if reputation systems were more fragmented (larger n). Assuming that the likelihood of a certain equilibrium behavior increases when the conditions for the equilibrium become less restrictive, we can conclude that a higher degree of information sharing (a smaller n) should lead to a higher rate of honored trust.

Inequality in how often different trustees are trusted is greater when n is smaller because fewer trustees can then occupy a monopoly position while more trustees are excluded, potentially including trustees who are equally trustworthy as those occupying a monopoly position. Also, if n is smaller, those lucky to occupy a monopoly position get trusted in a larger fraction of the rounds and, in addition, fewer individual trustees ever get trusted before the dynamics reach the state that some trustees hold a monopoly position while the other trustees are permanently excluded. Summarizing, the theoretical analysis predicts that the levels of honored trust and inequality among trustees increase with larger information pools.

6.4 The experiment

6.4.1 The case for an experiment

The predicted differentiation in transaction volume produced by reputation systems is empirically hard to distinguish from other sources of inequality between trustees. The distributional extremities others have observed in situations of reputation-enabled trust (Barwick & Pathak, 2015, p. 104; Diekmann et al., 2014, p. 72) may be consistent with our argument but can also be theoretically attributed to trustee variability in quality, visibility, or, in the case of economic exchange, price. A compelling demonstration of the significant role of social feedback vis-à-vis relevant trustee characteristics in the unfair allocation of exchange opportunities is enabled by a controlled experimental research design. Through random assignment, experiments can evaluate whether ceteris paribus inequality is indeed greater in the presence than in the absence of information sharing among trustors.

A key distinction between reputation cascades and other forms of social feedback—such as information cascades (Bikhchandani et al., 1992), social influence (Muchnik et al., 2013; Salganik et al., 2006; Van de Rijt et al., 2014), diffusion (Centola, 2010; Rogers, 1995), or network externalities (Ellison, 1993; Young, 1996)—is that they are propelled by avoidance of trust abuse. A test of our argument must therefore also show that a reputation system leads trustors to herd around a single trustee more so in the presence than in the absence of a trust problem. The laboratory permits such controlled comparisons between situations of trust and no trust that are necessary to critically evaluate reputation cascades as a generative mechanism for inequality.

6.4.2 Design of the experiment

In our laboratory experiment, 340 subjects played our game under different reputation conditions and in the presence or absence of a trust problem.

Experimental games and conditions: Subjects played in groups of 4 trustors and 4 trustees ($N_1 = N_2 = 4$) and earned points that converted to US dollar cents at the end of the experiment ($S_1 = 0$, $P_1 = P_2 = 30$, $R_1 = R_2 = 50$). Heterogeneity among trustees in the incentive to abuse trust was not induced by design ($T_j = T_2$). Instead we relied on intrinsic heterogeneity among subjects in trustworthiness, for example, due to heterogeneity in social preferences (Aksoy & Weesie, 2012). This strategy allowed us to keep the experiment simple and readily understandable to subjects, and prevented heterogeneity in trustworthiness from becoming artificially salient.

We varied the trustees' payoff for trust abuse, T_2, across conditions in order to study the role of trust in the production of reputation cascades. In the condition "Trust Problem", T_2 was 80 or 100 points for all trustees; in the control condition, "No Trust Problem", abusing trust was costly for trustees: T_2 was 0.

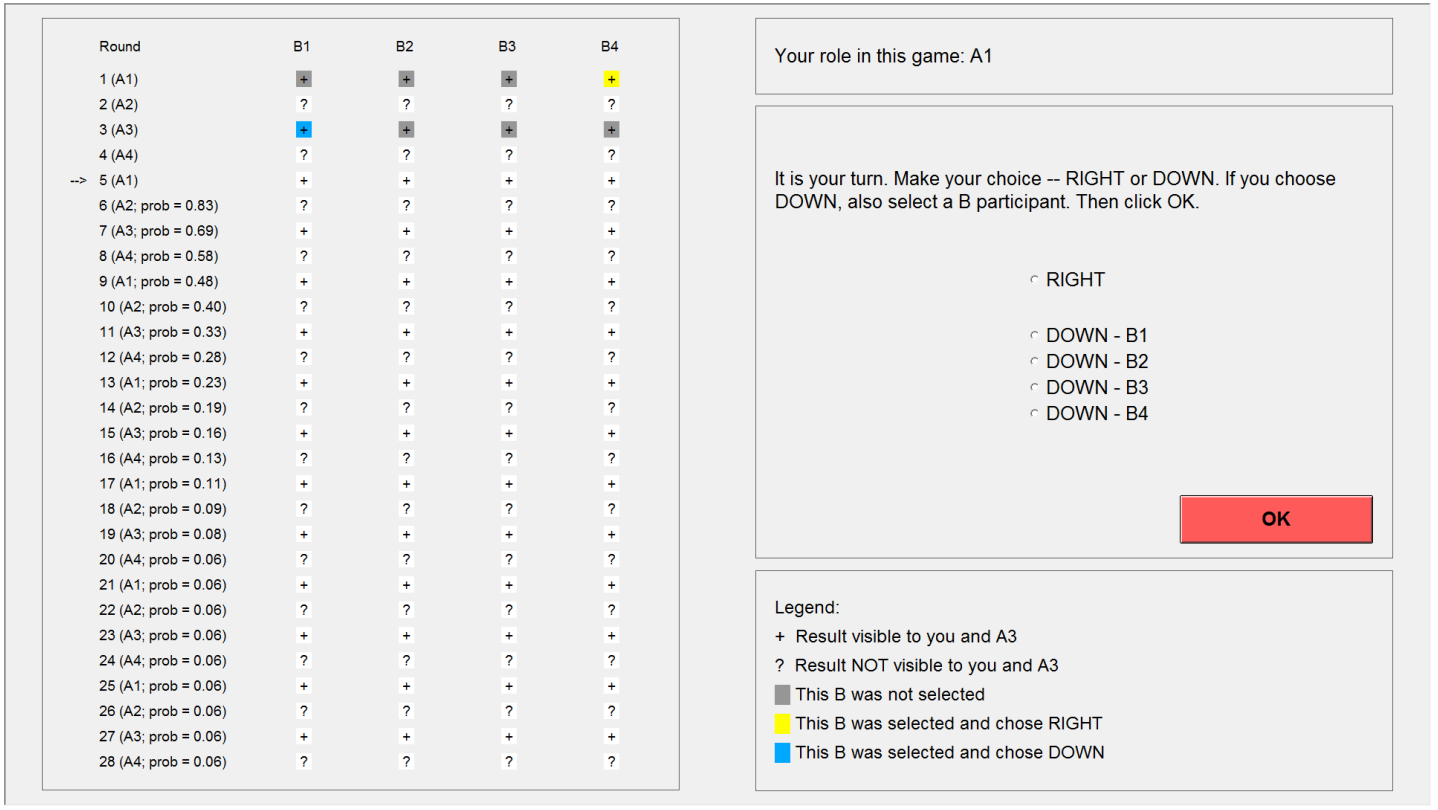
Games were played in three different reputation conditions. In the "Private" condition, the computer interface showed trustors only their own past choices and those made by trustees they placed trust in. In the "Partial" condition, the pair of even-numbered trustors (trustors 2 and 4) and the pair of odd-numbered trustors (trustors 1 and 3) could see the choices made by each other and corresponding trustee behavior but not by the other two trustors. In the "Full" condition, trustors could see the choices made by trustors and trustees in all prior rounds. We note that our Partial condition allows a direct demonstration of arbitrariness in the selection of trustees. The mutual exclusivity of reputation information available to even- and odd-numbered trustors makes it possible for two cascades to form involving two distinct trustees.[2] Information available to trustees was held constant across conditions; they always saw the entire history.

The computer interface: Figures 6.1a and 6.1b show screenshots from the Partial condition. The computer interface was inspired by that of Huck et al. (2012) and programmed in z-tree (Fischbacher, 2007). Figure 6.1a shows a screen of a trustor 1 (referred to as A1) in round 5. Trustor 1 is asked to choose to withhold trust (RIGHT) or place trust in a specific trustee (DOWN – B1, B2, B3, B4). The left-hand side of the screen shows a "history table" with 4 columns representing the 4 trustees (B1, B2, B3, and B4) and rows representing rounds. The current round, round 5, is indicated by the arrow "-->". In parentheses it is displayed which trustor (A) is at play in which round and for potential future rounds it is also displayed what the probability is that the game reaches that round.[3]

In the Partial condition (as well as the Private condition), a trustor saw + signs in the rounds for which he/she received or would receive information on trustor and trustee choices and question marks in the other rounds. Information on past choices was displayed by coloring the cells. Dark gray showed that a trustee was not chosen, yellow that a trustee was chosen and abused trust, and blue that a trustee was chosen and honored trust. In Figure 6.1a, trustor 1 sees that he/she placed trust in trustee 4

[2]The Private condition does in principle also allow a direct demonstration of arbitrariness in the selection of trustees. However, we expect to observe little repeated exchanges in the Private condition as we expect low levels of honored trust without information sharing and because the games lasted only a few rounds (see below).

[3]We acknowledge a mistake in the program: The probabilities for reaching the last eight displayed rounds were displayed incorrectly and were always the same as the probability of reaching round 20. Because this was identical in all conditions, we do not expect that differences in behavior across conditions are caused by this mistake.



(a) Screen for a trustor in the Partial condition.

Figure 6.1: Example screens from the experiment.

(b) Screen for a trustee in the Partial condition.

(B4) in round 1 and that trustee 4 abused trust. Trustor 1 also sees that, in round 3, trustor 3 placed trust in trustee 1 who then honored trust. Trustor 1 does not see what happened in rounds 2 and 4. We note that in the Private condition, trustor 1 would in round 5 only see the outcome of round 1. In the Full condition, trustor 1 would see the outcomes of all past rounds.

The screen was similar for trustees but trustees always saw what happened in all past rounds (see Figure 6.1b). In the Private and Partial conditions, trustees nevertheless also saw + signs and question marks, indicating about which past rounds the trustor currently at play has information and in which future rounds trustors will have information about the current round. Further information on the computer interface is found in the instructions that were distributed to participants and that are included in Appendix E.4.

Organization of the experiment: In each session of our experiment, subjects played 8 games subsequently. The first 4 games were played under one reputation condition and the second 4 games under a different reputation condition. The value of T_2 was the same throughout a session.

Each session had 24 subjects so that 3 groups of 8 subjects could be formed. For each of the 8 subsequent games, the subjects were randomly assigned to one of the 3 groups and the role of trustor and trustee.

We organized 14 sessions. For each of the values $T_2 = 80, 100$ (Trust Problem), we conducted 6 sessions: 2 with each reputation condition in games 1 to 4 (one for each of the other reputation conditions in games 5 to 8). For $T_2 = 0$ (No Trust Problem), we conducted 2 sessions: one with the Partial condition in games 1 to 4 and the Full condition in games 5 to 8, and one with the reversed order. The No Trust Problem and Private conditions were not used in conjunction as we expected that the absence of a trust problem alone as well as the absence of information sharing alone prevents the emergence of reputation cascades.

How many rounds the games lasted was determined in advance using a pseudo random algorithm (for a continuation probability $w = 5/6$) in order to make the length of games uniform across conditions. Subjects were informed that the length of the games was predetermined in this manner but not about the actual lengths of the games (see Dal Bó & Fréchette, 2011; Fréchette & Yuksel, 2013, on the implementation of indefinitely repeated games in experiments). The 8 games played in a session lasted 3, 5, 9, 5, 2, 8, 7, and 8 rounds.

The experiment was conducted at the CBPE laboratory at Stony Brook University. The participating subjects (overwhelmingly undergraduate students; average age = 20, 54% females) were recruited online using ORSEE (Greiner, 2004). Sessions

lasted up to about 90 minutes and subjects earned on average 23.6 US dollars.

6.5 Results

6.5.1 Exploratory data analysis

Data from all 336 games are visualized in Figure 6.2. Each game is represented as a small light gray grid consisting of some number of rounds (rows) and four trustees (columns), similar as in the computer interface used in the experiment. Black circles denote honored trust and gray triangles abused trust. The display of the games is organized by the experimental conditions.[4]

Figure 6.2 shows that cascade-like patterns—series of honored trust repeatedly placed by all four trustors in a single trustee—frequently obtained in the Full x Trust Problem condition, as demonstrated by vertical strings of black circles. In the Partial x Trust Problem condition, pairs of trustors often were locked in on two distinct trustees at the expense of the two other trustees, recognizable in Figure 6.2 as alternating diagonal patterns within a game. Such duopolies are not apparent in the Full x Trust Problem condition. In both the Partial x Trust Problem and Full x Trust Problem condition, the cascades often formed before other trustees were given any opportunity to honor trust. These observations suggest a substantial potential for cascading in reputation systems, with potentially trustworthy trustees being excluded from exchange in a fully arbitrary manner.

In Figure 6.2, cascade-like patterns—vertical series of black circles—are not apparent in the Private x Trust Problem condition, where trustors could not see the experiences of others. This indicates that shared preferences for a particular trustee identity were not a significant factor in the production of coordinated partner selection patterns.

One can also hardly spot any cascade-like patterns in either of the No Trust Problem conditions, in which abuse was costly for trustees. This suggests that cascades in the Trust Problem condition are not the result of a general tendency to imitate the choices of others, but rather the result of a desire to minimize the risk of abuse by selecting a transaction partner who honored trust before.

[4] In Figure 6.2, the games are furthermore arranged to blocks of three times four games separated by somewhat wider margins. Each of these blocks of three times four games shows data from the first or second sequence of four games played by three groups in a session.

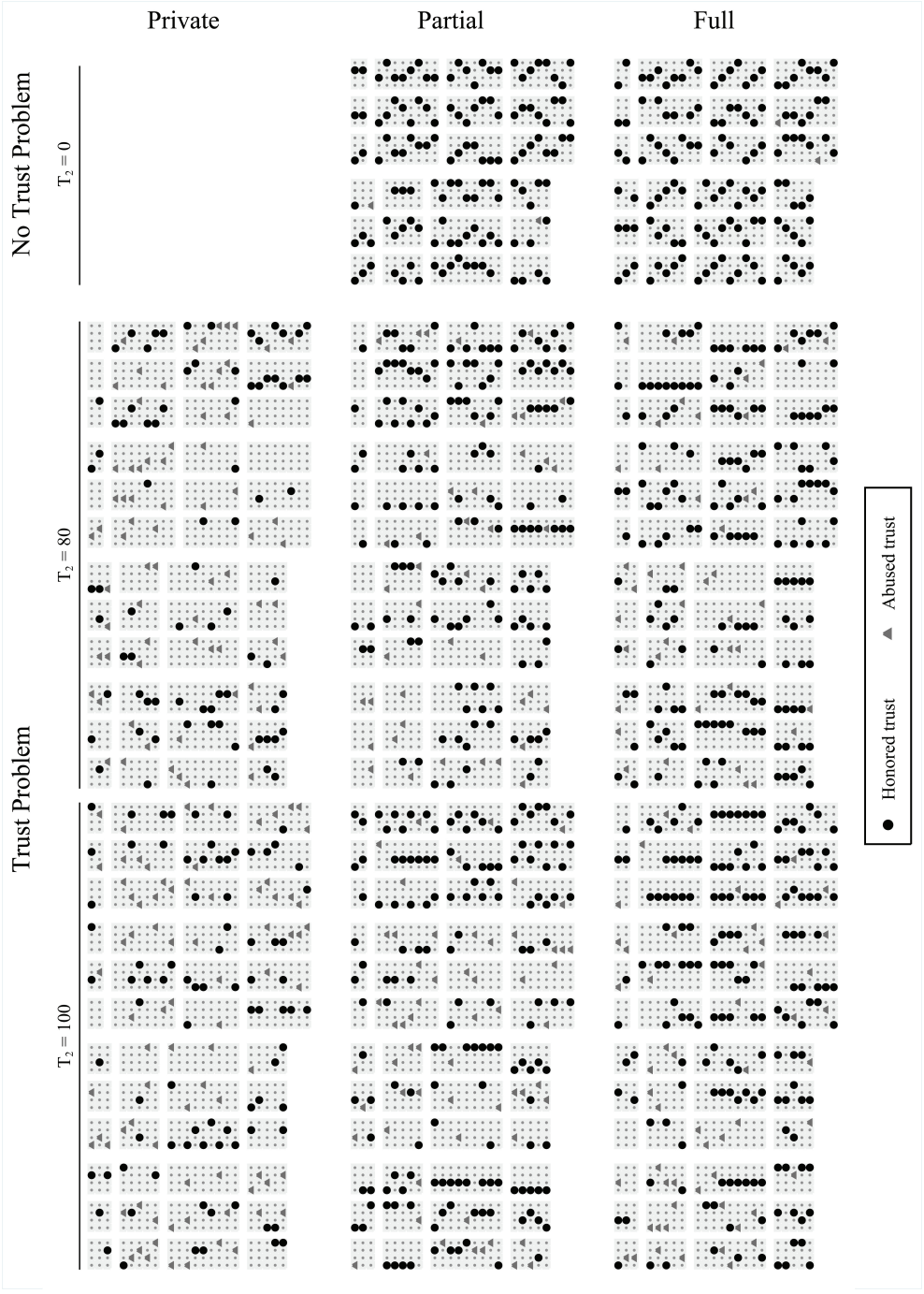


Figure 6.2: Overview of the data: Play histories of the observed 336 games.

6.5.2 Honored trust and inequality across reputation and trust conditions

To statistically investigate the effects of information sharing and the presence of a trust problem, we measure "efficiency" and "inequality" at the level of (repeated) games ($N = 336$). Efficiency is measured as the rate of honored trust—the number of rounds in which trust was placed and honored divided by the total number of rounds of a game. For inequality, we focus on how often each of the four trustees was trusted throughout the game and measure inequality in these counts with the Modified Coefficient Of Variation (MCOV; Allison, 1980b).[5]

Figure 6.3 and Table E.2 in Appendix E.3 report how information sharing affected the levels of honored trust and inequality in the Trust Problem conditions. The reported results were obtained from linear regressions of honored trust and the MCOV on dummy variables for the Partial and Full conditions, with standard errors adjusted for the clustering of games in sessions. Figure 6.3 illustrates the effects of information sharing across the three Trust Problem conditions by showing for each the estimate of both the level of honored trust (black circles) as well as the level of inequality (gray squares). Levels of honored trust were significantly ($p < .01$) higher in the two conditions with information sharing—Partial and Full—than in the Private condition. As predicted, information sharing did also lead to greater inequality, with trustee differentiation being significantly ($p < .05$) higher in Full than in Private.

Figure 6.3 invites the speculation that partial information sharing may achieve high levels of honored trust while preventing overly large inequalities. It is conceivable theoretically that an intermediate degree of information sharing pools the sanctioning potential of a large enough number of trustors to let the level of honored

[5]The method-of-moments estimate of the MCOV, (variance – mean) / mean^2, is inefficient (Allison, 1980b) and we, therefore, use the maximum-likelihood estimation of the MCOV described in Allison (1980a). Specifically, performing for each game a negative binomial regression of the counts of trust placed in each trustee yields, for each repeated game, the maximum-likelihood estimate of over-dispersion (alpha) that is an efficient estimate of the MCOV. This measure of inequality is estimated only for the 297 games in which trust was placed more than once.

Other common measures of inequality like the Gini Index or the Herfindahl Index lead to results that are mostly consistent with those reported here but are less suited for our purpose. The Gini index is downward-biased when calculated for small populations, with the size of the bias depending on the distribution of the variable of interest (Deltas, 2003; Van Ourti & Clarke, 2011). This is problematic because we expect the distribution of the number of times trustees were trusted to vary between the conditions. The Herfindahl Index calculated on count data where the number of counts is small tends to be downward-biased, too. Hall (2005) presents a method for correcting the bias but this correction is valid only if the total number of "events" (placements of trust) is large relative to the individual cell counts. This may be the case in some conditions of our experiment but not in others.

It is, furthermore, conceivable to measure inequality among trustees in payoffs rather than in how often they were trusted. Analyses of inequality in payoffs lead to results that resemble those reported here.

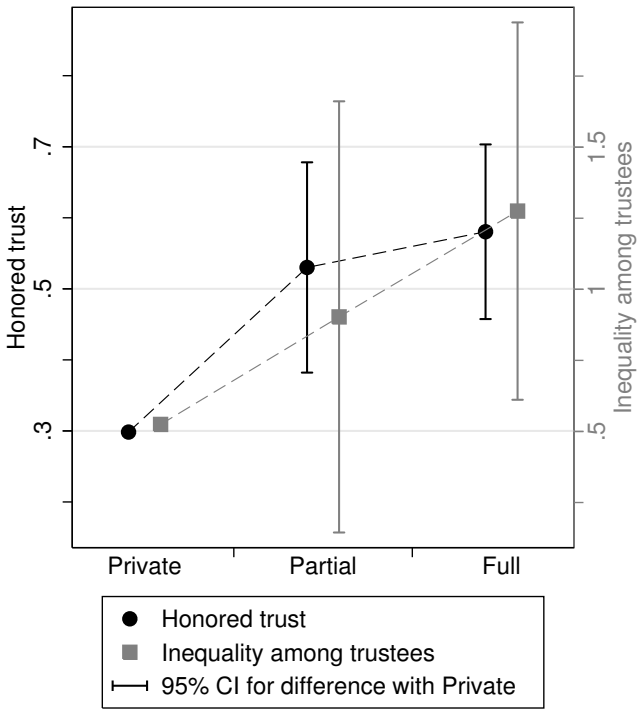


Figure 6.3: Honored trust (black circles; left vertical axis) and inequality (gray squares; right vertical axis) in the Trust Problem condition by degree of information sharing (Private, Partial, Full). 95% confidence intervals indicate the significance of differences in honored trust and inequality in the conditions Partial and Full compared to the Private condition.

trust approach its theoretical maximum. All that more comprehensive information sharing then does is creating "excess inequality" with no more than negligible gains in efficiency. Figure 6.3 shows that, in the Trust Problem condition, the uptick in honored trust occurred mainly from Private to Partial whereas inequality among trustees increased more continuously over the reputation conditions. However, second-order differences between effects on trust and inequality fall short of statistical significance.

Figure 6.4 illustrates the effects of the presence of a trust problem in the conditions with information sharing (Partial and Full), reporting results of linear regressions of honored trust and inequality on dummy variables for the experimental conditions (see also Table E.4 in Appendix E.3). The displayed 95% confidence intervals indicate the significance of differences in honored trust and inequality between the Trust Problem and No Trust Problem conditions (reported confidence intervals and standard errors

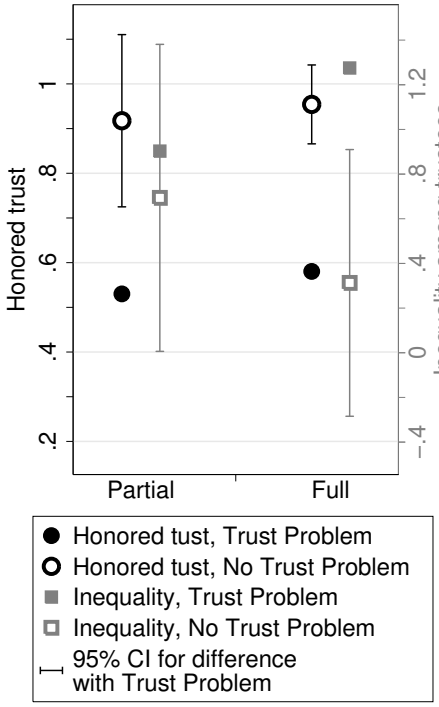


Figure 6.4: Honored trust (black circles; left vertical axis) and inequality (gray squares; right vertical axis) in the Partial and Full conditions by the Trust Problem condition. 95% confidence intervals indicate the significance of differences in honored trust and inequality between the Trust Problem and No Trust Problem condition.

are adjusted for the clustering of games in sessions). Figure 6.4 shows that inequality among trustees was significantly ($p < .05$) lower in the No Trust Problem condition than the Trust Problem condition with complete information sharing (Full), while the difference was in the same direction but not significant under Partial information sharing. This evidence supports the argument that inequality emerges due to the fear of trust abuse.

6.5.3 The sensitivity of trustors to trustee reputations

That inequality was higher in the Trust Problem conditions than the No Trust Problem conditions indicates that trustors reacted differently to trustee reputations depending on the presence of a trust problem. We used conditional logistic regressions to analyze how a trustee's reputation affected the odds of a trustor placing trust in that trustee instead of another trustee. These analyses were restricted to data from

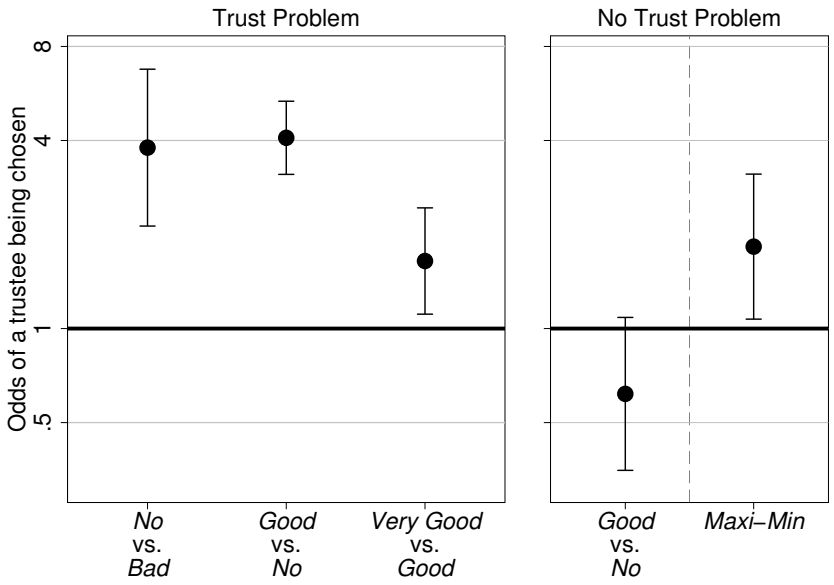


Figure 6.5: Effects of a trustee's reputation on the odds of a trustor who places trust choosing that trustee.

the Partial and Full conditions and performed separately for the Trust Problem and No Trust Problem conditions.[6] The two samples are 846 placements of trust (Trust Problem) and 262 placements of trust (No Trust Problem) by 237 and 48 individual trustors, respectively. Results are shown in Figure 6.5 and Table E.3 in Appendix E.3, with standard errors corrected for the clustering of choices in trustors.

For the Trust Problem conditions, our model includes three dummy variables that one after the other switch to 1 as a trustee's reputation improves. *No* is 1 if the trustor has no information of an abuse of trust by the focal trustee; *Good* is 1 if the trustor knows that the trustee honored trust at least once but not that the trustee ever abused trust; *Very Good* is 1 if the trustor knows that the trustee honored trust three or more times. The coefficient estimates inform about how much higher the odds of a trustee being chosen were if the trustee has one of these reputations rather than the next lower reputation (for trustees with "No" reputation, the reference category is trustees with a "Bad" reputation—trustees for which the trustor has observed an abuse of trust).

Figure 6.5 shows that in the Trust Problem conditions, trustors preferentially

[6]Including the Private condition does not change the results for the Trust Problem condition qualitatively.

chose trustees without a reputation (*No*) over trustees with at least one mark for abuse (*Bad*; $p < .001$). They also chose trustees with a "*Good*" reputation over trustees with "*No*" reputation ($p < .01$). These results are in line with the strategy for trustors assumed in our theoretical analysis. The results, furthermore, show that trustors significantly preferred trustees with 3 or more positive marks (*Very Good*) over trustees with a shorter history of honoring trust (*Good*; $p < .05$).

In the No Trust Problem conditions, such behavioral tendencies that lead to cascading and incentives for trustworthiness were absent. For the No Trust Problem conditions we omitted the variable *Very Good* because it rarely occurred that a trustor knew a trustee that honored trust three or more times. We, furthermore, excluded the only eight placements of trust by trustors who knew of an "irrational" abuse. The resulting regression with only one predictor variable reveals that in the No Trust Problem conditions trustors had a non-significant tendency to choose trustees with "*No*" reputation over trustees with a "*Good*" reputation ($p = .178$). This result, although not statistically significant, is in sharp contrast to what we observe in the Trust Problem conditions. It might indicate that, motivated by a preference for equality, trustors allocated exchanges in the No Trust Problem conditions in a manner that leads to equity among trustees.

A second conditional logistic regression with just one predictor variable supports the interpretation that trustors tried to reach equal outcomes in the No Trust Problem conditions. As illustrated in Figure 6.5, trustors did in the No Trust Problem conditions preferentially choose a trustee that has been trusted the least over the past rounds they observed (*Maxi-Min* = 1) over a trustee that has been trusted more often ($p < .05$).

We mention that Figure 6.4 in the preceding subsection also shows a pattern consistent with an interpretation that subjects tried to reduce inequality in the No Trust Problem conditions. Namely, compared to the Partial x No Trust Problem condition, inequality was lower in the Full x No Trust Problem condition, where subjects could better coordinate their actions to reach an equitable outcome (the difference is not statistically significant though).[7] That in the Trust Problem condition inequality increased (although not significantly) from the Partial to Full then suggest that in the Trust Problem conditions, the effect of an increased ability of trustors to distribute trust more evenly was overridden by the effect of an increased potential for cascades driven by the fear of trust abuse.

[7]Regressing the MCOV of the 48 games played in the No Trust Problem condition on a dummy for the Full condition yields a p-value of 0.417 for the difference in inequality between the Partial x No Trust Problem condition and the Full x No Trust Problem condition (adjusted for the clustering of games in two sessions).

6.6 Conclusions and discussion

It is well-known that reputation systems can resolve trust problems, especially if they span large numbers of actors (Milgrom et al., 1990; Buskens & Raub, 2013; Klein, 1997; Resnick et al., 2000). Here we argued that feedback effects in the reputation building process give rise to a form of cumulative advantage (DiPrete & Eirich, 2006; Merton, 1968; Van de Rijt et al., 2014) and an endogenous emergence of arbitrary inequality in transaction volume among trustees. Our theory also predicts that larger reputation systems lead to higher levels of inequality among trustees. The results of our experiment confirm the expectation that information sharing among trustors leads not only to higher levels of honored trust but also higher inequality among trustees. While inequality increased continuously over the conditions with Private, Partial, and Full information sharing, the increase in inequality between the Partial and Full conditions was not statistically significant. Hence, our experiment offers only weak evidence for the expectation that larger reputation systems lead to higher inequality. The results, furthermore, confirm that inequality emerged due to a fear of trust abuse, and not because of a general tendency to imitate the choices of others.

To the extent that unfounded inequality, with equally trustworthy trustees holding large market shares or being excluded from exchange, is a socially undesirable outcome, our study poses a system design challenge: Can reputation systems be restructured to allow more equitable exchange without large sacrifices in their ability to curb opportunistic tendencies? Although we have no solid evidence for the claim that larger reputation systems lead to higher inequality, our Partial reputation condition provides a tentative proof of principle that some fragmentation in reputation systems can decrease inequality while leading to only small losses in efficiency. Further theoretical studies and experiments with larger groups are needed to further investigate whether there is an ideal size of reputation systems that leads to high efficiency but prevents excess inequality.

Our theoretical model and laboratory experiment forms merely an ideal-type with various discrepancies with interactions in real-world reputation systems. Future research should investigate the potential for cascading in reputation systems when restrictive assumptions of our model are relaxed. First, when allowing for price competition, trustees who lack a good reputation could be able to attract trustors by offering low prices. However, we conjecture that if there is some lower bound for prices (e.g., if negative prices are infeasible), an established trustee can lower his price such that trustors prefer his offer over taking the risk of going for the cheaper offer of a newcomer. In line with this argument, Barwick & Pathak (2015, p. 104) report high inequality in transaction volumes among real estate agents in the Greater Boston

area, where agents can set prices competitively, trust in an agent's effort is clearly an issue, and reputations supposedly play a crucial role. Moreover, controlled laboratory experiments show that trustors are willing to forgo potentially more lucrative offers of untested parties in favor of the relative safety of exchanging in ongoing relationships with proven partners (Cook et al., 2004; Kollock, 1994; Yamagishi et al., 1998).

Second, we assumed that exchanges occur strictly sequentially and that information is available without any delay. When allowing for simultaneous exchanges (Huck et al., 2012) and delays in the transmission of information (Manapat & Rand, 2012), a group of trustors who share information might get locked in on more than one single trustee. However, trade would probably still concentrate on fewer trustees if information on past dealings is shared in larger reputation systems.

Third, if trustors do not always have the possibility to choose a trustee, previously excluded trustees can receive a chance to demonstrate their trustworthiness and attract future trustors. However, our finding that, in the Trust Problem condition, trustees with a longer history of trustworthiness were preferentially chosen over trustees with a shorter history of trustworthiness (Figure 6.5) suggests otherwise. This tendency for strong and weak reputations to diverge might be even more pronounced in settings where reputation information is not always accurate and trustees may be able to fake a few positive ratings.

Relaxing the restrictive assumptions of our model could also reveal inefficiencies related to the endogenously produced trustee differentiation. For example, the reputational advantage gives a market leader a monopoly position that can potentially be exploited e.g., by asking high premiums (although, the argument on price competition above suggests that the threat of entrants may discipline an established party). Also, when trustors differ in their preferences for different trustees—e.g., because trustees offer slightly different goods or services or due to variation in geographical proximity between trustor-trustee pairs—trustors might ignore personal preferences in their attempts to avoid the risk of trust abuse. Differences in trustee quality could also lead to inefficiencies. Namely, if trustees differ in the value they provide to trustors when honoring trust, lock-in on a "mediocre trustee" could occur because individual trustors prefer exploiting the possibility of safely exchanging with an established mediocre trustee over further exploration (cf. Mason & Watts, 2012).

Appendix A

Mathematical details for Chapter 2

This appendix summarizes the notation used in Chapter 2, outlines the proofs of the propositions of this chapter, and provides some additional technical remarks.

A.1 Overview of notation and assumptions

Table A.1 summarizes the notation and assumptions used in Chapter 2.

Table A.1: Notation and assumptions used in Chapter 2

Symbol	Description and assumptions
Dilemma game G:	
G	A social dilemma game with two actors $i = 1, 2$; examples: Trust Game, Prisoner's Dilemma, Investment Game, Public Goods Game
D_i	Maximin and minimax strategy of actor i (a pure strategy)
$D = (D_1, D_2)$	Pareto-suboptimal strategy combination; unique subgame perfect equilibrium of G
C_i	Cooperation (a pure strategy)
$C = (C_1, C_2)$	Mutual cooperation and Pareto-optimal strategy combination; C is not an equilibrium of G; hence, $C_i \neq D_i$, for at least one actor $i = 1, 2$
B_i	Actor i's best-reply strategy against C_j $(i \neq j)$; B_i is a pure strategy; $B_i \neq C_i$ for $i = 1$ or $i = 2$;
U_i	Actor i's payoff (cardinal utility) in G
$U_i(D) = P_i$	Actor i's payoff from D
$U_i(C) = R_i$	Actor i's payoff from mutual cooperation; $R_i > P_i$ for $i = 1, 2$ (C is thus a Pareto-improvement compared to D)
$U_i(B_i, C_j) = T_i$	Actor i's best-reply payoff against C_j $(i \neq j)$; $T_i \geq R_i$ for $i = 1, 2$; $T_i > R_i$ for $i = 1$ or $i = 2$; in our examples: $B_i = D_i$ for $T_i > R_i$ and $B_i = C_i$ for $T_i = R_i$
Repeated game Γ:	
Γ	A noncooperative game with $N \geq 3$ actors who play in rounds $0, 1, 2, \ldots$; structure of the game is common knowledge
Rounds $t = 1, 2, \ldots$	In each round t, each of the N_1 actors in the role of actor 1 in G plays G once with each of the N_2 actors in the role of actor 2 in G ($N_1 + N_2 = N$); throughout Γ, each actor always plays in the same role
w	Probability that round $t + 1$ of Γ is played after round $t = 1, 2, \ldots$ has been played; $0 < w < 1$
Round 0	Actors can invest in social capital in the sense of establishing an information network that connects all N actors
Γ^-	Subgame of Γ after the information network has *not* been established

Γ^+	Subgame of Γ after the information network has been established
c	Total costs of establishing the information network; $c > 0$
U_i	Actor i's payoff in Γ (cardinal utility); U_i equals the sum of realized costs in round 0 and the exponentially discounted payoffs in rounds $1, 2, \ldots$
N_i-institution	Institution that provides actors in role i the possibility to invest in establishing the information network
N-institution	Institution that provides all actors the possibility to invest in establishing the information network

A.2 Proof of Proposition 2.1: Cooperation without a network

Proposition 2.1 is an extension of the fundamental theorem on trigger strategy equilibria in indefinitely often repeated games. We sketch the proof and refer to Friedman (1986, see, e.g., pp. 88–89) for details. We first show that if all other actors play trigger strategies in Γ^-, then the condition in Proposition 2.1 is a necessary and sufficient condition that a trigger strategy is a best-reply strategy for each actor. We prove this indirectly and thus assume that there is an actor i_1 who has a strategy that yields a higher payoff than the trigger strategy. Using such a strategy, there must be some round $t \geq 1$ such that i_1 defects for the first time against some partner j_1 in game G (otherwise, the alternative strategy cannot yield a higher payoff for i_1 than the trigger strategy). Without loss of generality, we can assume that actor i_1 deviates from the trigger strategy in round 1 and that actor i_1 is an actor in the role for which $(T_i - R_i)/(T_i - P_i)$ is maximal. We can focus on actors in that role, because the equilibrium condition has to be exactly restrictive enough to discipline the actors who have the largest incentive to deviate from the trigger strategies. We can restrict us to round 1, because in equilibrium, when arriving at round t, the game is exactly equivalent to the situation in round 1. As indicated, we have assumed for this role that actor i_1 plays all his games directly after each other, which implies that the other actors in role j who interact with i_1 later in round 1 are not informed on any earlier defection when they play themselves with i_1. Thus, these actors cooperate in round 1 when interacting with i_1. Since i_1's strategy implies defection against some partner j_1 in round 1, i_1's strategy can as well imply defection against each partner j in round 1. Consequently, each actor j will always play D_j after the first encounter with i_1. It is now straightforward to calculate what should hold if an actor i_1 has an incentive to defect. Namely, receiving N_j times T_i in round 1 and P_i in all subsequent games should be more attractive than receiving R_i throughout Γ^-:

$$\begin{aligned} N_j T_i + \sum_{t=2}^{\infty} w^{t-1} N_j P_i &> \sum_{t=1}^{\infty} w^{t-1} N_j R_i \\ \Leftrightarrow\ T_i + \frac{w P_i}{1-w} &> \frac{R_i}{1-w} \\ \Leftrightarrow\ w &< \frac{T_i - R_i}{T_i - P_i} \end{aligned} \tag{A.1}$$

This is in contradiction with the condition in Proposition 2.1, implying that there is no actor who has an incentive to deviate from the trigger strategies under the condition in Proposition 2.1. This completes the proof that the condition in Proposition 2.1 is sufficient for the existence of a cooperation equilibrium. The necessity follows from the fact that we considered the largest $(T_i - R_i)/(T_i - P_i)$ under the assumption that D is a combination of minimax strategies. That assumption implies that actors in role j do not have punishment strategies at hand that deter defection if the condition in Proposition 2.1 is not fulfilled. To see that the equilibrium is also subgame perfect, it suffices to note that a combination of trigger strategies implies that each actor i will choose D_i unconditionally for the rest of Γ^- as soon as i has information that some actor has defected. However, if each actor chooses D_i unconditionally, this always constitutes an equilibrium in Γ^- as well as in all subgames of Γ^-. Hence, the combination of trigger strategies is a subgame perfect equilibrium under the condition in Proposition 2.1. This completes the sketch of the proof.

A.3 Remark: Interaction order and incentives for defection

Note why we assume that an actor i_1 with the highest value for $(T_i - R_i)/(T_i - P_i)$, when playing his first game with an actor in role j in a given round, plays the games with all other $N_j - 1$ actors j immediately afterwards in that round. The first actor j_1 who encounters i_1's defection, will play D_j against all actors in role i, since j_1 uses a trigger strategy. Hence, actors i_2, i_3, ... will start playing D_i at least after having played with that actor j_1. Consequently, other actors j will start playing D_j at least after having played with i_2 or i_3. Thus, if i_1 would not play the games with all actors j immediately after his first defection against j_1, it could happen that j_1 plays against i_2 and i_2 against j_2 before i_1 plays against j_2. Through this series of games, j_2 would defect when playing with i_1. This would decrease the incentive for i_1 to deviate from the trigger strategy in the first place. We have thus chosen the

sequence of games within a round so that the incentives for deviating from a trigger strategy are maximized. It is far from trivial to specify conditions for a cooperation equilibrium under assumptions that allow for the possibility that an actor with the highest value for $(T_i - R_i)/(T_i - P_i)$ does not play all of his games in a given round immediately after each other since the ordering can affect incentives to defect and can also affect optimal timing of the first defection.

A.4 Remark: The requirement of pure strategies

Note that the proof of Proposition 2.1 would be more complicated for games G in which payoffs are affected by probabilistic moves of Nature or if C_i or D_i would be mixed strategies so that payoffs could be expected rather than certain values.

A.5 Proof of Proposition 2.2: Cooperation in a network

To prove Proposition 2.2, we use a similar approach as for Proposition 2.1. The important difference with Proposition 2.1 is that after an actor i has defected once, all partners in the other role j are informed immediately. Because they use trigger strategies, they will immediately start playing D_j themselves in all future games G. Thus, payoff maximization requires that the actor i who started to defect, plays D_i in all future games G. We can focus again on an actor in role i with the largest $(T_i - R_i)/(T_i - P_i)$ and on deviations in round 1. If an actor in role i has an incentive to deviate from the trigger strategy against one of the other actors j^*, with $1 \leq j^* \leq N_j$, it should hold that (note that we now again exploit our assumption that payoffs are certain rather than expected values):

$$(j^* - 1)R_i + T_i + (N_j - j^*)P_i + \sum_{t=2}^{\infty} w^{t-1} N_j P_i > \sum_{t=1}^{\infty} w^{t-1} N_j R_i = N_j R_i + \frac{w N_j R_i}{1-w}. \quad \text{(A.2)}$$

Note that the left-hand side of Eq. (A.2) is maximized for $j^* = N_j$: the best the actor in role i can do, is to defect in the final game G that he plays in round 1 of Γ^+. This implies

$$\begin{aligned} (N_j - 1)R_i + T_i + \frac{w N_j P_i}{1-w} &> N_j R_i + \frac{w N_j R_i}{1-w} \\ \Leftrightarrow\ T_i + \frac{w N_j P_i}{1-w} &> R_i + \frac{w N_j R_i}{1-w}, \end{aligned} \quad \text{(A.3)}$$

which is also equivalent with

$$w < \frac{T_i - R_i}{T_i - P_i + (N_j - 1)(R_i - P_i)}. \tag{A.4}$$

This is in contradiction with the condition in Proposition 2.2, thus refuting our assumption that i has a better reply than a trigger strategy to trigger strategies of all other actors under the condition of Proposition 2.2. It follows that a trigger strategy is indeed a best-reply strategy for all actors, because actors in role j have even smaller incentives to deviate. This shows that trigger strategies played by all actors in Γ^+ constitute a Nash equilibrium under the condition in Proposition 2.2. To see that this Nash equilibrium is also subgame perfect, it suffices to note that a combination of trigger strategies implies that all actors i choose D_i unconditionally for the rest of Γ^+ as soon as some player defected. However, if each actor chooses D_i unconditionally, this always constitutes an equilibrium in Γ^+ as well as in all subgames of Γ^+. Hence, the combination of trigger strategies is a subgame perfect equilibrium under the condition in Proposition 2.2. This completes the sketch of the proof.

A.6 Proof of Proposition 2.3: Properties of w_i^+ and w^+

Proposition 2.3 follows directly from our assumptions on the parameters of G and Γ. First, since $T_i > R_i > P_i$ and $N_j \geq 2$ for actors in role i with an incentive to defect, it follows from $w_i^+ = (T_i - R_i)/(T_i - P_i + (N_j - 1)(R_i - P_i))$ that w_i^+ is larger than 0 and smaller than 1, decreases in N_j, and goes to 0 for $N_j \to \infty$. For $N_j = 1$, we have $w_i^+ = w_i^-$ and, hence, $w_i^+ < w_i^-$ for $N_j \geq 2$.

Second, since $T_i > R_i > P_i$ and $N_j \geq 2$ for at least one i, it also follows from $w^+ = \max_{i=1,2}(T_i - R_i)/(T_i - P_i + (N_j - 1)(R_i - P_i))$ that w_i^+ is larger than 0 and smaller than 1, weakly decreases in N_j when actors in role i have an incentive to defect, and goes to 0 if $N_j \to \infty$ when actors in role $i \neq j$ have an incentive to defect. Since $N_j \geq 2$, $w^+ < w^-$.

Note, too, that $w_i^- > w_j^-$, while $w_i^+ > w_j^+$ $(i \neq j)$ if actors in both roles i and j have an incentive to defect and if the number N_j of i's partners is sufficiently larger than the number N_i of j's partners.

A.7 Proof of Proposition 2.4: Value of social capital

Proposition 2.4 follows directly from the equilibrium payoffs in Γ^+ and Γ^- under the assumption that $w^+ \leq w < w^-$ so that cooperation equilibria exist only in Γ^+ and that D will be played throughout in Γ^-. Namely, the payoff associated with the trigger strategy equilibrium in Γ^+ is $N_j R_i/(1-w)$, while $N_j P_i/(1-w)$ is the payoff in Γ^-. The upper bound on the value of social capital is equal to the difference between these payoffs.

A.8 Remark: The specification of an upper bound on the value of social capital

Note that playing D throughout need not be the only equilibrium outcome of Γ^- when $w < w^-$. For example, assume that G is the Trust Game. When $w < w^-$ there can be equilibria of Γ^- such that sellers sometimes abuse trust, while buyers always place trust and each actor's equilibrium payoff is larger than $N_j P_i/(1-w)$. The value of social capital would then be smaller. Hence, Proposition 2.4 only specifies an upper bound on the value of social capital.

A.9 Proof of Propositions 2.5 and 2.6: Investments in social capital under the N_i and N-institution

Propositions 2.5 and 2.6 can be derived as follows. The subgames Γ^+ and Γ^- always have equilibria such that all actors always play D in all games G in all rounds $1, 2, \ldots$ of Γ^+ and Γ^-. Moreover, under conditions (1) of Propositions 2.5 and 2.6 there are cooperation equilibria in Γ^+ but not in Γ^-. Cooperation equilibria are associated with higher payoffs for each actor than payoffs that are obtained when all actors always play D in all games G in all rounds $1, 2, \ldots$. Under the conditions of Proposition 2.5 and Proposition 2.6, we thus obtain the following equilibrium for Γ. First, each actor who can invest in social capital does indeed invest in round 0. Second, all actors play D unconditionally in each subgame Γ^-. Third, all actors play trigger strategies in subgame Γ^+. Proposition 2.5 and Proposition 2.6 follow directly.

The question emerges why the conditions in Proposition 2.5 and Proposition 2.6 are sufficient but not necessary for the existence of equilibria such that actors invest in social capital and subsequently always cooperate. Assume that $w^+ < w^- \leq w$. Then, cooperation equilibria also exist for subgame Γ^-. Thus, investments in social

capital are not necessary in round 0 for ensuring that cooperation equilibria exist in the subgames after round 0. Nevertheless, actors could play D unconditionally in Γ^- while they play trigger strategies in Γ^+. Investments in social capital are then still consistent with equilibrium behavior under conditions (2) of Propositions 2.5 and 2.6.

A.10 Remark: Homogeneity in payoff functions

We have assumed homogeneity in the sense that actors (at least actors for whom $T_i > R_i$) in the same role have the same payoff function. We have also assumed that actors with the highest temptation to defect play their games with the actors in the other role directly after each other. Together, these assumptions ensure that it cannot happen that there is a cooperation equilibrium for some pairs of actors in Γ^- (namely, pairs for whom the condition in Proposition 2.1 is fulfilled), but not for other pairs (those for whom the condition in Proposition 2.1 is not fulfilled). Propositions 2.1 and 2.2 also cover the case without homogeneity. However, the analysis of investments in social capital would become more complicated precisely because it then could happen that some but not all pairs of actors can cooperate conditionally in Γ^- so that for those pairs who can cooperate conditionally in Γ^-, no Pareto-improvements are feasible through cooperating conditionally in Γ^+. We leave a detailed analysis of this issue for a future paper.

A.11 Remark: n-actor social dilemma games

We have considered social dilemma games involving two actors. It would be interesting to consider n-actor social dilemma games with $n < N$. One could then assume that in each period of Γ subsets of the N actors play such n-actor social dilemma games. What are conditions such that establishing an information network between all N actors facilitates conditional cooperation in the separate n-actor dilemmas? This is an interesting question for which the answer is far from trivial and has to be left for future research.

Appendix B

Mathematical details for Chapter 3

This appendix provides the proofs for the propositions presented in Chapter 3 and is structured as follows. We first provide a summary of notation and assumptions. Then, we provide a sketch of the proofs of Propositions 3.1 to 3.4 on the sequential equilibria of the continuation games Γ^- and Γ^+ and the associated expected payoffs. These proofs relate to earlier results on sequential equilibria in finitely repeated games and are taken together in one section (Section B.2). The proof of the condition for the existence of an investment equilibrium (Proposition 3.5) is found in Section B.3 and the proofs of the comparative statics results for changes in π, S_1, P_1, and R_1 are presented in the Sections B.4 to B.7. The latter proofs all proceed over the same three steps that we explain in Section B.4 for the effect of changes in π. Sections B.4 and B.5 additionally provide Lemmas B.1 and B.2, which, respectively, imply Proposition 3.6 and Proposition 3.7 and establish in more detail how r_1 depends on π and S_1. Finally, the last section provides the proof of the effect of changes in N (Proposition 3.10).

B.1 Overview of notation and assumptions

Table B.1 summarizes the notation and assumptions used in Chapter 3.

Table B.1: Notation and assumptions used in Chapter 3.

Symbol	Description and assumptions
Trust Game with incomplete information:	
P_i	Payoff from withheld trust for a trustor ($i = 1$) or trustee ($i = 2$)
R_i	Actor i's payoff from honored trust; $R_i > P_i$
S_1	Trustor's payoff from abused trust; $S_1 < P_1$
T_2	Payoff from abuse of trust for a trustee of the opportunistic type; $T_2 > R_2$
$T_2 - \theta$	Payoff from abuse of trust for a trustee of the friendly type; $T_2 - \theta < R_2$
π	Probability that the trustee is of the friendly type; the trustee knows his type; the trustors are only informed about π
$RISK$	Measure for the risk a trustor incurs when placing trust; $RISK = (P_1 - S_1)/(R_1 - S_1)$
$TEMP$	Measure for the temptation of an opportunistic trustee to abuse trust; $TEMP = (T_2 - R_2)/(T_2 - P_2)$
Repeated game Γ *with two trustors interacting with one trustee:*	
Period 0.1	Probabilistic determination of the trustee's type
Period 0.2	Trustors can invest in establishing a link for information exchange between one another
c	Total costs of establishing information exchange; $c > 0$
Γ^-	Continuation of Γ after information exchange has *not* been established
Γ^+	Continuation of Γ after information exchange has been established
N	Number of Trust Games played between each trustor and the trustee
Periods $n = 1, 2, \ldots, 2N$	Periods in which the Trust Games are played; at the start of every odd period it is determined whether trustor 1 plays a TG with the trustee in that period while trustor 2 plays a TG with the trustee in the subsequent even period or vice versa
π_n^i	Trustor i's belief at the start of period n that the trustee is of the friendly type; π_n^i follows from Bayes' rule from the observed history and the trustees' strategies
X_1	Expected payoff of a trustor for the period in which the opportunistic trustee starts abusing trust with positive probability

τ	Number of TGs that need to be left before the opportunistic trustee starts abusing trust; in Γ^-, τ is the number of TGs that need to be left between a focal trustor and the trustee separately; in Γ^+, τ is the number of TGs that need to be left in total
$U_1^{\Gamma^-}$, $U_1^{\Gamma^+}$	A trustor's payoff in Γ^-, Γ^+; equals the sum of payoffs the trustor realized in periods $1, 2, \ldots, 2N$
r_1	A trustor's expected return on establishing the information exchange link; $U_1^{\Gamma^+} - U_1^{\Gamma^-}$; maximum a rational trustor is willing to pay for establishing information exchange

B.2 Sketch of the proof of Propositions 3.1 to 3.4: Equilibria and payoffs in Γ^- and Γ^+

In the analysis of Γ^- and Γ^+, we restrict the focus to equilibria that satisfy sequential rationality. In Γ^- (where each trustor is only informed about the outcomes of her own interactions with the trustee), the sequential equilibrium of the interactions between some trustor i and the trustee is identical to the sequential equilibrium of the finitely repeated TG with incomplete information and only one trustor. What complicates Proposition 3.1 (leading to the rounding in the exponents) is that, in Γ^-, the number of TGs trustor i and the trustee have left to play together after a given TG is the same irrespectively of whether this TG is played in an odd or the subsequent even period. In Γ^+ (where each trustor receives information also about the TGs of the other trustor), the sequential equilibrium as specified in Proposition 3.3 is such that with a given number of TGs left in total, the strategies and beliefs of the trustor at play and the trustee are as in the sequential equilibrium of the game with only one trustor, in the sense that the trustor's belief is the same as if she had played in all past TGs and that the strategies are the same as if she played in all the remaining TGs (see also Camerer & Weigelt, 1988). Buskens (2003) provides a formulation of the sequential equilibrium of the game with only one trustor that is similar to the formulation of our propositions.[1] Bower et al. (1997) provide the proof of this equilibrium for $N = 2$ and also for $N > 2$, which follows by induction. The proof that this is (generically) the unique equilibrium that satisfies sequential rationality is likewise found in Bower et al. (1997) and follows from the sketch of the derivation of the sequential equilibrium of the "one-trustor game" by Anderhub et al. (2002, Section 2). We note that the uniqueness of the sequential equilibrium is conditional

[1] Note that we count periods forward starting with 1 counting up to $2N$, whereas in Buskens (2003), as in many of the related papers, periods are counted backward such that the last period is period 1.

on the assumption of some reasonable refinement of out-of-equilibrium beliefs such as the intuitive criterion (see Anderhub et al., 2002) and the assumption that the only type of incomplete information is that the trustors are incompletely informed about whether or not the trustee has a short-term incentive to abuse trust (see Fudenberg & Maskin, 1986). The formulas for the calculation of the expected payoff of a trustor in Γ^- and Γ^+ as specified in Propositions 3.2 and 3.4, respectively, are implied by the sequential equilibria of these continuation games and their derivation is described in the main text.

B.3 Proof of Proposition 3.5: r_1 and the existence of investment equilibria

As stated in Section 3.3.4, a trustor's potential return on investment can be calculated as $r_1 = U_1^{\Gamma^+} - U_1^{\Gamma^-}$. This calculation yields

$$r_1 = \begin{cases} \frac{\tau(R_1-P_1)+P_1-\left(S_1+\frac{\pi}{RISK^{\tau-1}}(R_1-S_1)\right)}{2} & \text{if } \tau \leq N \\ \frac{(2N-\tau)(R_1-P_1)+\left(S_1+\frac{\pi}{RISK^{\tau-1}}(R_1-S_1)\right)-P_1}{2} & \text{if } N < \tau \leq 2N \\ 0 & \text{if } \tau > 2N. \end{cases} \tag{B.1}$$

This precise specification of r_1 (which we use in the rest of our proofs) implies the intervals for r_1 that Proposition 3.5 provides. Specifically, because $P_1 \leq S_1 + \frac{\pi}{RISK^{\tau-1}}(R_1 - S_1) < R_1$ (i.e., because the payoff a trustor can expect for the TG in which she still places trust with probability 1 while the opportunistic trustee begins to randomize is smaller than R_1 and, by the equilibrium property, must be at least as large as P_1), Eq. (B.1) implies that if $\tau \leq N$, $\frac{\tau-1}{2}(R_1 - P_1) < r_1 \leq \frac{\tau}{2}(R_1 - P_1)$ and that if $N < \tau \leq 2N$, $\frac{2N-\tau}{2}(R_1 - P_1) \leq r_1 < \frac{2N-(\tau-1)}{2}(R_1 - P_1)$.

An investment equilibrium exists if and only if, for each trustor, $c/2 \leq r_1$. If $c/2 \leq r_1$, proposing to invest maximizes a trustor's expected payoff, given the other trustor proposes to invest, because the relation cannot be established by the other trustor alone. On the other hand, if $c/2 > r_1$, both trustors proposing to invest is *not* an equilibrium because given that the other trustor proposes to invest, proposing to invest leaves the focal trustor (expectedly) worse off than not proposing to invest.

B.4 Additional results for changes in π and proof of Proposition 3.6

Lemma B.1 provides additional details on how r_1 changes in π and implies Proposition 3.6. Lemma B.1 quantifies the effects of a change in π and establishes that if trust is possible in Γ^+ as well as in Γ^- (possible in Γ^+ but not in Γ^-), r_1 increases (decreases) in a stepwise linear manner as π decreases.

Lemma B.1. *Given the specification of* Γ *and the definitions of* τ *and* r_1*, it holds that:*

- *If* $\tau + 1 \leq N$*,* r_1 *increases as* π *decreases; more specifically,*
  - r_1 *increases by* $\frac{1}{2}(R_1 - P_1)$ *if* π *decreases from* $RISK^\tau$ *to* $RISK^{\tau+1}$*.*
  - r_1 *increases linearly as* π *decreases gradually from* $RISK^\tau$ *to* $RISK^{\tau+1}$*.*
- *If* $N < \tau + 1 \leq 2N$*,* r_1 *decreases as* π *decreases; more specifically,*
  - r_1 *decreases by* $\frac{1}{2}(R_1 - P_1)$ *if* π *decreases from* $RISK^\tau$ *to* $RISK^{\tau+1}$*.*
  - r_1 *decreases linearly as* π *decreases gradually from* $RISK^\tau$ *to* $RISK^{\tau+1}$*.*

Procedure for the proof of the comparative statics results: To prove Lemma B.1, as well as to prove the postulated effects of changes in S_1, P_1, and R_1, we proceed in three steps. The procedure is best explained with reference to Figure B.1. As Figure 3.2, Figure B.1 shows how r_1 depends on π in an example with $N = 3$ and $RISK = 0.5$.

In the following proofs, we fix, as an anchor, a situation such that the opportunistic trustee's randomization starts some given number of TGs before the end of the game and such that a trustor is indifferent between placing and withholding trust in the TG in which the randomization starts. That is, we fix a situation in which π and $RISK$ are such that for a given τ, to which we refer as τ_0, π is precisely at the right-hand border of the interval $(RISK^{\tau_0-1}, RISK^{\tau_0}]$, the "interval τ_0." Figure B.1 presents the example in which we fix the interval $(\pi = 0.5, \pi = 0.25]$, where $\tau = 2$, as the interval τ_0.

In *step 1*, we show that r_1 changes as postulated if the parameter under study changes such that τ increases by 1 and that after the change a trustor is again indifferent between placing and withholding trust in the TG in which the trustee's randomization starts. We thus show how r_1 changes if π is at the right border of

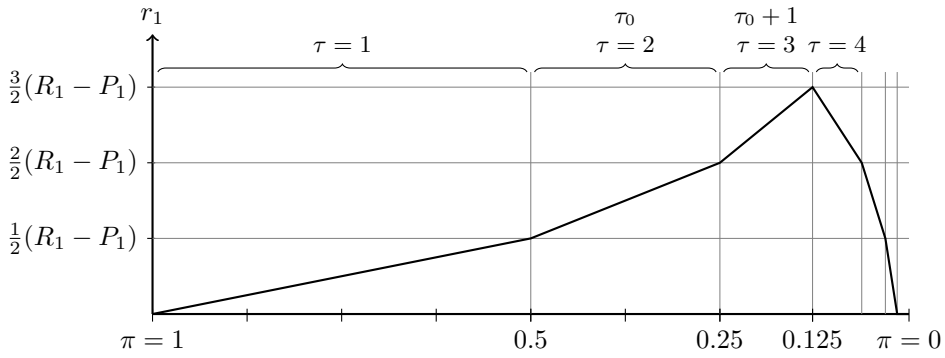


Figure B.1: Illustration for the visualization of the procedure used to prove the effect of changes in π on r_1 (example with $RISK = 0.5$ and $N = 3$).

the interval τ_0 before the parameter change, whereas after the parameter change, π is at the right border of the adjacent "τ interval" on the right, i.e., the τ interval in which randomization starts one TG earlier. We refer to the latter interval as the interval $\tau_0 + 1$ and we let $r_1^{\tau_0]}$ and $r_1^{\tau_0+1]}$ denote r_1 for the case that π is precisely at the right border of the interval τ_0 and $\tau_0 + 1$, respectively. This notation is somewhat cumbersome and we stress that the superscripts "$\tau_0]$" and "$\tau_0 + 1]$" for r_1 are indexes and not exponents. For the example illustrated in Figure B.1, we thus show in step 1 that r_1 increases by $\frac{1}{2}(R_1 - P_1)$ if π changes from 0.25 to 0.125. More generally, as we do not actually fix a specific τ_0, step 1 in the proof of the effects of changes in π shows that for any τ_0, $r_1^{\tau_0]} = r_1^{\tau_0+1]} - \frac{1}{2}(R_1 - P_1)$ if $\tau_0 + 1 \leq N$ and $r_1^{\tau_0]} = r_1^{\tau_0+1]} + \frac{1}{2}(R_1 - P_1)$ if $N < \tau_0 + 1 \leq 2N$.

In *step 2*, we prove that r_1 changes as postulated if the parameter under study changes such that we move again back towards the right border of the interval τ_0. That is, we show how $r_1^{\tau_0+1}$ (where we now use the superscript "$\tau_0 + 1$" to denote r_1 for *any* case that π is in the interval $\tau_0 + 1$) changes if π increases within the interval $\tau_0 + 1$ or if R_1, P_1, or S_1 changes such that *RISK* decreases and, therefore, the τ intervals shift to the right. Specifically for changes in π, we show that an increase in π within the interval $\tau_0 + 1$ leads to a linear decrease (increase) in $r_1^{\tau_0+1}$ if $\tau_0 + 1 \leq N$ (if $N < \tau_0 + 1 \leq 2N$).

Finally, we show, in *step 3*, that the limit of $r_1^{\tau_0+1}$ as we approach the right border of the interval τ_0 is larger or equal to (smaller or equal to) $r_1^{\tau_0]}$ if $r_1^{\tau_0]} < r_1^{\tau_0+1]}$ (if $r_1^{\tau_0]} > r_1^{\tau_0+1]}$). Specifically, for the example illustrated in Figure B.1, we show in step 3 that if π increases and more and more closely approaches 0.25 (and, as established in step 2, $r_1^{\tau_0+1}$, consequently, decreases), $r_1^{\tau_0+1}$ always remains at least as large as $r_1^{\tau_0]}$.

Proof of Lemma B.1: *Step 1.* If $\pi = RISK^{\tau_0}$, the payoff a trustor can expect in the TG in which the trustee begins to randomize ($X_1 = S_1 + \frac{\pi}{RISK^{\tau_0-1}}(R_1 - S_1)$) equals P_1. Hence, for the case that π is at the right border of the interval τ_0, r_1 (as specified in Eq. B.1) reduces to $r_1^{\tau_0]} = \frac{\tau_0}{2}(R_1 - P_1)$ and $r_1^{\tau_0]} = \frac{2N-\tau_0}{2}(R_1 - P_1)$ for $\tau_0 \leq N$ and $N < \tau_0 \leq 2N$, respectively. For $\pi = RISK^{\tau_0+1}$, it likewise holds that $X_1 = P_1$ (i.e., $S_1 + \frac{\pi}{RISK^{\tau_0+1-1}}(R_1 - S_1) = P_1$) and, hence, $r_1^{\tau_0+1]} = \frac{\tau_0+1}{2}(R_1 - P_1)$ and $r_1^{\tau_0+1]} = \frac{2N-(\tau_0+1)}{2}(R_1 - P_1)$ for $\tau_0 + 1 \leq N$ and $N < \tau_0 + 1 \leq 2N$, respectively. Consequently, if π decreases from $RISK^{\tau_0}$ to $RISK^{\tau_0+1}$, r_1 increases (decreases) by $\frac{1}{2}(R_1 - P_1)$ if $\tau_0 + 1 \leq N$ (if $N < \tau_0 + 1 \leq 2N$).

Step 2. The derivative of r_1 in π, neglecting that τ is a function of π, is

$$\begin{array}{ll} \frac{\partial r_1}{\partial \pi} = -\frac{1}{2}\frac{R-S}{RISK^{\tau-1}} & \text{if } \tau \leq N, \\ \frac{\partial r_1}{\partial \pi} = \frac{1}{2}\frac{R-S}{RISK^{\tau-1}} & \text{if } N < \tau \leq 2N. \end{array} \tag{B.2}$$

This shows that a marginal increase in π by $\Delta\pi > 0$ that does not affect τ leads to a decrease (increase) in $r_1^{\tau_0+1}$ by $\frac{\Delta\pi}{2}\frac{R_1-S_1}{RISK^{\tau_0+1-1}} > 0$ if $\tau_0+1 \leq N$ (if $N < \tau_0+1 \leq 2N$). Thus, if $\tau_0 + 1 \leq N$ (if $N < \tau_0 + 1 \leq 2N$), $r_1^{\tau_0+1}$ decreases (increases) linearly if π increases within the interval $\tau_0 + 1$.

Step 3. From Eq. (B.1), it follows that within a τ interval, a change in π affects r_1 exclusively through a change in X_1 such that r_1 changes by $\frac{1}{2}\Delta X_1$. Hence, because $P_1 \leq X_1 < R_1$, a change in π within some τ interval leads at maximum to a change in r_1 by $\frac{1}{2}(R_1 - \mu - P_1)$ for some $\mu > 0$, which is smaller than the change in r_1 that results from a change in π as considered in step 1. This shows that $r_1^{\tau_0+1}$ cannot become smaller (larger) than $r_1^{\tau_0]}$ if $r_1^{\tau_0]} < r_1^{\tau_0+1]}$ (if $r_1^{\tau_0]} > r_1^{\tau_0+1]}$).

B.5 Additional results for changes in S_1 and proof of Proposition 3.7

Lemma B.2 establishes in detail how r_1 changes as S_1 decreases and implies Proposition 3.7. Recall that a decrease in S_1 leads to an increase in *RISK* and potentially to an earlier start of the randomization.

Lemma B.2. *Consider the specification of* Γ *and the definitions of* τ *and* r_1 *and define* $\hat{S}_1$ *and* $\check{S}_1$ *such that* $\left(\frac{P_1-\hat{S}_1}{R_1-\hat{S}_1}\right)^{\tau} = \left(\frac{P_1-\check{S}_1}{R_1-\check{S}_1}\right)^{\tau+1} = \pi$. *It holds that:*

- *If* $\tau + 1 \leq N$, r_1 *increases as* S_1 *decreases; more specifically,*
  - r_1 *increases by* $\frac{1}{2}(R_1 - P_1)$ *if* S_1 *decreases from* $\hat{S}_1$ *to* $\check{S}_1$.

  - *r_1 increases strictly monotonically as S_1 decreases gradually from $\hat{S}_1$ to $\breve{S}_1$.*

- *If $N < \tau + 1 \leq 2N$, r_1 decreases as S_1 decreases; more specifically,*
  - *r_1 decreases by $\frac{1}{2}(R_1 - P_1)$ if S_1 decreases from $\hat{S}_1$ to $\breve{S}_1$.*
  - *r_1 decreases strictly monotonically as S_1 decreases gradually from $\hat{S}_1$ to $\breve{S}_1$.*

We prove Lemma B.2 by going through the three steps introduced above.

Step 1. Given $\pi = \left(\frac{P_1 - \hat{S}_1}{R_1 - \hat{S}_1}\right)^{\tau_0}$, r_1 reduces to $\frac{\tau_0}{2}(R_1 - P_1)$ and $\frac{2N - \tau_0}{2}(R_1 - P_1)$ for $\tau_0 \leq N$ and $N < \tau_0 \leq 2N$, respectively. Given $\pi = \left(\frac{P_1 - \breve{S}_1}{R_1 - \breve{S}_1}\right)^{\tau_0+1}$, r_1 reduces to $\frac{\tau_0+1}{2}(R_1 - P_1)$ and $\frac{2N-(\tau_0+1)}{2}(R_1 - P_1)$ for $\tau_0 + 1 \leq N$ and $N < \tau_0 + 1 \leq 2N$, respectively. Hence, a decrease in S_1 from $\hat{S}_1$ to $\breve{S}_1$, leads to an increase (decrease) in r_1 by $\frac{1}{2}(R_1 - P_1)$ if $\tau_0 + 1 \leq N$ (if $N < \tau_0 + 1 \leq 2N$).

Step 2. The derivative of r_1 in S_1, neglecting that τ is a function of S_1, is

$$\begin{array}{ll} \frac{\partial r_1}{\partial S_1} = -\frac{1}{2}\big(1 - \frac{\pi}{RISK^{\tau-1}} + (\tau - 1)\frac{\pi}{RISK^{\tau}}(1 - RISK)\big) & \text{if } \tau \leq N, \\ \frac{\partial r_1}{\partial S_1} = \frac{1}{2}\big(1 - \frac{\pi}{RISK^{\tau-1}} + (\tau - 1)\frac{\pi}{RISK^{\tau}}(1 - RISK)\big) & \text{if } N < \tau \leq 2N. \end{array} \tag{B.3}$$

This shows that if $\tau_0 + 1 \leq N$, a marginal increase in S_1 by $\Delta S_1 > 0$ that does not affect τ leads to a change in $r_1^{\tau_0+1}$ by $\frac{-\Delta S_1}{2}\big(1 - \frac{\pi}{RISK^{\tau_0+1-1}} + (\tau_0 + 1 - 1)\frac{\pi}{RISK^{\tau_0+1}}(1 - RISK)\big)$. This must be smaller than 0 (i.e., a decrease in $r_1^{\tau_0+1}$) because π must be in the interval $(RISK^{\tau_0+1-1}, RISK^{\tau_0+1}]$, which implies that $\frac{\pi}{RISK^{\tau_0+1-1}} < 1$ and $\frac{\pi}{RISK^{\tau_0+1}} \geq 1$. If $N < \tau_0 + 1 \leq 2N$, a marginal increase in S_1 for which π remains in the interval $\tau_0 + 1$ leads to an equivalent increase in $r_1^{\tau_0+1}$. Thus, if $\tau_0 + 1 \leq N$ (if $N < \tau_0 + 1 \leq 2N$), $r_1^{\tau_0+1}$ decreases (increases) monotonically if S_1 increases such that π remains in the interval $\tau_0 + 1$.

Step 3. The argument provided in *Step 3* of the proof of Lemma B.1 holds also for changes in S_1 as it does there for changes in π. From Eq. (B.1), it follows that a change in S_1 for which π remains in the same τ interval affects r_1 exclusively through a change in X_1 such that r_1 changes by $\frac{1}{2}\Delta X_1$. Hence, because $P_1 \leq X_1 < R_1$, a change in S_1 for which π remains in the same τ interval leads at maximum to a change in r_1 by $\frac{1}{2}(R_1 - \mu - P_1)$ for some $\mu > 0$, which is smaller than the change in r_1 that results from a change in S_1 as considered in step 1. This shows that $r_1^{\tau_0+1}$ cannot become smaller (larger) than $r_1^{\tau_0]}$ if $r_1^{\tau_0]} < r_1^{\tau_0+1]}$ (if $r_1^{\tau_0]} > r_1^{\tau_0+1]}$).

B.6 Proof of Proposition 3.8: Changes in P_1

To prove Proposition 3.8, we go through the same three steps as in the preceding two proofs. Recall that an increase in P_1 leads to an increase in *RISK* and, hence,

potentially to an increase in τ.

Step 1. The change in P_1 that we consider here is an increase from some P_1 to $P_1 + \mu$ (where $0 < \mu < R_1 - P_1$) such that $\pi = \left(\frac{P_1 - S_1}{R_1 - S_1}\right)^{\tau_0} = \left(\frac{P_1 + \mu - S_1}{R_1 - S_1}\right)^{\tau_0 + 1}$, i.e., a change in P_1 that leads to an increase in τ by 1 and where before and after the change $X_1 = P_1$ and $X_1 = P_1 + \mu$, respectively. We treat the scenario that $\tau_0 + 1 \leq N$ in *step 1.a* and the scenario that $N < \tau_0 + 1 \leq 2N$ in *step 1.b*.

Step 1.a. Given $\tau_0 + 1 \leq N$, $r_1^{\tau_0]} = \frac{\tau_0}{2}(R_1 - P_1)$ and $r_1^{\tau_0+1]} = \frac{\tau_0+1}{2}\big(R_1 - (P_1 + \mu)\big)$. Subtracting $r_1^{\tau_0+1]}$ from $r_1^{\tau_0]}$, we see that r_1 changes by $\frac{1}{2}\big(\mu(\tau_0 + 1) - (R_1 - P_1)\big)$. If (as postulated) r_1 increases, it must hold that $\frac{1}{2}\big(\mu(\tau_0 + 1) - (R_1 - P_1)\big) < 0$, which requires that

$$\mu < \frac{R_1 - P_1}{\tau_0 + 1}. \tag{B.4}$$

To show that this is the case, we derive from $\left(\frac{P_1 - S_1}{R_1 - S_1}\right)^{\tau_0} = \left(\frac{P_1 + \mu - S_1}{R_1 - S_1}\right)^{\tau_0 + 1}$ that for the increase in P_1 by μ to lead to an increase in τ by 1 (with before and after the increase $P_1 = X_1$ and $P_1 + \mu = X_1$, respectively) it must hold that

$$\mu = (R_1 - S_1)\Big(\frac{P_1 - S_1}{R_1 - S_1}\Big)^{\frac{\tau_0}{\tau_0+1}} - (P_1 - S_1). \tag{B.5}$$

By replacing μ in Eq. (B.4) with the right-hand side of Eq. (B.5) and (without loss of generality) "normalizing" to $R_1 = 1$, $S_1 = 0$, and $0 < P_1 < 1$, we obtain

$$P_1^{\frac{\tau_0}{\tau_0+1}} - P_1 < \frac{1 - P_1}{\tau_0 + 1}. \tag{B.6}$$

Rearranging leads to

$$P_1^{\frac{\tau_0}{\tau_0+1}} < \frac{1}{\tau_0 + 1} + \frac{\tau_0}{\tau_0 + 1} P_1. \tag{B.7}$$

To establish that Eq. (B.7) holds, and r_1, thus, indeed increases, we now isolate the left-hand side of Eq. (B.7). We replace P_1 by $1 - w$ (so $P_1 = 1 - w$ and $w = 1 - P_1$), which allows rewriting $P_1^{\frac{\tau_0}{\tau_0+1}}$ (the left-hand side of inequality B.7) as a binomial series:

$$\begin{aligned}(1 - w)^{\frac{\tau_0}{\tau_0+1}} &= \sum_{k=0}^{\infty} \binom{\frac{\tau_0}{\tau_0+1}}{k} (-w)^k \\ &= 1 + \frac{\tau_0}{\tau_0 + 1}(-w) + \frac{\frac{\tau_0}{\tau_0+1}(\frac{\tau_0}{\tau_0+1} - 1)}{2!}(-w)^2 + \\ &\quad \ldots + \frac{\frac{\tau_0}{\tau_0+1}(\frac{\tau_0}{\tau_0+1} - 1)(\frac{\tau_0}{\tau_0+1} - 2)\ldots(\frac{\tau_0}{\tau_0+1} - k + 1)}{k!}(-w)^k + \ldots\end{aligned} \tag{B.8}$$

It can be seen that every element that is "added" to 1 in this series is smaller than 0. If k is even, the numerator is negative while $(-w)^k$ is positive. If k is uneven, the numerator is positive while $(-w)^k$ is negative. Thus, $(1-w)^{\frac{\tau_0}{\tau_0+1}}$ must be smaller than what we obtain when carrying out only one step of the summation. That is, Eq. (B.8) implies that

$$(1-w)^{\frac{\tau_0}{\tau_0+1}} < 1 + \frac{\tau_0}{\tau_0+1}(-w). \tag{B.9}$$

Replacing w again by $1-P_1$, we can thus assert that

$$P_1^{\frac{\tau_0}{\tau_0+1}} < 1 - \frac{\tau_0}{\tau_0+1}(1-P_1). \tag{B.10}$$

This can be rearranged to Eq. (B.7), which shows that Eq. (B.7) is true and, thus, proves that (as postulated) r_1 increases if P_1 increases as considered and $\tau_0+1 \leq N$, i.e., that $r_1^{\tau_0]} < r_1^{\tau_0+1]}$ if $\tau_0+1 \leq N$.

Step 1.b. Given $N < \tau_0+1 \leq 2N$, $r_1^{\tau_0]} = \frac{2N-\tau_0}{2}(R_1-P_1)$ and $r_1^{\tau_0+1]} = \frac{2N-(\tau_0+1)}{2}(R_1-(P_1+\mu))$. Subtracting $r_1^{\tau_0+1]}$ from $r_1^{\tau_0]}$, we see that r_1 changes by $\frac{1}{2}\big(\mu(2N-(\tau_0+1))+R_1-P_1\big)$. It holds that $\frac{1}{2}\big(\mu(2N-(\tau_0+1))+R_1-P_1\big) > 0$, implying that (as postulated) $r_1^{\tau_0]} > r_1^{\tau_0+1]}$, i.e., that r_1 decreases.

Step 2. The derivative of r_1 in P_1, neglecting that τ is a function of P_1, is

$$\begin{array}{ll} \frac{\partial r_1}{\partial P_1} = \frac{\tau-1}{2}\big(\frac{\pi}{RISK^\tau}-1\big) & \text{if } \tau \leq N, \\ \frac{\partial r_1}{\partial P_1} = -\frac{1}{2}\big(2N+(\tau-1)(\frac{\pi}{RISK^\tau}-1)\big) & \text{if } N < \tau \leq 2N. \end{array} \tag{B.11}$$

This shows that a marginal *decrease* in P_1 by $\Delta P_1 > 0$ leads to a change in $r_1^{\tau_0+1}$ by $\frac{-\Delta P_1(\tau_0+1-1)}{2}\big(\frac{\pi}{RISK^{\tau_0+1}}-1\big) \leq 0$ if $\tau_0+1 \leq N$ and to a change in $r_1^{\tau_0+1}$ by $\frac{\Delta P_1}{2}\big(2N+(\tau_0+1-1)(\frac{\pi}{RISK^{\tau_0+1}}-1)\big) > 0$ if $N < \tau_0+1 \leq 2N$. Thus, if $\tau_0+1 \leq N$ (if $N < \tau_0+1 \leq 2N$), $r_1^{\tau_0+1}$ decreases (increases) monotonically if P_1 decreases such that π remains in the interval τ_0+1.

Step 3. Finally, we consider an increase in P_1 by ϵ, where $0 < \epsilon \leq \mu$ (with μ as specified in Eq. B.5), such that before the change, *RISK* is such that π is at the right border of the interval τ_0 while after the change (which leads to an increase in *RISK*), *RISK* is such that π is in the interval τ_0+1 (i.e., before the change $\pi = \big(\frac{P_1-S_1}{R_1-S_1}\big)^{\tau_0}$ and after the change $\big(\frac{P_1+\epsilon-S_1}{R_1-S_1}\big)^{\tau_0+1} \leq \pi < \big(\frac{P_1+\epsilon-S_1}{R_1-S_1}\big)^{\tau_0+1-1}$). We know from step 2 that if ϵ is smaller, $r_1^{\tau_0+1}$ is smaller (larger) if $\tau_0+1 \leq N$ (if $N < \tau_0+1 \leq 2N$). In this step, we prove that as ϵ goes to 0, $r_1^{\tau_0+1}$ cannot get smaller than $r_1^{\tau_0]}$ if $\tau_0+1 \leq N$ and $r_1^{\tau_0+1}$ cannot get larger than $r_1^{\tau_0]}$ if $N < \tau_0+1 \leq 2N$.

Step 3.a. For $\tau_0 + 1 \leq N$, it follows from Eq. (B.1) that $r_1^{\tau_0]} = \frac{\tau_0}{2}(R_1 - P_1)$ and that

$$r_1^{\tau_0+1} = \frac{1}{2}\Big((\tau_0+1)\big(R_1 - (P_1+\epsilon)\big) + P_1 + \epsilon - \big(S_1 + \frac{\pi}{(\frac{P_1+\epsilon-S_1}{R_1-S_1})^{\tau_0+1-1}}(R_1 - S_1)\big)\Big). \tag{B.12}$$

Because as ϵ goes to 0, $\pi/(\frac{P_1+\epsilon-S_1}{R_1-S_1})^{\tau_0+1-1}$ goes to 1, it holds that

$$\begin{aligned}\lim_{\epsilon\downarrow 0} r_1^{\tau_0+1} =& \frac{1}{2}\Big((\tau_0+1)(R_1-(P_1+\epsilon)) + P_1 + \epsilon \\ & - \big(S_1 + \frac{\pi}{(\frac{P_1+\epsilon-S_1}{R_1-S_1})^{\tau_0+1-1}}(R_1-S_1)\big)\Big) \\ =& \frac{1}{2}\big((\tau_0+1)(R_1-P_1) + P_1 - (S + R_1 - S_1)\big) \\ =& \frac{\tau_0}{2}(R_1 - P_1).\end{aligned} \tag{B.13}$$

This proves that as ϵ goes to 0 (and, consequently, $r_1^{\tau_0+1}$ becomes smaller), $r_1^{\tau_0+1}$ decreases towards $r_1^{\tau_0]}$ but cannot get smaller than $r_1^{\tau_0]}$.

Step 3.b. For $N < \tau_0 + 1 \leq 2N$, $r_1^{\tau_0]} = \frac{2N-\tau_0}{2}(R_1 - P_1)$ and

$$r_1^{\tau_0+1} = \frac{1}{2}\Big(\big(2N-(\tau_0+1)\big)\big(R_1-(P_1+\epsilon)\big) + S_1 + \frac{\pi}{(\frac{P_1+\epsilon-S_1}{R_1-S_1})^{\tau_0+1-1}}(R_1-S_1) - (P_1+\epsilon)\Big). \tag{B.14}$$

Because as ϵ goes to 0, $\pi/(\frac{P_1+\epsilon-S_1}{R_1-S_1})^{\tau_0+1-1}$ goes to 1, it holds that

$$\begin{aligned}\lim_{\epsilon\downarrow 0} r_1^{\tau_0+1} =& \frac{1}{2}\Big(\big(2N-(\tau_0+1)\big)\big(R_1-(P_1+\epsilon)\big) + S_1 + \frac{\pi}{(\frac{P_1+\epsilon-S_1}{R_1-S_1})^{\tau_0+1-1}}(R_1-S_1) \\ & - (P_1+\epsilon)\Big) \\ =& \frac{1}{2}\Big(\big(2N-(\tau_0+1)\big)(R_1-P_1) + S_1 + R_1 - S_1 - P_1\Big) \\ =& \frac{1}{2}(2N-\tau_0)(R_1-P_1).\end{aligned} \tag{B.15}$$

This proves that as ϵ goes to 0 (and, consequently, $r_1^{\tau_0+1}$ becomes larger), $r_1^{\tau_0+1}$ increases towards $r_1^{\tau_0]}$ but cannot get larger than $r_1^{\tau_0]}$.

B.7 Proof of Proposition 3.9: Changes in R_1

To prove that if R_1 decreases r_1 changes as postulated in Proposition 3.9, we proceed similarly as in the proof of the effects of changes in P_1. Recall that if R_1 is smaller, *RISK* is larger and the randomization tends to start earlier.

Step 1. Here we consider a decrease in R_1 by μ (where $0 < \mu < R_1 - P_1$) such that $\pi = \big(\frac{P_1-S_1}{R_1-S_1}\big)^{\tau_0} = \big(\frac{P_1-S_1}{R_1-\mu-S_1}\big)^{\tau_0+1}$.

Step 1.a. Given $\tau_0 + 1 < N$, $r_1^{\tau_0]} = \frac{\tau_0}{2}(R_1 - P_1)$ and $r_1^{\tau_0+1]} = \frac{\tau_0+1}{2}(R_1 - \mu - P_1)$. Subtracting $r_1^{\tau_0+1]}$ from $r_1^{\tau_0]}$, we see that r_1 changes by $\frac{1}{2}\big(\mu(\tau_0+1) - (R_1 - P_1)\big)$. If

(as postulated) r_1 decreases due to the considered decrease in R_1, it must hold that $\frac{1}{2}\big(\mu(\tau_0+1)-(R_1-P_1)\big)>0$, which requires that

$$\mu > \frac{R_1-P_1}{1+\tau_0}. \tag{B.16}$$

To show that this is the case, we derive from $\big(\frac{P_1-S_1}{R_1-S_1}\big)^{\tau_0} = \big(\frac{P_1-S_1}{R_1-\mu-S_1}\big)^{\tau_0+1}$ that

$$\mu = (R_1-S_1)\bigg(1-\big(\frac{P_1-S_1}{R_1-S_1}\big)^{\frac{1}{\tau_0+1}}\bigg). \tag{B.17}$$

By replacing μ in Eq. (B.16) with the right-hand side of Eq. (B.17) and "normalizing" to $R_1=1$, $S_1=0$, and $0<P_1<1$, we obtain

$$1-P_1^{\frac{1}{\tau_0+1}} > \frac{1-P_1}{\tau_0+1}. \tag{B.18}$$

Rearranging leads to

$$P_1^{\frac{1}{\tau_0+1}} < 1-\frac{1-P_1}{\tau_0+1}. \tag{B.19}$$

Now we replace P_1 by $1-w$ (so $P_1=1-w$ and $w=1-P_1$), which allows rewriting $P_1^{\frac{1}{\tau_0+1}}$ as a binomial series:

$$\begin{aligned}(1-w)^{\frac{1}{\tau_0+1}} &= \sum_{k=0}^{\infty}\big(\tfrac{\frac{1}{\tau_0+1}}{k}\big)(-w)^k \\ &= 1+\frac{1}{\tau_0+1}(-w)+\frac{\frac{1}{\tau_0+1}(\frac{1}{\tau_0+1}-1)}{2!}(-w)^2+ \\ &\quad \ldots+\frac{\frac{1}{\tau_0+1}(\frac{1}{\tau_0+1}-1)(\frac{1}{\tau_0+1}-2)\ldots(\frac{1}{\tau_0+1}-k+1)}{k!}(-w)^k+\ldots\end{aligned} \tag{B.20}$$

Every element that is "added" to 1 in this series is smaller than 0. If k is even, the numerator is negative while $(-w)^k$ is positive. If k is uneven, the numerator is positive while $(-w)^k$ is negative. Thus, $(1-w)^{\frac{1}{\tau_0+1}}$ must be smaller than what we obtain when carrying out only one summation step. That is, Eq. (B.20) implies that

$$(1-w)^{\frac{1}{\tau_0+1}} < 1+\frac{1}{\tau_0+1}(-w). \tag{B.21}$$

Replacing w in Eq. (B.21) again by $1-P_1$ yields Eq. (B.19), which shows that Eq. (B.19) holds. This proves that (as postulated) r_1 decreases if R_1 decreases as considered and $\tau_0+1\leq N$.

Step 1.b. Given $N<\tau_0+1\leq 2N$, $r_1^{\tau_0]} = \frac{2N-\tau_0}{2}(R_1-P_1)$ and $r_1^{\tau_0+1]} =$

$\frac{2N-(\tau_0+1)}{2}(R_1-\mu-P_1)$. Subtracting $r_1^{\tau_0+1]}$ from $r_1^{\tau_0}$, we see that r_1 changes by $\frac{1}{2}\big(\mu(2N-(\tau_0+1))+R_1-P_1\big)>0$, which proves that (as postulated) r_1 decreases if R_1 decreases as considered and $N<\tau_0+1\leq 2N$.

Step 2. The derivative of r_1 in R_1, neglecting that τ is a function of R_1, is

$$\begin{array}{ll}\frac{\partial r_1}{\partial R_1}=\frac{\tau}{2}\big(1-\frac{\pi}{RISK^{\tau-1}}\big) & \text{if } \tau\leq N,\\ \frac{\partial r_1}{\partial R_1}=\frac{1}{2}\big((2N-\tau(1-\frac{\pi}{RISK^{\tau-1}})\big) & \text{if } N<\tau\leq 2N.\end{array} \tag{B.22}$$

This shows that a marginal increase in R_1 by $\Delta R_1>0$ that does not affect τ leads to a change in $r_1^{\tau_0+1}$ by $\frac{\Delta R_1(\tau_0+1)}{2}\big(1-\frac{\pi}{RISK^{\tau_0+1-1}}\big)>0$ if $\tau_0+1\leq N$ and to a change in $r_1^{\tau_0+1}$ by $\frac{\Delta R_1}{2}\big((2N-(\tau_0+1)(1-\frac{\pi}{RISK^{\tau_0+1-1}})\big)>0$ if $N<\tau_0+1\leq 2N$. Thus, if trust is possible at least in Γ^+, $r_1^{\tau_0+1}$ increases if R_1 increases such that π remains in the interval τ_0+1.

Step 3. Finally, we consider a decrease in R_1 by ϵ, where $0<\epsilon\leq\mu$ (with μ as specified in Eq. B.17), such that before the change, *RISK* is such that π is at the right border of the interval τ_0 while after the change (which leads to an increase in *RISK*), *RISK* is such that π is in the interval τ_0+1 (i.e., before the change $\pi=\big(\frac{P_1-S_1}{R_1-S_1}\big)^{\tau_0}$ and after the change $\big(\frac{P_1-S_1}{R_1-\epsilon-S_1}\big)^{\tau_0}\leq\pi<\big(\frac{P_1-S_1}{R_1-\epsilon-S_1}\big)^{\tau_0+1-1}$). We know from step 2 that $r_1^{\tau_0+1}$ is larger the smaller ϵ is. In this step we prove that also as ϵ goes to 0, $r_1^{\tau_0]}\geq r_1^{\tau_0+1}$ (i.e., $r_1^{\tau_0+1}$ cannot be larger than $r_1^{\tau_0]}$).

Step 3.a. For $\tau_0+1\leq N$, $r_1^{\tau_0]}=\frac{\tau_0}{2}(R_1-P_1)$ and

$$r_1^{\tau_0+1}=\frac{1}{2}\Big((\tau_0+1)(R_1-\epsilon-P_1)+P_1-\big(S_1+\frac{\pi}{(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}}(R_1-\epsilon-S_1)\big)\Big). \tag{B.23}$$

As ϵ goes to 0, $\pi/(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}$ goes to 1. Replacing $\pi/(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}$ by 1 and leaving ϵ out, we obtain

$$\begin{aligned}\lim_{\epsilon\downarrow 0} r_1^{\tau_0+1}&=\frac{1}{2}\Big((\tau_0+1)(R_1-\epsilon-P_1)+P_1\\&\quad-\big(S_1+\frac{\pi}{(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}}(R_1-\epsilon-S_1)\big)\Big)\\&=\frac{1}{2}\big((\tau_0+1)(R_1-P_1)+P_1-S-(R_1-S_1)\big)\\&=\frac{\tau_0}{2}(R_1-P_1).\end{aligned} \tag{B.24}$$

This proves that as ϵ goes to 0 (and, consequently, $r_1^{\tau_0+1}$ becomes larger), $r_1^{\tau_0+1}$ increases towards $r_1^{\tau_0]}$ but cannot get larger than $r_1^{\tau_0]}$.

Step 3.b. For $N<\tau_0+1\leq 2N$, $r_1^{\tau_0]}=\frac{2N-\tau_0}{2}(R_1-P_1)$ and

$$r_1^{\tau_0+1}=\frac{1}{2}\Big(\big(2N-(\tau_0+1)\big)(R_1-\epsilon-P_1)+S_1+\frac{\pi}{(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}}(R_1-\epsilon-S_1)-P_1\Big). \tag{B.25}$$

For ϵ going to 0, and, hence, $\pi/(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}$ going to 1, this gives

$$\begin{aligned}\lim_{\epsilon\downarrow 0} r_1^{\tau_0+1} &= \frac{1}{2}\Big(\big(2N-(\tau_0+1)\big)(R_1-\epsilon-P_1)+S_1 \\ &\quad + \frac{\pi}{(\frac{P_1-S_1}{R_1-\epsilon-S_1})^{\tau_0+1-1}}(R_1-\epsilon-S_1)-P_1\Big) \\ &= \frac{1}{2}\big((\tau_0+1)(R_1-P_1)+P_1-S-(R_1-S_1)\big) \\ &= \frac{\tau_0}{2}(R_1-P_1).\end{aligned} \tag{B.26}$$

This proves that also if $N < \tau_0 + 1 \leq 2N$ and as ϵ goes to 0 (and, consequently, $r_1^{\tau_0+1}$ becomes larger), $r_1^{\tau_0+1}$ increases towards $r_1^{\tau_0]}$ but cannot get larger than $r_1^{\tau_0]}$.

B.8 Proof of Proposition 3.10: Changes in N

Suppose that N changes from some N_0 to $N_0 + 1$. If $2(N_0 + 1) < \tau$, trust is also not even possible in Γ^+ after the increase in N; $U_1^{\Gamma^+}$ and $U_1^{\Gamma^-}$ both increase from N_0P_1 to $(N_0 + 1)P_1$ and, hence, r_1 does not change and remains 0. Similarly, if $N_0 \geq \tau$, trust is already possible in Γ^+ as well as in Γ^- before the increase in N; $U_1^{\Gamma^+}$ and $U_1^{\Gamma^-}$ both increase by R_1 and, consequently, r_1 does not change. This proves the "second part" of Proposition 3.10.

We have to consider three scenarios to prove the "first part" of Proposition 3.10, i.e., to establish that r_1 increases if, after the increase in N, trust is possible in Γ^+ but not in Γ^- or if at least before the increase, trust was not possible in Γ^-. First, if $2(N_0 + 1) = \tau$, trust would never be placed before the increase in N but with certainty in the first TG of Γ^+ after the increase in N; $U_1^{\Gamma^+}$ changes from $\frac{1}{2}(2N_0P_1)$ to $\frac{1}{2}(X_1 + P_1 + 2N_0P_1)$ and $U_1^{\Gamma^-}$ changes form N_0P_1 to $(N_0 + 1)P_1$. Consequently, r_1 increases by $\frac{1}{2}(X_1 - P_1)$. Second, if $N_0 + 1 < \tau \leq 2N_0$, trust would be placed with certainty in some TGs of Γ^+ before as well as after the change in N_0, whereas trust would never be placed in Γ^-; $U_1^{\Gamma^+}$ increases by R_1 while $U_1^{\Gamma^-}$ increases by P_1 and, consequently, r_1 increases by $R_1 - P_1$. Finally, if $N_0 + 1 = \tau$, trust would never be placed in Γ^- before the increase in N, whereas after the increase in N both trustors would place trust with certainty in their first TG of Γ^-; $U_1^{\Gamma^+}$ increases again by R_1 while $U_1^{\Gamma^-}$ increases by X_1 (changes from N_0P_1 to $X_1 + N_0P_1$) and, therefore, r_1 increases by $R_1 - X_1$.

Appendix C

Mathematical details for Chapter 4

This appendix provides a summary of the notation used in Chapter 4, the proofs for Propositions 4.3 through 4.6, and an additional remark.

C.1 Overview of notation and assumptions

Chapter 4 uses the notation and assumptions used in Chapter 3 and summarized in Table B.1. However, in Chapter 4, the trustee can invest in period 0.2, not the trustors. Furthermore, in Chapter 4 we analyze the game Γ for two types of friendly trustees. The superscript ga is used to let notation refer to the version of Γ with "guilt-avoiding" trustees who earn payoff $T_2 - \theta < R_2$ for abused trust. The superscript rs is used to let notation refer to the version of Γ with "reward-seeking" trustees who earn $R_2 + \upsilon > T_2$ if honoring trust. Table C.1 summarizes additional notation used in Chapter 4.

Table C.1: Notation and assumptions used in Chapter 4.

Symbol	Description and assumptions
ρ_F, ρ_O	Probabilities with which *F*riendly and *O*pportunistic trustees establish embeddedness in period 0.2
π^-, π^+	Beliefs of the trustors entering the continuation games Γ^- and Γ^+; π^- and π^+ are inferred by Bayes' rule and they are equal to π in case of a deviation from equilibrium for which Bayes' rule cannot be applied
τ^-, τ^+	τ as implied by π^- and π^+
$U_F^{\Gamma^-}$, $U_F^{\Gamma^+}$, $U_O^{\Gamma^-}$, $U_O^{\Gamma^+}$	*F*riendly and *O*pportunistic trustees' expected payoffs for the continuation games Γ^- and Γ^+
$\bar{c}$	Maximum cost of investment c for which there exists an equilibrium in which $\rho_F = \rho_O = 1$

C.2 Proof of Proposition 4.3: The condition for equilibria in which $\rho_F = \rho_O = 1$

In general, Γ has a sequential equilibrium in which $\rho_F = \rho_O = 1$ if and only if it holds for either type of trustee that $c \leq U^{\Gamma^+} - U^{\Gamma^-}$, i.e., if $c \leq \bar{c} = \min(U_F^{\Gamma^+} - U_F^{\Gamma^-}, U_O^{\Gamma^+} - U_O^{\Gamma^-})$. Otherwise, if $c > \bar{c}$, at least one type of trustee would be better off if he deviates from the conjectured equilibrium by not investing in establishing network embeddedness. To specify $\bar{c}$ explicitly, recall that in an equilibrium in which $\rho_F = \rho_O = 1$ it holds, by Bayes' rule, that $\pi^+ = \pi$. Furthermore, the out-of-equilibrium belief π^- must likewise equal π because we assume a passive conjecture—that the trustors do not change their belief about the trustee's type if they observe that the trustee deviates in period 0.2 from a conjectured equilibrium. So in an

equilibrium in which $\rho_F = \rho_O = 1$, $\pi^- = \pi^+ = \pi$ and, hence, $\tau^- = \tau^+$. We again use τ to denote τ^- and τ^+ simultaneously and we use, for example, $U_F^{\Gamma^+(\tau)}$ to denote the expected payoff of a friendly trustee for Γ^+ given the τ (i.e., τ^+) that results from $\pi^+ = \pi$. We can then formulate the condition for the existence of an equilibrium in which $\rho_F = \rho_O = 1$ as $c \leq \bar{c} = \min(U_F^{\Gamma^+(\tau)} - U_F^{\Gamma^-(\tau)}, U_O^{\Gamma^+(\tau)} - U_O^{\Gamma^-(\tau)})$.

Consider first the game Γ^{ga}. To calculate $U_F^{\Gamma^{ga+}(\tau)} - U_F^{\Gamma^{ga-}(\tau)}$ and $U_O^{\Gamma^{ga+}(\tau)} - U_O^{\Gamma^{ga-}(\tau)}$, we can use the formulas for the expected payoffs provided in Propositions 4.1 and 4.2. For the case that $\tau \leq N$, we obtain

$$\begin{aligned} &U_F^{\Gamma^{ga+}(\tau)} - U_F^{\Gamma^{ga-}(\tau)} = \tau(R_2 - P_2) - (T_2 - P_2)(1 - TEMP^\tau) \text{ and} \\ &U_O^{\Gamma^{ga+}(\tau)} - U_O^{\Gamma^{ga-}(\tau)} = \tau(R_2 - P_2) - (T_2 - P_2). \end{aligned} \tag{C.1}$$

As $0 < TEMP < 1$, it holds that $U_F^{\Gamma^{ga+}(\tau)} - U_F^{\Gamma^{ga-}(\tau)} > U_O^{\Gamma^{ga+}(\tau)} - U_O^{\Gamma^{ga-}(\tau)}$ and, hence,

$$\bar{c}^{ga} = \tau(R_2 - P_2) - (T_2 - P_2) \text{ if } \tau \leq N. \tag{C.2}$$

For the scenario that $N < \tau \leq 2N$ we have

$$\begin{aligned} &U_F^{\Gamma^{ga+}(\tau)} - U_F^{\Gamma^{ga-}(\tau)} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)(1 - TEMP^\tau) < \\ &U_O^{\Gamma^{ga+}(\tau)} - U_O^{\Gamma^{ga-}(\tau)} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2) \end{aligned} \tag{C.3}$$

and thus $\bar{c}^{ga} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)(1 - TEMP^\tau)$ if $N < \tau \leq 2N$.

Now consider the game Γ^{rs} and again assume that $\tau \leq N$. It is useful to introduce an alternative way of expressing the trustees' expected payoffs for Γ^{rs-} and Γ^{rs+}. Alternative to the formulas in Propositions 4.1 and 4.2, we express these payoffs for $\tau \leq N$ as follows.

$$U_F^{\Gamma^{rs-}} = 2\Bigg((N-\tau^-)(R_2+\upsilon) + \sum_{i=0}^{\tau^--1}\Big(TEMP^i(R_2+\upsilon) + (1-TEMP^i)P_2\Big)\Bigg) \tag{C.4a}$$

$$U_F^{\Gamma^{rs+}} = (2N-\tau^+)(R_2+\upsilon) + \sum_{i=0}^{\tau^+-1}\Big(TEMP^i(R_2+\upsilon) + (1-TEMP^i)P_2\Big) \tag{C.4b}$$

$$U_O^{\Gamma^{rs-}} = 2\Bigg((N-\tau^-)R_2 + \sum_{i=0}^{\tau^--2}\Big(TEMP^iR_2 + (1-TEMP^i)P_2\Big) + TEMP^{\tau^--1}T_2 + (1-TEMP^{\tau^--1})P_2\Bigg) \tag{C.4c}$$

$$U_O^{\Gamma^{rs+}} = (2N-\tau^+)R_2 + \sum_{i=0}^{\tau^+-2}\Big(TEMP^iR_2 + (1-TEMP^i)P_2\Big) + TEMP^{\tau^+-1}T_2 + (1-TEMP^{\tau^+-1})P_2 \tag{C.4d}$$

Eq. (C.4a) and (C.4b) for the friendly trustee are directly implied by the sequential equilibria specified in P_FBR 1 and P_FBR 3 (Propositions 1 and 3 of Frey et al. 2015b) in the way explained in the proof of Propositions 4.1 and 4.2. It is only that we do here *not* rearrange the summation. Eq. (C.4c) and (C.4d) for the opportunistic trustee are obtained when assuming that the trustee does—if trust gets placed—honor trust in the first $\tau-1$ TGs after the start of the randomization phase (in these TGs the trustee is indifferent between, on the one hand, honoring trust and maybe being trusted again and, on the other hand, abusing trust and certainly not being trusted again) and abuses trust in the very last TG if trust is still placed in that TG. In the first of these TGs, trust gets placed with certainty and from the second of these TGs on, the trustor(s) will place trust with constant probability *TEMP* as long as trust was always placed before. From Eq. (C.4a) and (C.4b) we obtain

$$U_F^{\Gamma^{rs+}(\tau)} - U_F^{\Gamma^{rs-}(\tau)} = (\tau - \sum_{i=0}^{\tau-1} TEMP^i)(R_2+\upsilon-P_2). \tag{C.5}$$

From Eq. (C.4c) and Eq. (C.4d) we obtain

$$U_O^{\Gamma^{rs+}(\tau)} - U_O^{\Gamma^{rs-}(\tau)} = (\tau - \sum_{i=0}^{\tau-1} TEMP^i)(R_2-P_2) - TEMP^{\tau-1}(T_2-R_2). \tag{C.6}$$

The comparison of Eq. (C.5) and Eq. (C.6) shows that $U_F^{\Gamma^{rs+}(\tau} - U_F^{\Gamma^{rs-}(\tau)} > U_O^{\Gamma^{rs+}(\tau)} -$

$U_O^{\Gamma^{rs-}(\tau)}$ because $(R_2 + \upsilon - P_2) > (R_2 - P_2)$ and $TEMP^{\tau-1}(T_2 - R_2) > 0$. Thus, if $\tau \leq N$, $\bar{c}^{rs} = U_O^{\Gamma^{rs+}(\tau)} - U_O^{\Gamma^{rs-}(\tau)}$. The specification of $\bar{c}^{rs}$ provided in Proposition 4.3, namely $\bar{c}^{rs} = \tau(R_2 - P_2) - (T_2 - P_2)$, follows from the expressions for $U_O^{\Gamma^{rs-}}$ and $U_O^{\Gamma^{rs+}}$ in Propositions 4.1 and 4.2 and can, of course, also be obtained from Eq. (C.6).

For the scenario that $N < \tau \leq 2N$ we derive $U_F^{\Gamma^{rs+}(\tau)} - U_F^{\Gamma^{rs-}(\tau)}$ using the same manner of expressing the expected payoffs as in Eq. (C.4a) and (C.4b). We have

$$\begin{aligned} U_F^{\Gamma^{rs+}(\tau)} - U_F^{\Gamma^{rs-}(\tau)} &= (2N - \tau)(R_2 + \upsilon) \\ &\quad + \sum_{i=0}^{\tau-1} \Big(TEMP^i(R_2 + \upsilon) + (1 - TEMP^i)P_2 \Big) - 2NP_2 \\ &= (2N - \tau + \sum_{i=0}^{\tau-1} TEMP^i)(R_2 + \upsilon - P_2). \end{aligned} \tag{C.7}$$

For the opportunistic trustee we obtain from Propositions 4.1 and 4.2 that if $N < \tau \leq 2N$

$$U_O^{\Gamma^{rs+}(\tau)} - U_O^{\Gamma^{rs-}(\tau)} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2). \tag{C.8}$$

It follows from Eq. (C.7) and (C.8) that $U_F^{\Gamma^{rs+}(\tau)} - U_F^{\Gamma^{rs-}(\tau)} > U_O^{\Gamma^{rs+}(\tau)} - U_O^{\Gamma^{rs-}(\tau)}$. Specifically, the inequality $U_F^{\Gamma^{rs+}(\tau)} - U_F^{\Gamma^{rs-}(\tau)} > U_O^{\Gamma^{rs+}(\tau)} - U_O^{\Gamma^{rs-}(\tau)}$ can be reduced to

$$(2N - \tau)\upsilon + \sum_{i=0}^{\tau-1} TEMP^i(R_2 + \upsilon - P_2) > (T_2 - P_2). \tag{C.9}$$

Eq. (C.9) must hold because $(R_2 + \upsilon - P_2) > (T_2 - P_2)$ and $\sum_{i=0}^{\tau-1} TEMP^i \geq 1$. This proves that $\bar{c}^{rs} = U_O^{\Gamma^{rs+}(\tau)} - U_O^{\Gamma^{rs-}(\tau)} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)$ if $N < \tau \leq 2N$.

C.3 Proof of Proposition 4.4: Comparative statics of the condition for equilibria in which $\rho_F = \rho_O = 1$

For the case that $N < \tau \leq 2N$, it will be useful to have an expression for $\bar{c}^{ga} = U_F^{\Gamma^{ga+}(\tau)} - U_F^{\Gamma^{ga-}(\tau)}$ derived using a formulation of $U_F^{\Gamma^{ga+}}$ that is equivalent to the formulation of $U_F^{\Gamma^{rs+}}$ in Eq. (C.4b). For $N < \tau \leq 2N$, we then have

$$\begin{aligned} \bar{c}^{ga} &= (2N - \tau)R + \sum_{i=0}^{\tau-1} \Big(TEMP^i \cdot R + (1 - TEMP^i)P \Big) - 2NP_2 \\ &= (2N - \tau + \sum_{i=0}^{\tau-1} TEMP^i)(R_2 - P_2). \end{aligned} \tag{C.10}$$

Effects of changes in π and the trustors' payoffs in the TG: For the case that $\tau \leq N$, it follows from the expressions in Proposition 4.3 for $\bar{c}^{ga}$ and $\bar{c}^{rs}$ (where $\bar{c}^{ga} = \bar{c}^{rs} = \tau(R_2 - P_2) - (T_2 - P_2)$) that $\bar{c}^{ga}$ and $\bar{c}^{rs}$ increase by $R_2 - P_2$ for every unit increase in τ. Now realize that $\tau = \lceil \log(\pi)/\log(RISK) \rceil$ decreases stepwise in π and increases stepwise in $RISK$. $RISK = \frac{P_1 - S_1}{R_1 - S_1}$, in turn, increases in P_1 and decreases in R_1 and S_1 (specifically, we have $\partial RISK/\partial P_1 = 1/(R_1 - S_1) > 0$, $\partial RISK/\partial R_1 = -(P_1 - S_1)/(R_1 - S_1)^2 < 0$, and $\partial RISK/\partial S_1 = -(R_1 - P_1)/(R_1 - S_1)^2 < 0$). Hence, if $\tau \leq N$, $\bar{c}^{ga}$ and $\bar{c}^{rs}$ increase if τ increases due to a decrease in π or an increase in $RISK$ caused by an increase in P_1 or a decrease in R_1 or S_1.

For the case that $N < \tau \leq 2N$, we need to consider Γ^{ga} and Γ^{rs} separately. For Γ^{ga} it follows from Eq. (C.10) that an increase in τ leads to a decrease in $\bar{c}^{ga}$. An increase in τ by 1 implies that "one additional $R_2 - P_2$" is subtracted while less than "one additional $R_2 - P_2$" is added in the summation (since $0 < TEMP < 1$). For Γ^{rs}, it follows from the expression in Proposition 4.3, namely, $\bar{c}^{rs} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)$, that if $N < \tau \leq 2N$, $\bar{c}^{rs}$ decreases by $R_2 - P_2$ for every unit increase in τ. Hence, if $N < \tau \leq 2N$, $\bar{c}^{ga}$ as well as $\bar{c}^{rs}$ decrease if τ increases due to a decrease in π or, alternatively, due to an increase in $RISK$ (caused by an increase in P_1 or a decrease in R_1 or S_1).

Effects of changes in the trustee's payoffs in the TG: Consider, first, an increase in T_2. It is easy to see that if $\tau \leq N$, $\bar{c}^{ga} = \bar{c}^{rs} = \tau(R_2 - P_2) - (T_2 - P_2)$ decreases in T_2. For the case that $N < \tau \leq 2N$, the effect of a change in T_2 on $\bar{c}^{ga}$ can be inferred from Eq. (C.10). If T_2 increases, $TEMP = \frac{T_2 - R_2}{T_2 - P_2}$ increases ($\partial TEMP/\partial T_2 = (R_2 - P_2)/(T_2 - P_2)^2 > 0$). Hence, if $N < \tau \leq 2N$, $\bar{c}^{ga}$ increases in T_2 because if T_2 is larger, every element that is added in the summation is larger. That, given $N < \tau \leq 2N$, $\bar{c}^{rs}$ likewise increases in T_2 follows directly from the respective expression in Proposition 4.3, namely, $\bar{c}^{rs} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)$.

Now consider an increase in R_2. For the case that $\tau \leq N$, the effect of an increase in R_2 follows again straightforwardly from $\bar{c}^{ga} = \bar{c}^{rs} = \tau(R_2 - P_2) - (T_2 - P_2)$; $\bar{c}^{ga}$ and $\bar{c}^{rs}$ increase in R_2. How $\bar{c}^{ga}$ changes in R_2 if $N < \tau \leq 2N$ follows from taking the derivative of $\bar{c}^{ga} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)(1 - TEMP^{\tau})$. This gives $\partial \bar{c}^{ga}/\partial R_2 = 2N - \tau(1 - TEMP^{\tau - 1})$, which, given $\tau \leq 2N$, is larger than 0. For Γ^{rs} we obtain from $\bar{c}^{rs} = (2N - \tau)(R_2 - P_2) + (T_2 - P_2)$ that, given $N < \tau \leq 2N$, $\partial \bar{c}^{rs}/\partial R_2 = 2N - \tau$. This shows that $\bar{c}^{rs}$ increases in R_2 if $N < \tau < 2N$ and does not change in R_2 if $\tau = 2N$.

Now consider an increase in P_2. For the scenario that $\tau \leq N$, we obtain from $\bar{c}^{ga} = \bar{c}^{rs} = \tau(R_2 - P_2) - (T_2 - P_2)$ that $\partial \bar{c}^{ga}/\partial P_2 = \partial \bar{c}^{rs}/\partial P_2 = 1 - \tau$. This shows that $\bar{c}^{ga}$ and $\bar{c}^{rs}$ decrease in P_2 if $\tau > 1$, while they are independent of P_2 if $\tau = 1$. For the

case that $N < \tau \leq 2N$, we obtain from $\bar{c}^{ga} = (2N-\tau)(R_2-P_2)+(T_2-P_2)(1-TEMP^{\tau})$ that $\partial\bar{c}^{ga}/\partial P_2 = (\tau-1)(1-TEMP)-2N$, which, given $\tau \leq 2N$, is smaller than 0. For Γ^{rs}, we derive from $\bar{c}^{rs} = (2N-\tau)(R_2-P_2)+(T_2-P_2)$ that $\partial\bar{c}^{rs}/\partial P_2 = \tau-1-2N$, which, given $\tau \leq 2N$ is likewise smaller than 0. This proves that $\bar{c}^{ga}$ and $\bar{c}^{rs}$ decrease in P_2 if $1 < \tau \leq 2N$.

Effects of changes in N***:*** It can directly be seen from the formulas provided in Proposition 4.3 that $\bar{c}^{ga}$ and $\bar{c}^{rs}$ do not depend on N if $\tau \leq N$, whereas $\bar{c}^{ga}$ and $\bar{c}^{rs}$ increase by $R_2 - P_2$ for every unit increase in N if $N < \tau \leq 2N$.

C.4 Proof of Proposition 4.5: The impossibility of an equilibrium in which $\rho_F = 1$ and $\rho_O = 0$ in Γ^{ga}

The combination $\rho_F = 1$ and $\rho_O = 0$ implies (by Bayes' rule) $\pi^- = 0$ and $\pi^+ = 1$ and, consequently, $\tau^- = \infty$ and $\tau^+ = 1$. Given this, a friendly trustee will indeed want to play $\rho_F = 1$ if $c \leq U_F^{\Gamma^{ga+}(\tau^+=1)} - U_F^{\Gamma^{ga-}(\tau^-=\infty)} = 2N(R_2 - P_2)$. But if this holds, an opportunistic trustee will want to play $\rho_O = 1$, too, because if $c \leq 2N(R_2-P_2)$, it also holds that $c < U_O^{\Gamma^{ga+}(\tau^+=1)} - U_O^{\Gamma^{ga-}(\tau^-=\infty)} = 2N(R_2 - P_2) + T_2 - R_2$. Consequently, the combination $\rho_F = 1$ and $\rho_O = 0$ cannot be part of a sequential equilibrium of Γ^{ga}.

C.5 Proof of Proposition 4.6: The condition for equilibria in which $\rho_F = 1$ and $\rho_O = 0$ in Γ^{rs}

Again, the combination $\rho_F = 1$ and $\rho_O = 0$ implies (by Bayes' rule) $\pi^- = 0$ and $\pi^+ = 1$ and, hence, $\tau^- = \infty$ and $\tau^+ = 1$. Given $\tau^- = \infty$ and $\tau^+ = 1$, $\rho_F = 1$ can be part of an equilibrium strategy for a friendly trustee if $c \leq U_F^{\Gamma^{rs+}(\tau^+=1)} - U_F^{\Gamma^{rs-}(\tau^-=0)} = 2N(R_2 + \upsilon) - 2NP_2$ while $\rho_O = 0$ can be part of an equilibrium strategy for an opportunistic trustee if $c \geq U_O^{\Gamma^{rs+}(\tau^+=1)} - U_O^{\Gamma^{rs-}(\tau^-=0)} = (2N-1)R_2 + T_2 - 2NP_2$. From this follows that the combination $\rho_F = 1$ and $\rho_O = 0$ is part of a sequential equilibrium of Γ^{rs} if and only if $2N(R_2-P_2)+T_2-R_2 \leq c \leq 2N(R_2+\upsilon-P_2)$. That there exists some c for which this holds is implied by the assumption that $T_2 < R_2+\upsilon$.

C.6 Remark: The order of interactions in periods 1 to $2N$

The assumption that it is determined probabilistically at the beginning of every odd period which trustor plays a TG with the trustee in that period and which trustor plays with the trustee in the subsequent even period is identical to the assumption in the model of FBR. A simpler assumption would be that the trustors take turns in interacting with the trustee such that trustor 1 always plays in the odd periods and trustor 2 always plays in the even periods. This alternative assumption would not change the sequential equilibria of Γ^- and Γ^+ as specified in P_FBR 1 and P_FBR 3 and a trustee's expected payoffs associated with these equilibria that are specified in Propositions 4.1 and 4.2. Hence, this simpler assumption would also not change the results on investments in network embeddedness presented in Propositions 4.3 through 4.6.

Appendix D

Empirical details for Chapter 5

This appendix provides additional information on the data analyses and the instructions used in the experiment.

D.1 Additional information on analyses and results

Figures D.1 and D.2 show the development of trustfulness and trustworthiness over the 6 TGs of an RTTG in the different experimental conditions. Table D.1 reports the regressions used for identifying the effects of embeddedness on trustfulness and trustworthiness.

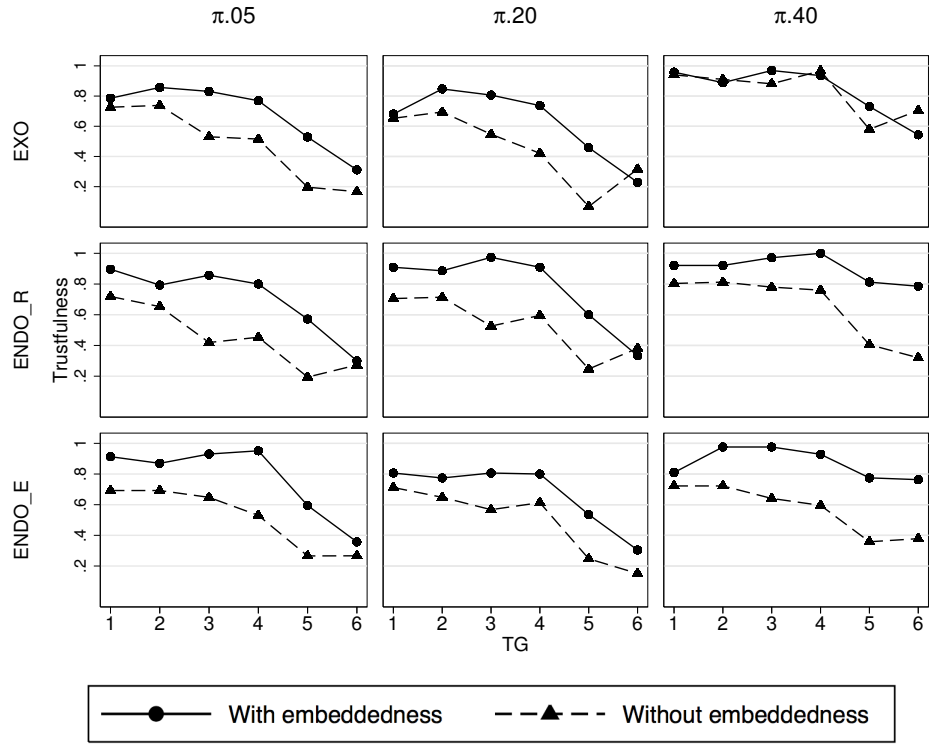


Figure D.1: Development of the average trustfulness over the six TGs of an RTTG with embeddedness and without embeddedness in the different conditions. Sample: TGs in which the trustor at play has not yet observed an abuse of trust in the current RTTG.

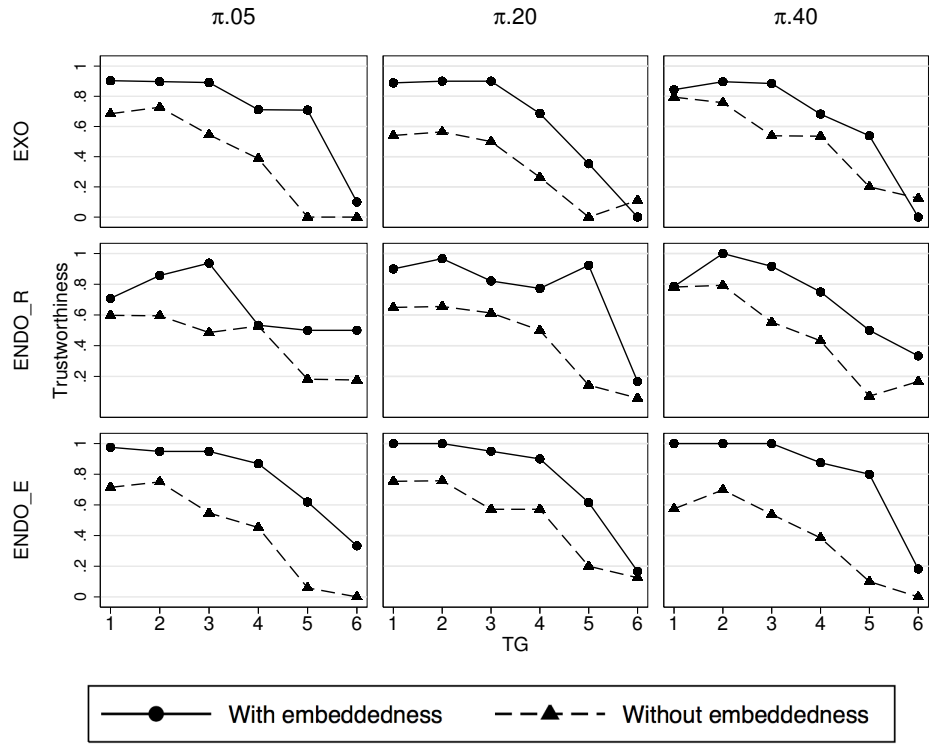


Figure D.2: Development of the average trustworthiness over the six TGs of an RTTG with embeddedness and without embeddedness in the different conditions. Sample: Behavior of opportunistic trustees in TGs in which trust was placed by a trustor that has not yet observed an abuse of trust in the current RTTG.

Table D.1: Multi-level logistic regressions of trustfulness and trustworthiness on experimental conditions. Random intercept at the subject level.

	Trustfulness		Trustworthiness	
EXO_NoNet, π.2 (Reference cat.)				
NET	1.13^{***}	(0.27)	2.91^{***}	(0.45)
π.05	0.27	(0.35)	0.71	(0.39)
π.4	2.46^{***}	(0.38)	1.19^{**}	(0.42)
ENDO_R	0.48	(0.31)	0.74^{*}	(0.37)
ENDO_E	0.23	(0.31)	1.22^{***}	(0.36)
NET X π.05	0.16	(0.26)	-0.37	(0.41)
NET X π.4	-0.85^{**}	(0.31)	-1.18^{*}	(0.48)
NET X ENDO_R	0.31	(0.30)	-0.37	(0.48)
π.05 X ENDO_R	-0.63	(0.43)	-1.04^{*}	(0.49)
π.4 X ENDO_R	-1.66^{***}	(0.46)	-0.90	(0.52)
NET X π.05 X ENDO_R	-0.04	(0.46)	-0.39	(0.66)
NET X π.4 X ENDO_R	1.17^{*}	(0.49)	0.64	(0.76)
NET X ENDO_E	0.07	(0.30)	-0.23	(0.53)
π.05 X ENDO_E	-0.09	(0.43)	-1.04^{*}	(0.49)
π.4 X ENDO_E	-1.95^{***}	(0.46)	-1.77^{***}	(0.53)
NET X π.05 X ENDO_E	0.43	(0.42)	0.78	(0.68)
NET X π.4 X ENDO_E	1.75^{***}	(0.46)	1.79^{*}	(0.81)
PERIOD	-1.17^{***}	(0.05)	-1.75^{***}	(0.11)
NET X PERIOD	-1.15^{***}	(0.08)	-1.80^{***}	(0.14)
TG2InPeriod	0.00	(0.08)	-0.12	(0.11)
NET X TG2InPeriod	-0.36^{**}	(0.12)	-1.06^{***}	(0.18)
Intercept	2.11^{***}	(0.27)	2.21^{***}	(0.33)
Variance at subject level	0.26^{***}	(0.05)	0.14	(0.09)
Number of decisions	6809		3053	
Number of subjects	342		336	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D.2 Instructions used in the experiment

This section provides an example of the instructions used in the experiment reported in Chapter 5. We reprint the instructions used for sessions with $\pi = 0.2$ and in which embeddedness could be established by trustees in the first six RTTGs and by trustors in the second six RTTGs. See Figure 5.2 in Chapter 5 for color reproductions of the screens.

- Instructions -

Welcome to this experiment. Please read the following instructions carefully. From now on you are not allowed to communicate with other participants. Please turn off your mobile phone and put it away. Also, you may not use any function of the computer that is not necessary to carry out the experiment. Thank you very much.

In this experiment, you can earn money by means of earning points. At the end of the experiment, you will be paid **1 Euro for every 150 points** that you earned. Other participants will not be able to see how much you earned.

These instructions are the same for all participants in the room. After everybody has read them, there will be a quiz in which you can make sure that you understand everything correctly. Then, we turn to the part during which you can earn points. In this part, you will play several "ABOPOLYs." Finally, you will be asked to fill in a questionnaire. Please remain seated after having filled in the questionnaire until the payment has taken place.

- The rules of ABOPOLY -

Figure 1 shows the basic interaction situation in which you can earn points in an ABOPOLY. A and B represent two participants. A starts and chooses between RIGHT and DOWN. If A chooses RIGHT, A and B both receive 30 points. If A chooses DOWN, B has to make a choice. If B also chooses DOWN, A and B both receive 50 points. If B chooses RIGHT, A receives 0 points and B receives either 100 or 0 points, depending on his/her type. B knows whether he/she is of the type who can get 100 points by choosing RIGHT. However, A does not know the type of B.

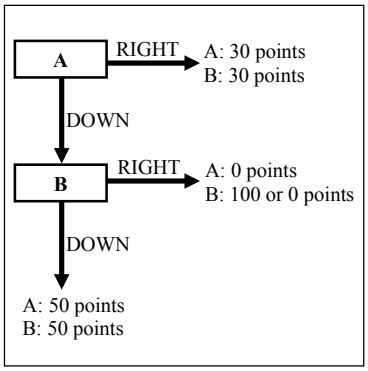


Figure 1

ABOPOLY is played in groups of three: two participants in the role of A – we call them A1 and A2 – and one a participant in the role of B. ABOPOLY proceeds over four steps that we explain in detail below. In Step 1, you are assigned the role of A1, A2, or B and you are grouped with two participants in complementary roles. In addition, in Step 1, B gets informed about his/her type. In Step 2, B can pay some points to make it possible for each A-player to see what choices are made in the interactions that the other A-player will have with B in Step 3. After that, in Step 3, A1 and A2 both have *three* interactions (as shown in Figure 1) with B. In Step 4, the points earned are transferred to each player's account.

During all these steps, you will see a screen similar to Figure 2. In the circle in the upper-right corner, you see in which step you are. In BOX 1, you receive

information or you are asked to take some action. BOX 2 shows the three interactions that A1 and B have together in Step 3. BOX 3 shows the interactions between A2 and B. The line CONNECTION BOX 2-3 indicates whether A1 and A2 can see what happens in each other's interactions. Note that Figure 2 shows a screen that the participant in the role of B sees. That is why it is written "You" instead of "B" in the small boxes that represent the B-player in the interactions shown in BOX 2 and BOX 3. Your screen would look a bit different if you were in the role of A1 or A2. To see this, compare BOX 2 and BOX 3 of Figure 2 to these boxes in Figure 3 that shows a screen of A1. More details on the elements of the screen will follow.

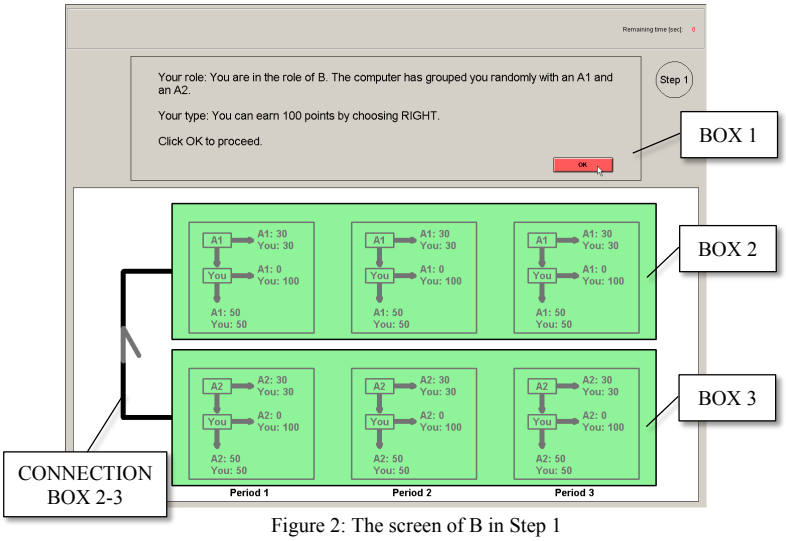


Figure 2: The screen of B in Step 1

Step 1 – The type of B: At the beginning of an ABOPOLY, you are assigned your role and grouped with two other participants. In addition, a random mechanism determines the type of the B-player of your group. The probability that B is of the type who can earn 100 points by choosing RIGHT is 0.8. The probability that B is of the type who earns 0 points by choosing RIGHT is 0.2. Put differently, in eight out of ten ABOPOLYs, B will be of the type who earns 100 points by choosing RIGHT and in two out of ten ABOPOLYs, B will be of the type who earns 0 points by choosing RIGHT.

Looking again at Figure 2, you see that, in Step 1, B sees his/her type in BOX 1. A1 and A2, however, do not get informed about B's type. As you can see on Figure 3, A1 can only read that B is now getting informed about his/her type. That an A-player does not know B's type is also visible in BOX 2 and BOX 3. On the screen of B (Figure 2), it is written "You: 100" next to the arrows that originate from the small boxes that represent B in the interactions and that point to the right (if B was of the type who earns 0 points by choosing RIGHT, it would be written "You: 0"). On the screen of A1

(Figure 3), it's written "B: 0 or 100" because A1 does not know whether B is of the type who earns 100 points by choosing RIGHT or whether B is of the type who earns 0 points by choosing RIGHT. Both A-players get the same information; so the screen of A2 looks similar to the screen of A1.

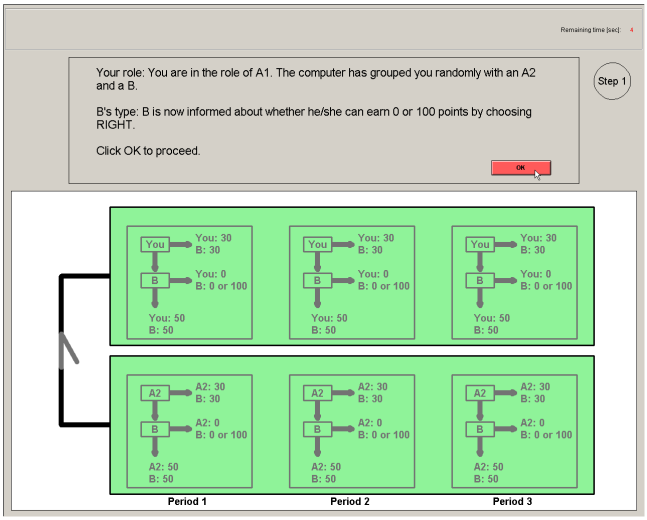


Figure 3: The screen of A1 in Step 1

Step 2 – Choosing what A1 and A2 can see: Before we explain Step 2, let's have a look at Figures 4 and 5. They show two screens that A1 might see in Step 3 when making a choice in his/her second interaction with B (Figures 4 and 5 assume the same fictive history of play in the earlier interactions). On both screens, the interaction that is taking place is highlighted yellow, the past interactions are shown in blue, and the numbers in the lower right corner of the boxes that show the past interactions inform about the order in which these interactions took place. A1 also sees on both screens what happened in her past interaction with B (A1 chose DOWN and B also chose DOWN). On the screen shown in Figure 4, A1 can likewise see what happened in the past interactions between A2 and B. However, on the screen shown in Figure 5, A1 does not see what choices were made in these interactions and only sees question marks instead.

Figure 4: A screen A1 might a see in Step 3 given that each A-player can observe what happens in the interactions of the other A-player.

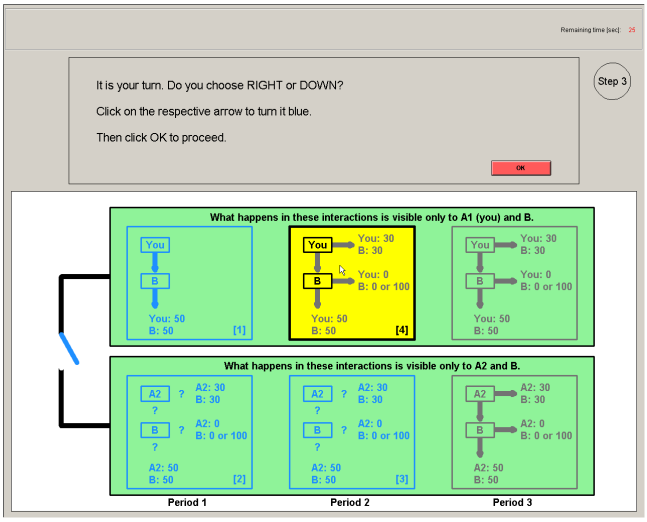


Figure 5: A screen A1 might see in Step 3 given that each A-player *cannot* observe what happens in the interactions of the other A-player.

It is determined in Step 2 whether the scenario shown in Figure 4 or the scenario shown in Figure 5 applies. More specifically, in Step 2, it is determined whether an A-player can observe what choices are made in the interactions of the other A-player. In Step 2, B is asked whether he/she wants to pay 40 points to make it possible for each A-player to see what choices are made in the interactions of the other A-player in Step 3. B has the option to click "Yes" or "No." If B chooses "Yes," A1 and A2 both see throughout Step 3 what happens in each other's interactions and B pays 40 points for making this possible (meaning that 40 points will be subtracted from the sum of points that B earns in Step 3). On the other hand, if B chooses "No," B does not pay anything and A1 and A2 will not be able to see what choices are made in each other's interactions in Step 3.

After B made a choice, all group members (A1, A2, and B) can read in BOX 1 of their screen whether or not it will be possible for A1 and A2 to see what happens in each other's interactions. In addition, on all subsequent screens everybody can see which scenario applies by reading the text at the top of BOX 2 and BOX 3 (compare this text on Figure 4 and Figure 5). Moreover, the "switch" in the line CONNECTION BOX 2-3 will be "closed" in the case that B chose "Yes" and, hence, A1 and A2 are able to observe what happens in each other's interactions (Figure 4), whereas this switch will be "open" otherwise (Figure 5). Note that because B participates in all interactions of Step 3, B can always see what choices were made in all past interactions.

Step 3 - The interactions between A1 and B and between A2 and B: In Step 3, there are three periods. In each period, there is an interaction between A1 and B and an interaction between A2 and B. Which interaction takes place first is chosen randomly every period. In each period, the chance is 50% that the interaction between A1 and B takes place first and the chance is likewise 50% that the interaction between A2 and B is played first. In the example shown in Figures 4 and 5, the interaction between A1 and B took place first in period 1, while in period 2, the interaction between A2 and B took place first (see the numbers in the lower-right corners of the boxes that show these interactions).

When it is your turn to make a choice, move your mouse to the yellow rectangle and click on the arrow that you want to choose to turn it blue. Then click on the OK button to confirm your choice. At the end of every interaction, B and the A-player who took part in the interaction can see the choices made in the interaction on their screen. What the A-player who was not involved in the interaction gets to see at that moment depends on the choice that B made in Step 2. If B clicked "Yes" in Step 2, this A-player will also see what choices were made; otherwise, this A-player only gets to know that the interaction is completed.

Step 4 – The points go to your account: At the end of an ABOPOLY, you can see in BOX 1 how many points you earned and these points are transferred to your virtual account. Your earnings equal the sum of the points you earned in the interactions in Step 3 minus 40 points if you are B and you chose "Yes" in Step 2.

- The repetitions of ABOPOLY -

Six ABOPOLYs: You will play six of these "four-step" ABOPOLYs. After every ABOPOLY you get a new role such that if you play the first ABOPOLY in the role of A1, you play the second ABOPOLY in the role of A2, the third in the role of B, and the fourth again in the role of A1 etc. In addition, for each ABOPOLY, you are anew grouped randomly with two participants.

And then everything again: After the six ABOPOLYs have been played, there will be a small change and you receive some further instructions. After reading these instructions, you play another six ABOPOLYs. You play these ABOPOLYs again in different groups and different roles.

Anonymity: It is very unlikely that you play more than one ABOPOLY with the same other participant (because the groups are determined randomly at the beginning of every ABOPOLY). Furthermore, even in the unlikely case that you are two times in a group with the same other participant, you and this other participant do not know this. Thus, the participants with whom you play an ABOPOLY do not know anything about the choices and experiences that you made in earlier ABOPOLYs. Note also that the participants that you play with are also sitting in this room but you will not find out during or after the experiment who they are.

You have finished reading the instructions. Feel free to have another look at the parts you found difficult to understand. If you have questions, please raise your hand. Otherwise, turn to the computer and click "OK."

- Instructions Part II -

In the remaining six ABOPOLYs, A1 and A2 rather than B decide in Step 2. In Step 2, A1 and A2 are both asked whether they want to pay 20 points each to see what choices are made in the interactions of the other A in Step 3. They both have the option to click "Yes" or "No." If they *both* choose "Yes," they both see throughout Step 3 what happens in each other's interactions and each A-player pays 20 points for making this possible (meaning that 20 points will be subtracted from the sum of points the A-player earns in Step 3). On the other hand, if one A-player chooses "No" or both A-players choose "No", they *both* do not pay anything and they will not be able to see what choices are made in the interactions of the other in Step 3. Everything else remains the same as in the six ABOPOLYs that you have already played.

If you have questions, please raise your hand. Otherwise click "Proceed with the second part of the experiment."

Appendix E

Theoretical and empirical details for Chapter 6

This appendix provides a summary of the notation used in Chapter 6, a formal analysis of the model presented in Section 6.2, additional information on the data analyses, and the instructions used in the experiment.

E.1 Overview of notation and assumptions

Table E.1 summarizes the notation and assumptions used in Chapter 6.

Table E.1: Notation and assumptions used in Chapter 6.

Symbol	Description and assumptions
N_1	Number of trustors
N_2	Number of trustees
Rounds $t = 1, 2, \ldots$	In each round t, one trustor can choose to place trust in one of the trustees or to withhold trust; trustors take turns such that trustor 1 makes a choice in rounds 1, $N_1 + 1$, $2N_1 + 1$, ...
P_i	Payoff for a trustor ($i = 1$) not placing trust or not being in turn and for a trustee ($i = 2$) who is not selected
R_i	Payoff from honored trust for the trustor in turn and the selected trustee; $P_i < R_i$
S_1	Trustor's payoff from abused trust; $S_1 < P_1$
T_j	Payoff from abuse of trust for a selected trustee j
$\mathbf{F}$	Distribution from which T_j is drawn for every trustee before round 1; $\mathbf{F}$ has unbounded density and while $\mathbf{F}$ is common knowledge, the manifestation of T_j is private information of trustee j
w	Probability that round $t + 1$ is played after round $t = 1, 2, \ldots$ has been played; $0 < w < 1$
n	Number of equally sized, disjoint information sharing communities into which the trustors are divided; trustors know only the choices made in past rounds in which a member of their information sharing community was at play; trustors of different information sharing communities alternate in taking decisions such that, for example, a trustor from community 1 plays in rounds 1, $n + 1$, $2n + 1$, ...

E.2 Game-theoretic analysis

This appendix provides a game-theoretic analysis of the model described in Section 6.2 in Chapter 6. We first introduce some notation to define a "*search & trigger strategy*" for trustors. We let H_{it} denote the set of trustworthy or **h**onorable trustees—trustees about whom trustor i at the beginning of round t has information on at least one instance of honoring trust and no instance of abusing trust. We let B_{it} denote the set of unknown or **b**lank trustees—trustees about whom i has no information on past

behavior at the beginning of round t.

Definition E.1. *The search & trigger strategy of a trustor.*

- *If* $H_{it} = \varnothing$ *and* $B_{it} \neq \varnothing$*, trustor* i *places trust with probability* $\gamma > 0$ *in a trustee that she chooses uniformly randomly from* B_{it} *and withholds trust with probability* $1 - \gamma$*.*
- *If* $H_{it} = \varnothing$ *and* $B_{it} = \varnothing$*, trustor* i *withholds trust.*
- *If* $H_{it} \neq \varnothing$*, trustor* i *places trust in a trustee that she chooses uniformly randomly from* H_{it}*.*

We restrict the analysis to symmetric equilibria in the sense that we require that γ is the same for all trustors. A remark at the end of this section discusses briefly the possibility of asymmetric equilibria. Allowing for asymmetric equilibria would not change any of the results relevant to the claims in Chapter 6.

A trustee best-responds to the search & trigger strategy of trustors either by always honoring trust or by always abusing trust when trusted. Lemma E.1 specifies the threshold for trustworthiness and the proportion of trustees that will be trustworthy.

Lemma E.1. *A trustee* j *best-responds to the search & trigger strategy of trustors by honoring trust when selected if and only if*

$$w^n \geq \frac{T_j - R_2}{T_j - P_2} \Leftrightarrow T_j \leq \frac{R_2 - w^n P_2}{1 - w^n}. \tag{E.1}$$

So the proportion ρ *of trustees who honor trust when selected is*

$$\rho = \int_0^{\frac{R_2 - w^n P_2}{1 - w^n}} d(T_j). \tag{E.2}$$

Proof. Lemma E.1 is a straightforward extension of a well-known result obtained from the analysis of indefinitely repeated trust games played between one trustor and one trustee. It can be shown that if, in such a game, the trustor plays a trigger strategy (places trust in every round as long as the trustee always honored trust, and never places trust again after a first abuse) the trustee's best response is to withstand the temptation of trust abuse and always honor trust if the continuation probability w is at least as large as $(T_j - R_2)/(T_j - P_2)$, i.e., if $w \geq (T_j - R_2)/(T_j - P_2)$, see e.g., Friedman (1986). If, in our model, the trustors play the search & trigger strategy, current trustworthiness also leads to future trust and abuse to no trust. However, a trustee's behavior vis-à-vis one trustor gets sanctioned exclusively in those future rounds in which that trustor or another trustors of the same information sharing

community is at play. Specifically, with a trustor from a given information sharing community being at play in every n^{th} round, a trustee will be punished or rewarded for his current behavior only in every n^{th} future round. To withstand the short-term incentive for trust abuse, a trustee needs to be accordingly stronger incentivized and it is a best-response for a trustee j to always honor trust only if $w^n \geq \frac{T_j - R_2}{T_j - P_2}$. □

Now Proposition E.1 specifies the condition under which there exists an equilibrium in which the trustors play the search & trigger strategy.

Proposition E.1. *There exists an equilibrium in which all trustors play the search & trigger strategy with the same probability γ and some portion ρ of trustees honor trust when selected if and only if*

$$\frac{P_1 - S_1}{R_1 - S_1} \leq \frac{\rho}{1 - w^{N_1}(1-\rho)}, \tag{E.3}$$

where ρ is the proportion of trustees who honor trust when selected as specified in Lemma E.1.

The proof for Proposition E.1 is provided further below together with the proof of Proposition E.2 (Proposition E.2 additionally specifies the equilibrium search probability γ). Eq. (E.3) specifies a search condition. If Eq. (E.3) holds, the search & trigger strategy induces a large enough proportion ρ of trustees to be trustworthy such that it is a best-response for each trustor i to "search" (place trust in a blank trustee) with the same probability $\gamma > 0$ as other trustors, if $H_{it} = \varnothing$ and $B_{it} \neq \varnothing$. If Eq. (E.3) does not hold, rational trustors will never place trust because the potential benefit of finding an honorable trustee does not warrant the risk of trust abuse when placing trust in a blank trustee. Before providing the proof of Proposition E.1, we discuss how the existence of an equilibrium involving honored trust depends on the number n of information sharing communities.

Our model reproduces the well-known result that information sharing can make honored trust possible when honored trust would not be possible without information sharing, as claimed in Section 6.3 in Chapter 6. Note, first, that the search & trigger strategy leads to the least restrictive possible condition for an equilibrium involving honored trust. It implies the most severe punishment for trust abuse (no more trust from any trustor of the information sharing community of an abused trustor) and the largest reward for honoring trust (continued trust from all trustors of the information sharing community of a focal trustor). Sanctioning of trustees by trustors outside of a focal information sharing community is impossible. Therefore, if the search & trigger strategy does not make it a best response for trustee j to honor trust, no other strategy of the trustors will. The search & trigger strategy thus maximizes

the proportion of trustworthy trustees and, hence, minimizes the likelihood of trust abuse when placing trust in a blank trustee. By playing the search & trigger strategy, trustors also take maximum advantage of knowing a trustworthy trustee and are thus maximally incentivized to search in the first place. Hence, if Eq. (E.3) does not hold, there cannot be an equilibrium involving honored trust.

The claim that information sharing facilitates trust then derives from the fact that Eq. (E.3) may hold for a small number n of information sharing communities but not for a larger n. The condition in Eq. (E.3) depends on n indirectly via ρ. The proportion of trustworthy trustees ρ increases weakly if n is smaller (compare Eq. (E.1) in Lemma E.1). As ρ increases, Eq. (E.3) becomes less restrictive ($\partial(\rho/(1-w^{N_1}(1-\rho)))/\partial\rho = (1-w^{N_1})/(1-w^{N_1}(1-\rho))^2 > 0$), which reflects that, as ρ increases, the likelihood of trust abuse when searching for a trustworthy trustee becomes smaller. Thus, there may be a critical value for the number of information sharing communities, n^*, such that an equilibrium involving honored trust exists if $n \leq n^*$ but not if the trustors are fragmented into more information sharing communities ($n > n^*$).

To complete the analysis, we provide Proposition E.2, which together with Lemma E.1 fully characterizes an equilibrium in which the trustors play the search & trigger strategy.

Proposition E.2.

- *If $n = N_1$, there exists an equilibrium in which all trustors play the search & trigger strategy with $\gamma = 1$ if and only if Eq. (E.3) holds.*
- *If $n = N_1$, there exists an equilibrium in which all trustors play the search & trigger strategy with the same probability $0 < \gamma < 1$ if and only if Eq. (E.3) holds with equality.*
- *If $n < N_1$, there exists an equilibrium in which all trustors play the search & trigger strategy with $\gamma = 1$ if and only if*

$$\frac{P_1 - S_1}{R_1 - S_1} \leq \frac{\rho}{1 - w^{N_1}(1-\rho)^{\frac{N_1}{n}}}. \tag{E.4}$$

- *If $n < N_1$, there exists an equilibrium in which all trustors play the search & trigger strategy with the same probability $0 < \gamma < 1$ if and only if*

$$\frac{\rho}{1 - w^{N_1}(1-\rho)^{\frac{N_1}{n}}} \leq \frac{P_1 - S_1}{R_1 - S_1} \leq \frac{\rho}{1 - w^{N_1}(1-\rho)}. \tag{E.5}$$

In such an equilibrium, γ is such that

$$\gamma = \frac{1 - \sqrt[\frac{N_1}{n}-1]{\frac{1-\frac{\rho}{(P_1-S_1)/(R_1-S_1)}}{(1-\rho)w^{N_1}}}}{\rho}. \tag{E.6}$$

Proof of Propositions E.1 and E.2: Proposition E.2 implies Proposition E.1. To prove Proposition E.2, we need to establish that playing the search & trigger with the specified probability γ is a best-response for a trustor i if the respective conditions in Proposition E.2 hold (and given the other trustors play the search & trigger strategy and trustees play their best-response to this strategy as specified in Lemma E.1). Three claims need to be established: (1) that it is a best-response for a trustor i to search with probability γ if the respective inequality in Proposition E.2 holds and i does not know an honorable trustee but knows blank trustees ($H_{it} = \varnothing$ and $B_{it} \neq \varnothing$), (2) that i should place trust in an honorable trustee if i knows such a trustee ($H_{it} \neq \varnothing$), and (3) that i should withhold trust if i has information on a past abuse of all N_2 trustees ($H_{it} = \varnothing$ and $B_{it} = \varnothing$).

We omit the proof of claims (2) and (3). That trustees fall apart into trustees who always honor trust and trustees who always abuse trust (compare Lemma E.1) straightforwardly implies that claims (2) and (3) hold irrespectively of the specific parameters of the game.

For claim (1), trustees falling apart into trustworthy and untrustworthy trustees implies that it can never pay off for a trustor i to deviate from searching (placing trust in a blank trustee) by placing trust in a trustee of which i knows of a past abuse. For claim (1) to hold, it must additionally not pay off for a trustor to deviate from searching with probability γ by strictly withholding trust. Such a deviation could pay off as it allows avoiding the risk of trust abuse. For a rational trustor to search with positive probability, the expected short-term costs of searching must not exceed the expected long-term benefit of searching. To establish the condition under which this is the case, we can apply the one-shot deviation principle (Mailath & Samuelson, 2006, Chap. 2), i.e., calculate the costs and benefits of searching in round t under the assumption that trustor i plays the search & trigger strategy with the same probability γ as all other trustors in all rounds following round t. Now we, first, calculate these costs and benefits without specifying whether $\gamma = 1$ or $0 < \gamma < 1$, assuming only that γ is larger than 0 and the same for all trustors. From the general search condition that this yields, we then derive the conditions and search probabilities in Proposition E.2.

Consider first the costs of searching. In round t, i's expected payoff from searching (placing trust in a blank trustee) with probability γ is $\gamma(\rho R_1 + (1-\rho)S_1) + (1-\gamma)P_1$ while i's payoff from strictly withholding trust is P_1. If $(P_1 - S_1)/(R_1 - S_1) < \rho$,

there are no search costs: i's expected payoff for round t is larger if i searches than if i strictly withholds trust. If $(P_1 - S_1)/(R_1 - S_1) > \rho$, on the other hand, i has an immediate expected search cost in round t from searching with probability γ of size

$$\gamma(P_1 - S_1) - \gamma\rho(R_1 - S_1). \tag{E.7}$$

Searching can then still be beneficial as it increases the chance of knowing an honorable trustee in the future. If i strictly withholds trust in round t and then reverts to playing the search & trigger strategy, the probability that i knows an honorable trustee in round $t + fN_1$, the f^{th} future round in which i would be at play, is $1 - (1 - \gamma\rho)^{f\frac{N_1}{n}-1}$. In the calculation of this probability, $(1 - \gamma\rho)$ is the probability that an honorable trustee is *not* found in a round in which the trustor at play searches with probability γ; $(1-\gamma\rho)^{f\frac{N_1}{n}-1}$ is the probability that no honorable trustee is found over the $f\frac{N_1}{n} - 1$ rounds in which one of the $\frac{N_1}{n}$ trustors of i's information sharing community is at play between round t and round $t + fN_1$. On the other hand, if i does search with probability γ in round t, the probability that i knows an honorable trustee in round $t + fN_1$ is $1 - (1 - \gamma\rho)^{f\frac{N_1}{n}-1}(1 - \gamma\rho) = 1 - (1 - \gamma\rho)^{f\frac{N_1}{n}}$. Hence, if i searches in round t with probability γ, this increases the probability that i knows an honorable trustee in round $t + fN_1$ by

$$\left(1 - (1 - \gamma\rho)^{f\frac{N_1}{n}}\right) - \left(1 - (1 - \gamma\rho)^{f\frac{N_1}{n}-1}\right) = \gamma\rho(1 - \gamma\rho)^{f\frac{N_1}{n}-1}. \tag{E.8}$$

In round $t + fN_1$, i receives R_1 if knowing an honorable trustee and $\gamma\big(\rho R_1 + (1 - \rho)S_1\big) + (1 - \gamma)P_1$ if not knowing an honorable trustee, given i plays the search & trigger strategy in that round. Thus, i's benefit in round $t + fN_1$ of knowing an honorable trustee is

$$(1 - \gamma\rho)(R_1 - S_1) - (1 - \gamma)(P_1 - S_1). \tag{E.9}$$

The expected benefit that i has in round $t+fN_1$ from searching in round t is obtained from multiplying Eq. (E.8) with Eq. (E.9)—multiplying the increase in the probability of knowing an honorable trustee with the value of knowing an honorable trustee. The expected total long-term benefit that i derives from searching in round t is obtained from summing up the expected benefits over potential future rounds in which i would be at play, multiplied with the probability that these rounds are reached. This yields that i's expected long-term benefit from searching in round t with probability γ is

$$\sum_{f=1}^{\infty} w^{f N_1} \gamma\rho(1-\gamma\rho)^{f\frac{N_1}{n}-1}\Big((1-\gamma\rho)(R_1-S_1)-(1-\gamma)(P_1-S_1)\Big). \tag{E.10}$$

It is then a best-response for trustor i to search with probability γ in round t if and only if the immediate cost of searching (Eq. (E.7)) does not exceed the expected long-term benefit from searching (Eq. (E.10)). That is, if

$$\begin{aligned}
&\gamma(P_1-S_1)-\gamma\rho(R_1-S_1)\leq\\
&\quad\sum_{f=1}^{\infty} w^{f N_1} \gamma\rho(1-\gamma\rho)^{f\frac{N_1}{n}-1}\Big((1-\gamma\rho)(R_1-S_1)-(1-\gamma)(P_1-S_1)\Big)\\
\Leftrightarrow\ &(1-\gamma\rho)\Big(\gamma(P_1-S_1)-\gamma\rho(R_1-S_1)\Big)\leq\\
&\quad\gamma\rho\Big((1-\gamma\rho)(R_1-S_1)-(1-\gamma)(P_1-S_1)\Big)\sum_{f=1}^{\infty} w^{f N_1}(1-\gamma\rho)^{f\frac{N_1}{n}}\\
\Leftrightarrow\ &(1-\gamma\rho)\Big(\gamma(P_1-S_1)-\gamma\rho(R_1-S_1)\Big)\leq\\
&\quad\left(\gamma\rho\Big((1-\gamma\rho)(R_1-S_1)-(1-\gamma)(P_1-S_1)\Big)w^{N_1}(1-\gamma\rho)^{\frac{N_1}{n}}\right)\Big/\left(1-w^{N_1}(1-\gamma\rho)^{\frac{N_1}{n}}\right)\\
\Leftrightarrow\ &(P_1-S_1)\Big((1-\gamma\rho)\gamma(1-w^{N_1}(1-\gamma\rho)^{\frac{N_1}{n}})+\gamma\rho(1-\gamma)w^{N_1}(1-\gamma\rho)^{\frac{N_1}{n}}\Big)\leq\\
&\quad(R_1-S_1)\Big(\gamma\rho(1-\gamma\rho)w^{N_1}(1-\gamma\rho)^{\frac{N_1}{n}}+\gamma\rho(1-\gamma\rho)(1-w^{N_1}(1-\gamma\rho)^{\frac{N_1}{n}})\Big)\\
\Leftrightarrow\ &\frac{P_1-S_1}{R_1-S_1}\leq\frac{\rho}{1-w^{N_1}(1-\rho)(1-\gamma\rho)^{\frac{N_1}{n}-1}}.
\end{aligned}\tag{E.11}$$

If Eq. (E.11) holds, it is a best-response for each trustor i to search with the same probability $\gamma > 0$ as other trustors if $H_{it} = \varnothing$ and $B_{it} \neq \varnothing$. Thus, if Eq. (E.11) holds, claim (1) holds and there exists an equilibrium in which all trustors play the search & trigger strategy with the same probability γ. If Eq. (E.11) does not hold, no such equilibrium exists. A trustor i would deviate from searching with probability γ by strictly withholding trust because the expected gain of searching does not warrant the risk of trust abuse in the search. In the following, we derive from Eq. (E.11) the conditions in Proposition E.2 and the equilibrium search probability γ.

The condition for equilibria in which $\gamma = 1$ ***if*** $n = N_1$**:** For the case that all trustors are "isolates" who do not share information, $n = N_1$, $(1-\gamma\rho)^{\frac{N_1}{n}-1}$ in the denominator of Eq. (E.11) becomes 1 and, hence, Eq. (E.11) reduces to Eq. (E.3) in Proposition E.1, namely

$$\frac{P_1-S_1}{R_1-S_1}\leq\frac{\rho}{1-w^{N_1}(1-\rho)}. \tag{(E.3) reprinted}$$

Thus, if $n = N_1$ and Eq. (E.3) holds, the benefit of searching is at least as large as the cost of searching for each isolate trustor and there exists an equilibrium in which all trustors play the search & trigger strategy with $\gamma = 1$.

The condition for equilibria in which $0 < \gamma < 1$ ***if*** $n = N_1$***:*** An equilibrium in which the trustors search with probability $0 < \gamma < 1$ requires that each trustor i is indifferent between searching and not searching in a period t in which $H_{it} = \varnothing$ and $B_{it} \neq \varnothing$. Generally, this requires that Eq. (E.11) holds with equality. If $n = N_1$, this is equivalent to the requirement that Eq. (E.3) holds with equality. We mention that Eq. (E.3) can hold with equality only for specific parameters and, hence, if $n = N_1$, an equilibrium with $0 < \gamma < 1$ can exist only for specific parameters.

The condition for equilibria in which $\gamma = 1$ ***if*** $n < N_1$***:*** To establish the condition for the existence of an equilibrium in which $\gamma = 1$ if $n < N_1$ (information sharing), we postulate that $\gamma = 1$. For $\gamma = 1$, Eq. (E.11) reduces to Eq. (E.4) in Proposition E.2, namely

$$\frac{P_1 - S_1}{R_1 - S_1} \leq \frac{\rho}{1 - w^{N_1}(1-\rho)^{\frac{N_1}{n}}}. \qquad \text{((E.4) reprinted)}$$

If Eq. (E.4) holds, it is thus a best-response for each trustor i to play the search & trigger strategy with probability $\gamma = 1$.

The condition for equilibria in which $0 < \gamma < 1$ ***if*** $n < N_1$***:*** As noted above, for the existence of an equilibrium in which $0 < \gamma < 1$, Eq. (E.11) must hold with equality. The search probability γ is itself part of Eq. (E.11). There thus exists an equilibrium in which $0 < \gamma < 1$ if it is possible to choose γ such that Eq. (E.11) holds with equality (under the restriction that $0 < \gamma < 1$). The right-hand side of Eq. (E.11) decreases monotonically in γ (the denominator is always between 0 and 1 and increases in γ, given $n < N_1$). This reflects that a trustor's incentive to search decreases in the probability with which her information sharing partners search in potential future rounds (we elaborate this in a remark following this proof). Hence, letting γ in the right-hand side of Eq. (E.11) approach 0 and 1 yields, respectively, the maximum and minimum value for $(P_1 - S_1)/(R_1 - S_1)$ for which Eq. (E.11) can hold with equality.

For γ going to 0, the right-hand side of Eq. (E.11) reduces to $\rho/(1 - w^{N_1}(1-\rho))$, because $\lim_{\gamma \downarrow 0}(1-\gamma\rho)^{\frac{N_1}{n}-1} = 1$. That is, for γ going to 0, Eq. (E.11) reduces to Eq. (E.3) in Proposition E.1. Thus, the lower bound for the existence of equilibria in which $0 < \gamma < 1$ if $n < N_1$ is identical to the lower bound for the existence of any equilibrium in which the trustors play the search & trigger strategy.

For γ going to 1, the right-hand side of Eq. (E.11) reduces to $\rho/(1-w^{N_1}(1-\rho)^{\frac{N_1}{n}})$, because $\lim_{\gamma\uparrow 1} 1-w^{N_1}(1-\rho)(1-\gamma\rho)^{\frac{N_1}{n}-1} = 1-w^{N_1}(1-\rho)^{\frac{N_1}{n}}$. That is, for γ going to 1, Eq. (E.11) reduces to Eq. (E.4). Thus, the upper bound for the existence of equilibria in which $0 < \gamma < 1$ if $n < N_1$ is identical to the lower bound for the existence of equilibria in which $\gamma = 1$ if $n < N_1$.

This shows that if $n < N_1$, the range for the existence of an equilibrium in which $0 < \gamma < 1$ is

$$\frac{\rho}{1-w^{N_1}(1-\rho)^{\frac{N_1}{n}}} \leq \frac{P_1-S_1}{R_1-S_1} \leq \frac{\rho}{1-w^{N_1}(1-\rho)}. \qquad \text{((E.5) reprinted)}$$

The search probability γ in such an equilibrium can be calculated by requiring that Eq. (E.11) holds with equality and solving it for γ. This yields Eq. (E.6), namely

$$\gamma = \frac{1-\sqrt[\frac{N_1}{n}-1]{\frac{1-\frac{\rho}{(P_1-S_1)/(R_1-S_1)}}{(1-\rho)w^{N_1}}}}{\rho}. \qquad \text{((E.6) reprinted)}$$

□

Remark – Search as a Volunteer's Dilemma: It can be seen from Proposition E.2 that it is possible that there exists an equilibrium in which $\gamma = 1$ if there is no information sharing ($n = N_1$) while only equilibria in which $0 < \gamma < 1$ exist if there is information sharing ($n < N_1$). For $n < N_1$, Eq. (E.4) is more restrictive than Eq. (E.3). In addition, the condition for an equilibrium in which $\gamma = 1$ can become more restrictive as n decreases and trustors share information in larger groups (Eq. (E.4) becomes more restrictive if n decreases because $0 < (1-\rho) < 1$ and the denominator of the right-hand side of Eq. (E.4), hence, increases towards 1 if n decreases). One could intervene that the proportion of trustworthy trustees, ρ, tends to increase in information sharing. However, as we did not assume a specific distribution $\mathbf{F}$, ρ may increase only marginally or even remain constant.

That the condition for an equilibrium in which $\gamma = 1$ can become more restrictive with (more) information sharing reflects that the search efforts of information sharing partners of a trustor i diminish the effect of i's own search. If there is information sharing and an equilibrium in which $\gamma = 1$ would only exists if there was no information sharing, searching a trustworthy trustee resembles a Volunteer's Dilemma (Diekmann, 1985): A trustor prefers to search if none of her information sharing partners will search but if all her information sharing partners will search with probability 1, she prefers not to search. In that situation, an equilibrium exists such that each trustor i trustor searches with the same probability $0 < \gamma < 1$ (which is chosen such

that each trustor is indifferent between searching and not searching).

To illustrate the issue, assume a simple example with $\rho = 0.4$. If there is no information sharing, i's search effort in round t increases her probability of knowing an honorable trustee in round $t + N_1$ from 0 to 0.4. If i shares information with three others who will search with probability 1 before i gets to play again, i's search effort in round t increases the chance that she knows an honorable trustee in round $t + N_1$ only from $1 - (1 - 0.4)^3 = 0.78$ to $1 - (1 - 0.4)^4 = 0.87$, where $(1 - 0.4)^3$ is the probability that no honorable trustee is found after three searches. For suitable payoffs, it could then be worthwhile for i to search if i is an isolate but not if i shares information with other trustors who search with probability 1 if not knowing an honorable trustee.

In such a situation, information sharing partners can incentivize each other to search with positive probability by each searching with a small probability. For example, if i's three information sharing partners search with probability $\gamma = 0.25$, a search effort of i in round t increases her probability of knowing an honorable trustee in round $t + N_1$ from $1 - (1 - 0.25 \cdot 0.4)^3 = 0.27$ to $1 - (1 - 0.25 \cdot 0.4)^3 \cdot (1 - 0.4) = 0.563$. It may then be worthwhile for i to search with $\gamma = 0.25$, too, while i would prefer not to search if her information sharing partners searched with $\gamma = 1$.

We have restricted the analysis to symmetric equilibria in which all trustors search with the same probability γ. However, it is not difficult to see that if and only if the condition for equilibria in which $0 < \gamma < 1$ is met, there exist also asymmetric equilibria—equilibria in which, for example, one or some trustors search while the others do not search.

E.3 Additional information on data analyses

Tables E.2, E.3, and E.4 provide the regression results on which Figures 6.3, 6.5, and 6.4 are based, respectively.

Table E.2: Linear regressions of the rate of honored trust and inequality among trustees (MCOV) on dummy variables for reputation conditions. Sample: Games played in the Trust Problem condition ($T_2 = 80, 100$). Standard errors adjusted for clustering of games in 12 sessions.

	Honored trust	Inequality
Partial	0.23^{**} (0.07)	0.38 (0.35)
Full	0.28^{***} (0.05)	0.75^{*} (0.30)
Intercept [Private]	0.30^{***} (0.03)	0.53^{**} (0.16)
Number of games	288	249

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.3: Conditional logistic regressions of trustor choices on trustee reputations (odds ratios). Standard errors adjusted for clustering of choices in trustors.

	Trust Problem	No Trust Problem	
	M1	M1	M2
No	3.80^{***} (1.12)		
Good	4.08^{***} (0.56)	0.62 (0.18)	
Very Good	1.65^{*} (0.33)		
Maxi-Min			1.83^{*} (0.50)
Trustee 1 [baseline alternative]			
Trustee 2	0.96 (0.09)	1.00 (0.21)	0.98 (0.21)
Trustee 3	0.82 (0.08)	1.02 (0.19)	1.00 (0.18)
Trustee 4	0.74^{**} (0.08)	0.93 (0.17)	0.91 (0.17)
Number of placements of trust	846	262	262
Number of trustors	237	48	48

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.4: Linear regressions for identifying effects of the Trust Problem condition on the rate of honored trust and inequality among trustees (MCOV) in the Partial and Full conditions. Standard errors adjusted for clustering of games in 14 sessions.

	Honored trust		Inequality	
	M1	M2	M1	M2
Partial		−0.05 (0.08)		−0.37 (0.38)
Full	0.05 (0.08)		0.37 (0.38)	
No Trust Problem	0.39*** (0.09)	0.37*** (0.04)	−0.21 (0.44)	−0.96** (0.28)
Partial X No Trust Problem		0.014 (0.10)		0.75 (0.44)
Full X No Trust Problem	−0.01 (0.10)		−0.75 (0.44)	
Intercept [Partial, Trust Problem]	0.53*** (0.07)		0.90*** (0.28)	
Intercept [Full, Trust Problem]		0.58*** (0.04)		1.28*** (0.27)
Number of games	240	240	217	217

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E.4 Instructions used in the experiment

This appendix contains an example of the instructions used in the experiment reported in Chapter 6. Reprinted are the instructions from the Full x Trust Problem condition. In addition, we reprint instructions from the Private condition used when this condition was played in the second half of a session. The reprinted figures were distributed on separate sheets. See Figure 6.1 in Chapter 6 for color reproductions of the screens.

Welcome and thank you for coming here!

The purpose of this experiment is to study decision making. Please do not communicate with other participants. Turn off your phone and put it away. Thank you very much. If at any point you have questions, raise your hand and we will assist you.

In this experiment, you will earn "points" by making decisions in "games." How much you earn depends on your decisions, the decisions of others and on chance. At the end of the session, you will be paid **1.5 US Dollar cents for every point** you earned.

All of the payments, participants, and other information that you read about in this study are real. Although some studies in other laboratories use fictional information, research in this laboratory focuses on studies of real situations. The following instructions tell you everything you need to know to earn as many points as possible and they are precisely the same for everyone in the room. Finally, everything is anonymous. No other participant will be able to link your decisions to your identity or get to know your name or earnings.

Description of the Game

The game is played in groups of 8 participants, four participants in the role of A (A1, A2, A3 and A4) and four participants in the role of B (B1, B2, B3 and B4). Before the game starts, you and the other participants are randomly separated into groups of 8 and each participant is randomly assigned his/her role and number. Throughout the game, all participants keep their role and number.

The Basic Interaction

The game proceeds in rounds. In each round only one of the As is active and interacts with the Bs. In round 1, it is A1's turn to interact with the Bs; in round 2, it is A2's turn, and so on such that in round 5, it is again A1's turn.

The As who are not active in a round get 30 points in that round. How many points the active A and the Bs get depends on their choices in the interaction shown in Figure 1 (see extra sheet). Please examine Figure 1 now. The active A (simply called "A" in Figure 1) chooses either "RIGHT" or "DOWN." If A chooses RIGHT, A gets 30 points and all four Bs get 30 points as well. If A chooses DOWN, A must select one of the four Bs (A must "SELECT B"). If A chooses DOWN and selects one of the Bs, the selected B participant chooses "RIGHT" or "DOWN." If the selected B chooses DOWN, A and the selected B get 50 points each. If the selected B chooses RIGHT, A gets nothing (0 points) and the selected B gets 80 points. In either case, the other Bs – the ones that were not selected by the A at play – get 30 points each, just as the As who are not active in this round.

The Duration of the Game

How many rounds the game lasts is determined randomly. It is as if we would roll a regular 6-sided die after every round and end the game if the outcome is a "6" but continue for at least one round if the outcome is not a "6." We let the computer do the "rolling of the die." Computer algorithms are never truly random but depend on the starting value used. We used the phone number of one of the researchers as the starting value. Anyway, all you need to know is that, for instance, if the game is in round 1, the probability that there will be a second round is $5/6 = 0.83$ and if the game is in round 7, the probability that there will be another round is also $5/6 = 0.83$.

The Computer Interface

All four As and Bs get informed about all choices. Have a look at Figures 2 and 3 (see extra sheets) that show the computer interface. What you will see on the right-hand side will be self-explanatory. On the left-hand side you see a "history window." Each of the four columns of plus signs ("+") represents one B participant and each row represents a round. The current round – round 5 on the example screens – is indicated by the arrow "-->". In parentheses it is displayed which A participant is at play in which round. For potential future rounds it is furthermore displayed what the probability is that the game does not end before this round. For example, given that the game is in round 5, the chance that the game does not end before round 7 is $0.83 * 0.83 = 0.69$. Hence, it is written "prob = 0.69" in round 7.

The color of the background of the plus signs of past rounds shows what choices were made. Dark-gray means that a B was not selected. Yellow means that B was selected and chose RIGHT. Blue means that B was selected and chose DOWN. If in a round the A-participant chooses RIGHT, the signs of all Bs are shown on a dark-gray background, as in round 4.

Should a game last more than 20 rounds, the history of the first rounds will disappear but you will still see the history of the 19 most recent rounds. Thus, you will always see at least 8 possible future rounds.

Organization of the Session

You will participate in 4 games, each lasting expectedly several rounds, one game played after the other. Then a small change to the rules of the game will be announced and then you participate again in 4 games. For each game you get randomly assigned to a new group of 8 participants and to your role. It is possible that you are with the same other participant in more than one game. However, should this happen, neither you nor the other participant will be able to notice this. Information about decisions

in earlier games will not be available to participants in later games.

Consent Forms and Quiz

If you wish to participate in this study, please read and sign the accompanying consent form; it explains your rights as a subject and the rules of confidentiality we adhere to. After signing the consent form, turn to the computer and answer a few questions that help you evaluate your understanding of the game.

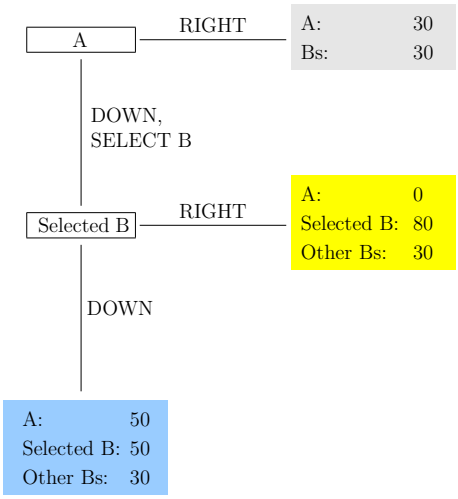


Figure 1: One round of play. The As who are not at play get 30 points each.

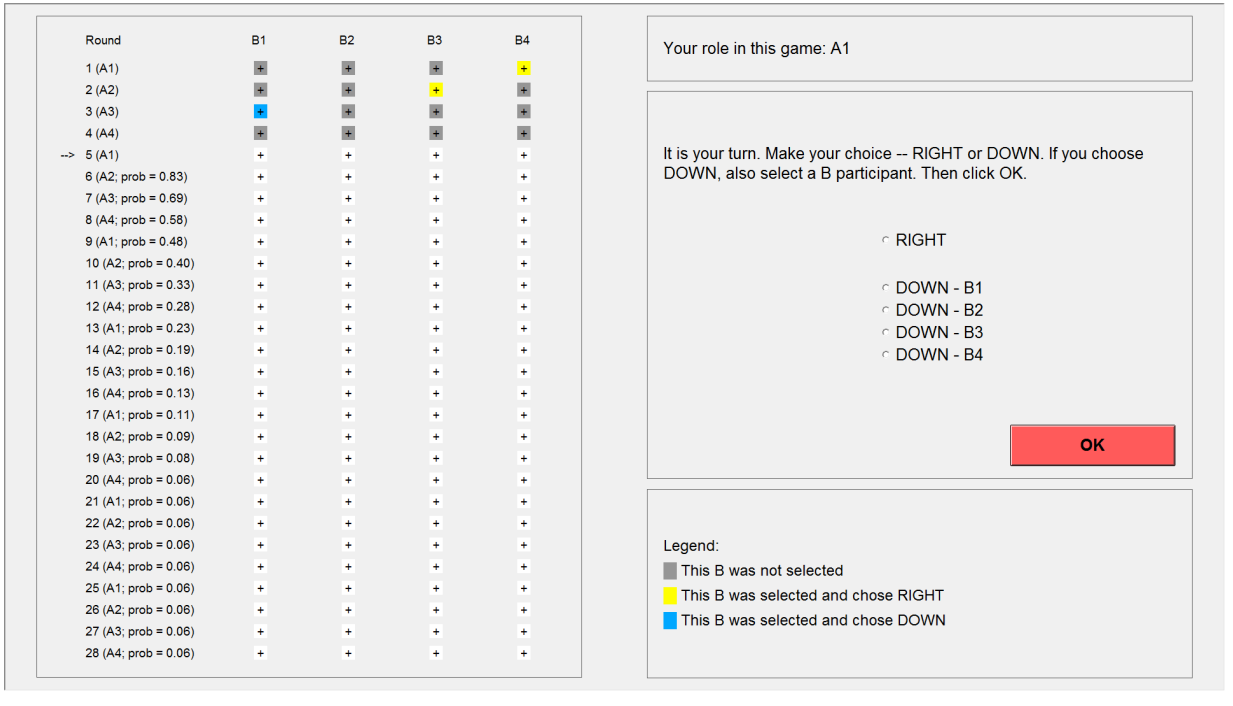


Figure 2: An Example Screen for an A

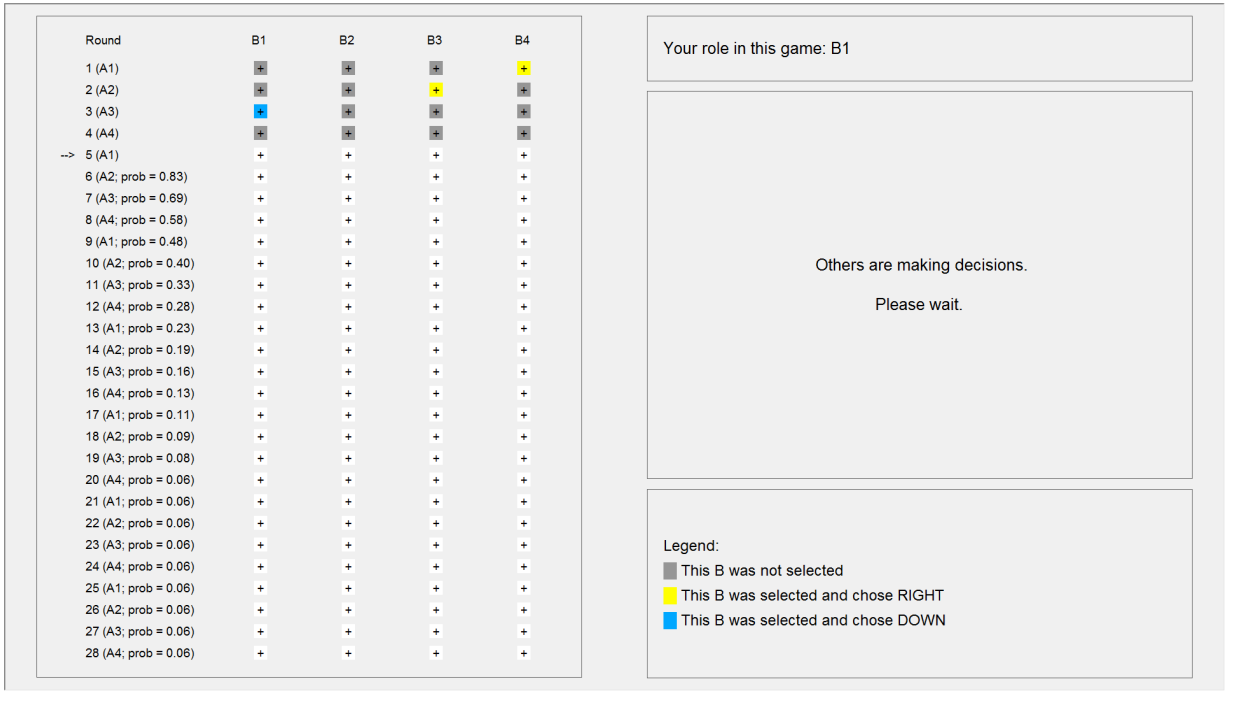


Figure 3: An Example Screen for a B

Instructions for Part 2

In the next four games, each A gets informed about the results of his/her own interactions but not about the results of the interactions of any other A. Take a moment to examine Figure 4, which shows an example screen for A1. A1 sees question marks ("?") in the rounds in which one of the other As is at play. The question marks of past rounds are not highlighted in color because A1 never receives information about the choices in these rounds.

The Bs still receive information about the outcome after every round and can see the entire history from the highlighting. To see this, examine Figure 5, which shows an example screen for B1. A participant in the role of a B sees plus signs ("+") and question marks in the same rows as the A who is at play (A1 in the example). The plus signs in past rounds indicate to a B which past rounds A has information about. The plus signs in potential future rounds show when it will be A's turn again. Note, the plus signs and question marks are placed in different rounds if it is a different A's turn.

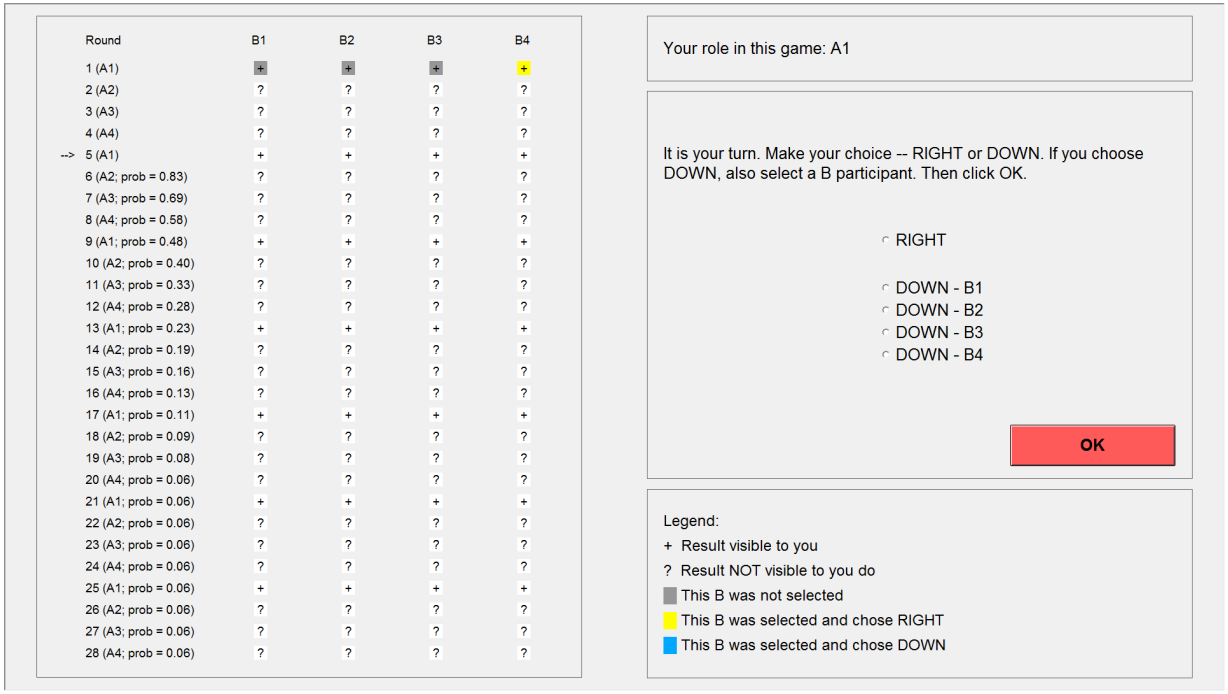


Figure 4: An Example Screen for an A

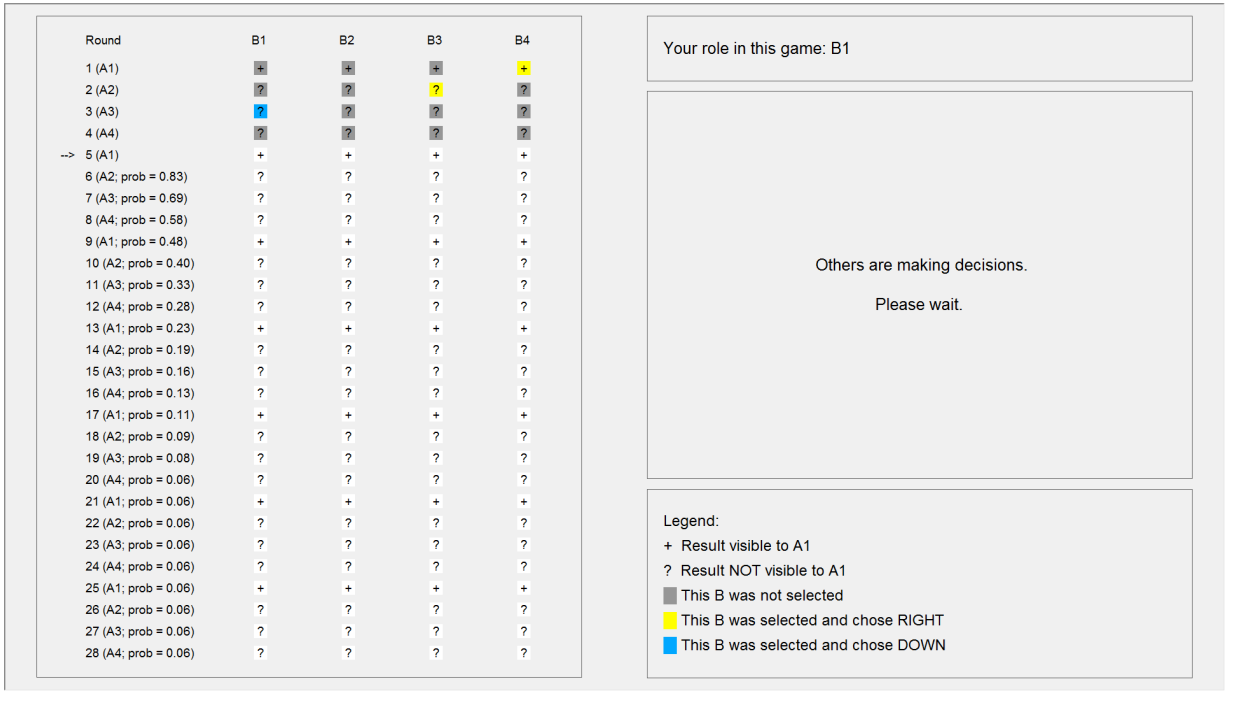


Figure 5: An Example Screen for a B

References

Abraham, M., Grimm, V., Ness, C., & Seebauer, M. (2014). Reputation formation in economic transactions. Working Paper, University of Erlangen-Nürnberg.

Aksoy, O., & Weesie, J. (2012). Beliefs about the social orientations of others: A parametric test of the triangle, false consensus, and cone hypotheses. *Journal of Experimental Social Psychology*, *48*(1), 45–54.

Allison, P. D. (1980a). Estimation and testing for a Markov model of reinforcement. *Sociological Methods & Research*, *8*(4), 434–453.

Allison, P. D. (1980b). Inequality and scientific productivity. *Social Studies of Science*, *10*(2), 163–179.

Anderhub, V., Engelmann, D., & Güth, W. (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior & Organization*, *48*(2), 197–216.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy*, *97*(6), 1447–1458.

Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics*, *110*(1), 1–21.

Aperjis, C., Zeckhauser, R. J., & Miao, Y. (2014). Variable temptations and black mark reputations. *Games and Economic Behavior*, *87*(1), 70–90.

Arrow, K. J. (1974). *The Limits of Organization*. New York: Norton.

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

Bacharach, M., & Gambetta, D. (2001). Trust in signs. In K. S. Cook (Ed.) *Trust in Society*, pp. 148–184. New York: Russell Sage.

Bagwell, K. (1990). Informational product differentiation as a barrier to entry. *International Journal of Industrial Organization*, *8*(2), 207–223.

Bain, J. S. (1956). *Barriers to New Competition: Their Character and Consequences in Manufacturing*. Cambridge, MA: Harvard University Press.

Banks, J. S., & Sobel, J. (1987). Equilibrium selection in signaling games. *Econometrica*, *55*(3), 647–661.

Barber, B. (1983). *The Logic and Limits of Trust*. New Brunswick, NJ: Rutgers University Press.

Barrera, D. (2014). Mechanisms of cooperation. In G. Manzo (Ed.) *Analytical Sociology*, pp. 169–195. Chichester: Wiley.

Barwick, P. J., & Pathak, P. A. (2015). The costs of free entry: An empirical study of real estate agents in Greater Boston. *RAND Journal of Economics*, *46*(1), 103–145.

Battigalli, P., & Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, *97*(2), 170–176.

Becker, G. S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature*, *14*(3), 817–826.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Binmore, K. G. (1998). *Game Theory and the Social Contract: Just Playing*. Cambridge, MA: MIT Press.

Blau, P. M. (1964). *Exchange and Power in Social Life*. Boston, MA: Wiley.

Bliege Bird, R., & Smith, E. A. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current Anthropology*, *46*(2), 221–248.

Bohnet, I., Harmgart, H., Huck, S., & Tyran, J.-R. (2005). Learning trust. *Journal of the European Economic Association*, *3*(2-3), 322–329.

Bohnet, I., & Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *American Economic Review*, *94*(2), 362–366.

Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, *50*(11), 1587–1602.

Bolton, G. E., & Ockenfels, A. (2009). The limits of trust in economic transactions: Investigations of perfect reputation systems. In K. S. Cook, C. Snijders, V. Buskens, & C. Cheshire (Eds.) *eTrust: Forming Relationships in the Online World*, pp. 15–36. New York: Russell Sage.

Bower, A. G., Garber, S., & Watson, J. C. (1997). Learning about a population of agents and the evolution of trust and cooperation. *International Journal of Industrial Organization*, *15*(2), 165–190.

Bowles, S., & Gintis, H. (2004). Persistent parochialism: Trust and exclusion in ethnic networks. *Journal of Economic Behavior & Organization*, *55*(1), 1–23.

Brandts, J., & Figueras, N. (2003). An exploration of reputation formation in exper-

imental games. *Journal of Economic Behavior & Organization*, *50*(1), 89–115.

Bronnenberg, B. J., Dhar, S. K., & Dubé, J. P. H. (2009). Brand history, geography, and the persistence of brand shares. *Journal of Political Economy*, *117*(1), 87–115.

Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica*, *72*(3), 747–780.

Brown, M., & Zehnder, C. (2007). Credit reporting, relationship banking, and loan repayment. *Journal of Money, Credit and Banking*, *39*(8), 1883–1918.

Buskens, V. (1998). The social structure of trust. *Social Networks*, *20*(3), 265–289.

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

Buskens, V. (2003). Trust in triads: Effects of exit, control, and learning. *Games and Economic Behavior*, *42*(2), 235–252.

Buskens, V., & Raub, W. (2002). Embedded trust: Control and learning. *Advances in Group Processes*, *19*, 167–202.

Buskens, V., & Raub, W. (2013). Rational choice research on social dilemmas: Embeddedness effects on trust. In R. Wittek, T. A. B. Snijders, & V. Nee (Eds.) *Handbook of Rational Choice Social Research*, pp. 113–150. Stanford, CA: Stanford University Press.

Buskens, V., Raub, W., & Van der Veer, J. (2010). Trust in triads: An experimental study. *Social Networks*, *32*(4), 301–312.

Buskens, V., & Van de Rijt, A. (2008). Dynamics of networks if everyone strives for structural holes. *American Journal of Sociology*, *114*(2), 371–407.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Camerer, C. F., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica*, *56*(1), 1–36.

Carbone, M., Nielsen, M., & Sassone, V. (2003). A formal model for trust in dynamic networks. *BRICS Report Series*, *10*(4), 21 pp.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, *329*(5996), 1194–1197.

Coleman, J. S. (1964). Collective decisions. *Sociological Inquiry*, *34*(2), 166–181.

Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology*, *91*(6), 1309–1335.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95–S120.

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

Cook, K. S., & Hardin, R. (2001). Norms of cooperativeness and networks of trust.

In M. Hechter, & K.-D. Opp (Eds.) *Social Norms*, pp. 327–347. New York: Russell Sage.

Cook, K. S., Rice, E., & Gerbasi, A. (2004). The emergence of trust networks under uncertainty: The case of transitional economies – insights from social psychological research. In J. Kornai, B. Rothstein, & S. Rose-Ackerman (Eds.) *Creating Social Trust in Post-Socialist Transition*, pp. 193–212. New York: Palgrave Macmillan.

Corten, R. (2014). *Computational Approaches to Studying the Co-evolution of Networks and Behavior in Social Dilemmas*. Chichester: Wiley.

Cripps, M. W., Mailath, G. J., & Samuelson, L. (2004). Imperfect monitoring and impermanent reputations. *Econometrica*, *72*(2), 407–432.

Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, *101*(1), 411–429.

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

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169–193.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, *49*(10), 1407–1424.

Deltas, G. (2003). The small-sample bias of the Gini coefficient: Results and implications for empirical research. *Review of Economics and Statistics*, *85*(1), 226–234.

Denrell, J., & Le Mens, G. (2007). Interdependent sampling and social influence. *Psychological Review*, *114*(2), 398–422.

Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, *2*(4), 265–279.

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution*, *29*(4), 605–610.

Diekmann, A., Jann, B., Przepiorka, W., & Wehrli, S. (2014). Reputation formation and the evolution of cooperation in anonymous online markets. *American Sociological Review*, *79*(1), 65–85.

Diekmann, A., & Lindenberg, S. (2015). Cooperation: Sociological aspects. In J. D. Wright (Ed.) *International Encyclopedia of the Social & Behavioral Sciences*, pp. 862–866. Amsterdam: Elsevier, 2^{nd} ed.

DiMaggio, P., & Louch, H. (1998). Socially embedded consumer transactions: For what kinds of purchases do people most often use networks? *American Sociological Review*, *63*(5), 619–637.

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology*, *32*, 271–297.

Djankov, S., McLiesh, C., & Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, *84*(2), 299–329.

Durkheim, E. (1973 [1893]). *De la Division du Travail Social*. Paris: PUF, 9^{th} ed.

Dutta, B., & Jackson, M. O. (Eds.) (2003). *Networks and Groups: Models of Strategic Formation*. Heidelberg: Springer.

Eguíluz, V. M., Zimmermann, M. G., Cela-Conde, C. J., & San Miguel, M. (2005). Cooperation and the emergence of role differentiation in the dynamics of social networks. *American Journal of Sociology*, *110*(4), 977–1008.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica*, *61*(5), 1047–1071.

Farrell, J. (1986). Moral hazard as an entry barrier. *RAND Journal of Economics*, *17*(3), 440–449.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fehrler, S., & Przepiorka, W. (2013). Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. *Evolution and Human Behavior*, *34*(2), 139–145.

Fischbacher, U. (2007). z-tree: Zürich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Flap, H. (2004). Creation and returns of social capital: A new research program. In H. Flap, & B. Völker (Eds.) *Creation and Returns of Social Capital*, pp. 3–23. London: Routledge.

Fosco, C., & Mengel, F. (2011). Cooperation through imitation and exclusion in networks. *Journal of Economic Dynamics and Control*, *35*(5), 641–658.

Fréchette, G. R., & Yuksel, S. (2013). Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. Working Paper, New York University.

Frey, V. (2014). Embedding trust: Trustees' investments in network embeddedness as credible commitments and signals of trustworthiness. Working Paper, Utrecht University.

Frey, V., Buskens, V., & Corten, R. (2015a). Investments in and effects of embeddedness: An experiment with Trust Games. Working Paper, Utrecht University.

Frey, V., Buskens, V., & Raub, W. (2015b). Embedding trust: A game theoretic model for investments in and returns on network embeddedness. *Journal of Mathematical Sociology*, *39*(1), 39–72.

Frey, V., & Van de Rijt, A. (2015). Reputation cascades. Working Paper, Utrecht

University.

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

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *54*(3), 533–554.

Fudenberg, D., & Maskin, E. (1990). Evolution and cooperation in noisy repeated games. *American Economic Review*, *80*(2), 274–279.

Fudenberg, D., & Tirole, J. (2000). *Game Theory*. Cambridge, MA: MIT Press, 7^{th} ed.

Gambetta, D. (1993). *The Sicilian Mafia: The Business of Private Protection*. Cambridge, MA: Harvard University Press.

Gambetta, D. (2009). Signaling. In P. Hedström, & P. Bearman (Eds.) *Oxford Handbook of Analytical Sociology*, pp. 168–194. Oxford: Oxford University Press.

Gazzale, R. S. (2009). Giving gossips their due: Information provision in games with private monitoring. Working Paper, Williams College.

Gërxhani, K., Brandts, J., & Schram, A. (2013). The emergence of employer information networks in an experimental labor market. *Social Networks*, *35*(4), 541–560.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology*, *213*(1), 103–119.

Gould, R. (2002). The origins of status hierarchies: A formal theory and empirical test. *American Journal of Sociology*, *107*(5), 1143–1178.

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

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

Granovetter, M. (2002). A theoretical agenda for economic sociology. In M. F. Guillén, R. Collins, P. England, & M. Meyer (Eds.) *The New Economic Sociology: Developments in an Emerging Field*, pp. 35–60. New York: Russell Sage.

Greif, A. (1989). Reputation and coalitions in medieval trade: Evidence on the Maghribi traders. *Journal of Economic History*, *49*(4), 857–882.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer, & V. Macho (Eds.) *Forschung und Wissenschaftliches Rechnen. GWDG Bericht 63*, pp. 79–93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal study. *Administrative Science Quarterly*, *40*(4), 619–652.

Gulati, R., & Gargiulo, M. (1999). Where do interorganizational networks come from? *American Journal of Sociology*, *104*(5), 1439–1493.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2014). On cooperation in open communities. *Journal of Public Economics*, *120*, 220–230.

Guseva, A., & Rona-Tas, A. (2001). Uncertainty, risk, and trust: Russian and American credit card markets compared. *American Sociological Review*, *66*(5), 623–646.

Hall, B. H. (2005). A note on the bias in Herfindahl-type measures based on count data. *Revue d'Economie Industrielle*, *110*, 149–156.

Hardin, R. (1982). Exchange theory on strategic bases. *Social Science Information*, *21*(2), 251–272.

Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage.

Heyes, A., & Kapur, S. (2012). Angry customers, e-word-of-mouth and incentives for quality provision. *Journal of Economic Behavior & Organization*, *84*(3), 813–828.

Hobbes, T. (1991 [1651]). *Leviathan*. Cambridge: Cambridge University Press.

Huck, S., Lünser, G. K., & Tyran, J. R. (2010). Consumer networks and firm reputation: A first experimental investigation. *Economics Letters*, *108*(2), 242–244.

Huck, S., Lünser, G. K., & Tyran, J.-R. (2012). Competition fosters trust. *Games and Economic Behavior*, *76*(1), 195–209.

Ioannides, Y. M., & Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, *42*(4), 1056–1093.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jackson, M. O., & Zenou, Y. (Eds.) (2013). *Economic Analyses of Social Networks*. London: Edward Elgar.

James, H. S. (2002). The trust paradox: A survey of economic inquiries into the nature of trust and trustworthiness. *Journal of Economic Behavior & Organization*, *47*(3), 291–307.

Jappelli, T., & Pagano, M. (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance*, *26*(10), 2017–2045.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889.

Kirman, A. (2001). Market organization and individual behavior: Evidence from fish markets. In J. E. Rauch, & A. Casella (Eds.) *Networks and Markets*, pp. 155–195. New York: Russell Sage.

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

Kollock, P. (1994). The emergence of exchange structures: An experimental study

of uncertainty, commitment, and trust. *American Journal of Sociology*, *100*(2), 313–345.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, *24*, 183–214.

Kollock, P. (1999). The production of trust in online markets. *Advances in Group Processes*, *16*, 99–123.

Kreps, D. (1990a). Corporate culture and economic theory. In J. E. Alt, & K. E. Shepsle (Eds.) *Perspectives on Positive Political Economy*, pp. 90–143. Cambridge: Cambridge University Press.

Kreps, D. M. (1990b). *Game Theory and Economic Modeling*. Oxford: Clarendon Press.

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, *27*(2), 245–252.

Kreps, D. M., & Wilson, R. (1982a). Reputation and imperfect information. *Journal of Economic Theory*, *27*(2), 253–279.

Kreps, D. M., & Wilson, R. (1982b). Sequential equilibria. *Econometrica*, *50*(4), 863–894.

Kuwabara, K. (2015). Do reputation systems undermine trust? Divergent effects of enforcement type on generalized trust and trustworthiness. *American Journal of Sociology*, *120*(5), 1390–1428.

Ledyard, J. (1995). Public goods: A survey of experimental research. In J. H. Kagel, & A. E. Roth (Eds.) *The Handbook of Experimental Economics*, pp. 111–194. Princeton, NJ: Princeton University Press.

Lin, N. (2002). *Social Capital: A Theory of Social Structure and Action*. Cambridge: Cambridge University Press.

Lin, N., Cook, K. S., & Burt, R. S. (2001). *Social Capital: Theory and Research*. New York: De Gruyter.

Lindenberg, S. (1992). The method of decreasing abstraction. In J. S. Coleman, & T. J. Fararo (Eds.) *Rational Choice Theory: Advocacy and Critique*, pp. 6–20. Newbury Park, CA: Sage.

Lindenberg, S., & Steg, L. (2007). Normative, gain and hedonic goal frames guiding environmental behavior. *Journal of Social Issues*, *63*(1), 117–137.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American Sociological Review*, *28*(1), 55–67.

MacLeod, B. W. (2007). Reputations, relationships, and contract enforcement. *Jour-*

nal of Economic Literature, *45*(3), 595–628.

Macy, M. W., & Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences*, *99*(Suppl 3), 7229–7236.

Mailath, G. J., Okuno-Fujiwara, M., & Postlewaite, A. (1993). Belief-based refinements in signaling games. *Journal of Economic Theory*, *60*(2), 241–276.

Mailath, G. J., & Samuelson, L. (2006). *Repeated Games and Reputations*. Oxford: Oxford University Press.

Manapat, M. L., Nowak, M. A., & Rand, D. G. (2013). Information, irrationality, and the evolution of trust. *Journal of Economic Behavior & Organization*, *90*, S57–S75.

Manapat, M. L., & Rand, D. G. (2012). Delayed and inconsistent information and the evolution of trust. *Dynamic Games and Applications*, *2*(4), 401–410.

Manzo, G., & Baldassarri, D. (2015). Heuristics, interactions, and status hierarchies: An agent-based model of deference exchange. *Sociological Methods & Research*, *44*(2), 329–387.

Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*(3), 764–769.

Merton, R. K. (1968). The Matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Milgrom, P. R., North, D. C., & Weingast, B. R. (1990). The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics*, *2*(1), 1–23.

Milgrom, P. R., & Roberts, J. (1982). Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica*, *50*(2), 443–459.

Milinski, M., Semmann, D., & Krambeck, H. (2002). Donors to charity gain in both indirect reciprocity and political reputation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1494), 881–883.

Mistzal, B. (1996). *Trust in Modern Societies: The Search for the Bases of Social Order*. Cambridge, MA: Polity Press.

Mizruchi, M. S., Stearns, L. B., & Marquis, C. (2006). The conditional nature of embeddedness: A study of borrowing by large US firms, 1973–1994. *American Sociological Review*, *71*(2), 310–333.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, *341*(6146), 647–651.

Neral, J., & Ochs, J. (1992). The sequential equilibrium theory of reputation building: A further test. *Econometrica*, *60*(5), 1151–1169.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1), 91–112.

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

Nowak, M., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298.

Paik, A., & Woodley, V. (2012). Symbols and investments as signals: Courtship behaviors in adolescent sexual relationships. *Rationality and Society*, *24*(1), 3–36.

Parsons, T. (1937). *The Structure of Social Action*. New York: Free Press.

Patel, D. S. (2012). Concealing to reveal: The informational role of islamic dress. *Rationality and Society*, *24*(3), 295–323.

Pennisi, E. (2005). How did cooperative behavior evolve? *Science*, *309*(5731), 93.

Perc, M., & Szolnoki, A. (2010). Coevolutionary games: A mini review. *BioSystems*, *99*(2), 109–125.

Porter, R. H. (1983). Optimal cartel trigger-price strategies. *Journal of Economic Theory*, *29*(2), 313–338.

Portes, A., & Sensenbrenner, J. (1993). Embeddedness and immigration: Notes on the social determinants of economic action. *American Journal of Sociology*, *98*(6), 1320–1350.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, *37*(1), 7–63.

Przepiorka, W. (2013). Buyers pay for and sellers invest in a good reputation: More evidence from ebay. *Journal of Socio-Economics*, *42*, 31–42.

Przepiorka, W., & Diekmann, A. (2013). Temporal embeddedness and signals of trustworthiness: Experimental tests of a game theoretic model in the United Kingdom, Russia, and Switzerland. *European Sociological Review*, *29*(5), 1010–1023.

Pujol, J. M., Flache, A., Delgado, J., & Sangüesa, R. (2005). How can social networks ever become complex? Modelling the emergence of complex networks from local social exchanges. *Journal of Artificial Societies and Social Simulation*, *8*(4), 18 pp.

Putnam, R. D. (1993). The prosperous community: Social capital and public life. *The American Prospect*, *13*(2), 35–42.

Rapoport, A. (1974). Prisoners dilemma: Recollections and observations. In *Game Theory as a Theory of a Conflict Resolution*, pp. 18–34. Dordrecht: Reidel.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor, MI: University of Michigan Press.

Rasmusen, E. (1994). *Games and Information: An Introduction to Game Theory*. Oxford: Blackwell, 4^{th} ed.

Raub, W. (2004). Hostage posting as a mechanism of trust: Binding, compensation, and signaling. *Rationality and Society*, *16*(3), 319–365.

Raub, W., & Buskens, V. (2012). Speltheoretische modellen voor sociale netwerken en sociaalkapitaaltheorie. In B. Völker (Ed.) *Over Gaten, Bruggen en Witte Paters: Sociaal Kapitaal in Sociologisch Onderzoek*, pp. 27–40. Amsterdam: Rozenberg Publishers.

Raub, W., Buskens, V., & Corten, R. (2015). Social dilemmas and cooperation. In N. Braun, & N. J. Saam (Eds.) *Handbuch Modellbildung und Simulation in den Sozialwissenschaften*, pp. 597–626. Oxford: Springer.

Raub, W., Buskens, V., & Frey, V. (2012). Vertrouwen als opbrengst van investeringen in sociaal kapitaal: Een eenvoudig theoretisch model. In V. Buskens, & I. Maas (Eds.) *Samenwerking in Sociale Dilemma's: Voorbeelden van Nederlands Onderzoek*, pp. 17–44. Amsterdam: Amsterdam University Press.

Raub, W., Buskens, V., & Frey, V. (2013). The rationality of social structure: Cooperation in social dilemmas through investments in and returns on social capital. *Social Networks*, *35*(4), 720–732.

Raub, W., Buskens, V., & Van Assen, M. (2011). Micro-macro links and microfoundations in sociology. *Journal of Mathematical Sociology*, *35*(1-3), 1–25.

Raub, W., Frey, V., & Buskens, V. (2014). Strategic network formation, games on networks, and trust. *Analyse und Kritik*, *36*(1), 135–152.

Raub, W., & Weesie, J. (1990). Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, *96*(3), 626–654.

Resnick, P., Kuwabara, K., Zeckhauser, R., & Friedman, E. (2000). Reputation systems. *Communications of the ACM*, *43*(12), 45–48.

Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics*, *9*(2), 79–101.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, *62*(3), 381–422.

Robinson, D. T., & Stuart, T. E. (2007). Network effects in the governance of strategic alliances. *Journal of Law, Economics, and Organization*, *23*(1), 242–273.

Roca, C. P., Sánchez, A., & Cuesta, J. A. (2012). Individual strategy update and emergence of cooperation in social networks. *Journal of Mathematical Sociology*, *36*(1), 1–21.

Rockenbach, B., & Sadrieh, A. (2012). Sharing information. *Journal of Economic Behavior & Organization*, *81*(2), 689–698.

Rogers, E. M. (1995). *Diffusion of Innovations*. New York: Free Press.

Rosenthal, R. (1981). Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory*, *25*(1), 92–100.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443–452.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

Rubinstein, A. (1985). Choices of conjectures in a bargaining game with incomplete information. In A. E. Roth (Ed.) *Game-Theoretic Models of Bargaining*, pp. 99–114. Cambridge: Cambridge University Press.

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*(5762), 854–856.

Schelling, T. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schmalensee, R. (1982). Product differentiation advantages of pioneering brands. *American Economic Review*, *72*(3), 349–365.

Schneider, F., & Weber, R. A. (2013). Long-term commitment and cooperation. Working Paper 130, Department of Economics, University of Zürich.

Schroeder, K. D., & Rojas, F. G. (2002). A game theoretical analysis of sexually transmitted disease epidemics. *Rationality and Society*, *14*(3), 353–383.

Simpson, B., & McGrimmon, T. (2008). Trust and embedded markets: A multi-method investigation of consumer transactions. *Social Networks*, *30*(1), 1–15.

Skyrms, B., & Pemantle, R. (2000). A dynamic model of social network formation. *Proceedings of the National Academy of Sciences*, *97*(16), 9340–9346.

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

Snijders, C., & Buskens, V. (2001). How to convince someone that you can be trusted? The role of hostages. *Journal of Mathematical Sociology*, *25*(4), 355–383.

Snijders, C., & Keren, G. (2001). Do you trust? Whom do you trust? When do you trust? *Advances in Group Processes*, *18*, 129–160.

Snijders, T. (2013). Network dynamics. In R. Wittek, T. A. B. Snijders, & V. Nee (Eds.) *Handbook of Rational Choice Social Research*, pp. 252–279. Stanford, CA: Stanford University Press.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, *87*(3), 355–374.

Sutter, M., Haigner, S., & Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies*, *77*(4), 1540–1566.

Udehn, L. (2002). The changing face of methodological individualism. *Annual Review of Sociology*, *28*, 479–507.

Uzzi, B. (1996). The sources and consequences of embeddedness for the economic

performance of organizations: The network effect. *American Sociological Review*, *61*(4), 674–698.

Van de Rijt, A., Kang, S. M., Restivo, M., & Patil, A. (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, *111*(19), 6934–6939.

Van Miltenburg, N., Buskens, V., & Raub, W. (2012). Trust in triads: Experience effects. *Social Networks*, *34*(4), 425–428.

Van Ourti, T., & Clarke, P. (2011). A simple correction to remove the bias of the Gini coefficient due to grouping. *Review of Economics and Statistics*, *93*(3), 982–994.

Vega-Redondo, F. (2006). Building up social capital in a changing world. *Journal of Economic Dynamics and Control*, *30*(11), 2305–2338.

Vega-Redondo, F. (2007). *Complex Social Networks*. Cambridge: Cambridge University Press.

Voss, T. (1985). *Rationale Akteure und soziale Institutionen*. München: Oldenbourg.

Weber, M. (1976 [1921]). *Wirtschaft und Gesellschaft*. Tübingen: Mohr, 5^{th} ed.

Wehrli, S. (2014). Reputation and networks: Studies in reputation effects and network formation. Doctoral Dissertation, ETH-Zürich.

Williamson, O. E. (1983). Credible commitments: Using hostages to support exchange. *American Economic Review*, *73*(4), 519–540.

Williamson, O. E. (1993). Calculativeness, trust, and economic organization. *Journal of Law and Economics*, *36*(1), 453–486.

Wippler, R., & Lindenberg, S. (1987). Collective phenomena and rational choice. In J. C. Alexander, B. Giesen, R. Münch, & N. J. Smelser (Eds.) *The Micro-Macro Link*, pp. 135–152. Berkeley, CA: University of California Press.

Xu, H., Liu, D., Wang, H., & Stavrou, A. (2015). E-commerce reputation manipulation: The emergence of reputation escalation as a service. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1296–1306. International World Wide Web Conferences Steering Committee.

Yamagishi, T., Cook, K. S., & Watabe, M. (1998). Uncertainty, trust, and commitment formation in the United States and Japan. *American Journal of Sociology*, *104*(1), 165–194.

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129–166.

Young, P. H. (1996). The economics of convention. *Journal of Economic Perspectives*, *10*(2), 105–122.

Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of Theoretical Biology*, *53*(1), 205–214.

Zak, P. J., & Knack, S. (2001). Trust and growth. *Economic Journal*, *111*(470),

295–321.

Samenvatting / Summary in Dutch[1]

Inleiding

Alice is net verhuisd vanwege een nieuwe baan. Twee punten op haar actielijstje zijn 'regel een schoonmaakster' en 'bestel online een laser-pointer'. Bij beide voornemens wordt Alice geconfronteerd met een vertrouwensprobleem. Een schoonmaakster kan, wanneer ze alleen in huis is, spullen stelen en dan met de noorderzon vertrekken. En de eigenaar van een webwinkel kan een bestelde laser-pointer niet verzenden of iets sturen dat niet aan de beschrijving voldoet.

Zulke vertrouwensproblemen zien we veel in het sociale en economische verkeer. In principe kunnen beide partijen profiteren als ze genoeg vertrouwen hebben en betrouwbaar zijn. Formele vangnetten zoals wetten en contracten bieden niet altijd voldoende zekerheid tegen onbetrouwbaar gedrag. De kosten van juridische vervolging van een online verkoper kunnen hoog zijn en ook als de rechter een schoonmaakster schuldig acht, word je als slachtoffer niet altijd volledig gecompenseerd voor het verlies. Er is dan ook vertrouwen vereist om een transactie mogelijk te maken. Het is in deze zin dat Arrow (1974) vertrouwen beschrijft als een 'belangrijk smeermiddel van de samenleving' (p. 23; vrij vertaald).

Vertrouwen kan worden bevorderd door de inbedding van transacties in sociale structuren waarin informatie over gedrag uit het verleden kan worden gedeeld. Als de nieuwe buren van Alice zich lovend uitlaten over hun schoonmaakster, vertrouwt Alice er wellicht op dat deze schoonmaakster niet zal stelen, en positieve beoordelingen die klanten van een webwinkel op een reputatiesysteem achterlaten, kunnen Alice voldoende vertrouwen geven om de laser-pointer te bestellen en vooraf te betalen. Sociale structuren voor verspreiding van informatie kunnen dus transacties mogelijk maken die niet mogelijk of moeilijker zouden zijn zonder dergelijke structuren.

In dit proefschrift stellen we dat als vertrouwensproblemen opgelost kunnen worden met sociale structuren voor het delen van informatie—denk aan mond-tot-mond-netwerken of online reputatiesystemen—actoren gemotiveerd kunnen worden om der-

[1]I thank Hans van Wijk for language editing.

gelijke sociale structuren op te zetten, om zo de voordelen van vertrouwen en betrouwbaarheid te realiseren. Alice nodigt wellicht haar nieuwe buren uit voor een diner, niet alleen uit beleefdheid, maar ook om hen te vragen of ze een betrouwbare schoonmaker kunnen aanbevelen. De eigenaar van een webwinkel investeert wellicht in een reputatiesysteem, in de verwachting dat deze investering het vertrouwen van potentile klanten bevordert. We gaan hier in hoofdstuk 2 tot en met 5 op in. Daarbij onderzoeken we in een geïntegreerd kader investeringen in de opzet van sociale structuren voor informatie-uitwisseling, en de effecten van die structuren op vertrouwen en betrouwbaarheid.

Bovendien stellen we dat vertrouwensproblemen en de verspreiding van informatie gevolgen kunnen hebben voor de structuur van transactienetwerken in vertrouwenssituaties. De angst voor misbruik van vertrouwen kan ertoe kan leiden dat grote aantallen mensen met maar heel weinig of slechts één andere partij transacties aangaan. Stel dat Alice twee verkopers vindt die laser-pointers aanbieden. De ene heeft veel goede beoordelingen gekregen, terwijl de andere nog geen enkele beoordeling heeft. Het lijkt logisch dat Alice dan voor de verkoper met een 'gevestigde reputatie' kiest. Een volgende klant zal waarschijnlijk hetzelfde doen, enzovoort. De gevestigde verkoper krijgt steeds meer klanten, terwijl de ander niet de kans krijgt om zijn betrouwbaarheid te laten zien. We betogen daarom dat de uitwisseling van informatie over eerder gedrag niet alleen vertrouwen en betrouwbaarheid kan bevorderen, maar dat het ook tot willekeurige ongelijkheid in transactievolumes kan leiden. In hoofdstuk 6 werken we dit argument uit.

In dit proefschrift onderzoeken we dus hoe sociale structuren voor de verspreiding van informatie gedrag rond vertrouwensproblemen beïnvloeden en, tegelijkertijd, hoe sociale structuren ontstaan in de aanwezigheid van vertrouwensproblemen. De studies in dit boek leveren daarmee een bijdrage aan het onderzoek naar de omstandigheden waaronder actoren vertrouwensproblemen kunnen overwinnen. Meer algemeen dragen deze studies bij aan de literatuur die mechanismen bestudeert die samenwerking mogelijk maken, ook al hebben actoren reden om misbruik van elkaar te maken. Door het bestuderen van het tot stand komen van informatienetwerken en transactienetwerken in de context van vertrouwensproblemen draagt dit proefschrift bovendien bij aan de literatuur over de vorming van sociale netwerken.

Onderzoeksvragen

In hoofdstuk 2 tot en met 5 bestuderen we '*investeringen in en opbrengsten van netwerken voor de uitwisseling van informatie*'. In de literatuur over sociaal kapitaal is het geen nieuw idee dat actoren doelgericht sociale relaties aangaan om er profijt

uit te trekken. Expliciete theoretische modellen voor dergelijke 'investeringen in sociaal kapitaal' zijn echter schaars en er zijn geen modellen bekend die inzicht geven in investeringen in informatie-uitwisselingsnetwerken als middel om vertrouwen en betrouwbaarheid te bevorderen. En hoewel er aanwijzingen zijn dat actoren doelgericht sociale structuren voor het delen van informatie opzetten om vertrouwensproblemen te overwinnen, is niet bekend onder welke omstandigheden dit het meest waarschijnlijk is. We proberen deze hiaten in de literatuur te vullen. De overkoepelende vraag die we in hoofdstuk 2 tot en met 5 theoretisch en empirisch onderzoeken is: *Welke omstandigheden stimuleren actoren het meest om sociale structuren voor de uitwisseling van informatie op te zetten met als oogmerk vertrouwensproblemen te overwinnen?*

Doelgerichte actoren zijn eerder geneigd een netwerk voor de verspreiding van informatie op te zetten als ze mogen verwachten dat dit vertrouwen en betrouwbaarheid bijzonder sterk bevordert. Daarom leidt de bovenstaande vraag ook tot de vraag: *Onder welke omstandigheden bevordert een informatie-uitwisselingsnetwerk vertrouwen en betrouwbaarheid het meest?* Het onderzoeken van deze vraag levert een bijdrage aan de literatuur over de effecten van netwerken. Deze literatuur suggereert dat het van de context afhangt in welke mate netwerken helpen vertrouwensproblemen op te lossen. De literatuur biedt echter geen systematische analyses van de contextafhankelijkheid van netwerkeffecten.

Terwijl we in hoofdstuk 2 tot en met 5 het ontstaan van structuren voor informatie-uitwisseling onderzoeken en relaties voor sociale of economische transacties in vertrouwenssituaties als gegeven aannemen, beschouwen we in hoofdstuk 6 structuren voor het delen van informatie als gegeven en endogeniseren we transactierelaties. Eerder onderzoek laat zien dat vertrouwensproblemen ertoe leiden dat actoren langdurige, dyadische transactierelaties vormen. Dit onderzoek was gericht op situaties waar netwerken of instituties voor de verspreiding van informatie over eerder gedrag ontbreken. Het is dus niet bekend welke uitwisselingsstructuren ontstaan in vertrouwenssituaties als actoren die vertrouwen kunnen geven informatie uitwisselen over hun ervaringen met actoren die vertrouwen kunnen misbruiken of belonen. Om in deze leemte in de literatuur te voorzien, onderzoeken we de vraag: *Hoe beïnvloedt het delen van informatie in vertrouwensproblemen transactienetwerken?*

Speltheoretische modellen en bevindingen

Investeringen in en opbrengsten van informatie-uitwisselingsnetwerken

Het doel van hoofdstuk 2 tot en met 5 is regelmatigheden op macroniveau te identificeren. Onder welke maatschappelijke condities op macroniveau is de kans het grootst dat vertrouwensproblemen tot de opzet van de sociale structuren voor informatie-uitwisseling leiden? En onder welke macro-omstandigheden bevorderen dergelijke structuren vertrouwen en betrouwbaarheid het meest? We behandelen deze vragen theoretisch in hoofdstuk 2 tot en met 4. In de speltheoretische modellen die we in deze hoofdstukken analyseren, kunnen actoren investeren in de opzet van een netwerk voor informatie-uitwisseling alvorens in vertrouwenssituaties (vertrouwensspellen) te participeren. In hoofdstuk 2 modelleren we de invloed van het informatienetwerk op gedrag in vertrouwenssituaties, gebruikmakend van het paradigma van een onbepaald aantal spelherhalingen. We analyseren investeringen in en opbrengsten van informatienetwerken in grote populaties. Het model is ook te generaliseren naar interacties in andere sociale dilemma's dan vertrouwenssituaties, zoals samenwerkingsproblemen in het Prisoner's Dilemma.

Het model dat we in hoofdstuk 3 en 4 analyseren, is beperkt tot de studie van interacties in vertrouwenssituaties. Het is ook veel eenvoudiger dan het model in hoofdstuk 2, omdat het gericht is op een scenario waarin twee actoren die vertrouwen kunnen geven ("kopers") herhaaldelijk met één actor interacteren die vertrouwen kan belonen ("verkoper"). Aan de andere kant is het model complexer, omdat het onvolledige informatie vooronderstelt—de vertrouwensnemer kan van een type zijn dat geen motief of mogelijkheid heeft om vertrouwen te misbruiken, maar de kopers kunnen niet direct zien of de verkoper van dit "betrouwbare" type is. Technisch gesproken bestuderen we in hoofdstuk 3 en 4 een eindig herhaald vertrouwensspel met onvolledige informatie. De aanname van onvolledige informatie maakt een gedetailleerdere modellering van de effecten van informatie-uitwisseling op gedrag mogelijk. In een model met volledige informatie, waarbij de koper de gedragsalternatieven en motieven van de verkoper met zekerheid kent (zoals in hoofdstuk 2), beïnvloedt informatie-uitwisseling gedrag uitsluitend omdat die uitwisseling het mogelijk maakt dat één geval van vertrouwensmisbruik door meerdere kopers gesanctioneerd wordt. In een model met onvolledige informatie kan informatie-uitwisseling vertrouwen en betrouwbaarheid ook bevorderen omdat het kopers in staat stelt van elkaars ervaringen met een verkoper te leren. Bovendien kan in een model waarin kopers onzeker zijn over het type van de verkoper (onvolledige informatie), worden aangetoond dat een inves-

tering van een verkoper in een informatie-uitwisselingsnetwerk als een geloofwaardig signaal kan dienen dat de verkoper van het betrouwbare type is. We besteden twee hoofdstukken aan de analyse van het model met onvolledige informatie: We onderzoeken het scenario waarin de kopers in een informatie-uitwisselingsnetwerk kunnen investeren in hoofdstuk 3 en we kijken in hoofdstuk 4 naar het scenario waarin de verkoper informatie-uitwisseling mogelijk kan maken.

De resultaten van de twee modellen zijn in veel gevallen hetzelfde. Daarmee stelt het gebruik van verschillende modellen ons in staat om de robuustheid van onze conclusies met betrekking tot specifieke modelaannames te onderzoeken. Elk model leidt echter ook tot een aantal voorspellingen die niet uit het andere model volgen.

Een belangrijke voorspelling van onze theoretische modellen is dat informatie-uitwisseling vertrouwen en betrouwbaarheid bijzonder sterk bevordert als het vertrouwensprobleem van gemiddelde grootte is, niet te klein en niet te groot. Dat wil zeggen, we verwachten dat de grootte van het netwerkeffect zich als een 'omgekeerde U' verhoudt tot de grootte van het vertrouwensprobleem. Bovendien suggereren onze theoretische analyses dat de kans op investeringen in de vorming van informatienetwerken zich ook als een omgekeerde U verhoudt tot de grootte van het vertrouwensprobleem. Zowel het model voor onbepaald vaak herhaalde spellen als het model voor eindig herhaalde spellen met onvolledige informatie impliceert deze effecten van de grootte van het vertrouwensprobleem. De twee modellen verschillen in zoverre, dat de motieven van verkopers de belangrijkste drijvende kracht zijn in het model voor onbepaalde vaak herhaalde spellen, terwijl de motieven van de kopers de belangrijkste drijvende kracht zijn in het model voor eindig herhaalde spellen.

We kunnen als volgt intuïtie ontwikkelen voor de mate waarin netwerkeffecten en de kans op investeringen in informatie-uitwisselingsnetwerken afhangen van de grootte van het vertrouwensprobleem. Stel dat Alice naast 'regel een schoonmaakster' nog twee andere punten op haar actielijst heeft die haar voor vertrouwensproblemen stellen, te weten 'regel een timmerman om een dakraam te plaatsen' en 'regel een hovenier'. Aan beide opdrachten zijn risico's verbonden: Als een timmerman probeert tijd en geld te besparen, zou bij zware regen lekkage kunnen optreden; en het hoveniersbedrijf zou incompetente medewerkers kunnen sturen die de tuin meer kwaad dan goed doen. Maar Alice is niet echt gehecht aan haar nieuwe tuin. Ze wil niet de moeite nemen informatie over tuinmannen te verzamelen en vertrouwt de taak aan de lokale tuinman toe. Het aanstellen van een schoonmaakster ligt gevoeliger. Alice is van mening dat er een aanzienlijk diefstalrisico is. Ze zou geen schoonmaakster aannemen over wie ze geen goede verhalen heeft gehoord. Daarom vraagt ze in haar buurt en op haar werk of iemand een schoonmaakster kan aanbevelen. En tot slot: Alice heeft weinig vertrouwen in timmerlieden en weet dat een lekkend dakraam een hoop

problemen kan veroorzaken. Ze zet 'regelen van timmerman voor dakraam' onderaan qua prioriteit, en laat het daar staan. Ze weet dat ze een timmerman zelfs niet zou vertrouwen als anderen positief over hem zijn, en daarom neemt ze niet eens de moeite om verdere informatie in te winnen. Uiteindelijk geeft ze het idee van een dakraam en een zonverlichte werkkamer op en koopt ze, met enige aarzeling, een grote lamp.

Hoofdstuk 5 doet verslag van een laboratoriumexperiment dat voorspellingen test die uit hoofdstuk 2 tot en met 4 resulteren. Er is tot nu toe geen duidelijk empirisch bewijs dat actoren doelgericht instituties of netwerken voor de informatieverspreiding opzetten om vertrouwen en betrouwbaarheid te bevorderen. Dat er veel transacties plaatsvinden in dergelijke omgevingen, kan verband houden met het opbloeien van vertrouwen in zulke contexten. Het kan ook zijn dat er vaak vertrouwensproblemen voorkomen in omgevingen waar informatienetwerken om andere redenen zijn opgezet (dus niet met het doel om vertrouwen en betrouwbaarheid te bevorderen). In het laboratorium kunnen we onderzoeken of het optreden van een vertrouwensprobleem daadwerkelijk aanzet tot de vorming van informatienetwerken. Bovendien biedt het laboratorium een zekere controle over theoretisch relevante parameters die van invloed zijn op de grootte van het vertrouwensprobleem en die buiten het laboratorium moeilijk te meten zijn. Bijvoorbeeld de kans dat een verkoper een financieel motief heeft om vertrouwen te misbruiken of de uitbetalingen in vertrouwensinteracties.

Het experiment implementeert de theoretische modellen uit hoofdstuk 3 en 4: twee kopers interacteren herhaaldelijk in vertrouwensspellen met één verkoper. De grootte van het vertrouwensprobleem wordt gemanipuleerd via de kans dat de verkoper een financieel motief heeft om vertrouwen te misbruiken. Informatie-uitwisseling tussen de twee kopers is exogeen of kan, tegen kosten, door de kopers of de verkoper voorafgaand aan de interacties worden mogelijk gemaakt. De resultaten laten zien dat een aanzienlijk deel van de kopers en verkopers investeert in informatie-uitwisseling, als ze daartoe de kans krijgen. De resultaten bevestigen ook dat informatie-uitwisseling vertrouwen en betrouwbaarheid bevordert, en ze tonen aan dat het effect van endogene informatie-uitwisseling sterker is dan van exogene informatie-uitwisseling. We vinden echter weinig bewijs voor de omgekeerde U-hypothese: er was geen sprake van systematische variatie in investeringen in of effecten van informatie-uitwisseling gerelateerd aan de grootte van het vertrouwensprobleem.

Het domino-effect van reputatie

In hoofdstuk 6 beschouwen we structuren voor het delen van informatie tussen kopers als exogeen en onderzoeken we de endogene vorming van de transactierelaties tussen kopers en verkopers. We gaan ervan uit dat de kopers hun transactiepartners (de

verkopers) kunnen kiezen. De vraag die we onderzoeken, is: Hoe beïnvloedt het delen van informatie in vertrouwensproblemen transactienetwerken?

Onze speltheoretische analyse voorspelt dat er sterke ongelijkheid in transactievolumes tussen verkopers kan ontstaan. Om het risico van misbruik zo klein mogelijk te houden zullen rationele kopers verkopers zonder transactiegeschiedenis mijden en de voorkeur geven aan verkopers met een goede reputatie. Als binnen een groep kopers die informatie delen voor het eerst een verkoper als betrouwbaar wordt aangemerkt, zullen *alle* kopers van deze groep in de daaropvolgende uitwisselingen vertrouwen stellen in die verkoper, die simpelweg het geluk had dat hij de eerste was die als betrouwbaar werd gekwalificeerd. Dit domino-effect leidt er echter wel toe dat potentieel betrouwbare nieuwkomers geen kans krijgen om een reputatie op te bouwen, waardoor de willekeurige aanvankelijke ongelijkheid wordt versterkt. Onze theoretische analyse suggereert dus dat vertrouwensproblemen en het delen van informatie tot een vorm van cumulatief voordeel leidt met als gevolg willekeurige ongelijkheid in transactievolumes tussen verkopers. De theorie impliceert dat de ongelijkheid tussen verkopers groter is als de opdrachtgevers informatie in grotere groepen delen. Dit betekent dat meer informatie-uitwisseling enerzijds kan helpen vertrouwensproblemen te overwinnen, maar anderzijds onbedoeld ook tot grote ongelijkheid kan leiden.

De resultaten van een laboratoriumexperiment ondersteunen de theorie. Als kopers in grote groepen informatie delen, is het percentage succesvolle transacties groter, maar er is ook meer ongelijkheid tussen verkopers in hoe vaak ze vertrouwd worden. Er is sprake van enige willekeur in deze ongelijkheid: groepen van kopers sloten vaak herhaald transacties met één betrouwbare verkoper, terwijl ze geen informatie hadden over de betrouwbaarheid van de andere verkopers. Ons experiment toont ook aan dat deze scheefgroei veroorzaakt wordt door de angst voor misbruik van vertrouwen (en niet, bijvoorbeeld, door een algemene tendens om de keuzes van anderen te imiteren). In een controlegroep waarin verkopers geen redenen hadden om vertrouwen te misbruiken—en waar dus geen vertrouwensproblemen speelden—hadden koper geen voorkeur voor verkoper met een goede reputatie boven onbekende verkopers.

Samenvattend laat dit proefschrift dus zien dat vertrouwensproblemen tot het ontstaan van bepaalde sociale structuren kunnen leiden. Actoren die met vertrouwensproblemen geconfronteerd zijn, kunnen gemotiveerd zijn informatie-uitwisselingsnetwerken op te zetten. Als er een risico van vertrouwensmisbruik is, kan informatie-uitwisseling leiden tot willekeurige ongelijkheid in transactievolumes.

Acknowledgments

This book is the result of an extensive process, a process that brought me to many places, intellectually as well as physically. It would not have been completed without the support of many people. Foremost among the many debts I have accumulated while working on this PhD dissertation is the gratitude I owe to my supervisors Vincent Buskens, Werner Raub, and Rense Corten for their academic support and guidance from the beginning to the end. My experience with Rense and Vincent as supervisors of my master thesis was a critical factor that led me to embark on this specific project in the first place. As my main PhD supervisor, Vincent was available for many joint math sessions on the whiteboard. He always maintained his admirably cheerful but careful manner, also when I despaired. Werner's merciless eye for arguments that are not watertight helped me to improve my work substantially and, over the years, taught me to avoid many pitfalls. Werner also took the lead in developing the theory presented in Chapter 2 as well as in several spinoff projects. Rense brought into the project his broad expertise in social networks research and thanks to an idea of his I could resolve the computer network problems that were holding back the experiment reported in Chapter 5. Vincent, Werner, and Rense, I have no doubt that the difficulties I sometimes had also caused difficulties for you on occasion. I thank you very much for having stood by me and for your patience. On a walk I had with Arnout van de Rijt on the campus of the State University of New York (SUNY) at Stony Brook, he said that you could not thank your PhD supervisors enough; they have invested so much into your development that you would be indebted to them forever. Arnout has himself invested so much into my development during my stay at SUNY and our continued collaboration that I am indebted to him as if he had been a supervisor. In my collaboration with Arnout I could look at network dynamics in situations of trust from a complementary perspective and I learned how much pondering questions related to the framing of research can actually contribute to designing sound research and finding the right way to look at the data.

The Interuniversity Center for Social Science Theory and Methodology (ICS) and the Department of Sociology at Utrecht University provided a very supportive and

stimulating home for my research. Institutions like the Forum Days, the Cooperative Relations Seminars, the PhD-student mentor (namely, Mariëlle Bedaux), and the daily department lunches (most frequently initiated by Ineke Maas) were invaluable assets, just as the very smooth administrative support provided by, among others, Bärbel Barendrecht, Tineke Edink, Marjet Janmaat, Ellen Jansen, Pim Sangers, and Babs van den Born. The research of a fellow ICS member shows that strong social ties at the workplace can negatively impact performance. While I also had my unproductive days, I do not blame this on having had many colleagues that I appreciate very much on a personal level. Antonie, Lieselotte, Sanne, and Sara deserve a special mention at this place.

Among my Utrecht colleagues, I am especially grateful also to the members of the research line Cooperation in Social and Economic Relations. This group included, in addition to my supervisors, Ozan Aksoy, Nikki van Gerwen, David Macro, Nynke van Miltenburg, Dominik Morbitzer, Wojtek Przepiorka, Charlotte Rutten, and Jeroen Weesie. They have provided an inspiring and critical environment and served as internal reviewers of many of the studies presented in this book. I am indebted also to Jacob Dijkstra, Andreas Flache, Thomas Gautschi, André Grow, Manuel Muñoz-Herrera, and Victor Stoica for having reviewed my work at ICS Forum Days. I cordially thank the members of the manuscript committee Andreas Diekmann, Peter van der Heijden, Arthur Schram, Chris Snijders, and Thomas Voss for their time to read the manuscript and for their comments.

Thanks to the ICS and my supervisors I could reach many places I could not have reached alone. At Cornell University in Ithaca I enjoyed the hospitality of Michael Macy and the members of the Social Dynamics Laboratory. On Long Island at SUNY at Stony Brook, I had the pleasure to work with Arnout van de Rijt and to discuss research as well as related and unrelated stuff also with other people of the cluster Big Data for the Social Sciences and the Center for Behavioral Political Economy (CBPE). The CBPE also provided a brand new experimental laboratory for my research. Shorter trips to scientific conferences furthermore brought me to places as close by as the University Museum in Utrecht and as far away as Hong Kong. These conference visits were a great inspiration, showed me that there are people out there who care about the work I do, and allowed me to build up a broader scientific network. These longer research visits and shorter trips to conferences were valuable for the work presented in this book but also on a more personal level. I would, for example, not want to miss the memories I have from my trip to Xi'an and Hong Kong together with Vincent and Werner.

After days of hard work, and especially after days where things did not go well, I could often cool my head with wind or water during training sessions with Hellas

Triathlon. The friendships and acquaintances I have at this association have played a crucial role in Utrecht becoming home for me and it was a distinct pleasure for me to race the last season in a team with Chris, Ferdinand, Peter, and Rick. I am grateful to Maike for having brought me to Hellas, which would probably not have happened if there had not been a tradition of participating in running races at the ICS Utrecht—a tradition that peaked for me with running a marathon with Vincent on Terschelling.

During my prolonged stay abroad, Switzerland remained a place where I could go home to. For this I cordially thank my parents Max and Edith and my sister Selina as well as my friends Anna, David, Fabian, Mario, Markus, Remo, and Roman.

Curriculum Vitae

Vincenz Frey was born on April 3, 1983 in Ehrendingen, Switzerland. He obtained a bachelor's degree in Sociology with a minor in Economics from the University of Bern in 2008. He studied for his masters degree at Utrecht University, The Netherlands. After graduating *cum laude* from the research master's program *Sociology and Social Research* in 2010, he became a PhD candidate at the Interuniversity Center for Social Science Theory and Methodology (ICS) in Utrecht. His PhD project was funded by the Netherlands Organization for Scientific Research (NWO) and completed in 2015. In 2013 and 2014, he was a visiting scholar for two and four months, respectively, at the Departments of Sociology at Cornell University and State University of New York (SUNY), Stony Brook. Currenty, he is post-doctoral researcher at Utrecht University.

Publications and working papers of the author

Frey, V. (2014). Embedding trust: Trustees' investments in network embeddedness as credible commitments and signals of trustworthiness. Working Paper, Utrecht University. (Chapter 4 of this thesis)

Frey, V., Buskens, V., & Corten, R. (2015a). Investments in and returns on embeddedness: An experiment with Trust Games. Working Paper, Utrecht University. (Chapter 5 of this thesis)

Frey, V., Buskens, V., & Raub, W. (2015b). Embedding trust: A game theoretic model for investments in and returns on network embeddedness. *Journal of Mathematical Sociology*, 39(1), 39–72. (Chapter 3 of this thesis)

Frey, V., Corten, R., & Buskens, V. (2010). Network and information effects in the emergence of efficient conventions: Theory and experimental findings. ISCORE Discussion Paper 271, Utrecht University.

Frey, V., Corten, R., & Buskens, V. (2012). Equilibrium selection in network coordination games: An experimental study. *Review of Network Economics*, 11(3), 27 pp.

Frey, V., & Van de Rijt, A. (2015). Reputation cascades. Working Paper, Utrecht University. (Chapter 6 of this thesis)

Raub, W., Buskens, V., & Frey, V. (2012). Vertrouwen als opbrengst van investeringen in sociaal kapitaal: Een eenvoudig theoretisch model. In V. Buskens, & I. Maas (Eds.) *Samenwerking in Sociale Dilemma's: Voorbeelden van Nederlands Onderzoek*, (pp. 17–44). Amsterdam: Amsterdam University Press.

Raub, W., Buskens, V., & Frey, V. (2013). The rationality of social structure: Cooperation in social dilemmas through investments in and returns on social capital. *Social Networks*, 35(4), 720–732. (Chapter 2 of this thesis)

Raub, W., Frey, V., & Buskens, V. (2014). Strategic network formation, games on networks, and trust. *Analyse und Kritik*, 36(1), 135–152.

Van der Lippe, T., Frey, V., & Tsvetkova, M. (2013). Outsourcing of domestic tasks: A matter of preferences? *Journal of Family Issues*, 34(12), 1574–1597.

ICS Dissertation series

The ICS-series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state-of-the-art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS expanded to the University of Nijmegen. Most of the projects are financed by the participating universities or by the Netherlands Organization for Scientific Research (NWO). The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS-series itself.

1. C. van Liere. (1990). *Lastige leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen.* Amsterdam: Thesis Publishers.
2. Marco H.D. van Leeuwen. (1990). *Bijstand in Amsterdam, ca. 1800–1850. Armenzorg als beheersings- en overlevingsstrategie.* ICS-dissertation, Utrecht.
3. I. Maas. (1990). *Deelname aan podiumkunsten via de podia, de media en actieve beoefening. Substitutie of leereffecten?.* Amsterdam: Thesis Publishers.
4. M.I. Broese van Groenou. (1991). *Gescheiden netwerken. De relaties met vrienden en verwanten na echtscheiding.* Amsterdam: Thesis Publishers.
5. Jan M.M. van den Bos. (1991). *Dutch EC policy making. A model-guided approach to coordination and negotiation.* Amsterdam: Thesis Publishers.
6. Karin Sanders. (1991). *Vrouwelijke pioniers. Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan.* Amsterdam: Thesis Publishers.
7. Sjerp de Vries. (1991). *Egoism, altruism, and social justice. Theory and experiments on cooperation in social dilemmas.* Amsterdam: Thesis Publishers.
8. Ronald S. Batenburg. (1991). *Automatisering in bedrijf.* Amsterdam: Thesis Publishers.
9. Rudi Wielers. (1991). *Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang.* Amsterdam: Thesis Publishers.
10. Gert P. Westert. (1991). *Verschillen in ziekenhuisgebruik.* ICS-dissertation, Groningen.
11. Hanneke Hermsen. (1992). *Votes and policy preferences. Equilibria in party systems.* Amsterdam: Thesis Publishers.
12. Cora J.M. Maas. (1992). *Probleemleerlingen in het basisonderwijs.* Amsterdam: Thesis Publishers.

13. Ed A.W. Boxman. (1992). *Contacten en carrière. Een empirisch-theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities.* Amsterdam: Thesis Publishers.
14. Conny G.J. Taes. (1992). *Kijken naar banen. Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs.* Amsterdam: Thesis Publishers.
15. Peter van Roozendaal. (1992). *Cabinets in multi-party democracies. The effect of dominant and central parties on cabinet composition and durability.* Amsterdam: Thesis Publishers.
16. Marcel van Dam. (1992). *Regio zonder regie. Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid.* Amsterdam: Thesis Publishers.
17. Tanja van der Lippe. (1993). *Arbeidsverdeling tussen mannen en vrouwen.* Amsterdam: Thesis Publishers.
18. Marc A. Jacobs. (1993). *Software: Kopen of kopiëren? Een sociaal-wetenschappe-lijk onderzoek onder PC-gebruikers.* Amsterdam: Thesis Publishers.
19. Peter van der Meer. (1993). *Verdringing op de Nederlandse arbeidsmarkt. Sector- en sekseverschillen.* Amsterdam: Thesis Publishers.
20. Gerbert Kraaykamp. (1993). *Over lezen gesproken. Een studie naar sociale differentiatie in leesgedrag.* Amsterdam: Thesis Publishers.
21. Evelien Zeggelink. (1993). *Strangers into friends. The evolution of friendship networks using an individual oriented modeling approach.* Amsterdam: Thesis Publishers.
22. Jaco Berveling. (1994). *Het stempel op de besluitvorming. Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen.* Amsterdam: Thesis Publishers.
23. Wim Bernasco. (1994). *Coupled careers. The effects of spouse's resources on success at work.* Amsterdam: Thesis Publishers.
24. Liset van Dijk. (1994). *Choices in child care. The distribution of child care among mothers, fathers and non-parental care providers.* Amsterdam: Thesis Publishers.
25. Jos de Haan. (1994). *Research groups in Dutch sociology.* Amsterdam: Thesis Publishers.
26. K. Boahene. (1995). *Innovation adoption as a socio-economic process. The case of the Ghanaian cocoa industry.* Amsterdam: Thesis Publishers.
27. Paul E.M. Ligthart. (1995). *Solidarity in economic transactions. An experimental study of framing effects in bargaining and contracting.* Amsterdam: Thesis Publishers.
28. Roger Th. A.J. Leenders. (1995). *Structure and influence. Statistical models for the dynamics of actor attributes, network structure, and their interdependence.* Amsterdam: Thesis Publishers.
29. Beate Völker. (1995). *Should auld acquaintance be forgot...? Institutions of communism, the transition to capitalism and personal networks: The case of East Germany.* Amsterdam: Thesis Publishers.
30. A. Cancrinus-Matthijsse. (1995). *Tussen hulpverlening en ondernemerschap. Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West-Europese landen.* Amsterdam: Thesis Publishers.
31. Nardi Steverink. (1996). *Zo lang mogelijk zelfstandig. Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen.* Amsterdam: Thesis Publishers.
32. Ellen Lindeman. (1996). *Participatie in vrijwilligerswerk.* Amsterdam: Thesis Publishers.

33. Chris Snijders. (1996). *Trust and commitments.* Amsterdam: Thesis Publishers.
34. Koos Postma. (1996). *Changing prejudice in Hungary. A study on the collapse of state socialism and its impact on prejudice against gypsies and Jews.* Amsterdam: Thesis Publishers.
35. Jooske T. van Busschbach. (1996). *Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties.* Amsterdam: Thesis Publishers.
36. René Torenvlied. (1996). *Besluiten in uitvoering. Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten.* Amsterdam: Thesis Publishers.
37. Andreas Flache. (1996). *The Double edge of networks. An analysis of the effect of informal networks on cooperation in social dilemmas.* Amsterdam: Thesis Publishers.
38. Kees van Veen. (1997). *Inside an internal labor market: Formal rules, flexibility and career lines in a Dutch manufacturing company.* Amsterdam: Thesis Publishers.
39. Lucienne van Eijk. (1997). *Activity and well-being in the elderly.* Amsterdam: Thesis Publishers.
40. Róbert Gál. (1997). *Unreliability. Contract discipline and contract governance under economic transition.* Amsterdam: Thesis Publishers.
41. Anne-Geerte van de Goor. (1997). *Effects of regulation on disability duration.* ICS-dissertation, Utrecht.
42. Boris Blumberg. (1997). *Das Management von Technologiekooperationen. Partnersuche und Verhandlungen mit dem Partner aus Empirisch-Theoretischer Perspektive.* ICS-dissertation, Utrecht.
43. Marijke von Bergh. (1997). *Loopbanen van oudere werknemers.* Amsterdam: Thesis Publishers.
44. Anna Petra Nieboer. (1997). *Life-events and well-being: A prospective study on changes in well-being of elderly people due to a serious illness event or death of the spouse.* Amsterdam: Thesis Publishers.
45. Jacques Niehof. (1997). *Resources and social reproduction: The effects of cultural and material resources on educational and occupational careers in industrial nations at the end of the twentieth century.* ICS-dissertation, Nijmegen.
46. Ariana Need. (1997). *The kindred vote. Individual and family effects of social class and religion on electoral change in the Netherlands, 1956–1994.* ICS-dissertation, Nijmegen.
47. Jim Allen. (1997). *Sector composition and the effect of education on wages: An international comparison.* Amsterdam: Thesis Publishers.
48. Jack B.F. Hutten. (1998). *Workload and provision of care in general practice. An empirical study of the relation between workload of Dutch general practitioners and the content and quality of their care.* ICS-dissertation, Utrecht.
49. Per B. Kropp. (1998). *Berufserfolg im Transformationsprozeß, Eine theoretisch-empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland.* ICS-dissertation, Utrecht.
50. Maarten H.J. Wolbers. (1998). *Diploma-inflatie en verdringing op de arbeidsmarkt. Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland.* ICS-dissertation, Nijmegen.
51. Wilma Smeenk. (1998). *Opportunity and marriage. The impact of individual resources and marriage market structure on first marriage timing and partner choice in the Netherlands.* ICS-dissertation, Nijmegen.

52. Marinus Spreen. (1999). *Sampling personal network structures: Statistical inference in ego-graphs.* ICS-dissertation, Groningen.
53. Vincent Buskens. (1999). *Social networks and trust.* ICS-dissertation, Utrecht.
54. Susanne Rijken. (1999). *Educational expansion and status attainment. A cross-national and over-time comparison.* ICS-dissertation, Utrecht.
55. Mérove Gijsberts. (1999). *The legitimation of inequality in state-socialist and market societies, 1987–1996.* ICS-dissertation, Utrecht.
56. Gerhard G. Van de Bunt. (1999). *Friends by choice. An actor-oriented statistical network model for friendship networks through time.* ICS-dissertation, Groningen.
57. Robert Thomson. (1999). *The party mandate: Election pledges and government actions in the Netherlands, 1986–1998.* Amsterdam: Thela Thesis.
58. Corine Baarda. (1999). *Politieke besluiten en boeren beslissingen. Het draagvlak van het mestbeleid tot 2000.* ICS-dissertation, Groningen.
59. Rafael Wittek. (1999). *Interdependence and informal control in organizations.* ICS-dissertation, Groningen.
60. Diane Payne. (1999). *Policy making in the European Union: An analysis of the impact of the reform of the structural funds in Ireland.* ICS-dissertation, Groningen.
61. René Veenstra. (1999). *Leerlingen – klassen – scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs.* Amsterdam: Thela Thesis.
62. Marjolein Achterkamp. (1999). *Influence strategies in collective decision making. A comparison of two models.* ICS-dissertation, Groningen.
63. Peter Mühlau. (2000). *The governance of the employment relation. A relational signaling perspective.* ICS-dissertation, Groningen.
64. Agnes Akkerman. (2000). *Verdeelde vakbeweging en stakingen. Concurrentie om leden.* ICS-dissertation, Groningen.
65. Sandra van Thiel. (2000). *Quangocratization: Trends, causes and consequences.* ICS-dissertation, Utrecht.
66. Rudi Turksema. (2000). *Supply of day care.* ICS-dissertation, Utrecht.
67. Sylvia E. Korupp (2000). *Mothers and the process of social stratification.* ICS-dissertation, Utrecht.
68. Bernard A. Nijstad (2000). *How the group affects the mind: Effects of communication in idea generating groups.* ICS-dissertation, Utrecht.
69. Inge F. de Wolf (2000). *Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers.* ICS-dissertation, Utrecht.
70. Jan Kratzer (2001). *Communication and performance: An empirical study in innovation teams.* ICS-dissertation, Groningen.
71. Madelon Kroneman (2001). *Healthcare systems and hospital bed use.* ICS/NIVEL-dissertation, Utrecht.
72. Herman van de Werfhorst (2001). *Field of study and social inequality. Four types of educational resources in the process of stratification in the Netherlands.* ICS-dissertation, Nijmegen.
73. Tamás Bartus (2001). *Social capital and earnings inequalities. The role of informal job search in Hungary.* ICS-dissertation, Groningen.
74. Hester Moerbeek (2001). *Friends and foes in the occupational career. The influence of sweet and sour social capital on the labour market.* ICS-dissertation, Nijmegen.
75. Marcel van Assen (2001). *Essays on actor perspectives in exchange networks and social dilemmas.* ICS-dissertation, Groningen.

76. Inge Sieben (2001). *Sibling similarities and social stratification. The impact of family background across countries and cohorts.* ICS-dissertation, Nijmegen.
77. Alinda van Bruggen (2001). *Individual production of social well-being. An exploratory study.* ICS-dissertation, Groningen.
78. Marcel Coenders (2001). *Nationalistic attitudes and ethnic exclusionism in a comparative perspective: An empirical study of attitudes toward the country and ethnic immigrants in 22 countries.* ICS-dissertation, Nijmegen.
79. Marcel Lubbers (2001). *Exclusionistic electorates. Extreme right-wing voting in Western Europe.* ICS-dissertation, Nijmegen.
80. Uwe Matzat (2001). *Social networks and cooperation in electronic communities. A theoretical-empirical analysis of academic communication and internet discussion groups.* ICS-dissertation, Groningen.
81. Jacques P.G. Janssen (2002). *Do opposites attract divorce? Dimensions of mixed marriage and the risk of divorce in the Netherlands.* ICS-dissertation, Nijmegen.
82. Miranda Jansen (2002). *Waardenoriëntaties en partnerrelaties. Een panelstudie naar wederzijdse invloeden.* ICS-dissertation, Utrecht.
83. Anne Rigt Poortman (2002). *Socioeconomic causes and consequences of divorce.* ICS-dissertation, Utrecht.
84. Alexander Gattig (2002). *Intertemporal decision making.* ICS-dissertation, Groningen.
85. Gerrit Rooks (2002). *Contract en conflict: Strategisch management van inkooptransacties.* ICS-dissertation, Utrecht.
86. Károly Takács (2002). *Social networks and intergroup conflict.* ICS-dissertation, Groningen.
87. Thomas Gautschi (2002). *Trust and exchange, effects of temporal embeddedness and network embeddedness on providing and dividing a surplus.* ICS-dissertation, Utrecht.
88. Hilde Bras (2002). *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850–1950.* Amsterdam: Aksant Academic Publishers.
89. Merijn Rengers (2002). *Economic lives of artists. Studies into careers and the labour market in the cultural sector.* ICS-dissertation, Utrecht.
90. Annelies Kassenberg (2002). *Wat scholieren bindt. Sociale gemeenschap in scholen.* ICS-dissertation, Groningen
91. Marc Verboord (2003). *Moet de meester dalen of de leerling klimmen? De invloed van literatuuronderwijs en ouders op het lezen van boeken tussen 1975 en 2000.* ICS-dissertation, Utrecht.
92. Marcel van Egmond (2003). *Rain falls on all of us (but some manage to get more wet than others): Political Context and Electoral Participation.* ICS-dissertation, Nijmegen.
93. Justine Horgan (2003). *High performance human resource management in Ireland and the Netherlands: Adoption and effectiveness.* ICS-dissertation, Groningen.
94. Corine Hoeben (2003). *LETS' be a community. Community in local exchange trading systems.* ICS-dissertation, Groningen.
95. Christian Steglich (2003). *The framing of decision situations. Automatic goal selection and rational goal pursuit.* ICS-dissertation, Groningen.
96. Johan van Wilsem (2003). *Crime and context. The impact of individual, neighborhood, city and country characteristics on victimization.* ICS-dissertation, Nijmegen.
97. Christiaan Monden (2003). *Education, inequality and health. The impact of partners and life course.* ICS-dissertation, Nijmegen.

98. Evelyn Hello (2003). *Educational attainment and ethnic attitudes. How to explain their relationship.* ICS-dissertation, Nijmegen.
99. Marnix Croes en Peter Tammes (2004). *Gif laten wij niet voortbestaan. Een onderzoek naar de overlevingskansen van joden in de Nederlandse gemeenten, 1940–1945.* Amsterdam: Aksant Academic Publishers.
100. Ineke Nagel (2004). *Cultuurdeelname in de levensloop.* ICS-dissertation, Utrecht.
101. Marieke van der Wal (2004). *Competencies to participate in life. Measurement and the impact of school.* ICS-dissertation, Groningen.
102. Vivian Meertens (2004). *Depressive symptoms in the general population: A multifactorial social approach.* ICS-dissertation, Nijmegen.
103. Hanneke Schuurmans (2004). *Promoting well-being in frail elderly people. Theory and intervention.* ICS-dissertation, Groningen.
104. Javier Arregui (2004). *Negotiation in legislative decision-making in the European Union.* ICS-dissertation, Groningen.
105. Tamar Fischer (2004). *Parental divorce, conflict and resources. The effects on children's behaviour problems, socioeconomic attainment, and transitions in the demographic career.* ICS-dissertation, Nijmegen.
106. René Bekkers (2004). *Giving and volunteering in the Netherlands: Sociological and psychological perspectives.* ICS-dissertation, Utrecht.
107. Renée van der Hulst (2004). *Gender differences in workplace authority: An empirical study on social networks.* ICS-dissertation, Groningen.
108. Rita Smaniotto (2004). *'You scratch my back and I scratch yours' versus 'Love thy neighbour'. Two Proximate Mechanisms of Reciprocal Altruism.* ICS-dissertation, Groningen.
109. Maurice Gesthuizen (2004). *The life-course of the low-educated in the Netherlands: Social and economic risks.* ICS-dissertation, Nijmegen.
110. Carlijne Philips (2005). *Vakantiegemeenschappen. Kwalitatief en kwantitatief onderzoek naar gelegenheid- en refreshergemeenschap tijdens de vakantie.* ICS-dissertation, Groningen.
111. Esther de Ruijter (2005). *Household outsourcing.* ICS-dissertation, Utrecht.
112. Frank van Tubergen (2005). *The integration of immigrants in cross-national perspective: Origin, destination, and community effects.* ICS-dissertation, Utrecht.
113. Ferry Koster (2005). *For the time being. Accounting for inconclusive findings concerning the effects of temporary employment relationships on solidary behavior of employees.* ICS-dissertation, Groningen.
114. Carolien Klein Haarhuis (2005). *Promoting anti-corruption reforms. Evaluating the implementation of a World Bank anti-corruption program in seven African countries (1999–2001).* ICS-dissertation, Utrecht.
115. Martin van der Gaag (2005). *Measurement of individual social capital.* ICS-dissertation, Groningen.
116. Johan Hansen (2005). *Shaping careers of men and women in organizational contexts.* ICS-dissertation, Utrecht.
117. Davide Barrera (2005). *Trust in embedded settings.* ICS-dissertation, Utrecht.
118. Mattijs Lambooij (2005). *Promoting cooperation. Studies into the effects of long-term and short-term rewards on cooperation of employees.* ICS-dissertation, Utrecht.
119. Lotte Vermeij (2006). *What's cooking? Cultural boundaries among Dutch teenagers of different ethnic origins in the context of school.* ICS-dissertation, Utrecht.

120. Mathilde Strating (2006). *Facing the challenge of rheumatoid arthritis. A 13-year prospective study among patients and cross-sectional study among their partners.* ICS-dissertation, Groningen.
121. Jannes de Vries (2006). *Measurement error in family background variables: The bias in the intergenerational transmission of status, cultural consumption, party preference, and religiosity.* ICS-dissertation, Nijmegen.
122. Stefan Thau (2006). *Workplace deviance: Four studies on employee motives and self-regulation.* ICS-dissertation, Groningen.
123. Mirjam Plantinga (2006). *Employee motivation and employee performance in child care. The effects of the introduction of market forces on employees in the Dutch child-care sector.* ICS-dissertation, Groningen.
124. Helga de Valk (2006). *Pathways into adulthood. A comparative study on family life transitions among migrant and Dutch youth.* ICS-dissertation, Utrecht.
125. Henrike Elzen (2006). *Self-management for chronically ill older people.* ICS-dissertation, Groningen.
126. Ayşe Güveli (2007). *New social classes within the service class in the Netherlands and Britain. Adjusting the EGP class schema for the technocrats and the social and cultural specialists.* ICS-dissertation, Nijmegen.
127. Willem-Jan Verhoeven (2007). *Income attainment in post-communist societies.* ICS-dissertation, Utrecht.
128. Marieke Voorpostel (2007). *Sibling support: The exchange of help among brothers and sisters in the Netherlands.* ICS-dissertation, Utrecht.
129. Jacob Dijkstra (2007). *The effects of externalities on partner choice and payoffs in exchange networks.* ICS-dissertation, Groningen.
130. Patricia van Echtelt (2007). *Time-greedy employment relationships: Four studies on the time claims of post-Fordist work.* ICS-dissertation, Groningen.
131. Sonja Vogt (2007). *Heterogeneity in social dilemmas: The case of social support.* ICS-dissertation, Utrecht.
132. Michael Schweinberger (2007). *Statistical methods for studying the evolution of networks and behavior.* ICS-dissertation, Groningen.
133. István Back (2007). *Commitment and evolution: Connecting emotion and reason in long-term relationships.* ICS-dissertation, Groningen.
134. Ruben van Gaalen (2007). *Solidarity and ambivalence in parent-child relationships.* ICS-dissertation, Utrecht.
135. Jan Reitsma (2007). *Religiosity and solidarity - Dimensions and relationships disentangled and tested.* ICS-dissertation, Nijmegen.
136. Jan Kornelis Dijkstra (2007) *Status and affection among (pre)adolescents and their relation with antisocial and prosocial behavior.* ICS-dissertation, Groningen.
137. Wouter van Gils (2007). *Full-time working couples in the Netherlands. Causes and consequences.* ICS-dissertation, Nijmegen.
138. Djamila Schans (2007). *Ethnic diversity in intergenerational solidarity.* ICS-dissertation, Utrecht.
139. Ruud van der Meulen (2007). *Brug over woelig water: Lidmaatschap van sportverenigingen, vriendschappen, kennissenkringen en veralgemeend vertrouwen.* ICS-dissertation, Nijmegen.
140. Andrea Knecht (2008). *Friendship selection and friends' influence. Dynamics of networks and actor attributes in early adolescence.* ICS-dissertation, Utrecht.

141. Ingrid Doorten (2008). *The division of unpaid work in the household: A stubborn pattern?*. ICS-dissertation, Utrecht.
142. Stijn Ruiter (2008). *Association in context and association as context: Causes and consequences of voluntary association involvement.* ICS-dissertation, Nijmegen.
143. Janneke Joly (2008). *People on our minds: When humanized contexts activate social norms.* ICS-dissertation, Groningen.
144. Margreet Frieling (2008). *'Joint production' als motor voor actief burgerschap in de buurt.* ICS-dissertion, Groningen.
145. Ellen Verbakel (2008). *The partner as resource or restriction? Labour market careers of husbands and wives and the consequences for inequality between couples.* ICS-dissertation, Nijmegen.
146. Gijs van Houten (2008). *Beleidsuitvoering in gelaagde stelsels. De doorwerking van aanbevelingen van de Stichting van de Arbeid in het CAO-overleg.* ICS-dissertation, Utrecht.
147. Eva Jaspers (2008). *Intolerance over time. Macro and micro level questions on attitudes towards euthanasia, homosexuality and ethnic minorities.* ICS-dissertation, Nijmegen.
148. Gijs Weijters (2008). *Youth delinquency in Dutch cities and schools: A multilevel approach.* ICS-dissertation, Nijmegen.
149. Jessica Pass (2009). *The self in social rejection.* ICS-dissertation, Groningen.
150. Gerald Mollenhorst (2009). *Networks in contexts. How meeting opportunities affect personal relationships.* ICS-dissertation, Utrecht.
151. Tom van der Meer (2009). *States of freely associating citizens: Comparative studies into the impact of state institutions on social, civic and political participation.* ICS-dissertation, Nijmegen.
152. Manuela Vieth (2009). *Commitments and reciprocity in trust situations. Experimental studies on obligation, indignation, and self-consistency.* ICS-dissertation, Utrecht.
153. Rense Corten (2009). *Co-evolution of social networks and behavior in social dilemmas: Theoretical and empirical perspectives.* ICS-dissertation, Utrecht.
154. Arieke J. Rijken (2009). *Happy families, high fertility? Childbearing choices in the context of family and partner relationships.* ICS-dissertation, Utrecht.
155. Jochem Tolsma (2009). *Ethnic hostility among ethnic majority and minority groups in the Netherlands. An investigation into the impact of social mobility experiences, the local living environment and educational attainment on ethnic hostility.* ICS-dissertation, Nijmegen.
156. Freek Bucx (2009). *Linked lives: Young adults' life course and relations with parents.* ICS-dissertation, Utrecht.
157. Philip Wotschack (2009). *Household governance and time allocation. Four studies on the combination of work and care.* ICS-dissertation, Groningen.
158. Nienke Moor (2009). *Explaining worldwide religious diversity. The relationship between subsistence technologies and ideas about the unknown in pre-industrial and (post-)industrial societies.* ICS-dissertation, Nijmegen.
159. Lieke ten Brummelhuis (2009). *Family matters at work. Depleting and enriching effects of employees' family lives on work outcomes.* ICS-dissertation, Utrecht.
160. Renske Keizer (2010). *Remaining childless. Causes and consequences from a life course perspective.* ICS-dissertation, Utrecht.

161. Miranda Sentse (2010). *Bridging contexts: The interplay between family, child, and peers in explaining problem behavior in early adolescence.* ICS-dissertation, Groningen.
162. Nicole Tieben (2010). *Transitions, tracks and transformations. Social inequality in transitions into, through and out of secondary education in the Netherlands for cohorts born between 1914 and 1985.* ICS-dissertation, Nijmegen.
163. Birgit Pauksztat (2010). *Speaking up in organizations: Four studies on employee voice.* ICS-dissertation, Groningen.
164. Richard Zijdeman (2010). *Status attainment in the Netherlands, 1811-1941. Spatial and temporal variation before and during industrialization.* ICS-dissertation, Utrecht.
165. Rianne Kloosterman (2010). *Social background and children's educational careers. The primary and secondary effects of social background over transitions and over time in the Netherlands.* ICS-dissertation, Nijmegen.
166. Olav Aarts (2010). *Religious diversity and religious involvement. A study of religious markets in Western societies at the end of the twentieth century.* ICS-dissertation, Nijmegen.
167. Stephanie Wiesmann (2010). *24/7 negotiation in couples transition to parenthood.* ICS-dissertation, Utrecht.
168. Borja Martinovic (2010). *Interethnic contacts: A dynamic analysis of interaction between immigrants and natives in Western countries.* ICS-dissertation, Utrecht.
169. Anne Roeters (2010). *Family life under pressure? Parents' paid work and the quantity and quality of parent-child and family time.* ICS-dissertation, Utrecht.
170. Jelle Sijtsema (2010). *Adolescent aggressive behavior: Status and stimulation goals in relation to the peer context.* ICS-dissertation, Groningen.
171. Kees Keizer (2010). *The spreading of disorder.* ICS-dissertation, Groningen.
172. Michael Mäs (2010). *The diversity puzzle. Explaining clustering and polarization of opinions.* ICS-dissertation, Groningen.
173. Marie-Louise Damen (2010). *Cultuurdeelname en CKV. Studies naar effecten van kunsteducatie op de cultuurdeelname van leerlingen tijdens en na het voortgezet onderwijs.* ICS-dissertation, Utrecht.
174. Marieke van de Rakt (2011). *Two generations of crime: The intergenerational transmission of convictions over the life course.* ICS-dissertation, Nijmegen.
175. Willem Huijnk (2011). *Family life and ethnic attitudes. The role of the family for attitudes towards intermarriage and acculturation among minority and majority groups.* ICS-dissertation, Utrecht.
176. Tim Huijts (2011). *Social ties and health in Europe. Individual associations, cross-national variations, and contextual explanations.* ICS-dissertation, Nijmegen.
177. Wouter Steenbeek (2011). *Social and physical disorder. How community, business presence and entrepreneurs influence disorder in Dutch neighborhoods.* ICS-dissertation, Utrecht.
178. Miranda Vervoort (2011). *Living together apart? Ethnic concentration in the neighborhood and ethnic minorities' social contacts and language practices.* ICS-dissertation, Utrecht.
179. Agnieszka Kanas (2011). *The economic performance of immigrants. The role of human and social capital.* ICS-dissertation, Utrecht.
180. Lea Ellwardt (2011). *Gossip in organizations. A social network study.* ICS-dissertation, Groningen.

181. Annemarije Oosterwaal (2011). *The gap between decision and implementation. Decision making, delegation and compliance in governmental and organizational settings.* ICS-dissertation, Utrecht.
182. Natascha Notten (2011). *Parents and the media. Causes and consequences of parental media socialization.* ICS-dissertation, Nijmegen.
183. Tobias Stark (2011). *Integration in schools. A process perspective on students' interethnic attitudes and interpersonal relationships.* ICS-dissertation, Groningen.
184. Giedo Jansen (2011). *Social cleavages and political choices. Large-scale comparisons of social class, religion and voting behavior in Western democracies.* ICS-dissertation, Nijmegen.
185. Ruud van der Horst (2011). *Network effects on treatment results in a closed forensic psychiatric setting.* ICS-dissertation, Groningen.
186. Mark Levels (2011). *Abortion laws in European countries between 1960 and 2010. Legislative developments and their consequences for women's reproductive decision-making.* ICS-dissertation, Nijmegen.
187. Marieke van Londen (2012). *Exclusion of ethnic minorities in the Netherlands. The effects of individual and situational characteristics on opposition to ethnic policy and ethnically mixed neighbourhoods.* ICS-dissertation, Nijmegen.
188. Sigrid M. Mohnen (2012). *Neighborhood context and health: How neighborhood social capital affects individual health.* ICS-dissertation, Utrecht.
189. Asya Zhelyazkova (2012). *Compliance under controversy: Analysis of the transposition of European directives and their provisions.* ICS-dissertation, Utrecht.
190. Valeska Korff (2012). *Between cause and control: Management in a humanitarian organization.* ICS-dissertation, Groningen.
191. Maike Gieling (2012). *Dealing with diversity: Adolescents' support for civil liberties and immigrant rights.* ICS-dissertation, Utrecht.
192. Katya Ivanova (2012). *From parents to partners: The impact of family on romantic relationships in adolescence and emerging adulthood.* ICS-dissertation, Groningen.
193. Jelmer Schalk (2012). *The performance of public corporate actors: Essays on effects of institutional and network embeddedness in supranational, national, and local collaborative contexts.* ICS-dissertation, Utrecht.
194. Alona Labun (2012). *Social networks and informal power in organizations.* ICS-dissertation, Groningen.
195. Michał Bojanowski (2012). *Essays on social network formation in heterogeneous populations: Models, methods, and empirical analyses.* ICS-dissertation, Utrecht.
196. Anca Minescu (2012). *Relative group position and intergroup attitudes in Russia.* ICS-dissertation, Utrecht.
197. Marieke van Schellen (2012). *Marriage and crime over the life course. The criminal careers of convicts and their spouses.* ICS-dissertation, Utrecht.
198. Mieke Maliepaard (2012). *Religious trends and social integration: Muslim minorities in the Netherlands.* ICS-dissertation, Utrecht.
199. Fransje Smits (2012). *Turks and Moroccans in the Low Countries around the year 2000: determinants of religiosity, trend in religiosity and determinants of the trend.* ICS-dissertation, Nijmegen.
200. Roderick Sluiter (2012). *The diffusion of morality policies among Western European countries between 1960 and 2010. A comparison of temporal and spatial diffusion patterns of six morality and eleven non-morality policies.* ICS-dissertation, Nijmegen.

201. Nicoletta Balbo (2012). *Family, friends and fertility.* ICS-dissertation, Groningen.
202. Anke Munniksma (2013). *Crossing ethnic boundaries: Parental resistance to and consequences of adolescents' cross-ethnic peer relations.* ICS-dissertation, Groningen.
203. Anja Abendroth (2013). *Working women in Europe. How the country, workplace, and family context matter.* ICS-dissertation, Utrecht.
204. Katia Begall (2013). *Occupational hazard? The relationship between working conditions and fertility.* ICS-dissertation, Groningen.
205. Hidde Bekhuis (2013). *The popularity of domestic cultural products: Cross-national differences and the relation to globalization.* ICS-dissertation, Utrecht.
206. Lieselotte Blommaert (2013). *Are Joris and Renske more employable than Rashid and Samira? A study on the prevalence and sources of ethnic discrimination in recruitment in the Netherlands using experimental and survey data.* ICS-dissertation, Utrecht.
207. Wiebke Schulz (2013). *Careers of men and women in the 19th and 20th centuries.* ICS-dissertation, Utrecht.
208. Ozan Aksoy (2013). *Essays on social preferences and beliefs in non-embedded social dilemmas.* ICS-dissertation, Utrecht.
209. Dominik Morbitzer (2013). *Limited farsightedness in network formation.* ICS-dissertation, Utrecht.
210. Thomas de Vroome (2013). *Earning your place: The relation between immigrants' economic and psychological integration in the Netherlands.* ICS-dissertation, Utrecht.
211. Marloes de Lange (2013). *Causes and consequences of employment flexibility among young people. Recent developments in the Netherlands and Europe.* ICS-dissertation, Nijmegen.
212. Roza Meuleman (2014). *Consuming the nation. Domestic cultural consumption: Its stratification and relation with nationalist attitudes.* ICS-dissertation, Utrecht.
213. Esther Havekes (2014). *Putting interethnic attitudes in context. The relationship between neighbourhood characteristics, interethnic attitudes and residential behaviour.* ICS-dissertation, Utrecht.
214. Zoltán Lippényi (2014). *Transitions toward an open society? Intergenerational occupational mobility in Hungary in the 19th and 20th centuries.* ICS-dissertation, Utrecht
215. Anouk Smeekes (2014). *The presence of the past: Historical rooting of national identity and current group dynamics.* ICS-dissertation, Utrecht.
216. Michael Savelkoul (2014). *Ethnic diversity and social capital. Testing underlying explanations derived from conflict and contact theories in Europe and the United States.* ICS-dissertation, Nijmegen.
217. Martijn Hogerbrugge (2014). *Misfortune and family: How negative events, family ties, and lives are linked.* ICS-dissertation, Utrecht.
218. Gina Potarca (2014). *Modern love. Comparative insights in online dating preferences and assortative mating.* ICS-dissertation, Groningen.
219. Mariska van der Horst (2014). *Gender, aspirations, and achievements: Relating work and family aspirations to occupational outcomes.* ICS-dissertation, Utrecht.
220. Gijs Huitsing (2014). *A social network perspective on bullying.* ICS dissertation, Groningen.
221. Thomas Kowalewski (2015). *Personal growth in organizational contexts.* ICS-dissertation, Groningen.

222. Manu Muñoz-Herrera (2015). *The impact of individual differences on network relations: Social exclusion and inequality in productive exchange and coordination games.* ICS-dissertation, Groningen.
223. Tim Immerzeel (2015). *Voting for a change. The democratic lure of populist radical right parties in voting behavior.* ICS-dissertation, Utrecht.
224. Fernando Nieto Morales (2015). *The control imperative: Studies on reorganization in the public and private sectors.* ICS-dissertation, Groningen.
225. Jellie Sierksma (2015). *Bounded helping: How morality and intergroup relations shape children's reasoning about helping.* ICS-dissertation, Utrecht.
226. Tinka Veldhuis (2015). *Captivated by fear. An evaluation of terrorism detention policy.* ICS-dissertation, Groningen.
227. Miranda Visser (2015). *Loyality in humanity. Turnover among expatriate humanitarian aid workers.* ICS-dissertation, Groningen.
228. Sarah Westphal (2015). *Are the kids alright? Essays on postdivorce residence arrangements and children's well-being.* ICS-dissertation, Utrecht.
229. Britta Rüschoff (2015). *Peers in careers: Peer relationships in the transition from school to work.* ICS-dissertation, Groningen.
230. Nynke van Miltenburg (2015). *Cooperation under peer sanctioning institutions: Collective decisions, noise, and endogenous implementation.* ICS-dissertation, Utrecht.
231. Antonie Knigge (2015). *Sources of sibling similarity. Status attainment in the Netherlands during modernization.* ICS-dissertation, Utrecht.
232. Sanne Smith (2015). *Ethnic segregation in friendship networks. Studies of its determinants in English, German, Dutch, and Swedish school classes.* ICS-dissertation, Utrecht.
233. Patrick Präg (2015). *Social stratification and health. Four essays on social determinants of health and wellbeing.* ICS-dissertation, Groningen.
234. Wike Been (2015). *European top managers' support for work-life arrangements.* ICS-dissertation, Utrecht.
235. André Grow (2016). *Status differentiation: New insights from agent-based modeling and social network analysis.* ICS-dissertation, Groningen.
236. Jesper Rözer (2016). *Family and personal networks. How a partner and children affect social relationships.* ICS-dissertation, Utrecht.
237. Kim Pattiselanno (2016). *At your own risk: The importance of group dynamics and peer processes in adolescent peer groups for adolescents' involvement in risk behaviors.* ICS- dissertation, Groningen.
238. Vincenz Frey (2016). *Network formation and trust.* ICS-dissertation, Utrecht.