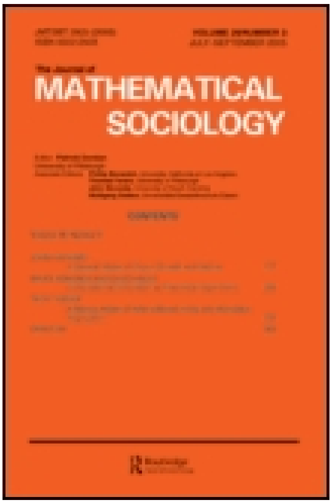


The Journal of Mathematical Sociology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmas20

# Embedding Trust: A Game-Theoretic Model for Investments in and Returns on Network Embeddedness

Vincenz Frey[a], Vincent Buskens[a] & Werner Raub[a]

[a] Department of Sociology/ICS, Utrecht University, Utrecht, The Netherlands

Published online: 20 Jan 2015.

*Journal of Mathematical Sociology*, 39: 39–72, 2015

ISSN: 0022-250X print/1545-5874 online
DOI: 10.1080/0022250X.2014.897947



# EMBEDDING TRUST: A GAME-THEORETIC MODEL FOR INVESTMENTS IN AND RETURNS ON NETWORK EMBEDDEDNESS

**Vincenz Frey, Vincent Buskens, and Werner Raub**
*Department of Sociology/ICS, Utrecht University, Utrecht, The Netherlands*

***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 article, 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 the information exchange relation is high particularly for trust problems that are neither too small nor too severe.***

***Keywords***: embeddedness, game theory, network formation, reputation, social dilemmas

## 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 article, 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

Address correspondence to Vincenz Frey, Department of Sociology/ICS, Utrecht University, Padualaan 14, Utrecht, 3584 CH, The Netherlands. E-mail: v.c.frey@uu.nl

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 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 (that is, 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, 1990), which we will introduce in Section 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, Raub, & Van der Veer; 2010; Coleman, 1990; DiMaggio & Louch, 1998; Huck, Lünser, & Tyran; 2010; Raub & Weesie, 1990; for a survey, see Buskens & Raub, 2013). Buskens and 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 article, 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—at costs—to establish a relation between one another. 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 building on the theory of reputation building in a sequential equilibrium of a finitely repeated game with incomplete information that was pioneered by Kreps, Milgrom, Roberts, and Wilson (1982) and Kreps and 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 (Raub & Weesie, 1990; Eguíluz, Zimmermann, Cela-Conde, & San Miguel, 2005; Vega-Redondo, 2006) or assume boundedly rational, backward-looking actors and consider only learning effects (Macy & Flache, 2002; Nowak & Sigmund, 2005; Roca, Sánchez, & Cuesta, 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 (Skyrms & Pemantle, 2000; Pujol, Flache, Delgado, & Sangüesa, 2005). 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, that is, the trustors' willingness to invest, varies in a nonmonotonic 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 article is organized as follows. In Section 2, we present the model in detail. In Section 3, we analyze the model and present our results. Finally, in Section 4, we conclude and point out directions for future research. Sections A.1 through A.7 provide additional results and the proofs.

## 2. THE MODEL

Before we introduce our model, we want to introduce its main building block, namely, the trust game (TG; Dasgupta, 1988).[1] The TG has two players—a trustor

[1]Throughout the article, we use standard game theory terminology and assumptions. See, e.g., Fudenberg and Tirole (2000) for a textbook.

(*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 compared with 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 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.

### 2.1. One-Shot Trust Games With Complete and Incomplete Information

In the standard trust game with complete information (Dasgupta, 1988; Kreps, 1990), 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.[2]

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 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 $\pi$ 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 $\pi$ 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 $\pi$. In Figure 1, the trustor does not know whether she has to move at the left or the right node. This is indicated by the dashed line that includes these

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

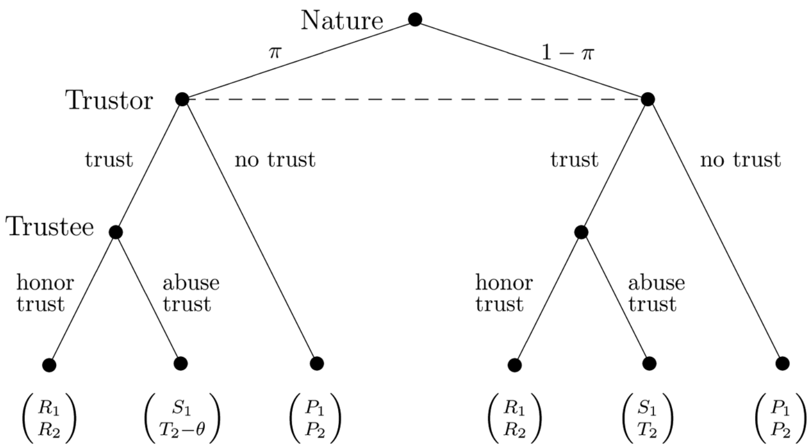


**FIGURE 1** Extensive form of a trust game with incomplete information. $S_1 < P_1 < R_1$, $P_2 < R_2 < T_2$, and $T_2 - \theta < R_2$.



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 $RISK := (P_1 - S_1)/(R_1 - S_1)$.

### 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 $\Gamma$ that we study, we 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.

**2.2.1. The Structure of $\Gamma$.** $\Gamma$ 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 $\pi$ and $1-\pi$, respectively. The probability $\pi$ 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,..., 2$N$. 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.[3]

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 outcomes of the previously played TGs in which the other trustor participated. 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.

**2.2.2. Further Assumptions on $\Gamma$ and an Illustrating Example.** We assume that the structure of $\Gamma$ is common knowledge and that $\Gamma$ 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 $\Gamma$ 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

[3]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.

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

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

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

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



## 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 3.1, 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 Subsections 3.2 and 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 3.4, 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 3.5 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.

### 3.1. 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, Garber, & Watson, 1997; Buskens, 2003), who apply the analysis of reputation building in sequential equilibrium pioneered by Kreps et al. (1982) and Kreps and 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, 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 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 $\Gamma$ 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 Camerer and Weigelt (1988), Bower et al. (1997), Anderhub, Engelmann, and Güth (2002) and Buskens (2003). We refer the reader to these studies as well as to Fudenberg and 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*) $\Gamma^+$ and (*continuation game*) $\Gamma^-$, 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 $\pi_n^i$ 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 $\pi_n^i$ also is the trustee's reputation; it indicates what type he is thought to be. Finally, similar to $RISK := \frac{P_1 - S_1}{R_1 - S_1}$, we define $TEMP := \frac{T_2 - R_2}{T_2 - P_2}$ as a second measure pertaining to the payoffs of the stage game. While *RISK* measures the risk a trustor incurs when placing trust (Section 2.1), *TEMP* measures an opportunistic trustee's temptation to abuse trust (see Buskens, 2002; Snijders, 1996).

### 3.2. Analysis of $\Gamma^-$

Our first theorem specifies the unique sequential equilibrium of $\Gamma^-$, that is, in periods 1 to $2N$ of $\Gamma$ after the trustors have not established the information exchange relation.[4] In $\Gamma^-$, 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.

**Theorem 1.** *The beliefs and strategies specified below constitute the unique sequential equilibrium of* $\Gamma^-$.

- *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* $\pi_n^i \geq RISK^{\lfloor \frac{2N-n+1}{2} \rfloor}$*, then* $h_n = 1$.
  - *If* $\pi_n^i < RISK^{\lfloor \frac{2N-n+1}{2} \rfloor}$*, then* $h_n = \frac{\pi_n^i}{1-\pi_n^i}\left(\frac{1}{RISK^{\lfloor \frac{2N-n+1}{2} \rfloor}} - 1\right)$.

Sections A.1 through A.7 provide the proofs of our theorems.

We describe the course of behavior and beliefs in the equilibrium defined in Theorem 1 focusing on the interactions between some trustor $i$ and the trustee. In $\Gamma^-$, 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; Camerer and 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.

[4] In the game $\Gamma$, 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 $\Gamma^+$ or $\Gamma^-$ is also a perfect Bayesian equilibrium of the respective continuation game (see Fudenberg & Tirole, 2000, Theorem 8.2).

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 $\pi$). 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.[5] Thereafter, the third phase starts: trustor $i$ does not place trust anymore.

We let $\tau$ 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 Theorem 1 that the integer $\tau$ is such that $\pi$ lies in the interval $[RISK^{\tau}, RISK^{\tau-1})$. This implies that

$$\tau = \left\lceil \frac{\log \pi}{\log RISK} \right\rceil. \tag{1}$$

It can be seen from Eq. (1) that $\tau$ increases stepwise in $RISK$ and decreases stepwise in $\pi$. That is, the second phase tends to start earlier if the risk associated with placing trust is higher and if the probability of playing with a friendly trustee is smaller.

Note furthermore that $\tau$ is independent of $N$. Hence, if $\tau$ 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.

Theorem 2 specifies a trustor's expected payoff associated with the unique sequential equilibrium of $\Gamma^-$.

**Theorem 2.** *The expected payoff for a trustor in $\Gamma^-$ is*

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

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 $\Gamma^-$ (and in $\Gamma^+$) is crucial; we therefore want to briefly sketch the intuition behind Theorem 2. If $\tau > N$, a trustor never places trust and receives

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

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 $R_1(\pi+(1-\pi)h_{N-\tau+1})+S_1(1-\pi)(1-h_{N-\tau+1})$. 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 $\Gamma^-$ depends essentially on $\pi$ and $RISK$ that together determine $\tau$. $N$ and $\tau$ 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 $\tau$, 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 $R_1-P_1$.[6]

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 $\tau(R_1-P_1)$, that is, 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 $\tau$ is larger (because $RISK$ is larger or $\pi$ smaller).

### 3.3. Analysis of $\Gamma^+$

An opportunistic trustee has a stronger incentive to build and maintain a reputation for being trustworthy in $\Gamma^+$ than in $\Gamma^-$. In $\Gamma^+$, 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 $\Gamma^+$ 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 $\Gamma^+$ than in $\Gamma^-$, whereas, obviously, the short-term incentive to abuse trust is the same in both continuation games. Our third theorem specifies the unique sequential equilibrium that results in the interactions between the trustors and the trustee in $\Gamma^+$.

[6]Note that $\pi$ 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.

**Theorem 3.** *The beliefs and strategies specified below constitute the unique sequential equilibrium of* $\Gamma^+$.

- *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 $\Gamma^+$ specified in Theorem 3, the trustee and the trustor at play in a given TG behave as if they played all $2N$ TGs together, that is, as if there was only one trustor playing $2N$ TGs with the trustee. The sequential equilibrium of $\Gamma^+$ evolves over the same three phases as the sequential equilibrium of the interactions between each trustor $i$ and the trustee in $\Gamma^-$. First, both trustors place trust and the trustee honors trust with probability 1 until and including the TG after which there are *in total* $\tau$ TGs left to be played, that is, until and including period $2N-\tau$, where $\tau$ is determined by $RISK$ and $\pi$ as specified in Eq. (1). 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.

Theorem 4 specifies a trustor's expected payoff associated with this equilibrium of the continuation game $\Gamma^+$.

**Theorem 4.** *In* $\Gamma^+$*, the expected payoff for a trustor is*

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

To understand how a trustor's expected payoff for $\Gamma^+$ 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.

### 3.4. 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 $\Gamma^+$ and $\Gamma^-$ and can be calculated as $r_1 = U_1^{\Gamma^+} - U_1^{\Gamma^-}$, where $r_1$ denotes a trustor's expected return on investment.[7] $r_1$ 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. Theorem 5 specifies $r_1$ and establishes that $\Gamma$ always has an investment equilibrium if the cost of investment per trustor is smaller than $r_1$.

**Theorem 5.** *In $\Gamma$, 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 ($r_1$), that is, iff $C/2 \leq r_1$, where $r_1$ falls in the following intervals:*

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

Theorem 5 distinguishes three scenarios. If $\tau \leq N$, there is an equilibrium phase in which trust is placed and honored with certainty in $\Gamma^+$ as well as $\Gamma^-$ but this phase is longer in $\Gamma^+$ because an opportunistic trustee remains trustworthy in $\Gamma^+$ until he has only half as many TGs left with each trustor compared to the situation in $\Gamma^-$.[8] If $N < \tau \leq 2N$, trust is placed with certainty in at least the first TG in $\Gamma^+$ but the trustors never place trust in $\Gamma^-$ because, given $\pi$ 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 $\pi$ or $N$ is very small or *RISK* very large such that $\tau > 2N$, trust is not even possible in $\Gamma^+$.

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

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

In the third scenario with trust being not even possible in $\Gamma^+$, 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, Theorem 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 $\Gamma^+$ but not in $\Gamma^-$ 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 $\Gamma^+$ but only until she has $\tau$ TGs left in $\Gamma^-$. If $N < \tau \leq 2N$, implying that trust is only possible in $\Gamma^+$, the number of *additional* TGs in which a trustor would profit from trust being placed and honored with certainty if the relation gets established simply equals the number of TGs in which she would expectedly do so in $\Gamma^+$ (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 Theorem 5 show that, for a given $N$ and $\tau$, $r_1$ 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 Theorem 5. Note, however, that the specification in Theorem 5 and the approximation of $r_1$ are sufficient to derive the main comparative statics.

Theorem 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 only gets 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.

### 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 $\Gamma$ 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 $\pi$, $S_1$, $P_1$, $R_1$, 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 $\pi$, $S_1$, $P_1$, 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 $\tau$ (an earlier start of the randomization). We thus, for example, establish how $r_1$ changes if $\pi$ *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 with 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 $\pi$, $S_1$, $P_1$, or $R_1$ triggers a change in $\tau$, that is, 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 $\pi$, $S_1$, $P_1$, or $R_1$ that does *not* trigger a change in $\tau$ 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 $\pi$, $S_1$, $P_1$, or $R_1$ irrespective of whether this change triggers a change in $\tau$ 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 $\tau$ and do not consider the effects of parameter changes that affect $X_1$ but not $\tau$.

**3.5.1. Changes in $\pi$.** Theorem 6 establishes how $r_1$ changes as the probability that the trustee is of the friendly type ($\pi$) decreases.

**Theorem 6.** *Given the specification of $\Gamma$ and the definitions of $\tau$ and $r_1$, it holds that:*

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

Theorem 6 states that as $\pi$ decreases the value of the relation first increases and then decreases again. To understand this result, recall that if $\pi$ is smaller, $\tau$ tends to be larger, that is, 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 $\Gamma^+$ from trust being placed and honored with certainty until she has $\tau/2$ TGs left, whereas she would do so in $\Gamma^-$ only until she has $\tau$ TGs left. Hence, if $\pi$ decreases such that $\tau$ increases and $\pi$ is large enough also after the decrease such that, given *RISK* and $N$, trust is possible also in $\Gamma^-$, 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 $\pi$ on $r_1$ is opposite if trust is not possible in $\Gamma^-$, both before and after the decrease in $\pi$, but possible in $\Gamma^+$, at least before the decrease in $\pi$. In this case, a decrease in $\pi$ that leads to an earlier start of randomization (an increase in $\tau$) leads to a decrease in the number of TGs in which a trustor could benefit from trust being placed and honored with certainty in $\Gamma^+$ while leaving $U_1^{\Gamma^-}$ unchanged at $NP_1$, and, consequently, it leads to a decrease in $r_1$. Note further that if $\pi$ is so small that, given *RISK* and $N$, trust is not even possible in $\Gamma^+$, $r_1 = 0$ irrespectively the precise $\pi$.

Figure 2 visualizes how $r_1$ depends on $\pi$ in an example with $N = 3$ and $RISK = 0.5$. It provides a more detailed picture than Theorem 6, illustrating also

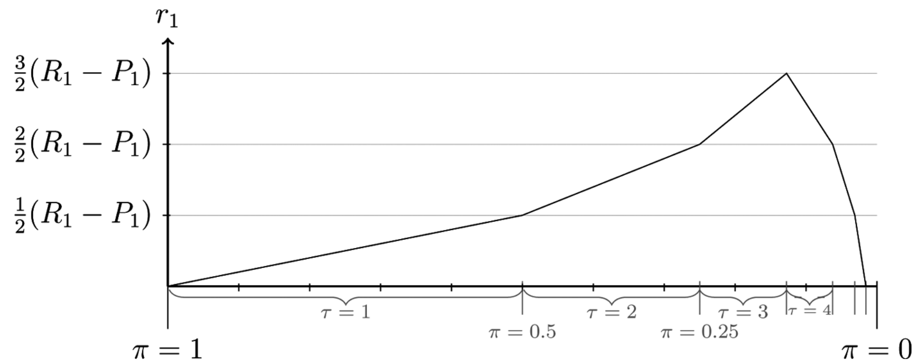


**FIGURE 2** The effect of changes in $\pi$ on $r_1$ in an example with $RISK = 0.5$ and $N = 3$.

the additional results established in Lemma 1 in Section A.3. Figure 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 would be placed and honored with certainty in half of the $2N$ TGs in $\Gamma^+$ while in $\Gamma^-$ the trustee's randomization would start in the first TG of each trustor and $\pi$ is just large enough such that the trustee does not want to start randomizing a TG earlier (in Figure 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.[9]

### 3.5.2. Changes in $S_1$.

Theorem 7 establishes how $r_1$ changes as the payoff a trustor receives if the trustee abuses her trust ($S_1$) decreases. If $S_1$ is smaller, *RISK* is larger and, consequently, $\tau$ tends to be larger, that is, randomization tends to start earlier.

**Theorem 7.** *Given the specification of* $\Gamma$ *and the definitions of* $\tau$ *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.*

Theorem 7 shows that the potential return on investment increases as $S_1$ decreases as long as trust remains possible also in $\Gamma^-$. As $S_1$ decreases further and trust is only possible in $\Gamma^+$, the potential return on investment decreases. It is also clear that as $S_1$ gets so small that trust is not even possible in $\Gamma^+$, $r_1 = 0$. Thus, $r_1$ changes in the same manner if $S_1$ decreases as it changes if $\pi$ decreases. This is not surprising. Neither a decrease in $S_1$ nor a decrease in $\pi$ 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 could benefit from trust being placed and honored with certainty if the relation gets established. We provide additional details on how $r_1$

[9] It can be seen from Figure 2 that the increase towards the maximum and the decrease thereafter is not linear even though $r_1$ depends linearly on $\pi$ for changes in $\pi$ that do not affect $\tau$ because the range of $\pi$ for which $\tau$ is constant is smaller, the smaller $\pi$.

depends on $S_1$ in Section A.4. and we note that a figure could be drawn for the dependence of $r_1$ on $S_1$ that is similar to Figure 2.

**3.5.3. Changes in $P_1$.** Theorem 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 $\tau$ tends to be larger, i.e., randomization tends to start earlier.

**Theorem 8.** *Given the specification of $\Gamma$ and the definitions of $\tau$ 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.*

Theorem 8 shows the maximum cost for which $\Gamma$ has an investment equilibrium changes in the same direction due to an increase in $P_1$ as due to a decrease in $\pi$ 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 $\Gamma^+$. 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 $\Gamma^-$ while (at least before the increase in $P_1$) trust is possible in $\Gamma^+$, these two effects have the same direction: both contribute to a decrease in $r_1$. However, if trust is also possible in $\Gamma^-$ 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).

**3.5.4. Changes in $R_1$.** Theorem 9 establishes how $r_1$ changes as the payoff a trustor gets if she places trust and the trustee honors trust ($R_1$) decreases. If $R_1$ is smaller, $\tau$ tends to be larger, i.e., randomization tends to start earlier.

**Theorem 9.** *Given the specification of $\Gamma$ and the definitions of $\tau$ 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 $\tau$ as well as a decrease in the probability that an opportunistic trustee honors trust in the interaction in which he starts to randomize. Theorem 9 shows that, as long as trust is possible in at least the first TG of $\Gamma^+$, the total of these effects is such that $r_1$ decreases as $R_1$ decreases.

**3.5.5. Changes in $N$.** Finally, Theorem 10 specifies how a change in the number of repetitions ($N$) affects $r_1$.

**Theorem 10.** *Given the specification of $\Gamma$ and the definitions of $\tau$ 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.*

Theorem 10 establishes that $r_1$ increases due to an increase in $N$ if trust is possible in $\Gamma^+$ but not in $\Gamma^-$ before and after the increase in $N$, whereas $r_1$ does not change if trust is also not possible in $\Gamma^+$ after the increase in $N$ or if trust is possible also in $\Gamma^-$ 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 $\Gamma^-$ but $R_1$ in $\Gamma^+$. In the latter case, a trustor's expected payoff for $\Gamma^+$ and $\Gamma^-$ 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 $N$.

**3.5.6. Summarizing.** Theorems 6 and 7 show that as $\pi$ 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 3.2 as $\tau(R_1 - P_1)$. Recall that if $\pi$ or $S_1$ decreases, randomization tends to start earlier (that is, $\tau$ tends to increase) and, hence, the trust problem becomes larger. Thus, Theorems 6 and 7 show that the maximum cost of investment per trustor for which $\Gamma$ has an investment equilibrium varies in a non-monotonic way if the size of the trust problem increases due to a decrease in $\pi$ 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 (Theorems 8 and 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 $\tau$ 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 $\pi$ or $S_1$ decreases or $P_1$ increases becomes small. In the example shown in Figure 2 with $RISK = 0.5$ and $N = 3$, the trustors' willingness to invest ($r_1$) increases if the proportion of friendly trustees ($\pi$) decreases from close to 100% to 12.5% ($0.5^3$) and it decreases if $\pi$ decreases from 12.5% to 1.6% ($0.5^6$). For $N = 4$, $r_1$ would decrease only if $\pi$ 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 $\pi$ 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.[10]

## 4. CONCLUSION 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

[10]We owe this remark to an anonymous reviewer.

"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 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 some 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, Cook, & Watabe, 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, Buskens, and Frey (2013; see also Raub, Frey, & Buskens, 2014), 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). The network does, however, 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 the current article, 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 article 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 $r_1 \geq C/2$).[11] 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 ($\tau$) 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 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"

[11]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.

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 (Schelling, 1960; Raub, 2004), 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.

## REFERENCES

Anderhub, V., Engelmann, D., & Güth, W. (2002). An experimental study of the repeated trust game with incomplete information. *Journal of Economic Behavior and Organization*, *48*, 197–216.

Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.

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*, 165–190.

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*, 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*, 301–312.

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*, 1–36.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Belknap Press of Harvard University Press.

Dasgupta, P. (1988). Trust as a commodity. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 49–72). Oxford, UK: Blackwell.

DiMaggio, P., & Louch, H. (1998). Socially embedded consumer transactions: For what kinds of purchases do people most often use networks? *American Sociological Review*, *63*, 619–637.

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*, 977–1008.

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, UK: Routledge.

Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *54*, 533–554.

Fudenberg, D., & Tirole, J. (2000). *Game theory* (7th ed.). Cambridge, MA: MIT Press.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, *78*, 1360–1380.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*, 481–510.

Huck, S., Lünser, G. K., & Tyran, J. R. (2010). Consumer networks and firm reputation: A first experimental investigation. *Economics Letters*, *108*, 242–244.

Ioannides, Y. M., & Loury, L. D. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, *42*, 1056–1093.

Jackson, M. O. (2008). *Social and economic networks*. Princeton, NJ: Princeton University Press.

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, NY: Russell Sage.

Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology*, *100*, 313–345.

Kreps, D. M. (1990). *Game theory and economic modelling*. Oxford, UK: 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*, 245–252.

Kreps, D. M., & Wilson, R. (1982a). Reputation and imperfect information. *Journal of Economic Theory*, *27*, 253–279.

Kreps, D. M., & Wilson, R. (1982b). Sequential equilibria. *Econometrica*, *50*, 863–894.

Lin, N. (2002). *Social capital: A theory of social structure and action*. Cambridge, UK: Cambridge University Press.

Macy, M. W., & Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences*, *99*(Suppl. 3), 7229–7236.

Nowak, M., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298.

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). Retrieved from http://jasss.soc.surrey.ac.uk/8/4/12.html

Raub, W. (2004). Hostage posting as a mechanism of trust: Binding, compensation, and signaling. *Rationality and Society*, *16*, 319–365.

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*, 720–732.

Raub, W., Frey, V., & Buskens, V. (2014). Strategic network formation, games on networks, and trust. *Analyse & Kritik*, *36*, 135–152.

Raub, W., & Weesie, J. (1990). Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, *96*, 626–654.

Roca, C. P., Sánchez, A., & Cuesta, J. A. (2012). Individual strategy update and emergence of cooperation in social networks. *Journal of Mathematical Sociology*, *36*, 1–21.
Schelling, T. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
Simpson, B., & McGrimmon, T. (2008). Trust and embedded markets: A multi-method investigation of consumer transactions. *Social Networks*, *30*, 1–15.
Skyrms, B., & Pemantle, R. (2000). A dynamic model of social network formation. *Proceedings of the National Academy of Sciences*, *97*, 9340–9346.
Snijders, C. (1996). *Trust and commitments*. Amsterdam, The Netherlands: Thela Thesis.
Uzzi, B. (1996). The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review*, *61*, 674–698.
Vega-Redondo, F. (2006). Building up social capital in a changing world. *Journal of Economic Dynamics and Control*, *30*, 2305–2338.
Yamagishi, T., Cook, K. S., & Watabe, M. (1998). Uncertainty, trust, and commitment formation in the United States and Japan. *American Journal of Sociology*, *104*, 165–194.



## APPENDIX: PROOFS AND ADDITIONAL RESULTS

This appendix provides the proofs for our Theorems and is structured as follows. We begin with a sketch of the proofs of Theorems 1 to 4 on the sequential equilibria of the continuation games $\Gamma^-$ and $\Gamma^+$ 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 A.1). The proof of the condition for the existence of an investment equilibrium (Theorem 5) is found in Section A.2, and the proofs of the comparative statics results for changes in $\pi$, $S_1$, $P_1$, and $R_1$ are presented in Sections A.3 to A.6. The latter proofs all proceed over the same three steps that we explain in Section A.3 for the effect of changes in $\pi$. Sections A.3 and A.4 additionally provide Lemmas 1 and 2, which, respectively, imply Theorem 6 and Theorem 7 and establish in more detail how $r_1$ depends on $\pi$ and $S_1$. Finally, the last section provides the proof of the effect of changes in $N$ (Theorem 10).

### A.1. Sketch of the Proof of Theorems 1 to 4: Equilibria and Payoffs in $\Gamma^-$ and $\Gamma^+$

In the analysis of $\Gamma^-$ and $\Gamma^+$, we restrict the focus to equilibria that satisfy sequential rationality. In $\Gamma^-$ (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 Theorem 1 (leading to the rounding in the exponents) is that, in $\Gamma^-$, 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 $\Gamma^+$ (where each trustor receives information also about the TGs of the other trustor), the sequential equilibrium as specified in Theorem 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 and 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 theorems.[12] 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 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 and Maskin, 1986). The formulas for the calculation of the expected payoff of a trustor in $\Gamma^-$and $\Gamma^+$ as specified in Theorems 2 and 4, respectively, are implied by the sequential equilibria of these continuation games and their derivation is described in the main text.

### A.2. Proof of Theorem 5: $r_1$ and the Existence of Investment Equilibria

As stated in Section 3.4, a trustor's potential return on investment can be calculated as $r_1 = U_1^{\Gamma^+} - U_1^{\Gamma^-}$. This calculation yields

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

This precise specification of $r_1$ (which we use in the rest of our proofs) implies the intervals for $r_1$ that Theorem 5 provides. Specifically, because $P_1 \leq S_1 + \frac{\pi}{RISK^{\tau-1}}(R_1 - S_1) < R_1$ (that is, 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. (2) 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.

[12]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.

### A.3. Additional Results for Changes in $\pi$ and Proof of Theorem 6

Lemma 1 provides additional details on how $r_1$ changes in $\pi$ and implies Theorem 6. Lemma 1 quantifies the effects of a change in $\pi$ and establishes that if trust is possible in $\Gamma^+$ as well as in $\Gamma^-$ (possible in $\Gamma^+$ but not in $\Gamma^-$), $r_1$ increases (decreases) in a stepwise linear manner as $\pi$ decreases.

**Lemma 1.** *Given the specification of $\Gamma$ and the definitions of $\tau$ and $r_1$, it holds that:*

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

**Procedure for the Proof of the Comparative Statics Results.** To prove Lemma 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 A1. As Figure 2, Figure A1 shows how $r_1$ depends on $\pi$ 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 randomization starts. That is, we fix a situation in which $\pi$ and

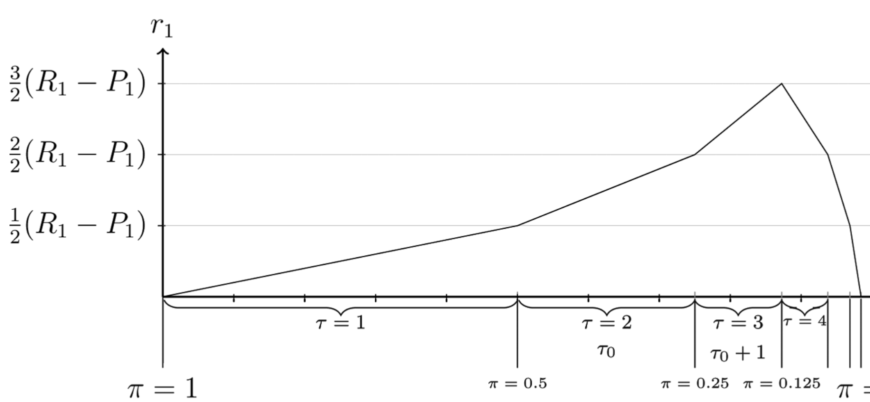


**FIGURE A1** Illustration for the visualization of the procedure used to prove the effect of changes in $\pi$ on $r_1$ (example with $RISK = 0.5$ and $N = 3$).

*RISK* are such that for a given $\tau$, to which we refer as $\tau_0$, $\pi$ is precisely at the right-hand border of the interval $(RISK^{\tau_0 - 1}, RISK^{\tau_0}]$, the "interval $\tau_0$." Figure A1 presents the example in which we fix the interval $(\pi = 0.5,\ \pi = 0.25]$, where $\tau = 2$, as the interval $\tau_0$.

In step 1, we show that $r_1$ changes as postulated if the parameter under study changes such that $\tau$ 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 $\pi$ is at the right border of the interval $\tau_0$ before the parameter change, whereas after the para meter change, $\pi$ is at the right border of the adjacent "$\tau$ interval" on the right, that is, the $\tau$ 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 $\pi$ is precisely at the right border of the interval $\tau_0$ 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 A1, we thus show in step 1 that $r_1$ increases by $\frac{1}{2}(R_1 - P_1)$ if $\pi$ changes from 0.25 to 0.125. More generally, as we do not actually fix a specific $\tau_0$, step 1 in the proof of the effects of changes in $\pi$ shows that for any $\tau_0$, $r_1^{\tau_0]} = r_1^{\tau_0+1]} - \frac{1}{2}(R_1 - P_1)$ 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 $\tau_0$. 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 $\pi$ is in the interval $\tau_0 + 1$) changes if $\pi$ increases within the interval $\tau_0 + 1$ or if $R_1$, $P_1$, or $S_1$ changes such that *RISK* decreases and, therefore, the $\tau$ intervals shift to the right. Specifically for changes in $\pi$, we show that an increase in $\pi$ 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 $\tau_0$ 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 A1, we show in step 3 that if $\pi$ 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 1.** *Step 1.* If $\pi = RISK^{\tau_0}$, the payoff a trustor can expect in the TG in which the trustee begins to randomize $\left(X_1 = S_1 + \frac{\pi}{RISK^{\tau_0 - 1}}(R_1 - S_1)\right)$ equals $P_1$. Hence, for the case that $\pi$ is at the right border of the interval $\tau_0$, $r_1$ (as specified in Eq. (2)) 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 $\pi$ 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 $\pi$, neglecting that $\tau$ is a function of $\pi$, is

$$\frac{\partial r_1}{\partial \pi} = -\frac{1}{2}\frac{R-S}{RISK^{\tau-1}} \quad \text{if } \tau \leq N,$$
$$\frac{\partial r_1}{\partial \pi} = \frac{1}{2}\frac{R-S}{RISK^{\tau-1}} \quad \text{if } N < \tau \leq 2N.$$

This shows that a marginal increase in $\pi$ by $\Delta\pi > 0$ that does not affect $\tau$ 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 $\pi$ increases within the interval $\tau_0 + 1$.

*Step 3.* From Eq. (2), it follows that within a $\tau$ interval, a change in $\pi$ 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 $\pi$ within some $\tau$ 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 $\pi$ 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 $r_1^{\tau_0]} > r_1^{\tau_0+1]}$).

### A.4. Additional Results for Changes in $S_1$ and Proof of Theorem 7

Lemma 2 establishes in detail how $r_1$ changes as $S_1$ decreases and implies Theorem 7. Recall that a decrease in $S_1$ leads to an increase in *RISK* and potentially to an earlier start of randomization.

**Lemma 2.** *Consider the specification of* $\Gamma$ *and the definitions of* $\tau$ *and* $r_1$ *and define* $\hat{S}_1$ *and* $\breve{S}_1$ *such that* $\left(\frac{P_1-\hat{S}_1}{R_1-\hat{S}_1}\right)^{\tau} = \left(\frac{P_1-\breve{S}_1}{R_1-\breve{S}_1}\right)^{\tau+1} = \pi$. *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* $\breve{S}_1$.
  - $r_1$ *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 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 $\tau$ is a function of $S_1$, is

$$\frac{\partial r_1}{\partial S_1} = -\frac{1}{2}\left(1 - \frac{\pi}{RISK^{\tau-1}} + (\tau-1)\frac{\pi}{RISK^{\tau}}(1-RISK)\right) \quad \text{if} \quad \tau \le N,$$
$$\frac{\partial r_1}{\partial S_1} = \frac{1}{2}\left(1 - \frac{\pi}{RISK^{\tau-1}} + (\tau-1)\frac{\pi}{RISK^{\tau}}(1-RISK)\right) \quad \text{if} \quad N < \tau \le 2N.$$

This shows that if $\tau_0 + 1 \le N$, a marginal increase in $S_1$ by $\Delta S_1 > 0$ that does not affect $\tau$ leads to a change in $r_1^{\tau_0+1}$ by $\frac{-\Delta S_1}{2}\left(1 - \frac{\pi}{RISK^{\tau_0+1-1}} + (\tau_0+1-1)\frac{\pi}{RISK^{\tau_0+1}}(1-RISK)\right)$. This must be smaller than 0 (that is, a decrease in $r_1^{\tau_0+1}$) because $\pi$ 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 $\frac{\pi}{RISK^{\tau_0+1}} \ge 1$. If $N < \tau_0 + 1 \le 2N$, a marginal increase in $S_1$ for which $\pi$ remains in the interval $\tau_0 + 1$ leads to an equivalent increase in $r_1^{\tau_0+1}$. Thus, if $\tau_0 + 1 \le N$ (if $N < \tau_0 + 1 \le 2N$), $r_1^{\tau_0+1}$ decreases (increases) monotonically if $S_1$ increases such that $\pi$ remains in the interval $\tau_0 + 1$.

*Step 3.* The argument provided in step 3 of the proof of Lemma 1 holds also for changes in $S_1$ as it did there for changes in $\pi$. From Eq. (2), it follows that a change in $S_1$ for which $\pi$ remains in the same $\tau$ 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 $P_1 \le X_1 < R_1$, a change in $S_1$ for which $\pi$ remains in the same $\tau$ 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 $r_1^{\tau_0]} > r_1^{\tau_0+1]}$).

## A.5. Proof of Theorem 8: Changes in $P_1$

To prove Theorem 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 $\tau$.

*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 $\pi = \left(\frac{P_1 - S_1}{R_1 - S_1}\right)^{\tau_0} = \left(\frac{P_1 + \mu - S_1}{R_1 - S_1}\right)^{\tau_0+1}$, that is, a change in $P_1$ that leads to an increase in $\tau$ 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 $\tau_0 + 1 \le N$ in step 1.a and the scenario that $N < \tau_0 + 1 \le 2N$ in step 1.b.

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

$$\mu < \frac{R_1 - P_1}{\tau_0 + 1}. \tag{3}$$

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 $\mu$ to lead to an increase in $\tau$ by 1 (with before and after the increase $P_1 = X_1$ and $P_1 + \mu = X_1$, respectively) it must hold that

$$\mu = (R_1 - S_1)\left(\frac{P_1 - S_1}{R_1 - S_1}\right)^{\frac{\tau_0}{\tau_0+1}} - (P_1 - S_1). \tag{4}$$

By replacing $\mu$ in Eq. (3) with the right-hand side of Eq. (4) and (without loss of generality) "normalizing" to $R_1 = 1$, $S_1 = 0$, and $0 < P_1 < 1$, we obtain

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

Rearranging leads to

$$P_1^{\frac{\tau_0}{\tau_0+1}} < \frac{1}{\tau_0 + 1} + \frac{\tau_0}{\tau_0 + 1} P_1. \tag{5}$$

To establish that Eq. (5) holds, and $r_1$, thus, indeed increases, we now isolate the left-hand side of Eq. (5). 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 Eq. (5)) as a binomial series:

$$\begin{aligned}(1-w)^{\frac{\tau_0}{\tau_0+1}} &= \sum_{k=0}^{\infty}\binom{\frac{\tau_0}{\tau_0+1}}{k}(-w)^k \\ &= 1 + \frac{\tau_0}{\tau_0+1}(-w) + \frac{\frac{\tau_0}{\tau_0+1}\left(\frac{\tau_0}{\tau_0^0+1} - 1\right)}{2!}(-w)^2 + \\ &\quad \cdots + \frac{\frac{\tau_0}{\tau_0+1}\left(\frac{\tau_0}{\tau_0+1} - 1\right)\left(\frac{\tau_0}{\tau_0+1} - 2\right)\cdots\left(\frac{\tau_0}{\tau_0+1} - k + 1\right)}{k!}(-w)^k + \ldots\end{aligned} \tag{6}$$

It can be seen that every element that is "added" to 1 in this series is smaller than 0 because if $k$ is even, the numerator is negative while $(w)^k$ is positive and 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. (6) implies that

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

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

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

This can be rearranged to Eq. (5), which shows that Eq. (5) is true and, thus, proves that (as postulated) $r_1$ increases if $P_1$ increases as considered and $\tau_0 + 1 \le N$, that is, that $r_1^{\tau_0]} < r_1^{\tau_0+1]}$ if $\tau_0 + 1 \le N$.

*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 $\frac{1}{2}(\mu(2N - (\tau_0 + 1)) + R_1 - P_1)$. It holds that $\frac{1}{2}(\mu(2N - (\tau_0 + 1)) + R_1 - P_1) > 0$, implying that (as postulated) $r_1^{\tau_0]} > r_1^{\tau_0+1]}$, that is, that $r_1$ decreases.

*Step 2.* The derivative of $r_1$ in $P_1$, neglecting that $\tau$ is a function of $P_1$, is

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

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 $\frac{-\Delta P_1(\tau_0+1-1)}{2}\left(\frac{\pi}{RISK^{\tau_0+1}} - 1\right) \leq 0$ if $\tau_0 + 1 \leq N$ and to a change in $r_1^{\tau_0+1}$ by $\frac{\Delta P_1}{2}\left(2N + (\tau_0 + 1 - 1)\left(\frac{\pi}{RISK^{\tau_0+1}} - 1\right)\right) > 0$ 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 $\pi$ remains in the interval $\tau_0 + 1$.

*Step 3.* Finally, we consider an increase in $P_1$ by $\epsilon$, where $0 < \epsilon \leq \mu$ (with $\mu$ as specified in Eq. (4)), such that before the change, *RISK* is such that $\pi$ is at the right border of the interval $\tau_0$ while after the change (which leads to an increase in *RISK*), *RISK* is such that $\pi$ is in the interval $\tau_0 + 1$ (i.e., before the change $\pi = \left(\frac{P_1-S_1}{R_1-S_1}\right)^{\tau_0}$ and after the change $\left(\frac{P_1+\epsilon-S_1}{R_1-S_1}\right)^{\tau_0+1} \leq \pi < \left(\frac{P_1+\epsilon-S_1}{R_1-S_1}\right)^{\tau_0+1-1}$). We know from step 2 that if $\epsilon$ 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 $\epsilon$ 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. (2) that $r_1^{\tau_0]} = \frac{\tau_0}{2}(R_1 - P_1)$ and that

$$r_1^{\tau_0+1} = \frac{1}{2}\left((\tau_0 + 1)(R_1 - (P_1 + \epsilon)) + P_1 + \epsilon - \left(S_1 + \frac{\pi}{\left(\frac{P_1+\epsilon-S_1}{R_1-S_1}\right)^{\tau_0+1-1}}(R_1 - S_1)\right)\right).$$

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

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

This proves that as $\epsilon$ 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

$$r_1^{\tau_0+1} = \frac{1}{2}\left((2N-(\tau_0+1))(R_1-(P_1+\epsilon))+S_1+\frac{\pi}{\left(\frac{P_1+\epsilon-S_1}{R_1-S_1}\right)^{\tau_0+1-1}}(R_1-S_1)-(P_1+\epsilon)\right).$$

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

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

This proves that as $\epsilon$ 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 $r_1^{\tau_0]}$.

### A.6. Proof of Theorem 9: Changes in $R_1$

To prove that $r_1$ changes as postulated in Theorem 9 if $R_1$ decreases, 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 randomization tends to start earlier.

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

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

$$\mu > \frac{R_1 - P_1}{1 + \tau_0}. \tag{7}$$

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

$$\mu = (R_1 - S_1)\left(1 - \left(\frac{P_1 - S_1}{R_1 - S_1}\right)^{\frac{1}{\tau_0+1}}\right). \tag{8}$$

By replacing $\mu$ in Eq. (7) with the right-hand side of Eq. (8) and "normalizing" to $R_1 = 1$, $S_1 = 0$, and $0 < P_1 < 1$, we obtain

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

Rearranging leads to

$$P_1^{\frac{1}{\tau_0+1}} < 1 - \frac{1 - P_1}{\tau_0 + 1}. \tag{9}$$

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:

$$\begin{aligned}(1 - w)^{\frac{1}{\tau_0+1}} &= \sum_{k=0}^{\infty} \binom{\frac{1}{\tau_0+1}}{k} (-w)^k \\ &= 1 + \tfrac{1}{\tau_0+1}(-w) + \frac{\frac{1}{\tau_0+1}\left(\frac{1}{\tau_0+1} - 1\right)}{2!}(-w)^2 + \\ &\quad \ldots + \frac{\frac{1}{\tau_0+1}\left(\frac{1}{\tau_0+1} - 1\right)\left(\frac{1}{\tau_0+1} - 2\right)\ldots\left(\frac{1}{\tau_0+1} - k + 1\right)}{k!}(-w)^k + \ldots\end{aligned} \tag{10}$$

Every element that is "added" to 1 in this series is smaller than 0 because if $k$ is even, the numerator is negative, whereas $(-w)^k$ is positive and if $k$ is uneven, the numerator is positive, whereas $(-w)^k$ 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. (10) implies that

$$(1 - w)^{\frac{1}{\tau_0+1}} < 1 + \frac{1}{\tau_0 + 1}(-w). \tag{11}$$

Replacing $w$ in Eq. (11) again by $1 - P_1$ yields Eq. (9), which shows that Eq. (9) 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 $\frac{1}{2}(\mu(2N - (\tau_0 + 1)) + R_1 - P_1) > 0$, 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 $\tau$ is a function of $R_1$, is

$$\begin{aligned}\frac{\partial r_1}{\partial R_1} &= \frac{\tau}{2}\left(1 - \frac{\pi}{RISK^{\tau-1}}\right) && \textit{if } \tau \leq N, \\ \frac{\partial r_1}{\partial R_1} &= \frac{1}{2}\left(2N - \tau(1 - \frac{\pi}{RISK^{\tau-1}})\right) && \textit{if } < N\tau \leq 2N.\end{aligned}$$

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

*Step 3.* Finally, we consider a decrease in $R_1$ by $\epsilon$, where $0 < \epsilon \leq \mu$ (with $\mu$ as specified in Eq. (8)), such that before the change, $RISK$ is such that $\pi$ is at the right border of the interval $\tau_0$ while after the change (which leads to an increase in *RISK*), *RISK* is such that $\pi$ is in the interval $\tau_0+1$ (i.e., before the change $\pi = \left(\frac{P_1-S_1}{R_1-S_1}\right)^{\tau_0}$ and after the change $\left(\frac{P_1-S_1}{R_1-\epsilon-S_1}\right)^{\tau_0} \le \pi < \left(\frac{P_1-S_1}{R_1-\epsilon-S_1}\right)^{\tau_0+1-1}$). We know from step 2 that $r_1^{\tau_0+1}$ is larger the smaller $\epsilon$ is. In this step we prove that also as $\epsilon$ goes to 0, $r_1^{\tau_0]} \ge r_1^{\tau_0+1}$ (that is, $r_1^{\tau_0+1}$ cannot be larger than $r_1^{\tau_0]}$).

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

$$r_1^{\tau_0+1} = \frac{1}{2}\left((\tau_0+1)(R_1-\epsilon-P_1) + P_1 - \left(S_1 + \frac{\pi}{\left(\frac{P_1-S_1}{R_1-\epsilon-S_1}\right)^{\tau_0+1-1}}(R_1-\epsilon-S_1)\right)\right).$$

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

$$\begin{aligned}\lim_{\varepsilon\downarrow 0} r_1^{\tau_0+1} &= \frac{1}{2}\left((\tau_0+1)(R_1-\epsilon-P_1) + P_1 - \left(S_1 + \frac{\pi}{\left(\frac{P_1-S_1}{R_1-\epsilon-S_1}\right)^{\tau_0+1-1}}(R_1-\epsilon-S_1)\right)\right)\\ &= \frac{1}{2}((\tau_0+1)(R_1-P_1) + P_1 - S - (R_1 - S_1))\\ &= \frac{\tau_0}{2}(R_1-P_1).\end{aligned}$$

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

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

$$r_1^{\tau_0+1} = \frac{1}{2}\left((2N-(\tau_0+1))(R_1-\epsilon-P_1) + S_1 + \frac{\pi}{\left(\frac{P_1-S_1}{R_1-\epsilon-S_1}\right)^{\tau_0+1-1}}(R_1-\epsilon-S_1) - P_1\right).$$

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

$$\begin{aligned}\lim_{\varepsilon\downarrow 0} r_1^{\tau_0+1} &= \frac{1}{2}\left((2N-(\tau_0+1))(R_1-\epsilon-P_1) + S_1 + \frac{\pi}{\left(\frac{P_1-S_1}{R_1-\varepsilon-S_1}\right)^{\tau_0+1-1}}(R_1-\epsilon-S_1) - P_1\right)\\ &= \frac{1}{2}((\tau_0+1)(R_1-P_1) + P_1 - S - (R_1 - S_1))\\ &= \frac{\tau_0}{2}(R_1-P_1).\end{aligned}$$

This proves that also if $N < \tau_0+1 \le 2N$ and as $\epsilon$ 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 $r_1^{\tau_0]}$.

## A.7. Proof of Theorem 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 $\Gamma^+$ 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 $\Gamma^+$ as well as in $\Gamma^-$ 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 Theorem 10.

We have to consider three scenarios to prove the "first part" of Theorem 10, that is, to establish that $r_1$ increases if, after the increase in $N$, trust is possible in $\Gamma^+$ but not in $\Gamma^-$ or if at least before the increase, trust was not possible in $\Gamma^-$. 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 $\Gamma^+$ 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 $\Gamma^+$ before as well as after the change in $N_0$, whereas trust would never be placed in $\Gamma^-$; $U_1^{\Gamma^+}$ 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 $\Gamma^-$ before the increase in $N$, whereas after the increase in $N$ both trustors would place trust with certainty in their first TG of $\Gamma^-$; $U_1^{\Gamma^+}$ 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 $R_1 - X_1$.