

# Literature Review

Embeddedness and cooperation: The effect of control

Thom Volker (5868777)

Supervisors: Prof. dr. ir. Vincent Buskens & Prof. dr. Werner Raub

*Sociology and Social Research Master's programme*

*Utrecht University*

June 15, 2021

Word count (excluding in-text references and appendix): 7131 (5956)

## Abstract

How cooperation between human beings emerges and prevails is one of the longest standing research questions in the social sciences. Game-theoretical predictions render cooperation between rational and selfish individuals in isolated social dilemmas impossible. However, taking relationships between the actors involved (dyadic embeddedness) and relationships between the actors involved and third-parties (network embeddedness) into account, cooperation between rational and selfish individuals is possible. Here, theoretical and experimental research on dyadic and network embeddedness is reviewed, with a specific focus on the mechanism called “control”, which reflects future sanction opportunities of the actors involved. It is shown that embeddedness in general, and control opportunities specifically, indeed foster cooperation, although this effect is more pronounced and more consistently present in dyadically embedded interactions.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Embeddedness effects</b>	<b>4</b>
<b>3</b>	<b>Control effects under dyadic embeddedness</b>	<b>7</b>
3.1	Control effects in exogenously established dyadic relations . . . . .	7
3.2	Control in endogenously established dyadic relations . . . . .	11
<b>4</b>	<b>Control effects under network embeddedness</b>	<b>12</b>
4.1	Control in exogenously embedded interactions in networks . . . . .	13
4.2	Control in endogenously embedded interactions in networks . . . . .	14
<b>5</b>	<b>Conclusion and implications for future research</b>	<b>15</b>
<b>6</b>	<b>Literature</b>	<b>18</b>
<b>A</b>	<b>Appendix 1 - Thesis proposal</b>	<b>24</b>

# 1 Introduction

Social dilemmas are at the core of everyday life. Students may anticipate a good grade for group work with minimal effort by free-riding on the work of their peers, researchers could obtain another publication by letting their collaborators do the lion’s share of the required work (Corten, Buskens, & Rosenkranz, 2020) and a car dealer may maximize the returns of a transaction by hiding several vehicle defects when selling a second-hand car to a relatively uninformed customer (Buskens & Weesie, 2000a). Under the assumption of “social isolation”, reflecting that the interacting actors can be considered perfect strangers that do not anticipate any future interactions, actors can maximize their individual returns by behaving uncooperatively. However, if all actors involved would apply this strategy, the students will not obtain a good grade, the researchers will not publish a joint paper, and the intended transaction will not materialize. The collective returns that emerge from self-regarding, goal-directed behaviour would be lower than what could have been achieved under mutual cooperation: hence the term “social dilemma” (Kollock, 1998; Ostrom, 1998).

Theoretical predictions render a cooperative interaction between two actors in an isolated social dilemma impossible (e.g., see Luce & Raiffa, 1989 for the Prisoner’s Dilemma; and Buskens & Raub, 2002 for the Trust Game). Consider for example the standard one-shot Prisoner’s Dilemma, in which two actors simultaneously decide to cooperate or defect. Regardless of the choice of one’s partner, an individual actor obtains the highest payoff by acting uncooperatively. Defecting, rather than cooperating, leads to a higher payoff when the other player cooperates, as well as when the other player defects. Hence, the Nash equilibrium is mutual defection, even though mutual cooperation would yield a Pareto improvement, in the sense that all players would be better off. Contrary to these ominous theoretical predictions, in practice researchers generally find non-negligible rates of cooperation in one-shot games (e.g., see Cooper, DeJong, Forsythe, & Ross, 1996; Hayashi, Ostrom, Walker, & Yamagishi, 1999; Snijders & Keren, 2001), although these cooperation rates tend to decline when participants gain experience with these games (Dal Bó, 2005).

Multiple scholars, however, noted that most real-life interactions do not take place in social isolation, but are actually embedded (e.g., Axelrod, 1984; Granovetter, 1985). Embeddedness refers to the fact that the actors involved share a common environment that could foster cooperation (e.g., Yamagishi & Yamagishi, 1994; Buskens & Raub, 2002). The actors may have interacted in the past, and/or speculate on interacting in the future, which is referred to as *dyadic embeddedness* (Buskens & Raub, 2002). Additionally, the actors may be connected indirectly, through third parties that have

interacted with any of the two in the past, or speculate on doing so in the future, which is referred to as *network embeddedness* (Buskens & Raub, 2002). Both dyadic and network embeddedness, which may apply simultaneously, potentially foster cooperation.

In this review, it will be assessed how, from a game-theoretic perspective, embeddedness affects cooperation, with an explicit focus on experimental studies. First, it will be illustrated which hypotheses on the effects of embeddedness can be derived from game-theoretical models. These models assume rationality of the players, in the sense that the actors involved maximize their utility given the expectations on the behaviour of their partner, using the common knowledge that all actors know all elements of the game, and know that all other actors know all elements of the game. Notably, inferring the utilities of the actors involved is often complicated, because it is not explicit how actors value potential payoffs. Throughout this review it is generally assumed that actors are self-regarding, and thus only care about their own payoffs. In the experiments considered, these payoffs yield points that are translated into money. If required, these rather strict assumptions can be relaxed to derive new hypotheses. Furthermore, the empirical findings discussed in this review generally originate from 2-person Prisoner’s Dilemma Games or Trust Games. Specifically, in most experiments considered, the behaviour of an actor only affects oneself, as well as the single actor toward whom this behaviour is directed. When occasionally another type of game or an  $N$ -person game (i.e., a game in which the behaviour of an actor is directed toward more than one others) is discussed, this will be explicitly addressed. This review will be concluded with a discussion of the empirical findings, the implications of these findings and possible directions for future research.

## 2 Embeddedness effects

In the game-theoretic literature, there are two mechanisms through which embeddedness is considered to affect cooperation between actors: *control* and *learning* (e.g., Yamagishi & Yamagishi, 1994; Buskens & Raub, 2013). Control, also termed control through “voice” in the sense of Hirschman (1970), reflects the opportunity to sanction opportunistic behaviour by exerting control over one’s partner’s long-term returns and will be the focus of this review. Under dyadic embeddedness, one can punish the defection of one’s interaction partner in a previous interaction by refraining from cooperation in current and future interactions (i.e., an actor can rely on direct reciprocity; Nowak, 2012). Additionally, under network embeddedness, one might inform future transaction partners of a defecting actor, who can in turn refuse to cooperate with this actor (which is also called indirect

reciprocity; Nowak, 2012; Sigmund, 2012). Hence, the short-term benefits of acting uncooperatively come with the prospect of future retaliation, hanging over the head of opportunistic actors as the sword of Damocles. Notably, what Buskens & Raub (2013) termed “control” differs from what is called “the illusion of control” in the social psychological literature (e.g., Morris, Sim, & Giroto, 1998; Hayashi et al., 1999). Where “control” in the sense of Buskens & Raub (2013) refers to actual sanctioning opportunities, “the illusion of control” concerns the idea that one might act as if it is possible to control the behaviour of one’s partner in one-shot games, if one’s decision is made prior to the decision of one’s partner, even though no information is transferred between the players (Morris et al., 1998).

Learning refers to the situation where the actors involved have interacted in the past, when they are embedded dyadically, and hereby gained information about each other’s past behaviour (see Buskens & Raub, 2013 for a more thorough discussion and reflections on how learning fits into a game-theoretic paradigm). If one’s partner behaved cooperatively in the past, one may infer that this partner will behave cooperatively again in the current and future transactions. When the actors involved are embedded in a common network, the actors may learn from others how their transaction partners behaved in interactions with third parties. If one’s current partner behaved cooperatively during past interactions with third parties, one might infer that the transaction partner will cooperate in the current transaction as well. Obviously, when an actor’s partner has abused cooperative behaviour of this actor, or of a third party, the actor may not be willing to take the risk of getting exploited, and will defect in the current interaction.

A second distinction relates to the nature of the embeddedness of a transaction. Namely, in research settings, transaction partners are often matched by researchers for a prespecified duration, a situation that is commonly referred to as *exogenous embeddedness*. However, in real life, people often deliberately choose their partners, at least to a certain extent (Chaudhuri, 2009; Yamagishi et al., 1994). Some researchers tried to incorporate this characteristic of real-life encounters in experiments by letting participants choose their transaction partners or specify the interaction duration, which is referred to as *endogenous embeddedness*. However, under the assumption of rationality, no differences between exogenously and endogenously established relationships are to be expected. The formal requirements of a relationship that allow for cooperation to be supported in equilibrium do not depend on whether the relationship is established exogenously or endogenously, and the actors should theoretically only care about whether these prerequisites are met. However, for other reasons that will be discussed in later sections, endogenously formed relations may have

a larger effect on cooperation than exogenously formed relationships (e.g., Chaudhuri, 2011; Frey, Buskens, & Corten, 2019; Schneider & Weber, 2013; Wang, Suri, & Watts, 2012).

Table 1

*Two forms of embeddedness separated by how these can be established.*

Two forms of embeddedness		
Two ways of establishment	Dyad	Network
Exogenous	Actors are matched by the researcher for a prespecified number of rounds.	Actors are matched by the researcher with a new partner every round, but obtain information on their partner's past behaviour from third parties.
Endogenous	Actors are matched by the researchers but deliberately choose their interaction duration.	Actors choose an interaction partner every round, potentially based on information from third parties.

The subsequent sections will be concentrated around the question to what extent control affects cooperation. Hence, the effect of learning, as well as alternative explanations for the emergence of cooperation such as inequity aversion or altruism (e.g., Fehr & Schmidt, 1999; Carpenter, Connolly, & Myers, 2008; Dreber, Fudenberg, & Rand, 2014) fall beyond the scope of this review. A distinction will be made between control under dyadic embeddedness and under network embeddedness, as well as between exogenously and endogenously formed relationships (Table 1).

Note that a great deal of the work published in this area has not distinguished between learning and control explicitly, but merely addresses the question how different forms of embeddedness in general affect cooperation. Nevertheless, it is often possible to assess the effect of control, either explicitly or implicitly. In general, there are two ways to disentangle learning and control. The first possibility is to study solely behaviour of participants in the first round of a given game, because then no learning could have taken place (e.g., Dal Bó, 2005; Dal Bó & Fréchette, 2011; Embrey, Fréchette, & Yuksel, 2018). The second way, which is often used to analyse behaviour in finitely repeated games, is to assess the effect of the number of rounds left, after controlling for any learning that could have taken place (e.g., Bolton, Katok, & Ockenfels, 2004; Buskens, Raub, & van der Veer, 2010). Specifically, previous actions by an actor's transaction partner are taken into account when analysing the behaviour of an actor in any given round and assessing the effect of the number of rounds to play.

### 3 Control effects under dyadic embeddedness

Applying the game-theoretic assumptions as outlined above on an isolated social dilemma renders cooperation theoretically impossible. Consider the Prisoner’s Dilemma and the Trust Game as two canonical examples. In a Prisoner’s Dilemma, equilibrium behaviour yields mutual defection, as defection is a strictly best response, even though mutual cooperation would yield a Pareto-improvement. Similar reasoning applies in a one-shot Trust Game. In a Trust Game, rather than mutually risking exploitation, only the trustor, who decides whether or not to trust the trustee, risks being exploited. The trustee on the other hand, has to decide whether to honour or abuse trust, if trust is placed, but cannot choose between these options if no trust is placed. Equilibrium behaviour implies that a trustee, if given the opportunity, would abuse trust. The trustor, realizing that trust will be abused if it is placed, withholds trust. Once again, the outcome where no trust is placed is Pareto-suboptimal, as both players would be better off if trust is placed and honoured. However, cooperation can emerge among selfish and rational actors by taking dyadic embeddedness into account.

#### 3.1 Control effects in exogenously established dyadic relations

##### 3.1.1 Infinitely repeated games

As an extension of a one-shot game, consider the infinitely repeated Prisoner’s Dilemma. The Prisoner’s Dilemma is now played indefinitely often. After each round, a new round will be played with continuation probability  $\delta$ , which Axelrod (1984) aptly termed “the shadow of the future”, and the game will end with probability  $(1 - \delta)$ . The introduction of a continuation probability allows the actors to condition their behaviour in future rounds on the outcome of the current round. If the continuation probability is sufficiently large with respect to the short-term incentives to behave uncooperatively, mutual cooperation can be supported in an equilibrium that cannot be payoff dominated by other equilibria (e.g., see Roth & Murnighan, 1978 for the technical details). The prospect of mutual cooperation allows for control, in the sense that cooperative behaviour of one’s partner in the current round can be rewarded with cooperative behaviour in future rounds, as well as opportunistic behaviour can be punished with defection. Hence, behaving cooperatively can be in the self-interest of rational players. Following Buskens & Raub (2013), qualitative predictions about the behaviour of the actors involved can be derived. Without going into the details, it can be expected that cooperation is more likely to emerge if the continuation probability increases, and

when the gains of a single, unilateral defection decrease. Namely, when the continuation probability increases, and when the gains of a single, unilateral defection decrease, future sanctions become more threatening, expectedly resulting in more control opportunities and hence more cooperation. Note that similar predictions apply to a Trust Game and a broader set of social dilemma games (e.g., Buskens & Raub, 2013).

Roth & Murnighan (1978) and Murnighan & Roth (1983), who initiated experimental research on cooperation in infinitely repeated Prisoner’s Dilemmas, showed that cooperation became more likely under a higher continuation probability than under a lower continuation probability. Nevertheless, the increase was small, and the authors concluded that the value of the continuation probability hardly affects cooperation. However, in these studies, the participants played against the experimenters, rather than against each other, and the points participants earned during the experiments were not linearly translated into the monetary rewards the subjects could earn, which potentially affected the incentives of the subjects. Dal Bó (2005) improved upon this initial work, and found that first round cooperation rates increased with the continuation probability. Particularly, the differences in cooperation between the continuation probabilities increased with the amount of experience the participants gained. Hence, people become more willing to cooperate when future transactions weigh more heavily on one’s total payoff.

Similar observations were made by Dal Bó and Fréchette (2011, 2018). These authors argued that the limited effect of increasing the continuation probability in earlier studies was due to the fact that participants have to gain experience to properly evaluate the meaning of this increase. With sufficient experience, participants cooperated more often when cooperation was supported in equilibrium, and defected when it was not. Specifically, the continuation probability had a positive effect on first round cooperation. Moreover, the authors showed that as the gains from unilateral defections increased, cooperation was less likely to occur. Hence, in infinitely repeated games, studies consistently show that cooperation increases with the expected length of the game and decreases with the temptation to unilaterally defect, consistent with the hypothesized effects of control.

### **3.1.2 Finitely repeated games**

Control effects in dyadic relations can also be studied in finitely repeated interactions, where actors play a game for a fixed number of rounds. However, assuming rationality as defined above, it follows from backward induction that no cooperation is possible in finitely repeated games (e.g., Luce & Raiffa, 1989; Selten, 1978). Namely, in the last round of the game, non-cooperative behaviour



cannot be punished in any subsequent round, and hence, defecting is a strictly dominant strategy. As actors defect in the final round, actions in the penultimate round do not affect behaviour in the last round, and again, no actor has an incentive to cooperate. This pattern repeats itself to the first round of the game, and hence, assuming rationality, no cooperation is predicted in any round.

However, when relaxing the assumption that actors have complete information, cooperation can be supported in equilibrium until the final rounds of the game (Kreps et al., 1982; Kreps & Wilson, 1982). If an actor believes with sufficiently high probability that a partner has no incentive to defect until this partner is defected on, the benefits of mutual cooperation outweigh the gains of exploiting such a conditionally cooperative partner in an early round and being punished with defection thereafter. The prospect of mutual cooperation directly allows for the introduction of control (Buskens & Raub, 2013). Namely, after an actor defects, the actor proved to be no conditional cooperator, and the finitely repeated game would continue with mutual defection hereafter. As long as both players cooperate, it cannot be known whether any of the two players is a conditional cooperator. The actors can thus control one another, because future payoffs depend on one's behaviour in the current round. It follows from this model that cooperation increases with the number of rounds to be played and decreases with the temptation to defect. Additionally, it is expected that in the final rounds of the finitely repeated game, when the long term benefits of mutual cooperation no longer outweigh the short-term costs of maintaining one's reputation (i.e., the opportunities to control future behaviour of one another diminish), actors will try to exploit their partners.

It appeared that subjects generally act cooperatively in initial rounds of a finitely repeated game, rendering the backward induction model less plausible than the sequential equilibrium model by Kreps & Wilson (1982; e.g., Camerer & Weigelt, 1988; Mao, Dworkin, Suri, & Watts, 2017; Neral & Ochs, 1992). Rapoport (1997, p. 122), consequently, concluded that in practice, subjects do not rely on, or are not capable of, backward induction. In line with the hypothesized effects of control, cooperation generally sharply decreases toward the end of the game (e.g., Buskens et al., 2010; Embrey et al., 2018; Mao et al., 2017; Van Miltenburg, Buskens, & Raub, 2012). However, this decline in cooperation cannot entirely be ascribed to the lack of control opportunities in the final rounds of the game. Participants may namely refrain from cooperation for three different reasons, of which the latter two reflect control. First, defection could be a response to defection of one's partner in earlier rounds. Second, an actor may figure (perhaps from previously played games) that in final rounds of the game, hardly any cooperation is possible, and hence defecting serves as

a protection against being exploited. Third, an actor may realize that the short-term benefits of defecting outweigh the possible returns of another round of mutual cooperation.

Buskens et al. (2010) explicitly studied the presence of control effects in a Trust Game, and found, under dyadic embeddedness, a positive effect of the number of rounds still to be played on cooperation, for both the trustor and the trustee, after controlling for learning effects. Other studies corroborated the finding that the number of rounds that are to be played affects cooperation rates. Embrey et al. (2018) performed a meta-study with data from multiple previously held experiments on finitely repeated Prisoner’s Dilemma games by Andreoni & Miller (1993), Dal Bó (2005), Cooper et al. (1996), Bereby-Meyer & Roth (2006) and Friedman & Oprea (2012). The combined evidence from these studies and a newly designed experiment, showed that first round cooperation rates, where no learning could have occurred, increased with the length of the game and that this effect increases with the experience of the subjects. Additionally, it was shown that cooperation increases when the temptation to unilaterally defect becomes smaller. Van Miltenburg et al. (2012) and Mao et al. (2017) also showed the importance of experience, in the sense that with experience, near perfect cooperation can be achieved in early rounds that remains above 80% until the final three rounds, but decreases to negligible levels shortly hereafter. However, learning and control effects were not explicitly distinguished in these analyses. Notably, the latter experiment is repeated on twenty consecutive days, with twenty 10-round games per day, and showed that the pattern of high initial cooperation and a strong end-game effect keeps repeating itself, regardless of the fact that actors experience that in final rounds cooperation seldom prevails. Conclusively, in line with the expectations, cooperation rates increase with the importance of future interactions and decrease in the final rounds of the game, showing the importance of future sanction opportunities for cooperation.

Overall, both studies on infinitely and finitely repeated games showed the importance of the prospect of future interactions for cooperation. However, it also appeared that cooperation rates in first rounds of infinitely repeated games are systematically higher than cooperation rates in the first rounds of finitely repeated games of the same expected length (i.e., an expected length of 2 and 4 rounds), especially after the subjects have gained experience with the game they play (Dal Bó, 2005; Dal Bó & Fréchette, 2011). It seems plausible that in finitely repeated games, subjects understand that cooperation in final rounds no longer outweighs the short-term gains of defecting, or the risk of being defected upon, and thus defect themselves. Hence, control opportunities are hardly present in the final stage(s) of a finitely repeated game, and the overall sanctioning opportunities are thus

smaller in a finitely repeated game than in an infinitely repeated game of the same expected length. Consequently, first round cooperation in finitely repeated games may decrease, due to this reason.

### 3.2 Control in endogenously established dyadic relations

The previous section displayed a relatively consistent effect of control in exogenously embedded dyadic interactions. However, in many real-life interactions, dyadic embeddedness is not imposed exogenously, but established intentionally by the actors involved (Kollock, 1994). Under the assumption of rationality, no differences between exogenously and endogenously established relationships are to be expected, and hence, similar hypotheses on control effects apply as outlined in the previous section. Yet, it is often remarked that selection may play a role when actors can set up relationships endogenously. For example, it may be that actors who realize that, under embeddedness, cooperation is more likely to flourish, will establish embeddedness and behave cooperatively, while actors who do not see such benefits do not engage in embeddedness and are more likely to defect (for a theoretical model, see, e.g., Schneider & Weber, 2013). Following these lines of reasoning, cooperation may be higher under endogenously established dyadic relationships than under exogenously established ones. A more thorough discussion of (empirical findings concerning) such hypotheses is left for future work; the following section is limited to the discussion of control effects under endogenously established dyadic embeddedness.

Empirical studies show that dyadic embeddedness also fosters first-round cooperation when it is chosen by the actors involved in an interaction, without having prior information with respect to the subsequent transaction partner (Brown, Falk, & Fehr, 2004; Schneider & Weber, 2013; Sokolova, Buskens, & Raub, 2021). Consistent with the hypothesized effects of control, cooperation increases with the, endogenously chosen, length of an interaction (Schneider & Weber, 2013). However, note that when embeddedness is established endogenously, actors are no longer randomly assigned to the embeddedness condition, and hence, differences in cooperation under different, endogenously chosen, interaction lengths can no longer be attributed to only these different lengths. Indeed, people who *a priori* indicated a higher willingness to cooperate were more likely to establish a long-term relationship, although this willingness could only partially explain differences in cooperation between one-shot and endogenously chosen long-term relationships (Schneider & Weber, 2013).

Additionally, only having the opportunity to engage in long-term interactions already fosters cooperation (Brown et al., 2004). These findings are in line with the idea that actors realize that having a good reputation may facilitate future lucrative interactions (Kollock, 1994), and that such

a reputation can be built by behaving cooperatively. In fact, even at the cost of a portion of subjects' payoffs, these are willing to establish a long-term relationship, and are subsequently more likely to behave cooperatively in this relationship compared to behaviour in one-shot games (Sokolova et al., 2021). Note that the studies in this review focused on control through “voice”, in the sense of Hirschman (1970). Studies on the effect of control through “exit”, in which actors can abandon an uncooperative relationship, are omitted from this review for the sake of brevity (interested readers are referred to the work by Schuessler, 1989; Hauk, 2003; Hauk & Nagel, 2001; Zhang et al., 2016). Although systematic testing of the effect of control on cooperation in endogenous dyadic embedded interactions is, to the best of my knowledge, fairly limited, the available evidence suggests that it likely is at least as strong as in exogenously dyadic embedded interactions. Therefore, both under exogenously and endogenously established dyadic relations, control opportunities foster cooperation between actors.

## 4 Control effects under network embeddedness

The research on control effects through dyadic embeddedness as discussed in the previous sections deliberately ignores an important factor that characterizes many real-life interactions: the involvement of third-parties (Granovetter, 1985). Even if an actor does not engage in a long-term relationship with the same partner, the behaviour of this actor may have consequences for future interactions, albeit with different partners (Buskens & Raub, 2002; Kandori, 1992). When actors only obtain information about their own interactions, there exists a cooperative Nash equilibrium, where each actor cooperates until the first defection is observed, if the expected number of future interactions is large enough relative to the gains of defection (Kandori, 1992; Nowak & Sigmund, 2005). Yet, the conditions under which such an equilibrium is supported are fairly restrictive, and the equilibrium itself is extremely unstable. These conditions can be relaxed by introducing information transmission through the network. When there is perfect information, in the sense that information is provided about all interactions, it is theoretically unimportant whether an interaction is embedded dyadically or in a network, as it is the content that matters. When the spread of information is not perfect, in the sense that only part of the information is disseminated, cooperation is expected to increase with the amount of reliable information that is disseminated through the network (for a more detailed discussion, see Raub & Weesie, 1990; Kandori, 1992; and Nowak & Sigmund, 2005), and with the importance of future interactions (i.e., the expected number of future interactions), because both

factors allow for more severe future sanctions and thus increase control opportunities (for infinitely repeated games, see Raub & Weesie, 1990; Buskens & Weesie, 2000b; Kandori, 1992; for finitely repeated games, see Buskens, 2003). Note however that making assumptions on the flow of reliable information about actors' behaviour may be problematic, because whether to share information on someone's behaviour is itself a (second-order) social dilemma, and because actors may face strategic incentives to share misleading information. These issues are deliberately ignored in the current review, because reliable information is provided in all experiments discussed.

#### 4.1 Control in exogenously embedded interactions in networks

Whereas dyadic embeddedness generally leads to substantially more cooperation due to control effects relative to atomized interactions, the effects of network embeddedness are not so clear-cut. Some studies found support for the hypothesis that cooperation increases with information exchange through the network, in the sense that network embeddedness yielded higher cooperation rates from the first round onwards (Pfeiffer, Tran, Krumme, & Rand, 2012; Seinen & Schram, 2006), or found control effects through an end-game effect after controlling for learning (Bolton et al., 2004). Corten, Rosenkranz, Buskens, & Cook (2016), however, have not found any effect of control through network embeddedness on cooperation. These studies, however, differed in the fact that in the former three studies, information was summarized (e.g., a reputation or the total number of cooperative and defective actions were shown to the subjects), whereas in the latter, all previous actions by all actors in the network were shown. Counterintuitively, more information thus does not necessarily result in more cooperation. Additionally, the former three studies differed from the last one in the size of the transaction network. Whereas in the former three studies the network size was at least sixteen, the latter study yielded networks containing six actors. Since the outcomes of own transactions were shown in the condition without network embeddedness in the latter study, the condition with network embeddedness was effectively compared to a condition with a substantial amount of dyadic embeddedness, which may attenuate the effect of network embeddedness.

Additionally, a number of studies focused on the effects of network embeddedness in Trust Games played in triads, with two trustors and one trustee. In these studies, it generally appeared that trustors seldom reap the benefits of network embeddedness, in the sense that network embeddedness does not foster trustfulness, when compared to games with only dyadic embeddedness (Buskens et al., 2010; Frey et al., 2019; Van Miltenburg et al., 2012). A similar observation was made when the focus was on Investment Games, rather than Trust Games (Barrera & Buskens, 2009). Hence, the

additional control opportunities offered by network embeddedness do not result in more trustfulness of the trustors. Trustees, on the other hand, seem to realize that network embeddedness equips trustors with more sanction opportunities, because the second trustor can sanction the abuse of trust against the first trustor, and vice versa. Buskens et al. (2010) found significant network control effects on the trustworthiness of the trustee, while Frey et al. (2019) showed, without testing, that first-round trustworthiness increased under exogenous network embeddedness. Van Miltenburg et al. (2012) could not find any effects of network control, but in this study, the condition with only dyadic embeddedness already yielded trustworthiness rates of above 80% until the last rounds, leaving little room for improvement under network embeddedness.

## 4.2 Control in endogenously embedded interactions in networks

Besides under exogenous embeddedness, network control effects can also be assessed under endogenous embeddedness. Note again that under the assumption of rationality, no differences between exogenously and endogenously embedded interactions are expected. Yet, it has often been remarked that assortment may play an important role in fostering cooperation under endogenously established network embeddedness (e.g., Nowak, 2012; Perc et al., 2017), because under endogenous network embeddedness, actors can select their partners based on the observation that a potential partner has behaved cooperatively in the past. This allows for new sanction opportunities, in the sense that actors can refuse to engage in a relationship with actors who defected in the past. Yet, relatively few studies assessed cooperation under endogenously established network embeddedness in two-person dilemmas. In fact, to the best of my knowledge, no studies exist that explicitly test control effects in these settings. However, some studies exist that allow to get an initial grasp of control effects under endogenously established network embeddedness.

Corten et al. (2020), who studied cooperation in dynamic networks of six actors found no support for an effect of network embeddedness, and no indication for a control effect. Note that in this study actors chose to engage in an interaction, and hence cooperation rates were defined as a proportion of the total number of possible interactions. When cooperation was assessed relative to the total number of actual interactions, cooperation was somewhat higher when actors received information about their partners through their network connections, relative to when people only gathered information through their own interactions. However, the effect of network control was not specifically assessed. Additionally, some studies assessed cooperation rates in  $N$ -person, rather than 2-person, Prisoner’s Dilemmas, which somewhat changes the nature of the game. Without

going into the specifics, it appeared that in  $N$ -person Prisoner’s Dilemma games, dynamic partner updating fostered cooperation, but not necessarily due to a control effect (Rand, Arbesman, & Christakis, 2011; Riedl & Ule, 2002; Wang et al., 2012). Rather, it appeared that cooperators were allowed to select other cooperators, maintaining high cooperation rates. When no partner updating was possible, these actors more often switched to defection to avoid getting exploited.

Overall, it appears that the findings on control effects through network embeddedness are mixed. When control effects were found, it mainly affected those actors that were in the position to abuse trust (i.e., both players in a Prisoner’s Dilemma game, but only trustees in a Trust Game). The actors that had to place trust generally did not seem to realize that network embeddedness would increase possible sanction opportunities and hence would allow for more trustfulness, or, alternatively, were pessimistic about the extent to which trustees would realize this. Additionally, a substantial part of the effect of endogenously established network embeddedness seems to be attributable to assortment rather than to control. Specifically, cooperators propose ties to other cooperators, while defectors have to change their strategy to be seen as suitable partners. Additionally, when ties with actors that abuse trust can be easily abandoned, cooperative relationships seem to occur somewhat more frequently.

## 5 Conclusion and implications for future research

In the simplest interactions, that is, one-shot interactions in a social vacuum, high levels of cooperation are difficult to sustain when games are played multiple times (Simpson & Willer, 2015). However, this review showed that dyadic embeddedness consistently fosters the establishment of cooperative relationships due to the possibility to control future payoffs of one’s interaction partner, both when embeddedness is established exogenously and endogenously. A longer interaction duration, as well as a smaller temptation to defect, are associated with higher cooperation rates, both indicative for a control effect. These effects appeared to increase with the experience of the actors involved. For network embeddedness, similar hypotheses can be formulated, although these were less often tested explicitly, and, if tests were performed, less often confirmed. First of all, the effects of network embeddedness, if present, seem smaller than the effects of dyadic embeddedness, and when both forms of embeddedness operated simultaneously, network embeddedness adds little to the effect of dyadic embeddedness. When actors obtain an easy to interpret reputation, network embeddedness seems to be more effective in promoting cooperation than when actors themselves have

to make sense of a long history of played games. Additionally, network embeddedness predominantly affected actors that had the opportunity to abuse trust, and seems to be more effective when the actors involved established their relations endogenously. Yet, in the latter case, the increased effect seems to be due to assortment, in the sense that cooperators end up playing with other cooperators.

A possible explanation for the fact that positive effects of network embeddedness on cooperation are not always found, could be that past research has often failed to take the importance of experience in playing games into account. Subjects may experience difficulties understanding the relatively complex dynamics of games in networks, especially because actors have to rely on third parties to sanction defectors (Milinski, Semmann, Bakker, & Krambeck, 2001). If actors are unsure whether unilateral defections of their partners will be sanctioned by third parties, they may overestimate the risk of cooperation, and protect themselves by defecting. Since, under network embeddedness, many actors rely on first-order assessment of defective behaviour (i.e., actors base their decision to cooperate on whether their partner cooperated in the previous round; Sigmund, 2012), a few initial defections may result in a quick collapse of cooperation (Kollock, 1994). It is not until participants understand the benefits offered by embeddedness, and hence defect less initially, that cooperation will flourish. Studies that allow for gaining experience throughout the experiment find relatively high cooperation rates (e.g., Wang et al., 2012; Frey et al., 2019), and show that cooperation increases with experience, when cooperation can be supported in equilibrium (e.g., Van Miltenburg et al., 2012; Dal Bó, 2005; Dal Bó & Fréchette, 2011). The introduction of simplified reputation scores may fasten this process, as it lays less of a cognitive burden on the participants.

The idea that the complexity of a game complicates the emergence of cooperation is also in line with the finding that trustors generally have more difficulties with reaping the benefits of network embeddedness than trustees. Trustees may relatively quickly understand that behaviour against a given trustor is observed by the second trustor, and hence untrustworthy behaviour can be sanctioned by both trustors. The trustors, however, have to anticipate how the trustee evaluates the presence of two potential sources of sanctioning (see Buskens et al., 2010 for a similar argument). Additionally, the importance of understanding the dynamics of the game potentially explain why endogenously, rather than exogenously, established embeddedness seems more effective in promoting cooperation. Actors who understand the benefits of embeddedness are more likely to establish embeddedness endogenously, and are subsequently more likely to reap these benefits by behaving cooperatively, while actors who do not see such benefits may be less willing to do so. However, possible effects of experience have been assessed predominantly under dyadic embeddedness. As



studies in networks yield more complex dynamics, experience may be even more important. Hence, a thorough assessment of the effects of experience and of the complexity of reputation measures on cooperation in networks could be of great value.

Methodologically, all empirical studies in the current review employed an experimental design. Multiple papers have been devoted to a discussion of this design (e.g., see Falk & Heckman, 2009; Jackson & Cox, 2013; Morris, 2014), and the advantages (e.g., the possibility to make causal claims and to disentangle the slightest sources of variation) and disadvantages (e.g., difficulties concerning generalizability to real-world settings, unnatural behaviour of participants due to the research setting) unequivocally apply to the papers discussed in this review. Hence, (longitudinal) observational studies are required to assess whether the experimental findings indeed apply in real-life encounters. Additionally, the studies reviewed are no different from the vast majority of experimental studies, in the sense that a highly selective sample is used. When the complexity of the experiment and the understanding of the participants is indeed related to the level of cooperation observed, the use of, inherently highly educated, university students may provide overly optimistic estimates of the effects of embeddedness. Yet, recent studies involving participants recruited from online labour markets that yield a more representative sample (e.g., Rand, 2012; Difallah, Filatova, & Ipeirotis, 2018), such as Amazon’s Mechanical Turk, do not report fundamentally different results (e.g., Rand et al., 2011; Mao et al., 2017; Pfeiffer et al., 2012; Wang et al., 2012).

Employing different research designs could possibly overcome some of the limitations inherent to the experimental method, but comes with new challenges. Regarding these different research designs as competitors is unlikely to foster the advancement of science. Rather, it may prove fruitful to integrate the evidence collected under a variety of research designs in a common framework (Buskens & Raub, 2013; Jackson & Cox, 2013; Munafò & Smith, 2018). Namely, a robust line of evidence is necessarily built by combining various ways of testing the same hypotheses, using different sources of data and different methodologies. Recent statistical advances allow for such a unifying approach. Using Bayesian Evidence Synthesis, the support over studies for a given theory, assessed by multiple, different hypotheses concerning the same underlying concept, can be quantified (Kuiper, Buskens, Raub, & Hoijtink, 2013). This approach allows to expand the classical meta-analytic way of thinking that focuses on statistical reliability, by combining a broader range of relevant studies, leading to a statistical synthesis of conclusions from diverse sources.

There are several directions in which one could proceed using this approach. Consider, for example, the effect of control under exogenously imposed network embeddedness. The studies

considered in this review operationalized network embeddedness rather differently, with different network sizes, possibly simultaneous with dyadic embeddedness, in different games, and with varying amounts of experience of the subjects. Hence, obtaining a single estimate of network embeddedness in the sense of a meta-analysis is infeasible. However, if control through network embeddedness fosters cooperation as expected, combined evidence from multiple, diverse sources should confirm this. Under this wide variety of circumstances, one could statistically assess whether an increase in first round cooperation due to network embeddedness adequately summarizes the empirical results, as well as whether network embeddedness leads to a more severe end-game effect. Because both hypotheses essentially represent the effect of control through network embeddedness, one could, using Bayesian Evidence Synthesis, even combine the evidence for both distinct hypotheses. That is, one could quantify the support for both distinct hypotheses over multiple studies into a single measure, to assess whether control through network embeddedness indeed affects cooperation overall. Variations of this approach, such as testing interactions between various hypotheses, are possible, as well as extending the set of studies under consideration with observational studies, field experiments and vignette studies, to get a broad, statistically substantiated quantification of the overall hypothesis that network embeddedness fosters cooperation.

## 6 Literature

- Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The Economic Journal*, 103(418), 570–585. <https://doi.org/10.2307/2234532>
- Axelrod, R. M. (1984). *The evolution of cooperation*. New York: Basic Books, Inc.
- Barrera, D., & Buskens, V. (2009). Third-party effects. In *ETrust: Forming relationships in the online world* (pp. 37–72). Retrieved from <http://www.jstor.org/stable/10.7758/9781610446082.6>
- Bereby-Meyer, Y., & Roth, A. E. (2006). The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. *American Economic Review*, 96(4), 1029–1042. <https://doi.org/10.1257/aer.96.4.1029>
- Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? An experimental investigation. *Management Science*, 50(11), 1587–1602. <https://doi.org/10.1287/mnsc.1030.0199>
- Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions.

- Econometrica*, 72(3), 747–780. <https://doi.org/10.1111/j.1468-0262.2004.00511.x>
- Buskens, V. (2003). Trust in triads: Effects of exit, control, and learning. *Games and Economic Behavior*, 42(2), 235–252. [https://doi.org/10.1016/S0899-8256\(02\)00563-8](https://doi.org/10.1016/S0899-8256(02)00563-8)
- Buskens, V., & Raub, W. (2002). Embedded trust: Control and learning. In *Advances in Group Processes: Vol. 19. Advances in Group Processes* (pp. 167–202). [https://doi.org/10.1016/S0882-6145\(02\)19007-2](https://doi.org/10.1016/S0882-6145(02)19007-2)
- 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.), *The handbook of rational choice social research* (pp. 113–150). Stanford, California: Stanford University Press.
- Buskens, V., Raub, W., & van der Veer, J. (2010). Trust in triads: An experimental study. *Social Networks*, 32(4), 301–312. <https://doi.org/10.1016/j.socnet.2010.05.001>
- Buskens, V., & Weesie, J. (2000a). An experiment on the effects of embeddedness in trust situations: Buying a used car. *Rationality and Society*, 12(2), 227–253. <https://doi.org/10.1177/104346300012002004>
- Buskens, V., & Weesie, J. (2000b). Cooperation via social networks. *Analyse & Kritik*, 22(1), 44–74. <https://doi.org/doi:10.1515/auk-2000-0103>
- Camerer, C., & Weigelt, K. (1988). Experimental tests of a sequential equilibrium reputation model. *Econometrica*, 56(1), 1–36. <https://doi.org/10.2307/1911840>
- Carpenter, J., Connolly, C., & Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics*, 11, 282–298. <https://doi.org/10.1007/s10683-007-9193-x>
- Chaudhuri, A. (2009). *Experiments in economics: Playing fair with money*. Routledge.
- Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics*, 14(1), 47–83. <https://doi.org/10.1007/s10683-010-9257-1>
- Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1996). Cooperation without reputation: Experimental evidence from prisoner’s dilemma games. *Games and Economic Behavior*, 12(2), 187–218. <https://doi.org/10.1006/game.1996.0013>
- Corten, R., Buskens, V., & Rosenkranz, S. (2020). Cooperation, reputation effects, and network dynamics: Experimental evidence. In V. Buskens, R. Corten, & C. Snijders (Eds.), *Advances in the sociology of trust and cooperation* (pp. 391–416). <https://doi.org/10.1515/9783110647495-017>
- Corten, R., Rosenkranz, S., Buskens, V., & Cook, K. S. (2016). Reputation effects in social networks do not promote cooperation: An experimental test of the raub & weesie model. *PLOS ONE*,

- 11(7), 1–17. <https://doi.org/10.1371/journal.pone.0155703>
- Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *The American Economic Review*, 95, 1591–1604. <https://doi.org/10.1257/000282805775014434>
- Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101(1), 411–429. <https://doi.org/10.1257/aer.101.1.411>
- Dal Bó, P., & Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56, 60–114. <https://doi.org/10.1257/jel.2016.0980>
- Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of mechanical turk workers. *Proceedings of the eleventh acm international conference on web search and data mining*, 135–143. <https://doi.org/10.1145/3159652.3159661>
- Dreber, A., Fudenberg, D., & Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, 98, 41–55. <https://doi.org/10.1016/j.jebo.2013.12.007>
- Embrey, M., Fréchette, G. R., & Yuksel, S. (2018). Cooperation in the Finitely Repeated Prisoner’s Dilemma. *The Quarterly Journal of Economics*, 133(1), 509–551. <https://doi.org/10.1093/qje/qjx033>
- Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535–538. <https://doi.org/10.1126/science.1168244>
- Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, 114(3), 817–868. <https://doi.org/10.1162/003355399556151>
- Frey, V., Buskens, V., & Corten, R. (2019). Investments in and returns on network embeddedness: An experiment with trust games. *Social Networks*, 56, 81–92. <https://doi.org/10.1016/j.socnet.2018.07.006>
- Friedman, D., & Oprea, R. (2012). A continuous dilemma. *American Economic Review*, 102(1), 337–363. <https://doi.org/10.1257/aer.102.1.337>
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510. Retrieved from <https://www.jstor.org/stable/2780199>
- Hauk, E. (2003). Multiple prisoner’s dilemma games with (out) an outside option: An experimental

- study. *Theory and Decision*, 54(3), 207–229. <https://doi.org/10.1023/A:1027385819400>
- Hauk, E., & Nagel, R. (2001). Choice of partners in multiple two-person prisoner’s dilemma games: An experimental study. *Journal of Conflict Resolution*, 45(6), 770–793. <https://doi.org/10.1177/0022002701045006004>
- Hayashi, N., Ostrom, E., Walker, J., & Yamagishi, T. (1999). Reciprocity, trust, and the sense of control: A cross-societal study. *Rationality and Society*, 11, 27–46. <https://doi.org/10.1177/104346399011001002>
- Hirschman, A. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states*. Cambridge, Massachusetts: Harvard Univ. Press.
- Jackson, M., & Cox, D. R. (2013). The principles of experimental design and their application in sociology. *Annual Review of Sociology*, 39(1), 27–49. <https://doi.org/10.1146/annurev-soc-071811-145443>
- Kandori, M. (1992). Social Norms and Community Enforcement. *The Review of Economic Studies*, 59(1), 63–80. <https://doi.org/10.2307/2297925>
- Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology*, 100(2), 313–345. <https://doi.org/10.1086/230539>
- Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 24(1), 183–214. <https://doi.org/10.1146/annurev.soc.24.1.183>
- Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners’ dilemma. *Journal of Economic Theory*, 27(2), 245–252. [https://doi.org/10.1016/0022-0531\(82\)90029-1](https://doi.org/10.1016/0022-0531(82)90029-1)
- Kreps, D. M., & Wilson, R. (1982). Reputation and imperfect information. *Journal of Economic Theory*, 27(2), 253–279. [https://doi.org/10.1016/0022-0531\(82\)90030-8](https://doi.org/10.1016/0022-0531(82)90030-8)
- Kuiper, R. M., Buskens, V., Raub, W., & Huijink, H. (2013). Combining statistical evidence from several studies: A method using bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods & Research*, 42(1), 60–81. <https://doi.org/10.1177/0049124112464867>
- Luce, R. D., & Raiffa, H. (1989). *Games and decisions: Introduction and critical survey* (Rev. ed.). Courier Corporation.
- Mao, A., Dworkin, L., Suri, S., & Watts, D. J. (2017). Resilient cooperators stabilize long-run cooperation in the finitely repeated prisoner’s dilemma. *Nature Communications*, 8(1), 1–10.

<https://doi.org/10.1038/ncomms13800>

- Milinski, M., Semmann, D., Bakker, T. C. M., & Krambeck, H.-J. (2001). Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1484), 2495–2501. <https://doi.org/10.1098/rspb.2001.1809>
- Morris, M. W., Sim, D. L. H., & Girotto, V. (1998). Distinguishing sources of cooperation in the one-round prisoner’s dilemma: Evidence for cooperative decisions based on the illusion of control. *Journal of Experimental Social Psychology*, 34(5), 494–512. <https://doi.org/10.1006/jesp.1998.1361>
- Morris, Z. (2014). Chapter 8 - laboratory experiments in sociology. In M. Webster & J. Sell (Eds.), *Laboratory experiments in the social sciences (second edition)* (Second Edition, pp. 183–197). <https://doi.org/10.1016/B978-0-12-404681-8.00008-X>
- Munafò, M. R., & Smith, G. D. (2018). Robust research needs many lines of evidence. *Nature*, 553(7689), 399–401. <https://doi.org/10.1038/d41586-018-01023-3>
- Murnighan, J. K., & Roth, A. E. (1983). Expecting continued play in prisoner’s dilemma games: A test of several models. *Journal of Conflict Resolution*, 27(2), 279–300. <https://doi.org/10.1177/0022002783027002004>
- Neral, J., & Ochs, J. (1992). The sequential equilibrium theory of reputation building: A further test. *Econometrica*, 60, 1151–1169. <https://doi.org/10.2307/2951542>
- Nowak, M. A. (2012). Evolving cooperation. *Journal of Theoretical Biology*, 299, 1–8. <https://doi.org/10.1016/j.jtbi.2012.01.014>
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291–1298. <https://doi.org/10.1038/nature04131>
- Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action. *The American Political Science Review*, 92(1), 1–22. <https://doi.org/10.2307/2585925>
- Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., & Szolnoki, A. (2017). Statistical physics of human cooperation. *Physics Reports*, 687, 1–51. <https://doi.org/10.1016/j.physrep.2017.05.004>
- Pfeiffer, T., Tran, L., Krumme, C., & Rand, D. G. (2012). The value of reputation. *Journal of the Royal Society Interface*, 9(76), 2791–2797. <https://doi.org/10.1098/rsif.2012.0332>
- Rand, D. G. (2012). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172–179. <https://doi.org/10.1016/j.jtbi.2012.01.014>

- Rand, D. G., Arbesman, S., & Christakis, N. A. (2011). Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, 108(48), 19193–19198. <https://doi.org/10.1073/pnas.1108243108>
- Rapoport, A. (1997). Order of play in strategically equivalent games in extensive form. *International Journal of Game Theory*, 26(1), 113–136. <https://doi.org/10.1007/BF01262516>
- Raub, W., & Weesie, J. (1990). Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, 96(3), 626–654. <https://doi.org/10.1086/229574>
- Riedl, A., & Ule, A. (2002). *Exclusion and cooperation in social network experiments*. Retrieved from [https://www.researchgate.net/profile/Arno-Riedl/publication/237666617\\_Exclusion\\_and\\_Cooperation\\_in\\_Social\\_Network\\_Experiments/links/02e7e52d48a2e40046000000/Exclusion-and-Cooperation-in-Social-Network-Experiments.pdf](https://www.researchgate.net/profile/Arno-Riedl/publication/237666617_Exclusion_and_Cooperation_in_Social_Network_Experiments/links/02e7e52d48a2e40046000000/Exclusion-and-Cooperation-in-Social-Network-Experiments.pdf)
- Roth, A. E., & Murnighan, J. (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, 17(2), 189–198. [https://doi.org/10.1016/0022-2496\(78\)90030-5](https://doi.org/10.1016/0022-2496(78)90030-5)
- Schneider, F., & Weber, A., Roberto. (2013). *Long-term commitment and cooperation* (Working Paper No. 130). Retrieved from University of Zurich Department of Economics website: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2334376](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2334376)
- Schuessler, R. (1989). Exit threats and cooperation under anonymity. *Journal of Conflict Resolution*, 33(4), 728–749. <https://doi.org/10.1177/0022002789033004007>
- Seinen, I., & Schram, A. (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review*, 50(3), 581–602. <https://doi.org/10.1016/j.euroecorev.2004.10.005>
- Selten, R. (1978). The chain store paradox. *Theory and Decision*, 9(2), 127–159. <https://doi.org/10.1007/BF00131770>
- Sigmund, K. (2012). Moral assessment in indirect reciprocity. *Journal of Theoretical Biology*, 299, 25–30. <https://doi.org/10.1016/j.jtbi.2011.03.024>
- Simpson, B., & Willer, R. (2015). Beyond altruism: Sociological foundations of cooperation and prosocial behavior. *Annual Review of Sociology*, 41(1), 43–63. <https://doi.org/10.1146/annurev-soc-073014-112242>
- Snijders, C., & Keren, G. (2001). *Do you trust? Whom do you trust? When do you trust?* In *Advances in Group Processes: Vol. 18* (pp. 129–160). <https://doi.org/10.1016/S0882->



- Sokolova, A., Buskens, V., & Raub, W. (2021). *Cooperation through investments in repeated interactions and contractual agreements: An experimental study*. Unpublished manuscript.
- Van Miltenburg, N., Buskens, V., & Raub, W. (2012). Trust in triads: Experience effects. *Social Networks*, 34(4), 425–428. <https://doi.org/10.1016/j.socnet.2012.01.006>
- Wang, J., Suri, S., & Watts, D. J. (2012). Cooperation and assortativity with dynamic partner updating. *Proceedings of the National Academy of Sciences*, 109(36), 14363–14368. <https://doi.org/10.1073/pnas.1120867109>
- Yamagishi, T., Hayashi, N., & Jin, N. (1994). Prisoner’s dilemma networks: Selection strategy versus action strategy. In U. Schulz, W. Albers, & U. Mueller (Eds.), *Social Dilemmas and Cooperation* (pp. 233–250). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the united states and japan. *Motivation and Emotion*, 18(2), 129–166. <https://doi.org/10.1007/BF02249397>
- Zhang, B.-Y., Fan, S.-J., Li, C., Zheng, X.-D., Bao, J.-Z., Cressman, R., & Tao, Y. (2016). Opting out against defection leads to stable coexistence with cooperation. *Scientific Reports*, 6(1), 1–7. <https://doi.org/10.1038/srep35902>

## A Appendix 1 - Thesis proposal

In the past review, it became apparent that dyadic embeddedness consistently fosters cooperation. However, despite similar theoretical predictions (see Section ??), network embeddedness proves much less efficient in fostering embeddedness. In fact, it is questionable to what extent network embeddedness fosters cooperation at all. The first goal of the thesis will be to statistically summarize the available evidence on control effects through network embeddedness, using past experimental studies of which available data is present (Barrera & Buskens, 2009; Buskens et al., 2010; Corten et al., 2016; Frey et al., 2019; Van Miltenburg et al., 2012). Data from other studies (e.g., Bolton et al., 2004) can be a great addition, but the current project does not depend on the availability of additional data. In these studies, specific predictions that represent control effects through network embeddedness can be tested, such as an increase in first round cooperation rates under network embeddedness, a positive effect on cooperation of the number of rounds that are to be played (after controlling for learning) and an end-game effect that starts later under network embeddedness. In the thesis, I will use Bayesian Evidence Synthesis (Kuiper et al., 2013) to quantify the overall



support for these hypotheses separately, and I will combine the evidence for all these hypotheses into a single measure, as the distinct hypotheses all represent effectively the same mechanism.

Additionally, the characteristics of a relation that are argued to foster cooperation, such as the length of a game and the temptation to defect, are thoroughly tested under dyadic embeddedness (Dal Bó, 2005; Dal Bó & Fréchette, 2011; Embrey et al., 2018). Network embeddedness yields similar predictions, but such predictions, especially on the length of the supergame, are hardly tested. If time and resources allow to do so, it could be of great value to design a new experiment, in which it can be tested whether the conditions that are well-known to foster cooperation under dyadic embeddedness have a similar effect under network embeddedness.<sup>1</sup> Additionally, the current literature review provided a preliminary indication that how one's past action are translated into a reputation (a complete history of past actions, a reputation score depending on some rule that is updated every round, a summary of past actions) affects cooperation. If the complexity of the dynamics of the game indeed affect cooperation, it can be expected that more complex reputation measures foster cooperation less than simpler alternatives, even though simpler measures may come with additional complications (see e.g., Nowak & Sigmund, 2005).

This set-up might lead to a new experiment with a three-way full factorial design, in which the length of the game, the temptation to defect and the reputation measure are varied. One could think of an experiment in which three continuation probabilities (e.g.,  $\delta \in \{0, 1/2, 3/4\}$ ) are crossed with two different gains (losses) of unilateral defection (cooperation) and two measures of reputation. For the duration of the supergame, subjects can be matched randomly every round, but take their reputation with them, until the supergame ends. Subjects can play multiple supergames until, say, 1.5 hours passed. This set up allows subjects to gain familiarity with the dynamics of the game. The results from this experiment can be incorporated in the synthesis of all available evidence.

---

<sup>1</sup>This new experiment idea is not discussed at all with my supervisors, but it came up while writing the conclusion, after which I had no time to discuss any of it. The primary goal of the thesis will be a synthesis of network control effects, as outlined in the previous paragraph, and this section is a possible extension of the project.