Literature Review

Control effects on cooperation in embedded interactions

Thom Volker (5868777)

Supervisors: Vincent Buskens & Werner Raub

Sociology and Social Research Master's programme

Utrecht University

May 25, 2021

Word count: 6259

Contents

1 Key words

Key words:

- Dyadic + network embeddedness (focus on network embeddedness for thesis itself).
- Use of own (readily available) data
- Try to align the M&S project and SaSR project as much as possible.
- Endogenous + exogenous (might not be fruitful in the thesis itself).
- Note that the two theses are to be written in parallel, and have to be finished in May 2022.

2 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). In these situations, for all individuals involved it would be rational to behave opportunistically. However, the rational decision to behave opportunistically would yield lower collective returns than what could have been achieved under mutual cooperation: hence the term "social dilemma" (Kollock, 1998; Ostrom, 1998). Under the assumption of "social isolation", that is, the interacting actors can be considered perfect strangers that do not anticipate any future interactions, actors can maximize their individual returns by behaving uncooperatively.

Theoretically, it is well established that in isolated social dilemmas, it is generally hard to initiate a cooperative relationship (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. Regardless of the choice of one's partner, an individual actor obtains the highest payoff by acting opportunistically. Defecting, rather than cooperating, when the other player cooperates leads to a higher payoff, similarly to defecting when the other player defects as well. Hence, the Nash equilibrium is mutual defection, even though both players would be better off with mutual cooperation. Contrary to these ominous theoretical findings, in practice researchers generally find non-negligible rates of cooperation in one-shot games (e.g., see Hayashi, Ostrom, Walker, & Yamagishi, 1999; Cooper, DeJong, Forsythe, & Ross, 1996; 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., Buskens & Raub, 2002; Yamagishi & Yamagishi, 1994). These 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). Obviously, actors can be embedded both dyadically and in a network at the same time. Both forms of embeddedness potentially affect cooperative behaviour in social dilemmas substantially.

In the subsequent sections of this literature review, it will be outlined how embeddedness can affect trust, and cooperative behaviour in general, with a specific focus on control effects through network embeddedness from a game-theoretic perspective.[**IS IT TOO EARLY TO MENTION CONTROL HERE?**] Notably, game-theoretic assumptions yield that actors maximize their utilities. Yet, in research settings, it is often complicated to infer the utilities of the participants at hand, as it is not explicit how they value potential payoffs as specified in the study at hand. Generally, researchers try to overcome this problem by incentivizing the behaviour of participants by offering monetary rewards in the form of points that are translated into money at the end of the experiment. The amount of points one earns depend on the behaviour of oneself, as well as on the behaviour of the others one plays the game with. Although these monetary payoffs do not diminish the possibility that subjects may strive for non-materialistic goals, it is assumed throughout that incentivizing the payoffs allows to interpret the payoffs as the actors' utilities. Furthermore, the empirical findings discussed in this review generally relied on 2-person Prisoner's Dilemma games or Trust Games. That is, in general, experiments will be considered in which 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 of 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 current findings, the implications of these findings and possible directions for future research.

3 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., Buskens & Raub, 2013; Yamagishi & Yamagishi, 1994). Control, which will be the focus of this review, denotes the opportunity to sanction opportunistic behaviour by exerting control over one's partner's long-term returns. 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). This mechanism can also be referred to as control through "voice", in the sense of Hirschman (1970). Hence, the short-term benefits of acting opportunistically 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, & Girotto, 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 finding 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, on the contrary, refers to the situation where the actors involved have interacted in the past, when they are embedded dyadically, and hereby gained information about each others' past behaviour (Buskens & Raub, 2013). 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 behave cooperatively 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 defect in the current interaction as well.

A second distinction that can be made relates to the nature of the embeddedness of a transaction. Namely, rather often, researchers decide who will interact with whom in an experiment, a situation that is commonly referred to as *exogenous embeddedness*. However, in real life, people often choose with whom they wish to interact, at least to a certain extent (Chaudhuri, 2009; Yamagishi et al., 1994). Some researchers tried to incorporate this characteristic of real-life encounters in their research by letting participants choose their transaction partners, which is referred to as *endogenous embeddedness*. Overall, it appeared that endogenously formed relations tend to have a larger effect on cooperation rates than exogenously formed relationships (e.g., Chaudhuri, 2011; Frey, Buskens, & Corten, 2019; Gürerk, Irlenbusch, & Rockenbach, 2014; Schneider & Weber, 2013; Wang, Suri, & Watts, 2012).

The empirical findings that are to be discussed in the next section 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 outside the game-theoretic paradigm 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. 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. The second way, that 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. 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.

4 Control effects in dyadic relations

Control effects in dyadic relations can be studied in both finitely and infinitely repeated interactions. In finitely repeated interactions, however, it follows from backward induction that, under the assumption of game-theoretic rationality, no cooperation is possible (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 always yields a higher payoff than cooperation. As actors decided to defect in the final round, actions in the penultimate round do not affect behaviour in the last round, and again defection is the payoff maximizing strategy. This pattern repeats itself to the first round of the game, and hence, under the assumption of rationality, no cooperation is possible in any round of the game. A wide variety of experiments however showed that subjects act cooperatively in initial rounds, leading Rapoport (1997, p. 122) to conclude that in practice, subjects do not rely on, or are not capable of backward induction.

Kreps, Milgrom, Roberts, & Wilson (1982) proved that even populations of rational actors could maintain high levels of cooperation, under the assumption that these rational actors belief with sufficiently high probability that their transaction partners have no incentive to defect, until their partners are defected on themselves. If the rational actors believe that the probability to meet such a conditional cooperator is high enough, the benefits of mutual cooperation outweigh the gains of exploiting a conditional cooperator once 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, it is immediately known to the other player that the defecting actor is not a conditional cooperator, and the finitely repeated game would be one of mutual defection hereafter if the players are rational. Yet, as long as both players cooperate, it is not 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. In the final rounds of the finitely repeated game however, the long term benefits of mutual cooperation may no longer outweigh the short-term costs of maintaining one's reputation (i.e., the opportunities to control future behaviour of one another diminish), and hence rational actors will try to exploit their partners in these rounds.

In infinitely, or indefinitely, repeated games, there is no end-game effect, as there is no predetermined final round. Namely, a game will be played for another round with a certain continuation probability δ, which Axelrod (1984) aptly termed the "shadow of the future", and end with probability $(1 - \delta)$. In these games, cooperation can supported in equilibrium if the continuation probability is large enough (see e.g. Buskens & Raub, 2013 for the technical details). Comparing cooperation rates in the first round of an infinitely repeated game to cooperation rates in a one-shot game or in games with a different continuation probability allows to assess the effect of control. Namely, in a one-shot game, there is no possibility to sanction opportunistic behaviour, and thus there is no control, while in repeated games with different continuation probabilities, the control opportunities differ as well. Although the analysis of first round behaviour is merely a practical issue, that is, due to the continuation probability δ there is no guarantee that a second round will be played, it is beneficial to disentangle learning and control. Namely, in the first round of a finitely repeated game, no learning can have occurred yet.

4.1 Control in exogenously dyadically embedded interactions (repeated games)

In exogenously formed finitely repeated games, one can generally observe a pattern of high cooperation in the initial rounds of a finitely repeated game, with a sharp decrease in cooperation toward the end of the game (e.g., Buskens, Raub, & Van der Veer, 2010; Embrey, Fréchette, & Yuksel, 2018; Mao, Dworkin, Suri, & Watts, 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. First, defection could be a response to defection of one's partner in an earlier rounds. Second, an actor may have learned in previous games that in the final rounds of the repeated 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 study the presence of control effects in a Trust Game, which slightly changes the nature of the game relative to a Prisoner's Dilemma. Rather than mutually risking possible exploitation, in the Trust Game 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. Hence, the possible actions of the trustee depend on the initial choice of the trustor. Under dyadic embeddedness, Buskens et al. (2010) find that after controlling for learning effects, there is a positive effect of the number of rounds still to be played (i.e., control opportunities) on cooperation, for both the trustor and the trustee. Additionally, these authors study the effect of dyadic embeddedness when the dyadic relation is embedded in a small network, where two trustors are in a relationship with a single trustee (i.e., a triad). In this network, information can be exchanged between the first trustor and the second trustor, and hence, the trustors can learn about the trustee not only from their own game with the trustee, but also from the game of the other trustor in the triad with the same trustee. Also in this condition, the number of rounds that has to be played between a trustor and a trustee is positively related to cooperation for both trustor, and trustee, after controlling for learning effects in terms of past moves of the trustee in the game with the trustor at hand, as well as in terms of past moves of the trustee in the game with the other trustor.

Other studies corroborate the finding that the number of rounds that are to be played affect cooperation rates. Embrey et al. (2018) performs 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, show that first round cooperation rates, where no learning could have occurred, increases with the length of the game. Additionally, Van Miltenburg et al. (2012) shows that dyadically embedded actors learn to play cooperatively in initial rounds of a finitely repeated Trust Game of fifteen rounds, with near perfect cooperation in the first round of later played games that remains above 80% until the twelfth round, but decreases to negligible levels shortly hereafter. However, no explicit distinction between learning and control effects is made in the analyses. Mao et al. (2017) make the same observation that people learn to play cooperatively in early rounds of a finitely repeated Prisoner's Dilemma game, with near perfect cooperation in the first round that remains above 80% but decreases to negligible levels in the final two rounds, although again no explicit distinction between control and learning effects is made. Remarkably, this experiment is repeated on twenty consecutive days, with twenty 10-round games per day, and shows that this pattern keeps repeating itself throughout the study period, regardless of the fact that actors experience that in final rounds cooperation seldom prevails.

Research into cooperation in infinitely repeated games was initiated by Roth & Murnighan (1978) and Murnighan & Roth (1983), who found on average somewhat higher cooperation rates under a high continuation probability relative to under low continuation probabilities. However, this increase was generally small, leading the authors to conclude that the introduction of infinitely repeated interactions hardly fosters cooperation. However, these authors let the participants play against the experimenters, rather than against each other. Additionally, these studies did not translate the amount of points participants earned during the experiment linearly to a monetary reward, which potentially affects the incentives of the subjects to strive for a higher number of points. Dal Bó (2005) improves upon this initial work in infinitely repeated Prisoner's Dilemma games, and finds significant differences between three continuation probabilities (i.e., $\delta = 0, 1/2, 3/4$), such that first round cooperation rates increase with the continuation probability. Particularly, the differences become larger with the experience participants gained. Hence, the prospect of future benefits through mutual cooperation and the fear of missing out on these benefits after initiating defection seem to result in a higher willingness to cooperate.

Similar observations were made by Dal Bó & Fréchette (2011) and Dal Bó & Fréchette (2018). These authors argue that the limited effect of increasing the continuation probability in earlier studies is due to the fact that participants have to gain experience to properly evaluate the effect of this increase. With sufficient experience, participants generally cooperate when cooperation can be supported in equilibrium, and defect when it cannot. Moreover, 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, especially after the subjects have gained experience with the game they play (Dal Bó, 2005; Dal Bó & Fréchette, 2011, 2018). Namely, in the finitely repeated games, in earlier games the subjects experienced that cooperation will be abused in the final round of the game, and protect themselves against this behaviour by defecting in an earlier stage already. No such behaviour is present in infinitely repeated games; in fact, the cooperation rates tend to increase with the experience of the participants. Notably, because the defecting behaviour of most participants in final rounds of a finitely repeated game implies that future sanctions for behaviour in the penultimate round are not credible, subjects experience less control opportunities in finitely repeated games relative to infinitely repeated games of the same expected length. Consequently, first round cooperation in finitely repeated games may decrease, due to this reason.

Although the evidence consistently shows that first round cooperation increases with the (expected) number of rounds left, there remains substantial heterogeneity in cooperation rates, even in games of approximately the same length. Whereas Cooper et al. (1996) find cooperation rates in finitely repeated games of around 65%, Embrey et al. (2018) find above 80% cooperation and Mao et al. (2017) observes near perfect cooperation, all in the first round, even though all games have a fixed end at round 10 and differences in the payoff matrix are only minor. Differences in economic incentives, however, are likely to affect this difference. In the study by Cooper et al. (1996), only relative payoffs were important, as the player with the highest number of points in each game received a monetary payoff of $1 at the end of each game, while the losing partner received nothing. In Embrey et al. (2018) and Mao et al. (2017), players' monetary payoffs were linearly derived from the number of points they gathered, neglecting relative standings and emphasizing the importance of receiving a high number of points.

4.2 Control in endogenously dyadically embedded interactions (repeated games)

The previous section shows that in exogenously embedded dyadic interactions there is a relatively consistent effect of control on cooperation. However, a substantial portion of real-life interactions, dyadic embeddedness is not imposed exogenously, but rather established intentionally by the actors involved (Kollock, 1994). 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). 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 build by behaving cooperatively. In fact, even at the cost of a portion of subjects' payoffs, they 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).

When dyadic embeddedness is established endogenously, however, not necessarily the effect of control induces more cooperation, as there might also be a selection effect at work. Then, people who are more willing to cooperate are more likely to engage in repeated interactions, and hence cooperation in repeated interactions is higher. Schneider & Weber (2013) show that people who *a priori* indicate a higher willingness to cooperate were indeed more likely to engage in repeated interactions, although this characteristic only explained part of the increase in first round cooperation rates relative to one-shot games. Again, it was shown, however, that a longer, endogenously chosen, duration yields more first round cooperation, similarly to the findings under exogenous dyadic embeddedness (Schneider & Weber, 2013), which is indicative for a control effect. Additionally, Brown et al. (2004) find that, already in the first round, actors are more trustful and trustworthy when the possibility of endogenously chosen repeated interactions exist, relative to when continuing an interaction is not possible, regardless of whether an interaction is indeed continued. Note that in this experiment, actors are randomly assigned to a treatment where endogenously chosen repeated interactions are possible or a treatment when such interactions are not possible. Hence, the prospect of a mutually cooperative interaction already induces more cooperation. 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.

[Potentially add a section containing the studies by Schuessler (1989), Jin, Hayashi & Shinotsuka (1993), Hauk & Nagel (2001), Hauk (2003) and Zhang, Fan, Li, Zheng, Bao, Cressman, & Tao (2016), who additionally consider opting out of a repeated interaction. For now I focused on network embeddedness first, but given that there are still three weeks left, I might add this. At first, I thought that these experiments concerned dyadic embeddedness, because interacting actors obtain information only about there own interactions with a given partner and not about interactions by any other player, and also do not build any kind of reputation through which sanctions by third parties can be implemented. Yet, in Werner's farewell lecture I saw that this is discussed as a part of network embeddedness, because the actor's are embedded in a network in the sense that after being defected, they can approach a new transaction partner that "shares" the common network. Yet, I had the impression that sanctions were implemented dyadically, because defecting actors are left by an actor willing to cooperate.]

5 Control effects of network embeddedness

In the previous section, it is shown that in a wide variety of research settings, the potential to engage in future interactions with a given partner drastically improves cooperation rates. However, this research 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). If reliable information on an actor's behaviour spread through a network, a rational actor may be more inclined to behave cooperatively, because failing to do so would provide others with good reasons to refrain from cooperating, negatively affecting one's future returns. A rational actor would foresee that defecting in a given game may have severe drawbacks during later interactions, and hence will be more inclined to cooperate. That is, the actor can be *controlled*, due to the prospect of future interactions. Note however that making assumptions on the flow of reliable information about actors' behaviour may be problematic, as actors may face strategic incentives to share misleading information, and whether or not share information on someone's behaviour is itself a (second-order) social dilemma. These difficulties are generally ignored in research on network embeddedness.

Similarly to dyadic embeddedness, control effects under network embeddedness can be studied in finitely as well as in infinitely repeated games, where actors are matched exogenously or endogenously. Theoretical contributions showed that in an infinitely repeated game setting, a higher expected number of interactions and a faster flow of information (i.e., a denser network) render cooperation more likely (Buskens & Weesie, 2000b; Raub & Weesie, 1990). Yet, most experimental studies on network embeddedness employ a finitely repeated game design. Buskens (2003) developed a theoretical model that closely resembles the model by Kreps et al. (1982), including the assumption that actors belief with sufficiently large probability that some trustees will not abuse trust, but concerns interacting triads in which two trustors and one trustee play finitely repeated Trust Games. Although the model concerns a small network, the conclusion that network embeddedness fosters cooperation still holds. All these models rely on the idea that reliable information on misbehaving actors spread through the network, providing others with good reasons to refrain from cooperating with a defecting actor.

5.1 Control in exogenously embedded interactions in networks

Whereas dyadic embeddedness generally leads to substantial higher cooperation rates due to control effects relative to atomized interactions, the effects of network embeddedness seem at best mixed. Corten (2016) finds that having information about all interactions in a network of six actors does not improve cooperation rates among randomly matched actors at all, relative to a condition where subjects are only provided information on their own interactions. Even though no explicit distinction between control and learning is made, cooperation in initial rounds of the game is also not higher under network embeddedness. If there was an effect of network embeddedness on cooperation at all, this effect was more towards the negative side. Similar observations were made by Barrera & Buskens (2009), although here network embeddedness was studied relative to dyadic embeddedness in Investment Games played in triads with a focus on the behaviour of trustors. Especially the trustors do not seem to realize that network embeddedness allows for more severe punishments if trust is abused.

More rigorous testing on the effects of control under exogenous network embeddedness has been done by Buskens et al. (2010), who study Trust Games in triads. The authors find no statistically significant effect of exogenous network control on trustfulness of the trustors, although first round trustfulness rates are somewhat higher under network embeddedness. Studies by Van Miltenburg et al. (2012) and Frey et al. (2019) observe similar behaviour of trustors, although no explicit tests of control effects were made. Hence, trustors seem to have difficulties with realizing that sanction opportunities accumulate as there are additional trustors involved that can punish untrustworthy behaviour. Trustees, however, seem to clearly realize that sanctioning under network embeddedness may be more harmful than under dyadic embeddedness. Buskens et al. (2010) find significant network control effects on the trustworthiness of the trustee, while Frey et al. (2019) graphically show that first-round trustworthiness increases under exogenous network embeddedness. Van Miltenburg et al. (2012) does not seem to show any effects of network control, but in this study dyadic embeddedness on itself already resulted in rates of trustworthiness mainly above 0.80 until the last two rounds, after subjects gained experience. Hence, when network embeddedness was established exogenously, it hardly seemed to affect cooperation rates.

5.2 Control in endogenously embedded interactions in networks

Besides network control effects under exogenously embedded interactions, network control effects can be assessed under endogenously embedded interactions. 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. Frey et al. (2019) do study the effect of endogenous network embeddedness on cooperation in addition to (exogenous) dyadic embeddedness in Trust Games. Both network embeddedness established by the trustors, as well as network embeddedness established by the trustee, positively affects first-round trustfulness and trustworthiness rates, although no statistical test for this effect was performed. These choices likely represent the fact that these trustors and trustees understand the perks of embeddedness for establishing a cooperative relation. However, as either trustors or trustees had to invest a portion of their payoffs to establish embeddedness, there is likely to be a selection effect as well. Namely, the willingness to invest may signal once willingness to establish a cooperative relationship.

Corten et al. (2020), who studies the effect of dynamic networks of six actors on cooperation rates finds no support for an effect of network embeddedness, and hence no indication for a control effect. Note that in this study actors chose to engage in an interaction, and hence cooperation rates where defined as a proportion of the total number of games. When cooperation is assessed relative to the total number of 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, although the effect of network control was not specifically assessed. Additionally, Wang et al. (2012) shows that in N-person Prisoner's Dilemma games with dynamic partner updating, cooperation starts at higher levels than in fixed networks. In these interactions, actors can repeatedly select and deselect network members, resulting in the fact that having a poor reputation may result in ending up without (cooperative) interaction partners. Actors seem to realize this, and tend to start cooperatively, but start to defect in later stages.

READ Rapoport, Diekmann, and Franzen 1995

Overall, it appears that the findings on the effects of control through network embeddedness are, at best, mixed. Although some studies show some indication of control effects through network embeddedness, most studies do not find such effects. When control effects were found, it mainly affected those actors that were in the position to abuse trust (i.e., both players can abuse trust in a Prisoner's Dilemma game, but only trustees can abuse trust in a Trust Game). The positive effects of network embeddedness that are found seem to be more due to assortment than due to control. That is, cooperative actors may signal a willingness to cooperate, while actors that do not intend to cooperate do not. Additionally, when ties with actors that abuse trust can be easily abandoned, cooperative relationships seem to occur more frequently. An exception to this finding is the study by Corten et al. (2020). However, cooperation is often vulnerable, and an initial pattern of actors that try to exploit each other may easily result in a collapse of cooperation (Kollock, 1994). In such situations, the effect of experience may be particularly important. Ample studies show that initial games may fail to reach cooperative relationships, even though cooperation can be supported in equilibrium (e.g., Van Miltenburg et al., 2012; Dal Bó, 2005; Dal Bó & Fréchette, 2011). However, as subjects gain experience, cooperation may reach high levels, when it can be supported in equilibrium.

6 Implications for future research

In the simplest interactions, that is, one-shot interactions in a social vacuum, high levels of cooperation can hardly be sustained if games are repeated (Simpson & Willer, 2015). In this review, it became apparent that embeddedness can foster cooperation between actors due to the possibility to control future payoffs of one's interaction partner. The available evidence rather consistently shows that more control opportunities induce more cooperation. Although the introduction of embeddedness narrows the gap to real-life interactions, an overwhelming amount of the experiments conducted remains rather artificial in nature. This approach has clear advantages, as it allows to separate the slightest sources of variation that would be entangled in survey research and to assess the causal effect of changing certain features of an interaction on cooperation. However, these artificial experiments do little in explaining why humans behave as they behave in real-life encounters, as it is hardly assessed how taking people out of their social context affects their behaviour.

Axelrod Launching 'The Evolution of Cooperation' - "There were several limitations in using people - even skilled people - to study how to best play the Prisoner's Dilemma.

Why does dyadic embeddedness foster cooperation more than network embeddedness. Sigmund (2012) argues that under network embeddedness, many players seem to rely on first-order assessment. That is, actors base their decision to cooperate on the observed cooperativeness of their transaction partners in the past. While this assessment rule might work under dyadic embeddedness, where the reason of one's partner's action can be inferred, it does not do well under network embeddedness. However, tracking whether one's partner's past behaviour against third parties were the result earlier defections against this partner is generally difficult (Milinski, Semmann, Bakker, & Krambeck, 2001 is referred that higher order assessment is cognitively taxing).

Kollock (1994) on experiments: "[T]he typical social exchange experiment guarantees the terms of exchange. In one sense this is an understandable feature of an experimental paradigm if one wishes to study, for example, the dynamics of power in a controlled way, but in another sense it sidesteps much that is crucial in actual interaction (concerns of uncertainty, risk, and being taken advantage of) and thus much that may be central in explaining why social structure in the form of networks of commitment exists.

Most studies in this field of research are narrowed down to assess how small changes in conditions affect cooperation, providing insight in multiple potentially important mechanisms that affect cooperation. This has led to a wide variety of factors that in one way or another affect cooperation when studied in isolation. A numerous amount of studies focused on cooperation in dyadically embedded relations, formed either exogenously or endogenously, to investigate what aspects of an interaction may foster and complicate cooperation. Additionally, there are numerous studies that assess how network embeddedness fosters cooperation, most of them focusing on a relatively small aspect of the network. However, these studies suffer from two important limitations. First, most experiments rely on an artificial and often greatly simplified social context, regardless of whether the focus of the studies is dyadic or network embeddedness. Hence, the extent to which findings from these experiments can be translated to everyday life is often questionable. Second, focusing on relatively small changes between conditions may complicate an assessment of the bigger picture. Predominantly, factors that seem important for cooperation when studied in isolation, may appear trivial when multiple factors are assessed simultaneously.

The majority of experimental studies on dyadic embeddedness do not allow for leaving a defective relationship, with Brown et al. (2004) as one of the rare exceptions. See also Hauk (2003; https://link.springer.com/content/pdf/10.1023/A:1027385819400.pdf) Yet, this feature is often covered in studies on network embeddedness. See e.g. Buskens 2003 for a theoretical outline of the effect of leaving opportunities.

Most research on embeddedness assesses how different relations between actors, that is, dyadic versus through third parties and exogenous versus endogenous affects cooperation. However, in a substantial proportion of this research, the mechanisms that promote cooperation remain relatively implicit. Hence, it becomes clear that certain manipulations affect cooperative behavior, but not why exactly these manipulations work, or, alternatively, what aspects of these manipulations make these changes come about.

Rand (2012) from Nowok - Evolving cooperation (2012) suggests that online labor markets such as Amazon's Mechanical Turk provide a useful platform for conducting behavioural experiments.

7 Conclusion

Obviously, there are other explanations possible that allow to explain cooperative behavior, such as relaxing assumptions on the rationality and selfishness of individuals. However, this review deliberately aims to focus on the effect of embeddedness, and in particular on control, to explain cooperative behaviour.

8 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. Retrieved from http://www.jstor.org/stable/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

Brown, M., Falk, A., & Fehr, E. (2004). Relational contracts and the nature of market interactions. *Econometrica*, *72*(3), 747–780. https://doi.org/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/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

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*, 301–312.

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.

Buskens, V., & Weesie, J. (2000b). Cooperation via social networks. *Analyse & Kritik*, *22*(1), 44–74. https://doi.org/doi:10.1515/auk-2000-0103

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.

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.

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/https://doi.org/10.1515/9783110647495-017

Corten, S. A. B., Rense AND Rosenkranz. (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/https://doi.org/10.1257/jel.20160980

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/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

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.

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.

Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2014). On cooperation in open communities. *Journal of Public Economics*, *120*, 220–230. https://doi.org/https://doi.org/10.1016/j.jpubeco.2014.10.001

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*. Retrieved from http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+11303198X&sourceid=fbw_bibsonomy

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/https://doi.org/10.1016/0022-0531(82)90029-1

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.

Milinski, M., Semmann, D., Bakker, T. C., & 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.

Morris, M. W., Sim, D. L., & 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.

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

Nowak, M. A. (2012). Evolving cooperation. *Journal of Theoretical Biology*, *299*, 1–8. https://doi.org/https://doi.org/10.1016/j.jtbi.2012.01.014

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997. *The American Political Science Review*, *92*(1), 1–22. Retrieved from http://www.jstor.org/stable/2585925

Rapoport, A. (1997). Order of play in strategically equivalent games in extensive form. *International Journal of Game Theory*, *26*(1), 113–136.

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

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/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

Selten, R. (1978). The chain store paradox. *Theory and Decision*, *9*(2), 127–159.

Sigmund, K. (2012). Moral assessment in indirect reciprocity. *Journal of Theoretical Biology*, *299*, 25–30. https://doi.org/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/S0882-6145(01)18006-9

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.

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.

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.