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

April 29, 2021

Word count: 3163

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 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 with minimal effort by free-riding on the work of their peers, researchers could obtain yet another publication by letting their collaborators do the lion's share of the required work (Corten, Buskens, & Rosenkranz, 2020) and a car dealer may hide several vehicle defects when selling a second-hand car to a relatively uninformed customer (Buskens & Weesie, 2000). 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 opportunistically. Under these circumstances, mutual cooperation would yield higher collective returns, yet no individual actor has an incentive to deviate from behaving opportunistically; hence the term "social dilemma" (Kollock, 1998; Ostrom, 1998).

Theoretically, it is well established that in isolated social dilemmas, it is generally hard to establish 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 one-shot Prisoner's dilemma as depicted in Table 1. According to the pay-off scheme, regardless of the choice of the other player, an individual actor obtains the highest pay-off by acting opportunistically (i.e., choosing D). Defecting when the other player cooperates leads to an increase in pay-offs, 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.

Although some experiments find some cooperation in one-shot games (e.g., see Hayashi, Ostrom, Walker, & Yamagishi, 1999 for the Prisoner's Dilemma), most experiments have shown that the amount of cooperative behavior in such settings is limited, especially after players have gained

Table 1

Prisoner's Dilemma payoff matrix.

		Actor 2	
		C	D
Actor 1	C	$(7,7)$	$(0,10)$
	D	$(10,0)$	$(2,2)$

experience with one-shot games (e.g., Cooper, DeJong, Forsythe, & Ross, 1996; Dal Bó, 2005). Discouraged one might conclude that cooperation is not or hardly possible, because human beings would typically exploit one another whenever the opportunity arises. However, cooperative behaviour is not as seldom as this finding would make us believe, as illustrated by various real-life examples, ranging from the number of packages that are shipped and paid for each day to research projects that really are a joint effort of multiple scientists. The paramount question then is which conditions make that cooperative behaviour will not be exploited.

Critics of this initial theoretical model have rightly noted that most real-life interactions do not take place in social isolation, but are actually embedded (Axelrod, 1984; Granovetter, 1985). Embeddedness refers to the fact that the actors involved often share a common environment. These actors may have interacted in the past, and/or speculate on interacting again in the future, which is referred to as *dyadic embeddedness*. 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*. Obviously, actors can be embedded both dyadically as well as in a network at the same time.

Generally, there are two mechanisms through which embeddedness is likely to affect cooperation between actors: *control* and *learning* (e.g., Buskens & Raub, 2013; Yamagishi & Yamagishi, 1994). Control denotes the opportunity to sanction opportunistic behavior by exerting some control over one's partner's long-term returns. Under dyadic embeddedness, one can punish the defection of one's interaction partner in the previous interactions by refraining from cooperation in the current round and the future. Additionally, it might be possible to inform future transaction partners of a defecting actor, who can in turn refuse to cooperate with this actor. 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.

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 others' past behavior

(Buskens & Raub, 2013). If this behavior is cooperative, the actors may infer that they deal with a partner that will behave cooperatively again in the future. When the actors are embedded in a common network, this network may allow the actors to learn about each other's past behavior during interactions with third parties. Again, if both actors behaved cooperatively during interactions with each other or with third parties, both actors may infer that they deal with a transaction partner that will behave cooperatively in the current transaction as well. Obviously, when one of the actors has abused the cooperative behavior of the other actor, or of a third party, this other actor may decide that he will not take the risk of getting exploited, and defect in their 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). Some researchers tried to incorporate this characteristic of real-life encounters in their research, which is referred to as *endogenous embeddedness*. This distinction is considered because overall, it appeared that endogenously formed relations tend to have a larger effect on cooperation rates than exogenously formed relationships [**ADD REFERENCES**].

Substantively, the remainder of this paper will be concentrated around the the question to what extent control affects cooperation. A distinction will be made between control in dyadically embedded interactions and in network embedded interactions, as well as between exogenously and endogenously formed relationships. This question will be answered via a literature review, with a focus on experimental research that employed 2-person Prisoner's Dilemma games, Trust games or Dictator games. That is, in general, experiments will be considered in which the behavior of an actor only affects oneself, as well as the single actor towards whom this behavior is directed. When occasionally an N-person game (i.e., when the behavior of an actor is directed towards more than one others) is discussed, this will be explicitly mentioned.

Notably, the payoff schemes of the games discussed (e.g., Table 1) traditionally reflect the utilities of the actors. Yet, in research settings, it is often complicated to infer the utilities of the participants at hand, as it is unclear how they value potential payoffs as specified in the study at hand. To overcome this problem, all participants of the experiments incorporated in this review faced economic incentives in the form of points that they can earn. The amount of points an actor will earn is dependent both on the behavior of this actor, as well as on the behavior of the transaction partner of this actor. Hence, standard game-theoretic assumptions yield that these monetary payoffs represent the actors' utilities.

Also 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 behavior 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 behavior 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 behavior of an actor in any given round and assessing the effect of the number of rounds to play.

3 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 behavior 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 behavior 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 that there is a substantial number of players with no incentive to defect, until they 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 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 behavior in the current round. In the final rounds of the finitely repeated game however, the long term benefits of mutual cooperation do no longer outweigh the short-term costs of maintaining one's reputation (i.e., the opportunities to control future behavior of one another diminish), and hence rational actors will try to exploit their partners in these rounds.

In infinitely, or indefinitely, repeated games, there generally is no, or only a small, end-game effect. 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 $(1 - \delta)$. Obviously, there thus is no predetermined final round. However, comparing cooperation rates in the first round of an infinitely repeated game to cooperation rates the first round of a finitely repeated game or in a one-shot game (i.e., a finitely repeated game of length one) allows to assess the effect of control. Namely, in a one-shot game, there is no possibility to sanction opportunistic behavior, and thus there is no control. In a finitely repeated game of the same expected length as an infinitely repeated game, there are fewer control opportunities, because in the final round of a finitely repeated game, no control is possible. Although the analysis of first round behavior 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.

3.1 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 towards 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 possibly returns of another round of mutual cooperation. [**ADD EMBREY, FRECHETTE & YUKSEL (2018) ON FINITELY REPEATED GAMES**]

Additionally, Buskens et al. (2010) 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 whether to honor 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 in terms of past moves of one's transaction partner, 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, it is found that 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. [**STILL CHECK WHETHER, AND IF SO, HOW, TO INCORPORATE THE SECOND TRUST IN TRIADS STUDY BY MILTENBURG ET AL, BECAUSE THE ANALYSES ARE SOMEWHAT DIFFERENT AND I DONT HAVE THEM READILY AT THE TOP OF MY MIND**].

(**THIS PARAGRAPH IS JUST A NOTE FOR MYSELF**: Add the article by Mao et al., 2017 about resilient cooperators and conditional cooperators. Namely, this article show that repeating a finitely repeated games many times (20) with the same pool of participants, does not result in cooperation unravelling towards the first round of the game. On the contrary, it shows that the presence of conditional cooperators in the population allow other players to abuse trust in the final rounds of the game. In fact, they show that 60% of their samples plays with a strategy in which they abuse cooperation of their partner by defecting in the final rounds of the game, when there is relatively little to win by mutual cooperation.)

(**THIS ALSO IS JUST A NOTE TO MYSELF:** Research into infinitely repeated games started with the work of Roth & Murnighan (1978), who found that playing a Prisoner's Dilemma game indefinitely often resulted in higher cooperation rates than the amount of cooperation that can be observed in one-shot games. However, this increase was generally small, leading the authors to conclude that the introduction of infinitely repeated interactions does foster cooperation, but only to a limited extent. The same findings were found by [ADD REFERENCES]. However, Dal Bó, 2005 remarks that these studies did not translate the amount of points participants earned during the experiment linearly to a monetary reward. Additionally, these studies do not allow subjects to gain familiarity with the game, which might be important given the rather artificial circumstances participants are in [@dal_bó_cooperation_2005].).

Dal Bó (2005) builds upon this initial work in infinitely repeated games, and compares cooperation rates in infinitely repeated games with cooperation rates in finitely repeated games of the same expected length and in one-shot games under a wide variety of circumstances. It is found that first round cooperation rates are significantly higher in infinitely repeated games (with continuation probability $\delta = 1/2$ and $\delta = 3/4$) than in one-shot games. In fact, first-round cooperation increases with δ, implying that the prospect of future benefits through mutual cooperation, and the fear of missing out on these benefits after defection, results in a higher willingness to cooperate. Additionally, 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. 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 behavior by defecting in an earlier stage already. No such behavior is present in infinitely repeated games; in fact, the cooperation rates tend to increase with the experience of the participants.

[**I HAVE TO WORK THIS OUT STILL**:These findings have been corroborated by (Dal Bó & Fréchette, 2011, 2018). These subsequent studies additionally remark that subjects learn to play]

- Dal Bo & Frechette
- Buskens, Raub & Van der Veer - the condition without reputation shows a clear end-game effect, indicate for the presence of control effects.

3.2 Control in endogenously dyadically embedded interactions (repeated games)

- Sokolova, Buskens & Raub - show higher cooperation rates in dyadically embedded interactions than in one-shot games.

4 Control effects of network embeddedness

Research on repeated interactions in dyadic relations involve two actors forced into an iterated relation. However, in reality, interaction partners may deliberately choose to engage in a relationship and have the freedom to terminate the relationship (Yamagishi et al., 1994).

Sánchez (2018) remarks in a review that lattices or networks do not support cooperation at all. Even though cooperation in networks generally starts at a high level, with the majority of the actors choosing to cooperate, cooperation rates generally decrease to a fraction of only 20%.

4.1 Control in exogenously embedded interactions in networks

4.2 Control in endogenously embedded interactions in networks

5 Conclusion

6 Literature

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

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. (2000). An experiment on the effects of embeddedness in trust situations: Buying a used car. *Rationality and Society*, *12*(2), 227–253.

Chaudhuri, A. (2009). *Experiments in economics: Playing fair with money*. Routledge.

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

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

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

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*(3), 481–510.

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

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.

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.

Sánchez, A. (2018). Physics of human cooperation: Experimental evidence and theoretical models. *Journal of Statistical Mechanics: Theory and Experiment*, *2018*(2), 024001.

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

Van Miltenburg, N., Buskens, V., & Raub, W. (2012). Trust in triads: Experience effects. *Social Networks*, *34*(4), 425–428.

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.