



**Doctoral School of Sociology and
Communication Science**

THESIS SYNOPSIS

Flóra Samu

Reputation and cooperation in social dilemma games

titled Ph.D. dissertation

Supervisors:

Prof. Károly Takács, PhD

Szabolcs Számadó, PhD

Budapest, 2023

Table of contents

1. Research Topic and Literature Review	4
2. Research design and data collection	8
3. Results.....	9
3.1. Scarce and directly beneficial reputation support cooperation (RQ1).....	9
3.2. Mechanisms that could support credible reputations and cooperation: cross-checking and social bonding (RQ2).....	12
3.3. Conditions for honest signalling (RQ3).....	15
3.4. Social preferences and the evolution of friendship and negative relations (RQ4)	16
4. References.....	18
5. Publications.....	24

1. Research Topic and Literature Review

The dissertation examines how social communication and social network formation maintain cooperation in social dilemmas, conflicts, in which actions with the highest utility for individuals leads to the most disadvantageous outcome for the group (e.g. Luce and Raiffa 1957, Rapoport and Chammah 1970, Axelrod 1984). More precisely, previous literature suggests that reputational information can facilitate cooperation because it provides guidance for conditional cooperation (Nowak and Sigmund 1998, 2005). The risk of cooperating with an individual with good reputation is obviously lower than the risk when the partner has bad reputation, but overall cooperation remains a risky decision (e.g., Ohtsuki et al. 2009). Thus, reputation is not sufficiently answering the question why individuals would choose risky cooperation.

If cooperation in a cooperative society is based on rational calculation (cf. Mújdricza 2019), it has to be incentivised. Social motives, such as the “desire to please others” (Bicchieri 2005 p. 23), for instance, can change the structure of the conflict in a social dilemma (Bicchieri 2005 p. 26), in a way that, the value of social rewards for mutual cooperation exceeds the benefit of unilateral defection. As a result, cooperation is preferred over defection but only if the other player cooperates as well (Skyrms 2004), meaning that cooperation remains risky without a reliable signal of partner’s expected behaviour. This dilemma can be solved, if a reliable guidance exist that foretells how the other player behaves. Generosity (Fehrler and Przepiorka 2013), postpone gratification (Posner 2000), or trustworthiness can be an indicative signal of cooperative intentions. Trusting in such reputational information, that coordinates decisions, is essential to facilitate cooperative behaviour. How a signal receives such meaning, and how it remains informative, however, is still an open question. Conditions under which signals acquire meaning - formalized in economics (Spence 1973) and biology (Grafen 1990) - are empirically tested for the first time in this dissertation. In short, it is not the cost of the signal that ensure its informativeness, but the differentiated marginal outcome of costs and benefits of sending signals for players with different intentions (e.g. cooperation vs. defection) (Penn and Számádó 2020). This means that signal costs and benefits have to be calibrated in a way that deters potential non-cooperators from using the signal of cooperative intent, while co-operators should be encouraged to use it.

It is well known, that social sanctions are not just the main drive of cooperation (Bicchieri 2005, Janky 2022), but sanctions also contribute to the establishment of a signalling system that can reliably signal future behaviour of the potential cooperation partner (Nowak and Sigmund 1998, 2005), as experiences are shared via public (e.g., buyers evaluate sellers on an e-commerce platform) or private channels (when opinions are spread through gossip) (Sommerfeld et al. 2008, Feinberg 2014). This way, sanctions as rewards or punishment could be applied indirectly. This dissertation confirms previous results, about the role of gossip in reputation building.

Solving the cooperation problem by assuming a desire for social acceptance opens up new directions of research. Uncertainties remain concerning the nature of social rewards: it has not yet clarified whether social rewards are available for every member of a community or whether it is a scarce/relative resource (Paine 1967, Anderson et al., 2015). If it is the latter, competition is closely related to the proposed solution. The importance of relative status for individuals is well known in sociology and in social psychology (Merton 1968, Festinger 1954), but contemporary research also reveals brain processes that connects rewards to relative status (Fliessbach et al. 2007). The dissertation tests whether competition for reputation is a necessary condition for cooperation.

The development of informative signals is especially interesting under competition, because under conflict of interest individuals are not interested in being honest about rivals. The reliability of reputational information is questionable when individuals can improve their own position by damaging the reputation of others. Pulling others back by diminishing their merit can be a good strategy to avoid loss of relative status, when status depends on the position of other members in the group. If social status is a scarce resource ('whoever stays out misses out') dishonesty about rivals is incentivised. As a result of the competition, the spread of false information can harm individuals' reliance on gossip, and consequently, cooperation. This dissertation contributes to the scientific discussion about two structural conditions that could ensure the reliability of gossip and the maintenance of cooperation, even under competition. The first mechanism that could make gossip and reputation more reliable is cross-checking (verifying gossip from multiple sources). The second mechanism builds on the assumption that honest gossip could be a signal of cooperative intent, therefore it is expected to be more reliable

if cooperation partners can exchange gossip before they meet again in a game with higher conflict of interest (“shadow of the future”, Axelrod 1980).

Honest – and therefore reliable – signals of cooperative intentions not necessarily reflect unobservable personal traits (true altruism) but rather social status (Baldassarri and Grossman 2013) or social embeddedness (Uehara 1990). Structural incentives could facilitate cooperative actions (Willer 2009), if a structural position differentiates the cost of cooperation by network positions. For instance, the cost of cooperation can be lower for a popular individual (support will be returned more quickly through social ties), while defection would be more visible (due to cross-checking opportunities through dense and closed networks). The dissertation investigates whether cooperative behaviour correlates with positive and negative network ties among children.

The structural explanations tested in the thesis are only one of many theories justifying the existence of cooperative behaviour. It is important to say, that even if structural explanations are rather external forces that may trigger cooperative behaviour, the dissertation does not claim that cooperative intention can be explained exclusively by external drivers. This work focuses on the effects of society on individuals’ action for a very simple reason: it kept this research on track and gave clarity to at least one piece of a puzzle.

The following hypotheses were tested in the dissertation:

RQ1: Scarce and directly beneficial reputation support cooperation

H1: The possibility of gossip increases cooperation.

H1a: Gossip will be in line with partners’ previous decisions.

H1b: Reputations are updated in the direction of the valence of gossip received.

H2a: Competition for scarce reputations increases cooperation.

H2b: Competition for scarce reputations increases negative gossiping.

H3: Direct monetary stakes for reputation decrease cooperation.

RQ2: Mechanisms that could support credible reputations and cooperation: cross-checking and social bonding

H1: Competition decreases the reliability of gossip

H2: Cross-checking increases the reliability of gossip

H3: Social bonding increases the reliability of gossip

RQ3: Conditions for honest signalling¹

H1: The signal cost paid by high-quality signallers maintains the honesty of signals.

H2: The honesty of communication is maintained by signalling trade-offs, in such a way that for honest communication it has to be fulfilled that the marginal cost of high-intensity signal compared to the low-intensity signal for high-quality signallers must be cheaper than for low-quality signallers.

H2a: In the absence of trade-off, (a) the signallers are expected to select randomly from the two available signals, therefore (b) receivers respond also randomly to the signals they receive; (c) the accuracy of decisions made by receivers are not better than picking randomly without signals.

H2b: If both types of signallers have the same trade-off between the two signals, then (a) both signaller type will use the same signal, and (b) receivers decide randomly on the allocation of the source; (c) the accuracy of decisions made by receivers are not better off than picking randomly without signals.

H2c: If the two signaller types have a different trade-off between the use of two signals in such a way that high-quality signallers have larger marginal benefits for choosing the high-intensity signal than low-quality signallers, then (a) the two types will use different signals depending on their type, thus (b) the receivers will be able to determine the type of signallers correctly and respond differently; (c) the accuracy of decisions made by receivers are better off than picking randomly without signals, i.e. they can allocate the resources according to their interest.

RQ4: Social preferences and the evolution of friendship and negative relations

H1: Prosocial children are more likely to be nominated as good friends.

¹ In order to define honesty, there is a distinction between high and low-quality signallers, and between high and low intensity signals. In the signaling game, the receiver of the signal has resources, which he/she could share after observing a signal, not knowing the type of the signaller. The receiver's goal is to share the resource with high-quality signallers. Communication is honest if high-quality signallers send high-intensity signals and low-quality signallers use low-intensity signals, so the receiver can make a good decision based on the observed signal.

H2a: Children are more likely to create or maintain friendship relations with similarly prosocial others.

H2b: Prosocial children are more likely to create or maintain friendship relations with similarly prosocial others.

H3: Selfish children are more likely to receive or maintain negative ties.

H4a: Prosocial children are more likely to create or maintain negative relations to selfish others.

H4b: Negative relations are more likely to occur between children with different social preferences.

2. Research design and data collection

The dissertation relies heavily on laboratory experiment in order to test hypotheses about the relationship between gossip, reputation and cooperation. Experiments were conducted at the Corvinus University of Budapest. Participants were protected by anonymity and during the experiment they had the right to withdraw their consent if they wanted. Only adults were involved in the experiment. The experimental procedures were approved by the Ethics Review Committee of the Centre for Social Sciences at Hungarian Academy of Sciences.

Laboratory experiments are commonly used in psychology, economics, but not in sociology (Webster and Sell 2014). Despite the fact that the random allocation of participants suggests that the observed differences between treatments are causal effects of the manipulations. Critical remarks about lab experiments relate to the external validity of results founded in a quite artificial, controlled setting, and their generalizability beyond the laboratory context. Participants could behave differently in a real-life situation. For instance, communication in the experiments – sending emojis – may have led to a failure in achieving cooperation, which would have happened in real life with more complex communication opportunities. Moreover, decisions were incentivized using monetary rewards to ensure participants take the task seriously. Again, the link between monetary benefits and reputation not necessarily direct in reality. The fact that participants were students from one university also limits the generalizability of the results.

The last chapter applies a standard statistical methodology used to study network dynamics in small, closed groups, called Stochastic Actor Oriented Models (SAOMs) (Snijders et al. 2010). This method can be applied to test theories about social mechanisms influencing the formation, persistence, and termination of relationships. To estimate SAOMs, the R package RSiena has been developed and constantly maintained by researchers (Ripley et al. 2011). SAOMs model the evolution of social ties over time by simulating mechanisms that may be responsible for changes between the empirically observed snapshots of the networks. It has been tested, whether prosocial behaviour can influence the tendency of receiving a friendship or a negative tie, and whether individuals tend to befriend similarly prosocial others.

3. Results

3.1. Scarce and directly beneficial reputation support cooperation (RQ1)

In the first study the dissertation tests experimentally whether competition damage the reliance on reputation in a non-iterative two-player Prisoner's Dilemma game with a stranger. By manipulating (1) whether reputation is a scarce resource and (2) whether direct benefits are attached to it, we were able to analyse the different incentives for reputation building and its effect on cooperation at the group level. One of the main contributions of this chapter is the direct measurement of private evaluations and the involvement of gossip in the process of reputation-based cooperation. Since the target of gossip was freely chosen, people were allowed to spread false gossip about individuals with high reputation when competition was severe, and they were interested in improving their own chances of reputation building by degrading the reputation of others. In this chapter, in addition to answering the main hypotheses about the effect of competition on the use of reputation, conclusions can be drawn about the development of a reliable reputational system as well.

More precisely, this chapter investigates how cooperation can be sustained by private reputations formed by direct observations and gossip. A slight increase has been found in the level of cooperation in one condition when the institution of reputations and gossip have been introduced (H1). Cooperation has faded over time, which is a typical feature of Social Dilemma experiments. Preconditions of gossip to be effective are confirmed: gossip was in line with

previous cooperation choices (H1a) and gossip received has altered the private reputations of others (H1b).

Two mechanisms were proposed that could safeguard the credibility of gossip for informing the choice of cooperation. First, it was investigated whether the scarcity of reputational resources could increase reputation-based cooperation (H2). Second, it was tested whether additional monetary incentives connected to reputation would distort the credibility of reputations (H3). Neither the scarcity of reputational scores nor monetary incentives alone could maintain reputation on the long term. The dissertation shows, however, that at the intersections of these two manipulations, competition for scarce monetary rewards resulted in higher cooperation in the short run.

A reputational system is reliable if it appropriately reflects the potential behaviour of others that is otherwise hidden to new partners. It functions well if it helps individuals to cooperate with those who have a higher reputation and defect against those who have lower willingness to cooperate. Results suggest that motivation for building a proper reputation system increases if people find it easy (without competition) to credibly signal prosociality (non-monetary rewards) or if external incentives encourage everyone to participate in the competition - maybe because higher positions in the reputation hierarchy are more robust. Even though reputations have seemingly been well translated to gossip under these conditions, they did not increase strategic cooperation in a long run in the one-shot PD game with stranger matching.

Moreover, gossip was influenced by previous reputational scores not just by the last observed action. Since reputation scores are influenced by messages beyond actions, reputation scores could have been inflated in the informal communication process. Hence, positive gossip increased good reputations and negative gossip downgraded bad reputations. This has important implications for the whole dynamics of the development or the maintenance of reputational systems.

The analysis of social norms has revealed a similar effect: while the strongest predictor of reputational scores was the focal actor's behaviour (cooperation vs. defection), the previous reputation score also had a significant impact. In other words, there was a strong inheritance of reputational scores. Reputation updates were influenced by actions and previous reputations,

but in little alignment with the “leading eight” social norms (Ohtsuki and Iwasa, 2006). Most importantly, the reputation of the opponent had a little effect on the reputational update (see Table 1, Panel C). This has two important implications: both justified punishment and the identification of defectors might be missing from the system, i.e., defection against an opponent in good or bad state has very similar effects. While the scope of the investigation of higher order norms is limited in the experiment, yet it shows that the presence of leading eight social norms cannot be taken for granted.

The combined effect of the reputation of the focal player both on gossip and reputational updating could explain the lack of increase of cooperation in our model on the long term. Beyond the ineffectiveness of the examined reputation systems, the fact that there is no large impact of reputations on cooperation could be attributed to several other factors. Primarily, the two-person PD game was investigated with random reshuffling of partners and no publicly available information, which itself is the most severe social dilemma in which rational action is simply defection. The magnitude of conflict in the Prisoner’s Dilemma game could be so strong that even a well-functioning reputation system could not increase cooperative acts (Rapoport and Chammah 1965). Unfortunately, in this study it is impossible to assess whether the magnitude of the conflict is responsible for the low level of cooperation.

The PD game is an interdependent situation and hence it is not the most appropriate to test fundamental tenets of the theory of competitive altruism (Roberts 1998, Barclay 2004, Barclay and Willer 2007). Future studies could investigate if scarce reputations and direct reputational incentives could increase giving in the dictator game in the lab or in field settings. In the experiment, cooperation could have collapsed before the reputation system had been sufficiently developed, which leaves open the question of the coevolution of reputation and cooperation (Rand and Nowak 2013). It is also possible that reputation scores worked to a limited extent because they were not directly communicated to others or due to the abstract situation and scores in the experiment. Even more, participants potentially had problems to remember their earlier experiences and might have also mixed up other participants as they were identified with numbers that are harder to recall than names or faces. The direct correspondence of the study to the theoretical literature is uncertain, as we used private reputations that are realistic but contradict to the assumption of publicly available and perfect information on choices or reputations in showcased models of cooperation (Wedekind and

Milinski 2000, Milinski et al 2002, Ohtsuki and Iwasa 2004, 2006, 2007). As reputations were private in the experiment, they could be used only to a limited extent for strategic reasons, and they could be linked to cooperation only through simplified gossip communication. This also limits the connection of the study to the theory of competitive altruism (Roberts 1998, Barclay 2004, Barclay and Willer 2007) as privately stored reputations cannot be used by the recipients for acquiring diverse benefits such as status, power, or access to resources, and participants could not select their interaction partners (Sylwester and Roberts 2013, Herrmann et al. 2019).

Still, results bring us closer to understanding under which conditions reputations and gossip contribute to cooperation. Further research is needed to find out, however, under which conditions gossip is used strategically and in a dishonest way to undermine the reputation of others, and under which conditions it could be considered as altruistic punishment (Feinberg et al. 2012).

3.2. Mechanisms that could support credible reputations and cooperation: cross-checking and social bonding (RQ2)

A reputation system can effectively maintain cooperation only if it is based on reliable information spreading. Gossip—an evaluative communication about third parties—could be the channel of reliable information transmission and hence could contribute to the maintenance of cooperation (Wu et al. 2016 for review). There is significant doubt, however, about why gossip should be honest and reliable at all (Smith 2014). In Chapter 3, I report the results of another experiment where two mechanisms were tested that could strengthen the credibility of reputational information (aka gossip).

The first mechanism that could make gossip (and as a consequence reputation) more reliable is cross-checking. Once the opportunity is given, individuals actively seek and cross-check social information to condition their future actions on a better-informed ground, which improves the reliability of reputations they assign to others. In the experiment, participants could verify gossip from multiple sources, and it was designed as a single opportunity to ask a second opinion about the same target.

It is argued that social bonding motives could also increase the credibility of social information exchange and hence make reputations reliable. The second mechanism builds on the

assumption that honest gossip could be a signal of cooperative intent, therefore it is expected to be more reliable if cooperation partners can exchange gossip before they play, again, a non-iterative two-player Prisoner's Dilemma. Although it was not possible to create real social bonds between participants in the experiment, selecting a single characteristic that is typical of social bonding and friendship formation could also be introduced in an abstract experimental setting: reciprocity in communication before the conflict. Note that reciprocity in communication did not mean reciprocity in interactions as participants played PD games against different partners to follow the settings described in models of indirect reciprocity (Alexander 1987, Nowak and Sigmund 1998, Milinski et al. 2002, Panchanathan and Boyd 2004).

Again, competition was manipulated, to test whether higher confidence in gossip due to cross-checking or social bonding is able to keep reputation informative under rivalry. It is argued that strong direct rivalry for reputations could increase opportunistic use of gossip and hence decrease the reliability of the information received. Competition was designed in a way that direct rivalry with a set of other participants meant a distribution of monetary payoffs depending on relative reputations.

Even if gossip and reputation scores were mutually aligned with each other and with the PD decisions, cooperation did not emerge to a very high rate in any of the conditions. Competition for reputations had divergent effects in the experiment. On the one hand, messages about rivals were more negative, which diminished the reliability of assigned reputations. On the other hand, cooperation was affected positively by the strength of competition. In line with competitive helping theory, rivalry increased cooperation regardless of the reputation of partners (Roberts et al. 2021). Still, no escalation of cooperation was observed; only the decline of cooperation slowed down (cf. Fischbacher et al. 2001).

Though reputation scores grew more as a result of positive messages received, the possibility of social bonding did not cause significant improvement for cooperation. Results are consistent with the fact that people place more weight on positive information if it comes from a stronger social bond (Bozoyan and Vogt 2016). The integration of received information from trusted sources is important for a well-functioning reputation system, but as social bonding did not

improve significantly how reputations are used to condition behaviour, this treatment did not substantially improve cooperation overall.

In the cross-checking condition, a greater cautiousness of participants was observed. Participants were less courageous in sending positive gossip about trustworthy partners. Besides greater cautiousness, participants often received conflicting information about the same target, which may lower the reliability of communication even compared to no information (Kuttler et al. 2002). Mixed gossip could have an averaging (Sommerfeld et al. 2008) and a majority effect (Laidre et al. 2013) on reputations. Surprisingly, people inclined to doubt multiple negative opinions as well (Sommerfeld et al. 2008, Hess and Hagen 2006).

Participants in the cross-checking and social bonding conditions were assigned lower reputation scores in general. Lower reputation scores in these conditions—measured as trustworthiness—may have been caused by a general lack of trust caused by the inefficient (Zand 1972) and sometimes contradicting information participants received. Social information needs to be available in large amounts to assist cooperation (Giardini and Vilone 2016, Romano et al. 2021). Correspondingly, the reputation of gossip sources was eroded if they failed to provide information.

Confidence in gossip from trustworthy sources was higher (Kuttler et al. 2002, Pasquini et al. 2007). People seek information from sources considered trusted (Van de Bunt et al. 2005), probably because of their (perceived) good access to information. Therefore, gossip and the dynamics of reputation and cooperation should be considered from the perspective of the social network structure and the position of relevant individuals within (see Takács et al. 2021 for review, Dumas et al. 2021).

Results suggest that a reliable reputation system is not a sufficient condition for cooperation in situations of moderate conflict of interest. At the same time, relative competition seems to play an important role for cooperation, which could be linked with keeping up with others (loss avoidance) or achieving reputational benefits (status maximization) for the development of widespread human cooperation (Roberts 1998, Barclay 2004, Barclay and Willer 2007, Wu et al. 2016, Herrmann et al. 2021, Raihani and Smith 2015, Raihani and Bshary 2015).

Overall, none of the tested mechanisms in their abstract form and out of social context were able to sustain a high level of cooperation in the laboratory.

3.3. Conditions for honest signalling (RQ3)

Chapter 4 is crucial to understand how reputation signals could develop and under what conditions they provide a reliable guidance to conditional decisions in a social dilemma. In this study my supervisor invested almost the same amount of work as me to prove empirically that signals (aimed at changing the behaviour of others) do not have to be costly to be informative. We modified the costs of initially meaningless signals in a so-called signalling game in different ways proposed by two theories with different projections about the development of reliable signals. The conflicting predictions tested were made by the proponents of the Handicap Principle (Zahavi 1975; Grafen, 1990) and by subsequent game theoretic models (Hurd, 1995, Számadó, 1999, Bergstrom et al., 2002). These implications have been translated into two main hypotheses. First, we tested if signalling costs for high-quality signallers alone could establish honest communication (H1). Second, we tested whether signalling trade-offs could lead to honest signalling (H2). We expected that the lack of trade-offs (H2a) and trade-offs that are identical for low- and high-quality signallers (H2b) would not be sufficient for receivers to differentiate signaller types efficiently. Instead, condition-dependent trade-offs have been expected to be the guarantees of honest communication (H2c).

Our results support the last claim: informative signals emerged in our experiment under the condition-dependent trade-off condition regardless of the cost of signals for high-quality signallers. Signals with zero or even negative production cost (benefit) were honest if trade-offs were condition dependent. An intention to separate signaller types by using different signals can be observed in the other manipulations as well (no trade-off, trade-off). Local conventions emerged in the no trade-off condition, although the direction of these conventions were random, thus at the aggregate level no honesty can be observed. There was an intention to separate in the trade-off condition as well, however, it was only mildly successful. Thus, we can see that the intention to separate signaller types (be honest) was successful only in the condition-dependent trade-off manipulations as predicted by game-theoretical models.

Overall, it is important to emphasize that no strict separating equilibria were observed in any of the treatment conditions. There was always some level of mixing indicating some randomness of decisions even among our subjects who were university students well able to grasp the structure of the simple experimental signalling game. The level of mixing, however, was much reduced under the condition-dependent trade-off manipulation. Some differentiation evolved between low- and high-type signallers in the simple trade-off manipulation. Only the condition-dependent trade-off manipulation allowed receivers to make an informed decision (i.e. do better than random choice). There could be several factors influencing the level of mixing observed in the experiment. The first one is the length of the experimental sessions. As a result of a learning process, a higher differentiation is expected between signaller types in the use of signal over time. The equilibrium signaller and receiver strategies were straightforward, yet they were somewhat contra-intuitive with negative signal cost. In these later manipulations, it was naturally a longer learning process to find (or being close to) the equilibrium strategy pair.

The second factor is the roles played by the participants. Participants alternated between all three roles in our experiment (i.e., receiver, low- and high-quality signaller). There are advantages and disadvantages of this setup. While participants might have a better understanding of the experimental setup, they may not be playing (optimizing for) a single role. We observed an intention to separate (to signal honestly) in all manipulations, which could be due to the fact that players alternated between all three roles, which in turn created an incentive to be honest (since both receivers and high-quality signallers are better off this way). Playing a single role might force them to optimize their actions for that single role only, thus this could remove the above effect.

3.4. Social preferences and the evolution of friendship and negative relations (RQ4)

In the last study the dissertation steps away from the lab and analyse the dynamics of friendships and negative relations in school classrooms with social network analysis methods, to test reputation-based partner selection on the field. In this study, I examined the role of social preferences in shaping social networks in a primary school environment. I developed hypotheses along a theoretical framework that emphasizes the role of social preferences in partner choice. Considering friendship and negative relations within the classroom, I tested

assumptions about prosociality-based tie formation using the toolkit of social network analysis (SNA). By estimating the dynamics of friendship and negative relations between two waves of data collection, I tested whether prosocial individuals, who have made more generous decisions in economic games in the first wave, receive more friendship and less negative nominations. Unlike other studies (Shin et al. 2019, Dijkstra and Berger 2018), I found no evidence of a preference for more generous friends. Moreover, I fail to prove the existence of repulsive forces towards selfish individuals.

The fact that social preferences are hidden traits, and individuals are not aware of others' social preferences, can explain the lack of impact in terms of networks. Nevertheless, unobserved preferences can be signalled in an observable way. Such signal could be reputational information from others. Reputation and prosocial behaviour are already linked: it has been empirically proven that generosity leads to reputational gains (Macfarlan and Lyle 2015). It is therefore conceivable that the effects of prosociality are mediated by popularity. In the models, popularity, indeed, attracts more friendship and less negative nomination, and friendship nominations between popular students are more likely, while negative nominations are less likely. This reasoning, however, is less valid in this case because, while other studies find a strong correlation between popularity and prosociality (Logis et al. 2013), there is no correlation in this sample.

The lack of evidence of asymmetric relations suggests that class dynamics at this stage of life may no longer be driven by indirect link reciprocity, because students already spent enough time together to gain direct experiences. As opposed to indirect reciprocity, direct reciprocity rather infers symmetric relations. Therefore, we should be able to observe that individuals organize themselves into partnerships with similar peers and maintain negative relations with dissimilar others. Results do not prove the presence of assortative matching in the friendship network, but I found that dissimilar individuals are more likely to establish or maintain negative relations.

I document that differences in generosity increases the probability of negative relationships among students. Interestingly, I found strong and significant effect only in the dictator game (DG), whose design is better suited for measuring intrinsic social preferences, while the other games (public goods game (PGG), first decision in the trust game (TG)) require more strategic

thinking (Thielmann et al. 2020). The second decision of the trust game (TGB) is somewhat closer to the dictator game and shows a similar but non-significant effect. Mutually negative nominations between prosocial and selfish individuals may indicate the combined presence of prosocial and antisocial punishment (Herrmann et al. 2008).

Lastly, a counter-intuitive result has been found. Prosocial individuals are more likely to maintain a negative relationship with each other if we define prosociality based on the contribution in the PGG. Note that this game was played first, and it is more likely that children make more non-strategic decisions in the first game (Anderson et al. 2000). It is also known that intuitive decisions are more cooperative (Rand et al. 2012, Rand et al. 2014, Rand 2016). Along this line, the literature on strong reciprocity also assumes intuitive drivers behind punishment (Gintis 2000, Fehr and Fischbacher 2003, Bowles and Gintis 2004). It might be the case that intuitive individuals establish a negative relationship with each other more easily.

One of the main contributions of this study has been to examine whether social preferences have consequences on the negative networks. Positive networks, on their own, do not represent all the forces that could shape social relationships, and the examination of forces acting in the opposite directions is important (Stadtfeld et al. 2020). One of the limitations of this study is that it does not model these two forces simultaneously. The signs of the studied effects in the two networks suggest that these dynamics are not just the two sides of a coin but have distinct effects. Still negative dynamics are underrepresented in research. Hopefully, the dissertation contributed to fill this gap.

4. References

Alexander RD. 1987 *The biology of moral systems*. New York, NY: Routledge.

Anderson, C., Hildreth, J.A.D. and Howland, L., 2015. Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological bulletin*, 141(3), p.574.

Axelrod, R. 1984. *The Evolution of Cooperation* Basic Books, New York

Baldassarri, D. and Grossman, G., 2013. The effect of group attachment and social position on prosocial behavior. Evidence from lab-in-the-field experiments. *PloS one*, 8(3), p.e58750.

Barclay, P. and Willer, R., 2007. Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610), pp.749-753.

Barclay, P., 2004. Trustworthiness and competitive altruism can also solve the “tragedy of the commons”. *Evolution and Human Behavior*, 25(4), pp.209-220.

Bergstrom, C. T., Számádó, S., and Lachmann, M. 2002. Separating equilibria in continuous signalling games. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1427), 1595-1606.

Bicchieri, C., 2005. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C., 2016. *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bowles, S., & Gintis, H. 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical population biology*, 65(1), 17-28.

Bozoyan C, Vogt S. 2016 The impact of third-party information on trust: valence, source, and reliability. *PLoS ONE* 11, e0149542.

Dijkstra, Jan Kornelis and Christian Berger 2018. Friendship selection and influence processes for physical aggression and prosociality: differences between single-sex and mixed-sex contexts”. In: *Sex roles* 78.9, pp. 625–636.

Dumas M, Barker JL, Power EA. 2021 When does reputation lie? Dynamic feedbacks between costly signals, social capital and social prominence. *Phil. Trans. R. Soc. B* 376, 20200298.

Fehr, E., and Fischbacher, U. 2003. The nature of human altruism. *Nature*, 425(6960), 785-791.

Fehrler, S. and Przepiorka, W., 2013. Charitable giving as a signal of trustworthiness: Disentangling the signaling benefits of altruistic acts. *Evolution and Human Behavior*, 34(2), pp.139-145.

Feinberg M, Willer R, Schultz M. 2014 Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* 25, 656-664.

Feinberg, M., Wilier, R., Stellar, J. and Keltner, D. 2012. The virtues of gossip: reputational information sharing as prosocial behavior. *J.Pers. Soc. Psychol.* 102, 1015.

Festinger, L. 1954. A theory of social comparison processes. *Human relations*, 7(2), 117-140.

- Fischbacher U, Gächter S, Fehr E. 2001 Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* 71, 397-404.
- Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C.E. and Falk, A., 2007. Social comparison affects reward-related brain activity in the human ventral striatum. *science*, 318(5854), pp.1305-1308.
- Giardini, F. and Vilone, D. 2016. Evolution of gossip-based indirect reciprocity on a bipartite network. *Sci. Rep.* 6, 37931 .
- Gintis, H. 2000. Strong reciprocity and human sociality. *Journal of theoretical biology*, 206(2), 169-179.
- Grafen, A., 1990. Biological signals as handicaps. *Journal of theoretical biology*, 144(4), pp.517-546.
- Herrmann, Benedikt, Christian Thoni, and Simon Gächter 2008. Antisocial punishment across societies". In: *Science* 319.5868, pp. 1362–1367.
- Herrmann, E., Engelmann, J. M. and Tomasello, M. 2019. Children engage in competitive altruism. *J. Exp. Child Psychol.* 179, 176–189 .
- Hess, N. H. and Hagen, E. H. 2006. Psychological adaptations for assessing gossip veracity. *Hum. Nat.* 17, 337–354.
- Hurd, P. L. 1995. Communication in discrete action-response games. *Journal of Theoretical Biology*, 174(2), 217-222.
- Janky, B., 2022. *Erkölc, racionalitás, közösség: A morális dilemmák egy szociológiai elmélete*, Gondolat Kiadó
- Kuttler AF, Parker JG, La Greca AM. 2002 Developmental and gender differences in preadolescents' judgments of the veracity of gossip. *Merrill-Palmer Q.* 48, 105-132.
- Laidre ME, Lamb A, Shultz S, Olsen M. 2013 Making sense of information in noisy networks: human communication, gossip, and distortion. *J. Theor. Biol.* 317, 152-160.
- Logis, Handrea A et al. 2013. Popularity as an organizing factor of preadolescent friendship networks: Beyond prosocial and aggressive behavior". In: *Journal of Research on Adolescence* 23.3, pp. 413–423.
- Luce, R. D. and Raiffa, H. 1957. *Games and Decisions* Wiley, New York

- Macfarlan, Shane J and Henry F Lyle 2015. Multiple reputation domains and cooperative behaviour in two Latin American communities”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1683, p. 20150009.
- Merton, R.K. and Merton, R.C., 1968. *Social theory and social structure*. Simon and Schuster.
- Milinski, M., Semmann, D. and Krambeck, H. J. 2002. Reputation helps solve the 'tragedy of the commons'. *Nature* 415, 424–426 .
- Mújdricza, F. 2019. The roots of trust. *European Journal of Mental Health*, 14(1), 109-142.
- Nowak, M. A. and Sigmund, K. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577.
- Nowak, M. A. and Sigmund, K. 2005. Evolution of indirect reciprocity. *Nature* 437, 1291–1298.
- Ohtsuki, H. and Iwasa, Y. 2004. How should we define goodness?—reputation dynamics in indirect reciprocity. *J. Theor. Biol.* 231, 107–120.
- Ohtsuki, H. and Iwasa, Y. 2006. The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* 239,435–444.
- Ohtsuki, H. and Iwasa, Y. 2007. Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *J. Theor. Biol.* 244, 518–531.
- Ohtsuki, H., Iwasa, Y. and Nowak, M.A., 2009. Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature*, 457(7225), pp.79-82.
- Paine, R. 1967. What is gossip about? An alternative hypothesis. *Man* 2, 278–285.
- Panchanathan K, Boyd R. 2004 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432, 499-502.
- Pasquini ES, Corriveau KH, Koenig M, Harris PL. 2007 Preschoolers monitor the relative accuracy of informants. *Dev. Psychol.* 43, 1216.
- Penn, D.J. and Számádó, S., 2020. The Handicap Principle: how an erroneous hypothesis became a scientific principle. *Biological Reviews*, 95(1), pp.267-290.
- Posner, E.A., 2000. *Law and social norms*, Harvard University Press
- Raihani NJ, Bshary R. 2015 Why humans might help strangers. *Front. Behav. Neurosci.* 9, 39.

- Raihani, N.J. and Smith, S., 2015. Competitive helping in online giving. *Current Biology*, 25(9), pp.1183-1186.
- Rand, D. G. and Nowak, M. A. 2013. Human cooperation. *Trends Cogn. Sci.* 17, 413–425 .
- Rand, David G 2016. Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation”. In: *Psychological science* 27.9, pp. 1192–1206.
- Rand, David G, Alexander Peysakhovich, et al. 2014. Social heuristics shape intuitive cooperation”. In: *Nature communications* 5.1, pp. 1–12.
- Rand, David G, Joshua D Greene, and Martin A Nowak 2012. Spontaneous giving and calculated greed”. In: *Nature* 489.7416, pp. 427–430.
- Rapoport, A. and Chammah, A.M., 1970. *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan Press
- Ripley, R.M., Snijders, T.A., Boda, Z., Vörös, A. and Preciado, P., 2011. *Manual for SIENA version 4.0*. University of Oxford.
- Roberts G, Raihani N, Bshary R, Manrique HM, Farina A, Samu F, Barclay P. 2021 The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. *Phil. Trans. R. Soc. B* 376, 20200290.
- Roberts, G. 1998. Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1394), 427-431.
- Romano A, Giardini F, Columbus S, de Kwaadsteniet EW, Kisfalusi D, Triki Z, Snijders C, Hagel K. 2021 Reputation and socio-ecology in humans. *Phil. Trans. R. Soc. B* 376, 20200295.
- Shin, Huiyoung, Allison M Ryan, and Elizabeth North 2019. Friendship processes around prosocial and aggressive behaviors: The role of teacher–student relatedness and differences between elementary-school and middle-school classrooms”. In: *Merrill-Palmer Quarterly* 65.2, pp. 232–263.
- Skyrms, B., 2004. *The stag hunt and the evolution of social structure*. Cambridge University Press.
- Smith ER. 2014 Evil acts and malicious gossip: a multiagent model of the effects of gossip in socially distributed person perception. *Pers. Soc. Psychol. Rev.* 18, 311-325.
- Snijders, Tom AB, Gerhard G Van de Bunt, and Christian EG Steglich 2010. Introduction to stochastic actor-based models for network dynamics”. In: *Social networks* 32.1, pp. 44–60.

- Sommerfeld, R. D., Krambeck, H.-J. and Milinski, M. 2008. Multiple gossip statements and their effect on reputation and trustworthiness. *Proc. R. Soc. B Biol. Sci.* 275, 2529–2536 .
- Spence, M. 1973. Job market signalling. In *Uncertainty in Economics* pp. 281-306. Academic Press.
- Stadtfeld, Christoph, Károly Takács, and András Vörös 2020. The emergence and stability of groups in social networks”. In: *Social Networks* 60, pp. 129–145.
- Sylwester, K. and Roberts, G. 2013. Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evol. Hum. Behav.* 34, 201–206 .
- Számadó, S. 1999. The validity of the handicap principle in discrete action–response games. *Journal of Theoretical Biology*, 198(4), 593-602.
- Takács K, Gross J, Testori M, Letina S, Kenny AR, Power EA, Wittek RPM. 2021 Networks of reliable reputations and cooperation: a review. *Phil. Trans. R. Soc. B* 376, 20200297.
- Thielmann, Isabel, Giuliana Spadaro, and Daniel Balliet 2020. Personality and prosocial behavior: A theoretical framework and meta-analysis.” In: *Psychological Bulletin* 146.1, p. 30.
- Uehara, E., 1990. Dual exchange theory, social networks, and informal social support. *American Journal of Sociology*, 96(3), pp.521-557.
- Van de Bunt GG, Wittek RP, de Klepper MC. 2005 The evolution of intra-organizational trust networks: the case of a German paper factory: an empirical test of six trust mechanisms. *Int. Sociol.* 20, 339-369.
- Webster, M. and Sell, J. eds., 2014. *Laboratory experiments in the social sciences*. Elsevier.
- Wedekind, C. and Milinski, M. 2000. Cooperation through image scoring in humans. *Science (New York, N.Y.)* 288, 850–852 .
- Willer, Robb 2009. Groups reward individual sacrifice: The status solution to the collective action problem”. In: *American Sociological Review* 74.1, pp. 23–43.
- Wu J, Balliet D, Van Lange PA. 2016 Reputation management: why and how gossip enhances generosity. *Evol. Hum. Behav.* 37, 193-201.
- Zahavi, A. 1975. Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, 43(1), 205-214.
- Zand DE. 1972 Trust and managerial problem solving. *Adm. Sci. Q.* 17, 229-239.

5. Publications

Samu, F. and Takács, K., 2021. Evaluating mechanisms that could support credible reputations and cooperation: cross-checking and social bonding. *Philosophical Transactions of the Royal Society B*, 376(1838), p.20200302.

Samu, F., Számadó, S. and Takács, K., 2020. Scarce and directly beneficial reputations support cooperation. *Scientific reports*, 10(1), pp.1-12.

Számadó, S., Samu, F., & Takács, K. 2022. Condition-dependent trade-offs maintain honest signalling. *Royal Society Open Science*, 9(10), 220335.

Roberts, G., Raihani, N., Bshary, R., Manrique, H.M., Farina, A., Samu, F. and Barclay, P., 2021. The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. *Philosophical Transactions of the Royal Society B*, 376(1838), p.20200290.

Manrique, H.M., Zeidler, H., Roberts, G., Barclay, P., Walker, M., Samu, F., Fariña, A., Bshary, R. and Raihani, N., 2021. The psychological foundations of reputation-based cooperation. *Philosophical Transactions of the Royal Society B*, 376(1838), p.20200287.