OTKA grant K 112929

Gossip, Reputation, and Cooperation: Building Blocks of Social Order

Principal Investigators: Szabolcs Számadó, Károly Takács.

30/09/2019

Research Report

The maintenance of social order and the emergence and extent of cooperation and coordination in human societies are among the most fundamental puzzles of social science. Reputation systems are frequent, seemingly stable, and often self-emerging features in many social contexts of human life; and they generally contribute to the prevention of open conflicts and enhancement of cooperation. Despite of being so important, however, the transmission of reputational information is not necessarily honest given the conflicting strategic interests of group members. If the honesty of communication is questionable, how could possibly unreliable reputations legitimize social order at all? This constitutes the paradox that we attempted to solve.

We promised three main lines of investigation to highlight the role of gossip and reputation in the maintenance of human cooperation. First, we utilized empirical network data from school classes and organizations to test conditional and contextual hypotheses about the role that gossip plays for reputation and cooperation. Second, we investigated theoretically how reputation mechanisms can be established at all, and how do individual strategies such as gossip influence them. Last but not least, we tested simple hypotheses about the relationship between gossip, reputation, and cooperation in laboratory experiments using the classical Prisoner's Dilemma game.

In line of the first proposed investigation we conducted cross-sectional and longitudinal analyses of data collected previously on gossip in secondary schools (Pál et al, 2016). We identified how the perceived reputation order plays a prominent role. After controlling for several social and structural characteristics, students develop disliking towards those who they look down on (disdain) and conform to others by disliking those who they perceive as being looked down on by their peers (conformity). The inconsistency with perceived reputations leads to disliking, in particular when individuals do not look up to those who they perceive to be well reputed by peers (frustration). The analysis has revealed interesting relationships between reputations and social networks that have been continued to be in the focus of our interest in further ongoing research using stochastic actor oriented methods (SAOMs).

We have implemented three lines of research as part of our theoretical investigations: (i) we have investigated the effect of dishonesty on the seminal model of indirect reciprocity; (ii) we have investigated the maintenance of cooperation in a novel spatially explicit model of image building; (iii) we have constructed a simple conceptual model of gossip based on a simple model of communication (action-response game).

Indirect reciprocity is often claimed as one of the key mechanisms of human cooperation. It works only if there is reputational score keeping and each individual can tell with a high probability which other individuals were good or bad previously. Gossip is proposed as a key mechanism that can maintain such coherence of reputations in the face of errors of transmission. Random errors,

however, are not the only source of uncertainty in such situations. The possibility of deceptive communication, where signalers aim to misinform the receiver cannot be excluded. While there is plenty of evidence for deceptive communication in humans, the possibility of deception is not yet incorporated into models of indirect reciprocity. Our investigation (Számadó et al., 2015) shows that when deceptive strategies are allowed in the population, the coherence of reputations and hence cooperation collapses. This collapse is independent of the norms and the cost and benefit values. It is due to the fact that there is no selection for honest communication in the framework of indirect reciprocity. It follows that indirect reciprocity can be only proposed plausibly as a mechanism of human cooperation if additional mechanisms are specified in the model that maintains honesty.

We investigated the role of hidden opinions in opinion dynamics in a two-layered spatially explicit model (Concealed Voter Model, CVM, Gastner et al., 2018) in the second project. Opinions are not directly observable in this model (first layer), however individuals can give a signal (second layer), which may or may not correspond to the opinion of the signaller. The consensus time (the time it takes for the whole population to converge on one opinion) can be estimated in the thermodynamic limit in this model. Accordingly, one can investigate how much hidden opinions slow down the convergence of the system as opposed to public opinions (which is the standard assumption behind opinion dynamics).

Recognizing that importance of that publicly revealed opinions might be different from true positions, we built and analyzed a modified version of the voter model with hypocrisy in a complete graph with a neutral competition between two alternatives (Gastner et al., 2019). We compare the process from various initial conditions, varying the proportions between the two opinions in the external (revealed) and internal (hidden) layer. We find that the group-level opinion emerges in two steps: (1) a fast and directional process, during which the number of the two kinds of hypocrites equalizes; and (2) a slower, random drift of opinions. We furthermore find that the initial abundances of opinions, but not the initial prevalence of hypocrisy, predicts the mean consensus time and determines the opinions' probabilities of winning.

We study a three-strategy spatial evolutionary prisoner's dilemma game with imitation and logit update rules (Vukov et al., 2015). Players can follow the always-cooperating, always-defecting or the win-stay-lose-shift (WSLS) strategies and gain their payoff from games with their direct neighbors on a square lattice. For the imitation dynamics, we found that the effects of spatiality combined with the presence of two cooperative strategies are so strong that they suppress even substantial changes in the payoff matrix, thus the phase diagrams are independent of the cyclic component's intensity. At the same time, this type of strategy update mechanism supports the formation of cooperative clusters that results in a cooperative society in a wider parameter range compared to the logit dynamics.

Last but not least, we constructed a general schematic model of gossip based in a simple action-response game (Számadó & Takács, in preparation). The most prominent role of gossip suggested by empirical evidence is to inform other individuals about potential norm breaking behaviour. While this function is strongly supported by empirical evidence, gossip can only function in this role if it is honest. Here we investigate the conditions of honest gossip and show that both cross-checking and punishment is a necessary condition; however, their relation is inverse: strong

cross-checking allows weaker punishment and weak cross-checking necessitates strong action against individuals breaking social norms.

We have carried out three main experiments as part of our experimental investigations.

First we conducted a laboratory experiment to study the relationship between third-party gossip and cooperation in the Prisoner's Dilemma (Samu et al., in preparation). The experiments took place in the spring of 2016 at the computer labs of the Corvinus University of Budapest. 160 participants (university students) played the game repeatedly, gave reputation scores to others, and sent evaluative information (gossip) to other selected participants. The experiment lasted for 75 minutes and involved no deception. Main manipulations concerned whether there was a conflict of interest in distributing reputation scores or not; and whether reputations scores counted for payment or not.

In the second experiment we continued our laboratory experiments with a study that investigates the role of trade-offs and signal cost in the honesty of signals (Számadó et al, in preparation). The main manipulation in this experiment is the cost of signal compared to the potential benefits. We have carried out the experiments that investigates the role of trade-offs and signal cost in the honesty of signals (Károly Takács, Flóra Samu, Szabolcs Számadó). The participants were playing a two-person signalling game, with two types of signallers, two types of signals and two possible response. Signallers new their type, but it was not observable by the receivers. Sharing the resource with one signaler type was beneficial for the receiver, while sharing with the other type was not. We were interested in the emergence of honest signalling, where the signal reveals the type of the signaller, as a function of signal cost. We have varied the equilibrium cost of signals and the tradeoff between different signal types.

Gossip, an evaluative talk about others who are not present, is believed to be an informal device that can help to solve the problem of cooperation in humans. While communication about previous acts and passing on reputational information could indeed be valuable for partner selection and conditional action in cooperation problems and could pose a threat of punishment to defectors, it is a puzzle what kind of mechanisms can make gossip and reputational information honest and credible. We propose two mechanisms that could be responsible for the efficiency of gossip for cooperation. One is that gossip could create social bonding between the sender and the receiver as suggested by Dunbar (1996) that might increase their future trust and cooperation potential. Another is the possibility of voluntary checks of received evaluative information from different sources that can work as an institutional device to ensure the honesty and credibility of gossip. We tested these mechanisms in our third experiment. We investigated the role of bonding and cross-checking in the context of the prisoners' dilemma game with the option of reputation building via trust scores (Samu & Takács, in preparation). We tested the efficiency of simplified social bonding and cross-checking devices in a laboratory experiment where subjects played the Prisoner's Dilemma with gossip interactions.

On top of these three main line of investigations discussed above we have carried out several other investigations that can be tied to our goal to understand the role of communication and reputation building in the maintenance of human cooperation.

We built a simple model of an idealized labor market (Takács, Squazzoni, and Bravo 2015), in which there is no objective difference in average quality between groups and hiring decisions are not biased in favor of any particular group. Our results show that inequality in employment emerges necessarily also in such idealized situations due to the limited supply of high quality individuals and asymmetric information. As an innovation of similar studies of the field, we have incorporated informal communication among employers about workers in our model. Recommendations are specific forms of gossip as they evaluate the merits and skills of third parties and these evaluations determine their future employment (entrance into new collaborative interactions). Our findings help to corroborate empirical findings on higher employment discrepancies in high rather than low status jobs. Moreover, we found that a stronger reliance on social network information has increased discrimination.

Both classical social psychological theories and recent formal models of opinion differentiation and bi-polarization assign a prominent role to negative social influence. We examine positive and negative influence with controlled exposure to opinions of other individuals in one experiment and with opinion exchange in another study (Takács et al., 2016). Results confirm that similarities induce attraction, but results do not support that discrepancy or disliking entails negative influence. Instead, our findings suggest a robust positive linear relationship between opinion distance and opinion shifts.

We have investigated peer review as a special social dilemma situation (Righi and Takács, 2017). It may not be clear how peer review, can work based on voluntary contributions of reviewers. In the absence of direct benefits scientist are not motivated to write high quality reviews, which can be seen as costly act of cooperation. We examine how the increased relevance of public good benefits (journal impact factor), the editorial policy of handling incoming reviews, and the acceptance decisions that take into account reputational information, can help the evolution of high-quality contributions from authors. We show with agent-based simulations why certain self-emerged current practices, such as the increased reliance on journal metrics and the reputation bias in acceptance, work efficiently for scientific development.

We investigated security market pricing process by means of individual based simulation (Biondi & Righi, 2017). It features collective market pricing mechanisms based upon evolving heterogenous expectations that incorporate signals of security issuer fundamental performance over time. We investigated the effect of information diffusion that corresponds to different institutional mechanisms. Our simulation analysis shows that transient information shocks can have permanent effects through mismatching reactions and self-reinforcing feedbacks, involving mispricing in both value and timing relative to the efficient market price series. We illustrate our results through paradigmatic cases of stochastic news, before generalising them to autocorrelated news.

We study the optimal referral strategy of a seller and its relationship with the type of communication channels among consumers (Carroni, et al., 2019). The seller faces a partially uninformed population of consumers, interconnected through a directed social network. Rewards are needed to bear a communication cost and to induce word of mouth (WOM) either privately (cost per contact) or publicly (fixed cost to inform all friends). We investigate (1) the incentives for the seller to move to a denser network, inducing either Private or Public WOM, and (2) the optimal mix between the two types of communication.

We investigate Zahavi's so-called Handicap Principle (Penn & Számadó, in press), which proposes that signals are honest because they are costly to produce. Here we provide a critical review of the Handicap Principle and its theoretical development. We explain why this idea is erroneous, and how it nevertheless became widely accepted as the leading explanation for honest signalling. Rather than being wasteful over-investments, honest signals evolve in this scenario because selection favours efficient and optimal investment into signal expression and minimizes signalling costs. This idea is very different from the handicap hypothesis, but it has been widely misinterpreted and equated to the Handicap Principle. Theoretical studies have since shown that signalling costs paid at the equilibrium are neither sufficient nor necessary to maintain signal honesty, and that honesty can evolve through differential benefits, as well as differential costs. There is no theoretical or empirical support for the Handicap Principle and the time is long overdue to usher this idea into an "honorable retirement".

Last but not least, we investigated the coevolution of cooperation and communication in Threshold Public Good Games. Because cooperation is often a costly act, to coordinate such actions actors can attempt to communicate, which is much less costly and hence can be deceitful, to ensure that cooperative decisions are only made once the chances of reaching the goal, thus meeting the threshold is secured. Our investigations reveal that communication can promote cooperation, even if the signal is deceitful, given that the rate is signalling is high, and individuals tend to re-evaluate their decisions to cooperate consistently based on the observed signals. Successful and stable collective cooperation can often evolve for very high cost values, which is nearly impossible in the classical Threshold Public Good Games without communication, regardless if the signal is honest or deceitful. Deceitful signals with easily biased re-evaluation strategies, the gullibles, can result in cooperation as much as honest signals and rigorous re-evaluation strategies, the stubborns. This bi-stability leads to successful collective actions in a wide-range of parameters conditions. Our results demonstrate that communication in collective actions can change the outcome significantly, and that even deceitful signals can promote cooperation.

We promised 10 articles published in international refereed with the impact factor of 10. The project has 10 published papers and one in print with five more in preparation. These papers have a cumulative impact factor: 36,6 (*PLOS One* (3), *Journal of Artificial Societies and Social Simulation*, *Physical Review E*, *Journal of Research on Adolescence*, *Scientometrics*, *Journal of Economic Interaction and Coordination*, *Management Science*, *Journal of Statistical Mechanics*, *Biological Reviews*). Two of the manuscripts not yet submitted are available from bioRxiv (https://www.biorxiv.org/).

On top of these research Szvetelszky Zsuzsanna wrote a book on gossip (in hungarian): Rejtett szervezetek. Typotex, Budapest, 2017. She has a new book under preparation with Bodor-Eranus Eliza. Also, she carried out a number of interviews and public lectures (not a full list):

Én vagyok itt, november 21. adás, 21.00-22.30 (Szvetelszky Zsuzsanna)

http://mta.hu/tudomany_hirei/pletyka-nelkul-nehezebb-lenne-az-egyutteles-136075/

http://www.hirstart.hu/hk/20150403_a_pletyka_mint_a_tarsadalmi_rend_fontos_tenyezoje

http://www.origo.hu/itthon/20150403-pletykakutatas-az-mta-n.html

http://index.hu/tudomany/2015/04/03/ahol_sokat_pletykalnak_nagyobb_a_tarsadalmi_rend/

http://www.estihirlap.hu/technika/tudomany/2015/04/03/a-pletykat-mint-a-tarsadalmi-rend-fontos-tenyezojet-kutatjak

http://www.havasok.hu/cikk/a-pletyka-mint-a-tarsadalmi-rend-fontos-tenyezoje

http://www.prae.hu/index.php?route=news/news&aid=26025

http://gondola.hu/cikkek/95756-Takacs_a_pletykat_kutatja.html

Detailed reports

Takács, Károly and Squazzoni, Flaminio (2015) High Standards Enhance Inequality in Idealized Labor Markets. Journal of Artificial Societies and Social Simulation. IF: 1.773

We built a simple model of an idealized labor market, in which there is no objective difference in average quality between groups and hiring decisions are not biased in favor of any particular group. Our results show that inequality in employment emerges necessarily also in such idealized situations due to the limited supply of high quality individuals and asymmetric information. Inequalities are exacerbated when employers have high standards and keep only the best workers in house. We found that ambitious workers get higher quality jobs even if ambition does not correlate or even negatively correlates with internal quality. Our findings help to corroborate empirical findings on higher employment discrepancies in high rather than low status jobs.

Vukov, Jeromos; Varga, Levente; Allen, Benjamin; Nowak, Martin A. and Szabó, György (2015) Payoff components and their effects in a spatial three-strategy evolutionary social dilemma. Physical Review E. IF: 2.288

We study a three-strategy spatial evolutionary prisoner's dilemma game with imitation and logit update rules. Players can follow the always-cooperating, always-defecting or the win-stay-lose-shift (WSLS) strategies and gain their payoff from games with their direct neighbors on a square lattice. The friendliness parameter of the WSLS strategy—characterizing its cooperation probability in the first round—tunes the cyclic component of the game determining whether the game can be characterized by a potential. We measured and calculated the phase diagrams of the system for a wide range of parameters. When the game is a potential game and the logit rule is applied, the theoretically predicted phase diagram agrees very well with the simulation results. Surprisingly, this phase diagram can be accurate even in the nonpotential case if there are only two surviving strategies in the stationary state; this result harmonizes with the fact that all 2 × 2 games are potential games. For the imitation dynamics, we found that the effects of spatiality combined with the presence of two cooperative strategies are so strong that they suppress even substantial changes in the payoff matrix, thus the phase diagrams are independent of the cyclic component's intensity. At the same time, this type of strategy update mechanism supports the formation of cooperative clusters that results in a cooperative society in a wider parameter range compared to the logit dynamics.

Pál, Judit; Stadtfeld, Christoph; Grow, André, and Takács, Károly (2016) Status Perceptions Matter: Understanding Disliking among Adolescents. Journal of Research on Adolescence. IF: 2.480

The emergence of disliking relations depends on how adolescents perceive the relative informal status of their peers. This phenomenon is examined on a longitudinal sample using dynamic network analysis (585 students across 16 classes in five schools). As hypothesized, individuals dislike those who they look down on (disdain), and conform to others by disliking those who they perceive as being looked down on by their peers (conformity). The inconsistency between status perceptions also leads to disliking, when individuals do not look up to those who they perceive to be admired by peers (frustration). Adolescents are not more likely to dislike those who they look up to (admiration). The results demonstrate the role of status perceptions on disliking tie formation.

Szabolcs Számadó, Ferenc Szalai & István Scheuring (2016) Deception undermines the stability of cooperation in games of indirect reciprocity. PLOS One. IF: 3.540

Indirect reciprocity is often claimed as one of the key mechanisms of human cooperation. It works only if there is a reputational score keeping and each individual can inform with high probability which other individuals were good or bad in the previous round. Gossip is often proposed as a mechanism that can maintain such coherence of reputations in the face of errors of transmission. Random errors, however, are not the only source of uncertainty in such situations. The possibility of deceptive communication, where the signallers aim to misinform the receiver cannot be excluded. While there is plenty of evidence for deceptive communication in humans the possibility of deception is not yet incorporated into models of indirect reciprocity. Here we show that when deceptive strategies are allowed in the population it will cause the collapse of the coherence of reputations and thus in turn it results the collapse of cooperation. This collapse is independent of the norms and the cost and benefit values. It is due to the fact that there is no selection for honest communication in the framework of indirect reciprocity. It follows that indirect reciprocity can be only proposed plausibly as a mechanism of human cooperation if additional mechanisms are specified in the model that maintains honesty.

Takács, Károly, Flache, Andreas, and Mäs, Michael (2016) Discrepancy and Disliking Do Not Induce Negative Opinion Shifts. PLOS One. IF: 3.540

Both classical social psychological theories and recent formal models of opinion differentiation and bi-polarization assign a prominent role to negative social influence. Negative influence is defined as shifts away from the opinion of others and hypothesized to be induced by discrepancy with or disliking of the source of influence. There is strong empirical support for the presence of positive social influence (a shift towards the opinion of others), but evidence that large opinion differences or disliking could trigger negative shifts is mixed. We examine positive and negative influence with controlled exposure to opinions of other individuals in one experiment and with opinion exchange in another study. Results confirm that similarities induce attraction, but results do not support that discrepancy or disliking entails negative influence. Instead, our findings suggest a robust positive linear relationship between opinion distance and opinion shifts.

Simone Righi & Károly Takács (2017) The miracle of peer review and development in science: an agent-based model. Scientometrics. IF: 2.147

It is not easy to rationalize how peer review, as the current grassroots of science, can work based on voluntary contributions of reviewers. There is no rationale to write impartial and thorough evaluations. If reviewers are unmotivated to carefully select high quality contributions, there is no risk in submitting low-quality work by authors. As a result, scientists face a social dilemma: if everyone acts according to his or her own self-interest, the outcome is low scientific quality. We examine how the increased relevance of public good benefits (journal impact factor), the editorial policy of handling incoming reviews, and the acceptance decisions that take into account reputational information, can help the evolution of high-quality contributions from authors. High effort from the side of reviewers is problematic even if authors cooperate: reviewers are still best off by producing low-quality reviews, which does not hinder scientific development, just adds random noise and unnecessary costs to it. We show with agent-based simulations why certain self-emerged current practices, such as the increased reliance on journal metrics and the reputation bias in

acceptance, work efficiently for scientific development. Our results find no proper guidelines, however, how the system of voluntary peer review with impartial and thorough evaluations could be sustainable jointly with rapid scientific development.

Biondi, Y., & Righi, S. (2017). Much ado about making money: the impact of disclosure, news and rumors on the formation of security market prices over time. Journal of Economic Interaction and Coordination. IF: 0.45

This article develops an agent-based model of security market pricing process, capable to capture main stylised facts. It features collective market pricing mechanisms based upon evolving heterogenous expectations that incorporate signals of security issuer fundamental performance over time. Distinctive signaling sources on this performance correspond to institutional mechanisms of information diffusion. These sources differ by duration effect (temporary, persistent, and permanent), confidence, and diffusion degree among investors over space and time. Under full and immediate diffusion and balanced reaction by all the investors, the value content of these sources is expected to be consistently and timely integrated by the market price process, implying efficient pricing. By relaxing these quite heroic conditions, we assess the impact of distinctive information sources over market price dynamics, through financial systemic properties such as market price volatility, exuberance and errancy, as well as market liquidity. Our simulation analysis shows that transient information shocks can have permanent effects through mismatching reactions and self-reinforcing feedbacks, involving mispricing in both value and timing relative to the efficient market price series. This mispricing depends on both the information diffusion process and the ongoing information confidence mood among investors over space and time. We illustrate our results through paradigmatic cases of stochastic news, before generalising them to autocorrelated news. Our results are further corroborated by robustness checks over the parameter space and across several market trading mechanisms.

Gastner, M.T., Oborny, B. and Gulyás, M (2018) Consensus time in a voter model with concealed and publicly expressed opinions. J. Stat. Mech. IF: 2.404

The voter model is a simple agent-based model to mimic opinion dynamics in social networks: a randomly chosen agent adopts the opinion of a randomly chosen neighbour. This process is repeated until a consensus emerges. Although the basic voter model is theoretically intriguing, it misses an important feature of real opinion dynamics: it does not distinguish between an agent's publicly expressed opinion and her inner conviction. A person may not feel comfortable declaring her conviction if her social circle appears to hold an opposing view. Here we introduce the Concealed Voter Model where we add a second, concealed layer of opinions to the public layer. If an agent's public and concealed opinions disagree, she can reconcile them by either publicly disclosing her previously secret point of view or by accepting her public opinion as inner conviction. We study a complete graph of agents who can choose from two opinions. We define a martingale M that determines the probability of all agents eventually agreeing on a particular opinion. By analyzing the evolution of M in the limit of a large number of agents, we derive the leading-order terms for the mean and standard deviation of the consensus time (i.e. the time needed until all opinions are identical). We thereby give a precise prediction by how much concealed opinions slow down a consensus.

Carroni, E., Pin, P., & Righi, S. (2019). Bring a friend! Privately or Publicly? Management Science. IF: 4.219

We study the optimal referral strategy of a seller and its relationship with the type of communication channels among consumers. The seller faces a partially uninformed population of consumers, interconnected through a directed social network. In the network, the seller offers rewards to informed consumers (influencers) conditional on inducing purchases by uninformed consumers (influenced). Rewards are needed to bear a communication cost and to induce word of mouth (WOM) either privately (cost per contact) or publicly (fixed cost to inform all friends). From the seller's viewpoint, eliciting Private WOM is more costly than eliciting Public WOM. We investigate (1) the incentives for the seller to move to a denser network, inducing either Private or Public WOM, and (2) the optimal mix between the two types of communication. A denser network is found to be always better not only for information diffusion but also for seller's profits, as long as Private WOM is concerned. Differently, under Public WOM, the seller may prefer an environment with less competition between informed consumers, and the presence of highly connected influencers (hubs) is the main driver to make network density beneficial to profits. When the seller is able to discriminate between Private and Public WOM, the optimal strategy is to cheaply incentivize the more connected people to pass on the information publicly and then offer a high bonus for Private WOM.

Gastner, M.T., Takács, K., Gulyás, M., Szvetelszky Zs., and Oborny, B. (2019) The Impact of Hypocrisy on Opinion Formation: A Dynamic Model. PLOS One. IF: 3.540

Humans have a demonstrated tendency to copy or imitate the behavior and attitude of others and actively influence each other's opinions. In plenty of empirical contexts, publicly revealed opinions are not necessarily in line with internal opinions, causing complex social influence dynamics. We study to what extent hypocrisy is sustained during opinion formation and how hidden opinions change the convergence to consensus in a group. We build and analyze a modified version of the voter model with hypocrisy in a complete graph with a neutral competition between two alternatives. We compare the process from various initial conditions, varying the proportions between the two opinions in the external (revealed) and internal (hidden) layer. According to our results, hypocrisy always prolongs the time needed for reaching a consensus. In a complete graph, this time span increases linearly with group size. We find that the group-level opinion emerges in two steps: (1) a fast and directional process, during which the number of the two kinds of hypocrites equalizes; and (2) a slower, random drift of opinions. During stage (2), the ratio of opinions in the external layer is approximately equal to the ratio in the internal layer; that is, the hidden opinions do not differ significantly from the revealed ones at the group level. We furthermore find that the initial abundances of opinions, but not the initial prevalence of hypocrisy, predicts the mean consensus time and determines the opinions' probabilities of winning. These insights highlight the unimportance of hypocrisy in consensus formation under neutral conditions. Our results have important societal implications in relation to hidden voter preferences in polls and improve our understanding of opinion formation in a more realistic setting than that of conventional voter models.

Dustin J. Penn and Szabolcs Számadó (in press) The Handicap Principle: how an erroneous hypothesis became a scientific principle. Biological Reviews. IF: 10.288

The most widely cited explanation for the evolution of reliable signals is Zahavi's so-called Handicap Principle, which proposes that signals are honest because they are costly to produce. Here we provide a critical review of the Handicap Principle and its theoretical development. We explain why this idea is erroneous, and how it nevertheless became widely accepted as the leading explanation for honest signalling. In 1975, Zahavi proposed that elaborate secondary sexual characters impose "handicaps" on male survival, not due to inadvertent signalling trade offs, but as a mechanism that functions to demonstrate males' genetic quality to potential mates. His handicap hypothesis received many criticisms, and in response, Zahavi clarified his hypothesis and explained that it assumes that signals are wasteful as well as costly, and that they evolve because wastefulness enforces honesty. He proposed that signals evolve under "signal selection", a non-Darwinian type of selection that favours waste rather than efficiency. He maintained that the handicap hypothesis provides a general principle to explain the evolution of all types of signalling systems, i.e. the Handicap Principle. In 1977, Zahavi proposed a second hypothesis for honest signalling, which received many different labels and interpretations, although it was assumed to be another example of handicap signalling. In 1990, Grafen published models that he claimed vindicated Zahavi's Handicap Principle. His conclusions were widely accepted and the Handicap Principle subsequently became the dominant paradigm for explaining the evolution of honest signalling in the biological and social sciences. Researchers have subsequently focused on testing predications of the Handicap Principle, such as measuring the absolute costs of honest signals (and using energetic and other proximate costs as proxies for fitness), but very few have attempted to test Grafen's models. We show that Grafen's models do not support the handicap hypothesis, although they do support Zahavi's second hypothesis, which proposes that males adjust their investment into the expression of their sexual signals according to their condition and ability to bear the costs (and risks to their survival). Rather than being wasteful over-investments, honest signals evolve in this scenario because selection favours efficient and optimal investment into signal expression and minimizes signalling costs. This idea is very different from the handicap hypothesis, but it has been widely misinterpreted and equated to the Handicap Principle. Theoretical studies have since shown that signalling costs paid at the equilibrium are neither sufficient nor necessary to maintain signal honesty, and that honesty can evolve through differential benefits, as well as differential costs. There have been increasing criticisms of the Handicap Principle, but they have focused on the limitations of Grafen's model and overlooked the fact that it is not a handicap model. This model is better understood within a Darwinian framework of adaptive signalling trade-offs, without the added burden and confusing logic of the Handicap Principle. There is no theoretical or empirical support for the Handicap Principle and the time is long overdue to usher this idea into an "honorable retirement".

Flóra Samu, Szabolcs Számadó and Károly Takács (in preparation) Scarce and directly beneficial reputations support cooperation.

A human solution to the problem of cooperation is the construction and maintenance of informal reputation hierarchies. Reputational information contributes to cooperation by providing guidelines about previous group-beneficial or free-rider behavior of opponents in social dilemma interactions. How reputation information could be credible, however, when outcomes of interactions are not publicly known, remains a puzzle. In this study, we propose that credibility could be ensured if reputation is a scarce resource and it is not believed to be earned for direct benefits. We tested

these propositions in a laboratory experiment in which participants played two-person Prisoner's Dilemma games without partner selection, could observe some other interactions and could communicate reputational information about possible prospective opponents to each other. We found that scarcity is a necessary condition for reputation scores to be informative. While cooperation has not been sustained at a high level in any of the conditions, reputational information clearly influenced cooperation decisions. The possibility of exchanging third-party information was able to increase the level of cooperation the most if reputation was a scarce resource and contrary to our expectations, when reputational scores have been directly translated into monetary benefits.

Szabolcs Számadó, Flóra Samu and Károly Takács (in preparation) Condition-dependent trade-offs maintain honest signaling: A laboratory experiment.

How and why animals and humans signal reliably is a key issue in biology and social sciences. For many years the dominant paradigm in biology was the Handicap Principle. It claims a causal relationship between honesty and signal cost and thus predicts that honest signals have to be costly. Contrary to the Handicap Principle more recent models predict that honest signaling is maintained by condition dependent signaling trade-offs and honest signals need not be costly at the equilibrium. Due to the difficulties of manipulating signal cost and signal trade-offs in nature there is surprisingly little evidence to test these predictions. Here we conduct a human laboratory experiment with a two-factorial design to test the role of equilibrium signal cost vs. trade-offs in the maintenance of honest signaling. We have found that the trade-off condition has much higher influence on the reliability of communication than the equilibrium cost condition. The highest level of honesty was observed in the condition dependent trade-off condition as predicted by recent models. Negative cost, contrary to the prediction of the Handicap Principle, promoted even higher level of honesty than the other type of costs under this condition.

Flóra Samu and Károly Takács (in preparation) The role of bonding and cross-checking in reputation building: A laboratory experiment.

Gossip, an evaluative talk about others who are not present, is believed to be an informal device that can help to solve the problem of cooperation in humans. While communication about previous acts and passing on reputational information could indeed be valuable for partner selection and conditional action in cooperation problems and could pose a threat of punishment to defectors, it is a puzzle what kind of mechanisms can make gossip and reputational information honest and credible. We propose two mechanisms that could be responsible for the efficiency of gossip for cooperation. One is that gossip could create social bonding between the sender and the receiver as suggested by Dunbar (1996) that might increase their future trust and cooperation potential. Another is the possibility of voluntary checks of received evaluative information from different sources that can work as an institutional device to ensure the honesty and credibility of gossip. We tested these mechanisms in our third experiment. We investigated the role of bonding and cross-checking in the context of the prisoners' dilemma game with the option of reputation building via trust scores. We tested the efficiency of simplified social bonding and cross-checking devices in a laboratory experiment where subjects played the Prisoner's Dilemma with gossip interactions.

Szabolcs Számadó and Károly Takács (in preparation) The paradox of honest gossip: the role of cross-checking and punishment.

Gossip plays a crucial role in the maintenance of human cooperation. Both theory and empirical evidence strongly support this conclusion. On the one hand, experimental work has shown that even the bare threat of gossip can increase cooperation in social dilemma games; on the other hand, models shown that gossip plays a crucial role in explaining large scale cooperation. The most prominent role of gossip suggested by empirical evidence is is to inform other individuals about potential norm breaking behaviour. This is the role assumed in games of IR as well: i.e. to inform others about the reputation of the focal individual; where reputation describes the cooperativeness of that individual in the light of the social norm adopted in the population. While this function is strongly supported by empirical evidence, gossip can only function in this role if it is honest. It is easy to see if one allows dishonest gossip the coherence of reputation breaks down across individuals, which in turn undermines cooperation. This crucial issue has been avoided by empirical or theoretical investigations so far; models assumed honest gossip by default and experiments rarely investigated situations where gossipers had incentives to be dishonest. Here we investigate the conditions of honest gossip and show that both cross-checking and punishment is a necessary condition; however, their relation is inverse: strong cross-checking allows weaker punishment and weak cross-checking necessitates strong action against individuals breaking social norms.

Gergely Boza & Szabolcs Számadó (in preparation) The coevolution of communication and cooperation in non-linear collective actions

Societies face various collective actions, in which in conflict with the selfish interest of every individual, a certain number of group members must act cooperatively in order to reach a collective goal. Such social dilemmas are best described as Threshold Public Good Games, in which the collective goal is reached only if the number of cooperative decisions reaches a threshold. Because cooperation is often a costly act, to coordinate such actions actors can attempt to communicate, which is much less costly and hence can be deceitful, to ensure that cooperative decisions are only made once the chances of reaching the goal, thus meeting the threshold is secured. Our investigations reveal that communication can promote cooperation, even if the signal is deceitful, given that the rate is signalling is high, and individuals tend to re-evaluate their decisions to cooperate consistently based on the observed signals. Successful and stable collective cooperation can often evolve for very high cost values, which is nearly impossible in the classical Threshold Public Good Games without communication, regardless if the signal is honest or deceitful. Deceitful signals with easily biased re-evaluation strategies, the gullibles, can result in cooperation as much as honest signals and rigorous re-evaluation strategies, the stubborns. This bi-stability leads to successful collective actions in a wide-range of parameters conditions. Our results demonstrate that communication in collective actions can change the outcome significantly, and that even deceitful signals can promote cooperation.

Publications

Published

Takács, Károly and Squazzoni, Flaminio (2015) High Standards Enhance Inequality in Idealized Labor Markets. *Journal of Artificial Societies and Social Simulation*, 18 (4) 2.

https://doi.org/10.18564/jasss.2940

Vukov, Jeromos; Varga, Levente; Allen, Benjamin; Nowak, Martin A. and Szabó, György (2015) Payoff components and their effects in a spatial three-strategy evolutionary social dilemma. *Physical Review E*. 92, 012813

https://doi.org/10.1103/PhysRevE.92.012813

Pál, Judit; Stadtfeld, Christoph; Grow, André, and Takács, Károly (2016) Status Perceptions Matter: Understanding Disliking among Adolescents. *Journal of Research on Adolescence*.

https://doi.org/10.1111/jora.12231

Szabolcs Számadó, Ferenc Szalai & István Scheuring (2016) Deception undermines the stability of cooperation in games of indirect reciprocity. *PLOS One*. e0147623.

https://doi.org/10.1371/journal.pone.0147623

Takács, Károly, Flache, Andreas, and Mäs, Michael (2016) Discrepancy and Disliking Do Not Induce Negative Opinion Shifts. *PLOS One*. 11(6): e0157948.

https://doi.org/10.1371/journal.pone.0157948

Simone Righi & Károly Takács (2017) The miracle of peer review and development in science: an agent-based model. *Scientometrics*. 113:587–607

https://doi.org/10.1007/s11192-017-2244-y

Biondi, Y., & Righi, S. (2017). Much ado about making money: the impact of disclosure, news and rumors on the formation of security market prices over time. *Journal of Economic Interaction and Coordination*, 1-30.

https://doi.org/10.1007/s11403-017-0201-8

Gastner, M.T., Oborny, B. and Gulyás, M (2018) Consensus time in a voter model with concealed and publicly expressed opinions. *J. Stat. Mech*. 063401

https://doi.org/10.1088/1742-5468/aac14a

Carroni, E., Pin, P., & Righi, S. (2019). Bring a friend! Privately or Publicly? *Management Science*.

https://doi.org/10.1287/mnsc.2018.3282

Gastner, M.T., Takács, K., Gulyás, M., Szvetelszky Zs., and Oborny, B. 2019. The Impact of Hypocrisy on Opinion Formation: A Dynamic Model. *PLOS One*, 14 (6), e0218729.

https://doi.org/10.1371/journal.pone.0218729

In press

Dustin J. Penn and Szabolcs Számadó (in press) The Handicap Principle: how an erroneous hypothesis became a scientific principle. *Biological Reviews*.

In preparation

Flóra Samu, Szabolcs Számadó and Károly Takács (in preparation) Scarce and directly beneficial reputations support cooperation.

https://BIORXIV/2019/788836

Szabolcs Számadó, Flóra Samu and Károly Takács (in preparation) Condition-dependent trade-offs maintain honest signaling: A laboratory experiment.

https://BIORXIV/2019/788828

Flóra Samu and Károly Takács (in preparation) The role of bonding and cross-checking in reputation building: A laboratory experiment.

Szabolcs Számadó and Károly Takács (in preparation) The paradox of honest gossip: the role of cross-checking and punishment.

Gergely Boza & Szabolcs Számadó (in preparation) The coevolution of communication and cooperation in non-linear collective actions