Introduction

Author for correspondence:
S. Számadó
e-mail: szamado.szabolcs@tk.mta.hu

The language of cooperation: reputation and honest signalling

S. Számadó[1,2], D. Balliet[3], F. Giardini[4], E. A. Power[5] and K. Takács[6,2]

[1]Department of Sociology and Communication, BUTE, Budapest, Hungary
[2]CSS-RECENS, Centre for Social Sciences, Budapest, Hungary
[3]Department of Experimental and Applied Psychology, Vrije Universiteit, Amsterdam, The Netherlands
[4]Department of Sociology, University of Groningen, Groningen, The Netherlands
[5]Department of Methodology, London School of Economics and Political Science, London, UK
[6]Institute for Analytical Sociology, Linköping University, Linköping, Sweden

SS, 0000-0003-2204-9705; DB, 0000-0001-5365-9675; FG, 0000-0001-7512-0048;
EAP, 0000-0002-3064-2050; KT, 0000-0001-9126-3233

Large-scale non-kin cooperation is a unique ingredient of human success. This type of cooperation is challenging to explain in a world of self-interested individuals. There is overwhelming empirical evidence from different disciplines that reputation and gossip promote cooperation in humans in different contexts. Despite decades of research, important details of reputation systems are still unclear. Our goal with this theme issue is to promote an interdisciplinary approach that allows us to explore and understand the evolution and maintenance of reputation systems with a special emphasis on gossip and honest signalling. The theme issue is organized around four main questions: What are the necessary conditions for reputation-based systems? What is the content and context of reputation systems? How can reputations promote cooperation? And, what is the role of gossip in maintaining reputation systems and thus cooperation?

This article is part of the theme issue 'The language of cooperation: reputation and honest signalling'.

1. Introduction

Cooperation among unrelated members is one of the most prominent features of human societies that makes life together highly successful [1–4]. Assigning reputation to others, sharing reputations of others and conditioning behaviour on reputation are among the main driving forces of such cooperation [5–8]. While the effect and role of reputations on cooperation is widely studied [8–16], challenging questions are left unanswered about the evolution and conditions of reputation systems, the context and content of reputation, and the operation of reputation-based systems. Signalling and communication are essential for reputations being shared, and among humans, third-party sharing of reputations through gossip can amplify the potential of reputations to sustain cooperation. Signals and gossip, however, are not necessarily honest and the informal mechanisms that can safeguard the functioning of reputation systems must be uncovered. In this thematic issue, scholars coming from different backgrounds compare theories, methods and perspectives, and offer new empirical research in order to lay the foundations for an interdisciplinary and systematic research programme on reputation, gossip and honesty for sustainable cooperation. The articles in this special issue all draw our attention to these complexities of the workings of reputation systems, asking: (i) What are the necessary conditions for reputation-based systems? (ii) What is the content and context of reputation systems? (iii) How can reputations promote cooperation? and (iv) What is the role of gossip in maintaining reputation systems and thus cooperation?

This theme issue is based on the workshop with the same title held in the Lorentz Center in Leiden (Netherlands, September 2019). We are very grateful

to both the Lorentz Center and the Royal Society for providing the physical and intellectual space for the discussion of this topic.

2. Background

Cooperation is the production of mutual benefits at a cost to the individual. The collective benefits that cooperation can produce help explain the success of societies that establish and sustain cooperation. Yet, cooperation can be difficult to achieve, since there can be higher immediate personal costs of cooperation that curb individuals' motivation to cooperate voluntarily. This inherent puzzle of cooperation explains why it is so rare in nature and why it brings about major transitions in evolution once it has been established [1]. *Large-scale non-kin cooperation* is unique to the human species. While social living species can display an impressive level of coordinated effort, this cooperation typically is fostered by kin selection, where cooperation is promoted by sufficiently high levels of relatedness between individuals [1]. Human societies are uniquely able to achieve high levels of cooperation, even with low levels of biological relatedness between members of the society. Cooperation in human societies is supported by a complex set of tools, including informal and formal mechanisms. Some mechanisms are able to operate because humans have evolved advanced cognitive and social skills and a complex system of communication that includes language and an opportunity to exchange information about other individuals. A wealth of empirical data from different disciplines demonstrates that this information exchange is crucial in maintaining cooperation [13,17–19].

A *reputation* can be seen as an evaluation of an individual based on their past action(s). The importance of such a reputation is that it will determine how others will treat/react to them in future encounters. This reputation can be solely based on *individual* experiences (e.g. based on past interactions and/or observations of the ego), or it can be based on the *aggregated* experiences of a group. The resulting reputations can be *stored privately* (e.g. in the brains of the individuals within the group), allowing different individuals to have different reputational assessments of the same person; or they can be *stored publicly* (e.g. a public webpage of tax evaders).

A *reputation system* based on aggregated experiences includes (i) a set of decision rules to guide how individuals should be evaluated based on their actions, usually called *social norms*; (ii) a *communication system* that allows for the exchange of experiences and evaluations; and (iii) a set of decision rules that guides the actions towards individuals with different reputations, usually called *behavioural norms* or action rules [20–25].

3. This issue's contributions

A first and fundamental point at issue is the *evolution of reputation systems*. While some animals have reputation systems based on individual experiences—one famous example is the cleaner wrasse client system [26]—aggregated reputation systems are missing. This begs the question: if aggregated reputation systems are such a powerful tool to promote cooperation, why do they only occur in humans? The

answer to this question likely involves the human ability of language. Aggregated reputations are formed through communication, thus the lack of language seems to pose a barrier for other species. Reputation systems require other cognitive skills as well, such as the ability to judge the observed actions of others based on the context of the action. Manrique *et al.* [27] evaluate these cognitive requirements in their review. They argue that a large working memory enables humans to better understand others' mental states, such as through perspective-taking and the attribution of intentions, which also allows humans to create and follow norms. Therefore, a large working memory may provide the foundation for increasingly complex reputation-based cooperation in humans. They conclude that most non-human species lack a working memory capacity with sufficient computational power to support the formation of complex systems of reputation-based cooperation.

Inference about the mental states of others plays a key role in our judgements of their actions. For example, evolutionary models show that cooperation can be stabilized by distinguishing between different reasons for the same action, such as selfish defection versus altruistic punishment. However, anthropologists have suggested that some societies seem to disregard mental states when passing judgement. Barrett & Saxe [28] argue that this is likely a misapprehension. They propose instead that the extent to which mental states are considered in moral judgement depends on the context and the reasons for judgement. There is no society that would completely disregard mental states; instead, it is the importance of mental states in making moral judgement that varies from context to context. It follows that it is important to find out why different societies evaluate actors in the same context differently. They argue that in general there are three important determinants: 'Who are the actors?' e.g. individuals versus corporations; 'What are they doing?' e.g. doing inherently dangerous actions versus non-dangerous actions; and finally 'Why is the judgement being made?' e.g. truth making versus restoring social cohesion.

Models of indirect reciprocity (IR) attempt to fully specify and analyse a reputation system (IR; [5,7,9,20,29–32]). Models of IR consider interactions in a large population that prevent the evolution of direct reciprocity (I scratch your back, you scratch mine). Instead of such direct help, IR models assume a helping system mediated by reputation (good versus bad). Individuals can gain a good reputation by helping others and individuals of good reputation will be helped by other members of the group. While the idea is very simple, details of IR models can be very complex [33]. Santos *et al.* [34] quantify cognitive complexity in IR and evaluate recent advances and potential extensions of IR models. They identify three determinants of cooperation under IR that can affect both the level of cooperation and the level of complexity and information requirements for IR to promote cooperation. These determinants are: (i) costly reputation spread, where individuals have to pay a cost to share information; (ii) interaction observability, which will influence the consistency of private reputations; and (iii) empathy when judging others, which again could influence the consistency of reputations.

Another potential use of reputation is *partner-choice*. As in IR models, individuals gain a good reputation by helping others, but in this framework, that reputation is then used to inform decisions about the selection of partners. These

models are often labelled *competitive altruism* because competition to be selected by potential partners can lead to an escalation of helping behaviour [12,35–48]. Roberts *et al.* [49] offer a comparative perspective on models of IR and reputation-based partner choice, by analysing their assumptions. This comparison shows that despite superficial similarities, such as the success of both models in describing how reputation can promote cooperation, these models use vastly different assumptions. While IR models assume one-off interactions between randomly chosen individuals, reputation-based partner choice assumes active partner choice. The authors argue that this latter scenario is more relevant in interactions among humans, thus reputation-based partner choice models deserve more attention.

Reputation systems can be used in different contexts thus the *content of reputations* may greatly vary. While most models investigate a simplified system of 'good' versus 'bad' reputation (see IR), real-life reputation systems are more complex [50–59]. Romano *et al.* [60] investigate how differences across social ecologies, institutions and cultures influence the workings of reputation systems. Their interdisciplinary review also contrasts empirical insights from various disciplines about ecological factors that shape reputations with the way models address the questions of reputation-based cooperation. Garfield *et al.* [61] use the Human Relations Area Files ethnographic database to characterize the cross-cultural diversity in reputation domains. They find that while there is substantial variation (not only between societies but also between men and women), reputations for cultural conformity, prosociality, social status and neural capital are widespread.

Another important question is whether the *everyday use of gossip* is consistent with the leading theories of reputation and cooperation. Dores *et al.* [62] test several predictions from theories of IR and reputation-based partner selection about the content of everyday gossip. Using experience sampling methods, the researchers asked a Dutch community sample to report on recent gossip events they experienced in order to acquire information about the content and the consequences of gossip in everyday life. Their results lend support to the hypothesis that gossip is often about other's cooperativeness and that it has consequences on the target's reputation and the receiver's intentions to interact with them in the future. Hess & Hagen [63] used experimental and survey vignettes among MTurkers and California sorority women to investigate how the social context of gossip impacts its content and its perceived consequences. They found that people were more likely to gossip, and to share negative gossip, when situations were more competitive, and that having allies may help militate against negative gossip.

A crucial ingredient of aggregated reputation systems is *communication*. While communication is an important part of aggregated reputation systems, humans can communicate about a wide variety of things. *Signalling theory* investigates the conditions of deceptive versus honest (informative) interactions between signallers and receivers [64–66]. Signals may carry information about intentions, conditions or quality [15,66–69]. Seminal signalling models investigate *signals of quality* [65,70,71]—these models, often named as 'honest signalling' models, investigate the conditions when observable signals correlate with non-observable (but relevant) qualities (e.g. mate value). Dumas *et al.* [72] investigate how

additional information about an individual's social prominence or social capital can alter the signalling system, by using analytical and agent-based models. The inclusion of this information can generate a 'reputational shield' and a 'reputational poverty trap', where a 'reputational shield', involves low-quality individuals being mistaken as high quality due to their social standing, and a 'reputational poverty trap' implies that high-quality individuals are unable to improve their reputation due to a lack of social capital. Drawing on a case study of religious signalling in South India, they suggest that more attention should be paid to the consequences of the complex feedbacks between reputation, prominence and social capital.

Another important issue of signalling theory is the evolution of *conventional signals* [73]. The correlation between meaning and form is an emergent property of conventional signalling systems; the same form might have different meanings in different signalling systems. *Ritualization* is often proposed as a mechanism to explain the evolution of conventional signals. Unlike in other animals, in humans rituals can gain normative force through individuals' consensual understanding that these behaviours ought to be followed because they constitute markers of group distinction. Przepiorka & Diekmann [74] call such norms signalling norms and investigate the conditions under which they can emerge. They argue it is the uncertainty about in-group members' trustworthiness that drives the emergence of signalling norms [75–77]. As they identify the conditions for the emergence of seemingly wasteful and even harmful behaviours, they find that it is not the quality-revealing aspect but the in-group commitment such behaviours produce that stabilizes these very behaviours. That is, signalling norms promote parochial cooperation also because they impede cooperation with out-group members [78–80].

Crucially, humans are able to communicate about individuals who are not present. Evaluative communication of past experiences and observations about an absent third party is called *gossip*. A wealth of empirical studies shows that the 'threat of gossip' can promote cooperative behaviour [18,81–89]. This threat of gossip refers to situations where individuals know that they are being observed and that these observations can be communicated to other group members. A general conclusion is that individuals will behave more cooperatively under the threat of gossip than they would do otherwise. This clearly shows that humans are aware of the importance of reputations and of the importance of gossip as a mediator in reputation building in systems of aggregated reputation.

Gossip is often viewed as a cheap tool to maintain cooperation, but gossip alone will not maintain cooperation in the long term. The threat of gossip can promote a surge of cooperation, but such cooperation will collapse if the threat of gossip is not backed up by efficient sanctions against non-cooperators [89]. This threat may then be ineffective if efficient sanctions are not possible, or if gossip cannot consistently inform individuals of who was cooperative and who was not. This raises the issue of the *reliability of gossip*. While most models assume that gossip is honest by default (i.e. individuals will communicate their experiences and observations with perfect fidelity), this need not be the case. There is rich evidence from different disciplines that humans are strategic and even manipulative in their communication [90–92]. We may not lie as much as we could

(see [93] for review), but it does not change the fact that gossip may not always be honest [94–99]. Studies show that widespread dishonest gossip could result in the collapse of reliable reputations, and thus the collapse of cooperation [100].

So, what maintains the *honesty of gossip*? Several scholars took up the challenge to investigate the conditions which promote honest gossip. Fonseca & Peters [101] investigated the ability of people to *detect* dishonest gossip and the willingness to reward those who spread honest gossip in an experimental Trust game. They found that people are reasonably adept at detecting dishonest gossip, yet the valence of gossip may matter for how gossipers are treated. People are less willing to trust gossipers who share negative gossip, even if this gossip is honest or ultimately helpful. In a complementary study, Giardini *et al.* [102] investigated the effect of partner choice on gossip and cooperation in a sequential Public Goods game. Introducing partner choice increased both the frequency of cooperative choices and of honest gossip among cooperators. When participants competed to be chosen, there was a correlation between cooperation and honesty, thus suggesting that the opportunity for partner choice promotes both the honesty of gossip and cooperation. The way in which gossip was reported was also different between cooperators and defectors who preferred to pass information as coming from someone else as a way of distancing from the source and thus avoid retaliation. Samu & Takács [103] studied cooperative behaviour in a laboratory experiment where subjects played two-person Prisoner's Dilemma games with varying interaction partners and were allowed to gossip about their experiences. They investigated the role of two mechanisms that could support the honesty and credibility of gossip: they tested how cross-checking the validity of third-party information and allowing for social bonding between the sender and the receiver affected the reliability of reputations, and ultimately cooperation. They found that, while overall individuals trusted gossip, neither cross-checking nor bonding opportunities increased cooperation. However, when there was a stronger competition for reputation, this tended to increase cooperation. Wu *et al.* [104] use fitness interdependence theory (or stake, [105–108]) to model individuals' gossip strategies (honest versus dishonest gossip) in a gossip triad across situations where the gossiper has varying levels of fitness interdependence with the target and/or the receiver. They investigate gossip strategies across four seminal games (i.e. Stag-hunt game, Snowdrift game, Helping game and Punishment game) to illustrate how the gossiper's action is determined by (i) the gossiper's fitness interdependence with the receiver and the target, and (ii) the marginal cost/benefit in terms of payoff differences between two possible game actions for the receiver and the target (i.e. game type). These models suggest a simple 'matching rule' that gossipers can use to make optimal decisions even under noise: be honest when there is a perfect match between the valence of fitness interdependences and effects of honest gossip on receiver and target; be dishonest when there is a perfect mismatch between these two factors.

While most models of reputation (e.g. most IR models) assume a so-called well-mixed population, where individuals can interact with any other member of the population, this is not a realistic assumption for human groups. While the role of *population structure* (i.e. spatially explicit models, network

topology, network dynamics) is well investigated in two-person social dilemma games (e.g. Prisoner's Dilemma, Snowdrift game, [21,109–112]; see for reviews: [113,114]), this issue has received less attention in the study of reputation systems. Takács *et al.* [115] review how the structure and dynamics of interactions help or prevent the establishment and sustainability of reputation-based cooperation. They review both theoretical and empirical work on how fixed topologies and co-evolving networks can facilitate cooperation, how reputations are formed if interactions take place in a social network, and how these mechanisms impact cooperation in small- and large-scale societies. Besides the overview on how networks act as catalysts of cooperation, they also highlight how certain network topologies can lead to the dissemination of biased beliefs and false reputations, and limit cooperation to in-group encounters but foster conflicts with out-groups.

Punishment is an important mechanism that can promote cooperation [116–120]. Punishment can be part of a reputation system, but punishment can also maintain cooperation without reputations (i.e. based on experience or direct observation). Molho & Wu [121] investigate the differences between direct punishment tactics (i.e. physical and verbal confrontation) and indirect reputation-based tactics (i.e. gossip and ostracism). They describe the costs, in terms of potential retaliation, and benefits, in terms of recalibrating others' behaviour, of direct and indirect punishment tactics. Further, they review the different functions of punishment and reputation-based tactics, as well as their antecedents. Finally, they highlight the need to better understand how distinct tactics to intervene against offences impact reputations. Podder *et al.* [24,25] investigated cooperation problems in which individuals have a choice to opt out of the interaction (the optional Public Goods game). The assignment of reputations is more complex in these interactions as there is more than a collectively good (cooperation) and bad (defection) choice. They find that the social norm that assigns a more moderate reputational penalty for opting out than for defection is able to sustain cooperation. Furthermore, they investigate the role of reputational dynamics in a setting where opting out as well as punishment can simultaneously be present and find that the two institutions work synergistically and promote large-scale cooperation under social norms that do not punish opting out choices too harshly.

For cooperation to persist, cooperators, on average, must be better-off than non-cooperators. Tsvetkova [122] investigates how public and objective reputational information affects *payoff inequality* in repeated social dilemma interactions in large groups with the use of an agent-based model with simple decision heuristics and varying partner matching rules. She demonstrates that reputational information does not necessarily increase inequality in strategically updated networks, and actually decreases it in randomly rewired networks. More importantly, reputational information almost always results in high correlation between final payoffs and cooperative behaviour. Reputational information may, therefore, both improve cooperation and reduce inequality. These model results are validated with data from cooperation experiments showing similar correlations.

Another important ingredient of cooperation is *trust*. Trust promotes cooperation, but organisms often need to know how much to trust others to cooperate. Fitness interdependence (or stake, [105–108]) can promote cooperation.

Barclay *et al.* [123] investigate whether individuals are able to find out the level of fitness interdependence from the level of helping behaviour. In their mathematical model, agents condition their help towards others based on their stake in the recipient's welfare, and recipients use other's helping decisions to assess whom to trust. They demonstrate that recipients will demand signals from others when they assign less value to the signallers and under conditions involving a higher cost of betrayed trust. Signal costs, however, are higher when signallers have a stronger incentive to defect.

4. Concluding remarks

While reputation has been claimed to be one of the key drivers of the success of human large-scale cooperation, how humans achieve this success is less well understood. Humans are uniquely able to enjoy the benefits of reputation-based cooperation due to our advanced abilities for communication, and expansive cognitive and social skills (e.g. working memory and perspective-taking). The theme issue outlines promising research directions about the evolution and conditions of reputation systems, the context and content of reputation, and the operation of reputation-based systems in social networks.

With this theme issue and the workshop that originally sparked it, we have sought to bring together the different strands of research on reputation-based cooperation being carried out across the diversity of the behavioural sciences. The authors of the pieces in this issue come for different academic disciplines so inevitably bring to this field distinct sets of questions, theories and tools of inquiry. Indeed, if one were to survey the operationalizations of core concepts such as reputation, gossip, punishment or cooperation in the papers in this theme issue, slight differences could surely be found. Despite this, we have found there to be important areas of shared understanding and interest across these traditional divides. There is a shared foundation and a shared recognition that more work is needed to advance the field: to add necessary complexity, whether to our models or our experiments, and to expand our empirical evidence, whether through investigations in new and more diverse settings, or through the use of new observational tools and data sources (for a review of open questions and novel directions for the development of a new interdisciplinary theoretical framework of reputation-based cooperation, see [124]).

The theme issue also highlights that our knowledge is still fragmented and there is no unified framework that could outline the necessary conditions for honest gossip that could create and maintain reliable reputations. While models about reputation systems are useful, they do not always reflect human social networks and reputation systems. For example, empirical investigations show that reputation is often more complex than the assignment of a binary evaluation. Another example is that communication needs to be to a large extent honest to sustain reputation-based systems in these models, yet empirical studies show that gossip need not be always honest in real-life situations. Network structure can also influence the flow of communication and thus the level of cooperation.

Further interdisciplinary work is much needed to resolve the issues identified above. The success of human cooperation is strongly supported by reputation systems, yet we are only just beginning to understand the complexity of these systems. While this theme issue offers many novel insights, it identifies just as many open issues. We have still much to learn about the language of cooperation.

References

1. Maynard SJ, Szathmáry E. 1995 *The major transitions in evolution*. Oxford, UK: Oxford University Press.

2. Noë R, Hammerstein P. 1995 Biological markets. *Trends Ecol. Evol.* **10**, 336–339. (doi:10.1016/S0169-5347(00)89123-5)

3. Granovetter M. 1978 Threshold models of collective behavior. *Am. J. Sociol.* **83**, 489–515. (doi:10.1086/226707)

4. Kollock P. 1998 Social dilemmas: the anatomy of cooperation. *Annu. Rev. Sociol.* **24**, 183–214. (doi:10.1146/annurev.soc.24.1.183)

5. Nowak MA, Sigmund K. 1998 Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573. (doi:10.1038/31225)

6. Milinski M, Semmann D, Krambeck HJ. 2002 Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424. (doi:10.1038/415424a)

7. Panchanathan K, Boyd R. 2004 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499. (doi:10.1038/nature02978)

8. Giardini F, Wittek R. 2019 Gossip, reputation, and sustainable cooperation. Sociological foundations. In *The Oxford handbook of gossip and reputation* (eds F Giardini, R Wittek), pp. 23–46. New York, NY: Oxford Handbooks.

9. Nowak MA, Sigmund K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291. (doi:10.1038/nature04131)

10. Tennie C, Frith U, Frith CD. 2010 Reputation management in the age of the world-wide web. *Trends Cogn. Sci.* **14**, 482–488. (doi:10.1016/j.tics.2010.07.003)

11. Simpson B, Willer R. 2015 Beyond altruism: sociological foundations of cooperation and prosocial behavior. *Annu. Rev. Sociol.* **41**, 43–63. (doi:10.1146/annurev-soc-073014-112242)

12. Barclay P. 2012 Harnessing the power of reputation: strengths and limits for promoting cooperative behaviors. *Evol. Psychol.* **10**, 147470491201000500. (doi:10.1177/147470491201000509)

13. Wu J, Balliet D, Van Lange PA. 2016 Gossip versus punishment: the efficiency of reputation to promote and maintain cooperation. *Sci. Rep.* **6**, 23919. (doi:10.1038/srep23919)

14. Milinski M. 2016 Reputation, a universal currency for human social interactions. *Phil. Trans. R. Soc. B* **371**, 20150100. (doi:10.1098/rstb.2015.0100)

15. Bird RB, Ready E, Power EA. 2018 The social significance of subtle signals. *Nat. Hum. Behav.* **2**, 452–457. (doi:10.1038/s41562-018-0298-3)

16. Marsh C. 2018 Indirect reciprocity and reputation management: interdisciplinary findings from evolutionary biology and economics. *Public Relat. Rev.* **44**, 463–470. (doi:10.1016/j.pubrev.2018.04.002)

17. Smith EA. 2010 Communication and collective action: language and the evolution of human cooperation. *Evol. Hum. Behav.* **31**, 231–245. (doi:10.1016/j.evolhumbehav.2010.03.001)

18. Feinberg M, Willer R, Schultz M. 2014 Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* **25**, 656–664. (doi:10.1177/0956797613510184)

19. Shank DB, Kashima Y, Peters K, Li Y, Robins G, Kirley M. 2019 Norm talk and human cooperation: can we talk ourselves into cooperation? *J. Pers. Soc. Psychol.* **117**, 99. (doi:10.1037/pspi0000163)

20. Ohtsuki H, Iwasa Y. 2004 How should we define goodness? Reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**, 107–120. (doi:10.1016/j.jtbi.2004.06.005)

21. Ohtsuki H, Hauert C, Lieberman E, Nowak MA. 2006 A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505. (doi:10.1038/nature04605)

22. Sigmund K. 2012 Moral assessment in indirect reciprocity. *J. Theor. Biol.* **299**, 25–30. (doi:10.1016/j.jtbi.2011.03.024)

23. Okada I. 2020 A review of theoretical studies on indirect reciprocity. *Games* **11**, 27. (doi:10.3390/g11030027)

24. Podder S, Righi S, Takács K. 2021 Local reputation, local selection, and the leading eight norms. *Sci. Rep.* **11**, 16560. (doi:10.1038/s41598-021-95130-3)

25. Podder S, Righi S, Pancotto F. 2021 Reputation and punishment sustain cooperation in the optional public goods game. *Phil. Trans. R. Soc. B* **376**, 20200293. (doi:10.1098/rstb.2020.0293)

26. Bshary R, Grutter AS. 2006 Image scoring and cooperation in a cleaner fish mutualism. *Nature* **441**, 975–978. (doi:10.1038/nature04755)

27. Manrique HM, Zeidler H, Roberts G, Barclay P, Walker M, Samu F, Fariña A, Bshary R, Raihani N. 2021 The psychological foundations of reputation-based cooperation. *Phil. Trans. R. Soc. B* **376**, 20200287. (doi:10.1098/rstb.2020.0287)

28. Barrett HC, Saxe RR. 2021 Are some cultures more mind-minded in their moral judgements than others? *Phil. Trans. R. Soc. B* **376**, 20200288. (doi:10.1098/rstb.2020.0288)

29. Leimar O, Hammerstein P. 2001 Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753. (doi:10.1098/rspb.2000.1573)

30. Pacheco JM, Santos FC, Chalub F. 2006 Stern-judging: a simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comput. Biol.* **2**, 1634–1638. (doi:10.1371/journal.pcbi.0020178)

31. Ohtsuki H, Iwasa Y. 2006 The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **239**, 435–444. (doi:10.1016/j.jtbi.2005.08.008)

32. Roberts G. 2008 Evolution of direct and indirect reciprocity. *Proc. R. Soc. B* **275**, 173–179. (doi:10.1098/rspb.2007.1134)

33. Santos FP, Santos FC, Pacheco JM. 2018 Social norm complexity and past reputations in the evolution of cooperation. *Nature* **555**, 242. (doi:10.1038/nature25763)

34. Santos FP, Pacheco JM, Santos FC. 2021 The complexity of human cooperation under indirect reciprocity. *Phil. Trans. R. Soc. B* **376**, 20200291. (doi:10.1098/rstb.2020.0291)

35. Roberts G. 1998 Competitive altruism: from reciprocity to the handicap principle. *Proc. R. Soc. Lond. B* **265**, 427–431. (doi:10.1098/rspb.1998.0312)

36. Barclay P. 2004 Trustworthiness and competitive altruism can also solve the 'tragedy of the commons'. *Evol. Hum. Behav.* **25**, 209–220. (doi:10.1016/j.evolhumbehav.2004.04.002)

37. Barclay P. 2010 Altruism as a courtship display: some effects of third-party generosity on audience perceptions. *Br. J. Psychol.* **101**, 123–135. (doi:10.1348/000712609X435733)

38. Barclay P. 2011 Competitive helping increases with the size of biological markets and invades defection. *J. Theor. Biol.* **281**, 47–55. (doi:10.1016/j.jtbi.2011.04.023)

39. Barclay P. 2016 Biological markets and the effects of partner choice on cooperation and friendship. *Curr. Opin. Psychol.* **7**, 33–38. (doi:10.1016/j.copsyc.2015.07.012)

40. Hardy CL, Van Vugt M. 2006 Nice guys finish first: the competitive altruism hypothesis. *Pers. Soc. Psychol. Bull.* **32**, 1402–1413. (doi:10.1177/0146167206291006)

41. Barclay P, Willer R. 2007 Partner choice creates competitive altruism in humans. *Proc. R. Soc. B* **274**, 749–753. (doi:10.1098/rspb.2006.0209)

42. Van Vugt M, Roberts G, Hardy C. 2007 Competitive altruism: development of reputation-based cooperation in groups. In *Handbook of evolutionary psychology* (eds R Dunbar, L Barrett). Oxford, UK: Oxford University Press.

43. Sylwester K, Roberts G. 2010 Cooperators benefit through reputation-based partner choice in economic games. *Biol. Lett.* **6**, 659–662. (doi:10.1098/rsbl.2010.0209)

44. Macfarlan SJ, Remiker M, Quinlan R. 2012 Competitive altruism explains labor exchange variation in a Dominican community. *Curr. Anthropol.* **53**, 118–124. (doi:10.1086/663700)

45. Böhm R, Regner T. 2013 Charitable giving among females and males: an empirical test of the competitive altruism hypothesis. *J. Bioecon.* **15**, 251–267. (doi:10.1007/s10818-013-9152-x)

46. Roberts G. 2015 Human cooperation: the race to give. *Curr. Biol.* **25**, R425–R427. (doi:10.1016/j.cub.2015.03.045)

47. Herrmann E, Engelmann JM, Tomasello M. 2019 Children engage in competitive altruism. *J. Exp. Child Psychol.* **179**, 176–189. (doi:10.1016/j.jecp.2018.11.008)

48. Barclay P. 2013 Strategies for cooperation in biological markets, especially for humans. *Evol. Human Behav.* **34**, 164–175. (doi:10.1016/j.evolhumbehav.2013.02.002)

49. 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. (doi:10.1098/rstb.2020.0290)

50. Dukerich JM, Carter SM. 2000 Distorted images and reputation repair. In: *The expressive organization: linking identity, reputation, and the corporate brand* (eds M Schultz, MJ Hatch, MH Larsen), pp. 97–112. New York, NY: Oxford University Press.

51. Heugens PP, Van Riel CB, Van Den Bosch FA. 2004 Reputation management capabilities as decision rules. *J. Manage. Stud.* **41**, 1349–1377. (doi:10.1111/j.1467-6486.2004.00478.x)

52. Fombrun CJ, Rindova V. 1996 Who's tops and who decides? The social construction of corporate reputations. New York University, Stern School of Business, Working Paper, 5–13.

53. Rhee M, Haunschild PR. 2006 The liability of good reputation: a study of product recalls in the US automobile industry. *Organ. Sci.* **17**, 101–117. (doi:10.1287/orsc.1050.0175)

54. Macfarlan SJ, Quinlan R, Remiker M. 2013 Cooperative behaviour and prosocial reputation dynamics in a Dominican village. *Proc. R. Soc. B* **280**, 20130557. (doi:10.1098/rspb.2013.0557)

55. Dafoe A, Renshon J, Huth P. 2014 Reputation and status as motives for war. *Annu. Rev. Political Sci.* **17**, 371–393. (doi:10.1146/annurev-polisci-071112-213421)

56. Bartley T, Child C. 2014 Shaming the corporation: the social production of targets and the anti-sweatshop movement. *Am. Sociol. Rev.* **79**, 653–679. (doi:10.1177/0003122414540653)

57. Tadelis S. 2016 Reputation and feedback systems in online platform markets. *Annu. Rev. Econ.* **8**, 321–340. (doi:10.1146/annurev-economics-080315-015325)

58. Rogers T, Goldstein NJ, Fox CR. 2018 Social mobilization. *Annu. Rev. Psychol.* **69**, 357–381. (doi:10.1146/annurev-psych-122414-033718)

59. Barclay P, Barker JL. 2020 Greener than thou: people who protect the environment are more cooperative, compete to be environmental, and benefit from reputation. *J. Environ. Psychol.* **72**, 101441. (doi:10.1016/j.jenvp.2020.101441)

60. 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. (doi:10.1098/rstb.2020.0295)

61. Garfield ZH, Schacht R, Post ER, Ingram D, Uehling A, Macfarlan SJ. 2021 The content and structure of reputation domains across human societies: a view from the evolutionary social sciences. *Phil. Trans. R. Soc. B* **376**, 20200296. (doi:10.1098/rstb.2020.0296)

62. Dores Cruz TD *et al.* 2021 Gossip and reputation in everyday life. *Phil. Trans. R. Soc. B* **376**, 20200301. (doi:10.1098/rstb.2020.0301)

63. Hess NH, Hagen EH. 2021 Competitive gossip: the impact of domain, resource value, resource scarcity and coalitions. *Phil. Trans. R. Soc. B* **376**, 20200305. (doi:10.1098/rstb.2020.0305)

64. Spence M. 1973 Job market signaling. *Q. J. Econ.* **87**, 355–374. (doi:10.2307/1882010)

65. Grafen A. 1990 Biological signals as handicaps. *J. Theor. Biol.* **144**, 517–546. (doi:10.1016/S0022-5193(05)80088-8)

66. Maynard SJ, Harper D. 2003 *Animal signals*. Oxford, UK: Oxford University Press.

67. Searcy WA, Nowicki S. 2005 *The evolution of animal communication. Reliability and deception in signaling systems*. Princeton, NJ: Princeton University Press.

68. Bird RB, Smith EA. 2005 Signaling theory, strategic interaction, and symbolic capital. *Curr. Anthropol.* **46**, 221–248. (doi:10.1086/427115)

69. Barker JL, Power EA, Heap S, Puurtinen M, Sosis R. 2019 Content, cost, and context: a framework for understanding human signaling systems. *Evol. Anthropol. Issues News Rev.* **28**, 86–99. (doi:10.1002/evan.21768)

70. Godfray HC. 1991 Signalling of need by offspring to their parents. *Nature* **352**, 328–330. (doi:10.1038/352328a0)

71. Maynard SJ. 1991 Honest signalling: the Philip Sidney game. *Anim. Behav.* **42**, 1034–1035. (doi:10.1016/S0003-3472(05)80161-7)

72. 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. (doi:10.1098/rstb.2020.0298)

73. Guilford T, Dawkins MS. 1995 What are conventional signals? *Anim. Behav.* **49**, 1689–1695. (doi:10.1016/0003-3472(95)90090-X)

74. Przepiorka W, Diekmann A. 2021 Parochial cooperation and the emergence of signalling norms. *Phil. Trans. R. Soc. B* **376**, 20200294. (doi:10.1098/rstb.2020.0294)

75. Fehrler S, Przepiorka W. 2013 Charitable giving as a signal of trustworthiness: disentangling the signaling benefits of altruistic acts. *Evol. Hum. Behav.* **34**, 139–145. (doi:10.1016/j.evolhumbehav.2012.11.005)

76. Posner EA. 2000 *Law and social norms*. Cambridge, MA: Harvard University Press.

77. Przepiorka W, Diekmann A. 2013 Temporal embeddedness and signals of trustworthiness: experimental tests of a game theoretic model in the United Kingdom, Russia, and Switzerland. *Eur. Sociol. Rev.* **29**, 1010–1023. (doi:10.1093/esr/jcs078)

78. Aksoy O, Gambetta D. 2016 Behind the veil: the strategic use of religious garb. *Eur. Sociol. Rev.* **6**, 792–806. (doi:10.1093/esr/jcw035)

79. Iannaccone LR. 1994 Why strict churches are strong. *Am. J. Sociol.* **99**, 1180–1211. (doi:10.1086/230409)

80. Sosis R. 2005 Does religion promote trust? The role of signaling, reputation, and punishment. *Interdiscip. J. Res. Religion* **1**, 1–30.

81. Sommerfeld RD, Krambeck HJ, Semmann D, Milinski M. 2007 Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl Acad. Sci. USA* **104**, 17 435–17 440. (doi:10.1073/pnas.0704598104)

82. Ellingsen T, Johannesson M. 2008 Anticipated verbal feedback induces altruistic behavior. *Evol. Hum. Behav.* **29**, 100–105. (doi:10.1016/j.evolhumbehav.2007.11.001)

83. Piazza J, Bering JM. 2008 Concerns about reputation via gossip promote generous allocations in an economic game. *Evol. Hum. Behav.* **2008**, 172–178. (doi:10.1016/j.evolhumbehav.2007.12.002)

84. Beersma B, Gerben AVK. 2011 How the grapevine keeps you in line: gossip increases contributions to the group. *Soc. Psychol. Pers. Sci.* **2**, 642–649. (doi:10.1177/1948550611405073)

85. Beersma B, Van Kleef GA. 2012 Why people gossip: an empirical analysis of social motives, antecedents, and consequences. *J. Appl. Soc. Psychol.* **42**, 2640–2670. (doi:10.1111/j.1559-1816.2012.00956.x)

86. Feinberg M, Willer R, Stellar J, Keltner D. 2012 The virtues of gossip: reputational information sharing as prosocial behavior. *J. Pers. Soc. Psychol.* **102**, 1015. (doi:10.1037/a0026650)

87. Wu J, Balliet D, Van Lange PA. 2015 When does gossip promote generosity? Indirect reciprocity under the shadow of the future. *Soc. Psychol. Pers. Sci.* **6**, 923–930. (doi:10.1177/1948550615595272)

88. Wu J, Balliet D, Van Lange PA. 2016 Reputation management: why and how gossip enhances generosity. *Evol. Hum. Behav.* **37**, 193–201. (doi:10.1016/j.evolhumbehav.2015.11.001)

89. Samu F, Számadó S, Takács K. 2020 Scarce and directly beneficial reputations support cooperation. *Sci. Rep.* **10**, 11486. (doi:10.1038/s41598-020-68123-x)

90. DePaulo BM, Kashy DA, Kirkendol SE, Wyer MM, Epstein JA. 1996 Lying in everyday life. *J. Pers. Soc. Psychol.* **70**, 979. (doi:10.1037/0022-3514.70.5.979)

91. Kashy DA, DePaulo BM. 1996 Who lies? *J. Pers. Soc. Psychol.* **70**, 1037. (doi:10.1037/0022-3514.70.5.1037)

92. Mann H, Garcia-Rada X, Houser D, Ariely D. 2014 Everybody else is doing it: exploring social transmission of lying behavior. *PLoS ONE* **9**, e109591. (doi:10.1371/journal.pone.0109591)

93. Jacobsen C, Fosgaard TR, Pascual-Ezama D. 2018 Why do we lie? A practical guide to the dishonesty literature. *J. Econ. Surveys* **32**, 357–387. (doi:10.1111/joes.12204)

94. Buss DM, Dedden LA. 1990 Derogation of competitors. *J. Soc. Pers. Relat.* **7**, 395–422. (doi:10.1177/0265407590073006)

95. Diekmann A, Jann B, Przepiorka W, Wehrli S. 2014 Reputation formation and the evolution of cooperation in anonymous online markets. *Am. Sociol. Rev.* **79**, 65–85. (doi:10.1177/0003122413512316)

96. Hess NH, Hagen EH. 2006 Sex differences in indirect aggression: psychological evidence from young adults. *Evol. Hum. Behav.* **27**, 231–245. (doi:10.1016/j.evolhumbehav.2005.11.001)

97. McAndrew FT. 2014 The 'sword of a woman': gossip and female aggression. *Aggress. Violent Behav.* **19**, 196–199. (doi:10.1016/j.avb.2014.04.006)

98. Fonseca MA, Peters K. 2018 Will any gossip do? Gossip does not need to be perfectly accurate to promote trust. *Games Econ. Behav.* **107**, 253–281. (doi:10.1016/j.geb.2017.09.015)

99. Peters K, Fonseca MA. 2020 Truth, lies, and gossip. *Psychol. Sci.* **31**, 702–714. (doi:10.1177/0956797620916708)

100. Számadó S, Szalai F, Scheuring I. 2016 Deception undermines the stability of cooperation in games of indirect reciprocity. *PLoS ONE* **11**, e0147623. (doi:10.1371/journal.pone.0147623)

101. Fonseca MA, Peters K. 2021 Is it costly to deceive? People are adept at detecting gossipers' lies but may not reward honesty. *Phil. Trans. R. Soc. B* **376**, 20200304. (doi:10.1098/rstb.2020.0304)

102. Giardini F, Vilone D, Sánchez A, Antonioni A. 2021 Gossip and competitive altruism support cooperation in a Public Good game. *Phil. Trans. R. Soc. B* **376**, 20200303. (doi:10.1098/rstb.2020.0303)

103. Samu F, Takács K. 2021 Evaluating mechanisms that could support credible reputations and cooperation: cross-checking and social bonding. *Phil. Trans. R. Soc. B* **376**, 20200302. (doi:10.1098/rstb.2020.0302)

104. Wu J *et al.* 2021 Honesty and dishonesty in gossip strategies: a fitness interdependence analysis. *Phil. Trans. R. Soc. B* **376**, 20200300. (doi:10.1098/rstb.2020.0300)

105. Roberts G. 2005 Cooperation through interdependence. *Anim. Behav.* **70**, 901–908. (doi:10.1016/j.anbehav.2005.02.006)

106. Brown SL, Brown RM. 2006 Selective investment theory: recasting the functional significance of close relationships. *Psychol. Inq.* **17**, 1–29. (doi:10.1207/s15327965pli1701_01)

107. Aktipis A *et al*. 2018 Understanding cooperation through fitness interdependence. *Nat. Hum. Behav.* **2**, 429–431. (doi:10.1038/s41562-018-0378-4)

108. Cronk L, Steklis D, Steklis N, van den Akker OR, Aktipis A. 2019 Kin terms and fitness interdependence. *Evol. Hum. Behav.* **40**, 281–291. (doi:10.1016/j.evolhumbehav.2018.12.004)

109. Nowak MA, May RM. 1992 Evolutionary games and spatial chaos. *Nature* **359**, 826–829. (doi:10.1038/359826a0)

110. Nowak MA, May RM. 1993 The spatial dilemmas of evolution. *Int. J. Bifurcation Chaos* **3**, 35–78. (doi:10.1142/S0218127493000040)

111. Hauert C. 2002 Effects of space in 2×2 games. *Int. J. Bifurcation Chaos* **12**, 1531–1548. (doi:10.1142/S0218127402005273)

112. Számadó S, Szalai F, Scheuring I. 2008 The effect of dispersal and neighbourhood in games of cooperation. *J. Theor. Biol.* **253**, 221–227. (doi:10.1016/j.jtbi.2008.02.037)

113. Szabó G, Fath G. 2007 Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216. (doi:10.1016/j.physrep.2007.04.004)

114. Perc M, Szolnoki A. 2010 Coevolutionary games—a mini review. *Biosystems* **99**, 109–125. (doi:10.1016/j.biosystems.2009.10.003)

115. 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. (doi:10.1098/rstb.2020.0297)

116. Yamagishi T. 1986 The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110. (doi:10.1037/0022-3514.51.1.110)

117. Fehr E, Gächter S. 2000 Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994. (doi:10.1257/aer.90.4.980)

118. Fehr E, Gächter S. 2002 Altruistic punishment in humans. *Nature* **415**, 137–140. (doi:10.1038/415137a)

119. Balliet D, Mulder LB, Van Lange PA. 2011 Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* **137**, 594. (doi:10.1037/a0023489)

120. Ostrom E. 2012 Experiments combining communication with punishment options demonstrate how individuals can overcome social dilemmas. *Behav. Brain Sci.* **35**, 33. (doi:10.1017/S0140525X11001282)

121. Molho C, Wu J. 2021 Direct punishment and indirect reputation-based tactics to intervene against offenses. *Phil. Trans. R. Soc. B* **376**, 20200289. (doi:10.1098/rstb.2020.0289)

122. Tsvetkova M. 2021 The effects of reputation on inequality in network cooperation games. *Phil. Trans. R. Soc. B* **376**, 20200299. (doi:10.1098/rstb.2020.0299)

123. Barclay P, Bliege Bird R, Roberts G, Számadó S. 2021 Cooperating to show that you care: costly helping as an honest signal of fitness interdependence. *Phil. Trans. R. Soc. B* **376**, 20200292. (doi:10.1098/rstb.2020.0292)

124. Giardini F, Balliet D, Power AE, Számadó S, Takács K. In press. Four puzzles of reputation-based cooperation: content, process, honesty, and structure. *Human Nat.*