Cooperation and Reputation in Primitive Societies

Fernando P. Santos[1,2], Francisco C. Santos[1,2] and Jorge M. Pacheco[1,3,4]

[1]INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal
[2]ATP-Group, Lisboa, Portugal
[3]Centro de Biologia Molecular e Ambiental da Universidade do Minho, 4710-057 Braga, Portugal
[4]Departamento de Matemática e Aplicações da Universidade do Minho, 4710-057 Braga, Portugal
fernando.pedro@tecnico.ulisboa.pt

Indirect Reciprocity (IR) is possibly the most elaborated and cognitively demanding mechanism of cooperation discovered so far. It involves status and reputations and has been heralded as providing the biological basis of our morality (Nowak and Sigmund, 2005). Whereas under direct reciprocity one expects to receive help from someone we have helped before, under IR one expects a return, not from someone we helped, but from someone else; in this sense, helping the "right" individuals may contribute to a reputation uplift that increases the chance of being helped, by someone else, at a later stage. This reputation shift depends on the socially adopted norm that defines what actions (and under which contexts) are reckoned as good or bad. Most theoretical models employed to date have studied how IR can lead to the emergence and sustainability of cooperation in infinite populations (Ohtsuki and Iwasa, 2006; Nowak and Sigmund, 2005). However, it is known that cooperation, norms, reciprocity and the art of managing reputations, are features that date back to primitive, small-scale societies While different features characterize a primitive society, here we take into consideration three of the most important: The evidence that interactions occur within small tribes, the central role played by reputations and the ease with which reputation information spreads within the tribe. In small populations, stochastic finite size effects are not only important, but may even render infinite populations analyses misleading (Imhof et al., 2005). Thus, it remains an open question which norms prevail in small-scale societies and their influence in the evolutionary dynamics of IR.

With the current extended abstract, we would like to summarize a new analysis of this problem. In Santos et al. (2016) we show that population size strongly influences the merits of each social norm, while proposing a new formal tool to assess the evolutionary dynamics of reputation-based systems in finite populations. We investigate to which extent norms found to promote cooperation in large populations will remain effective in small societies, and also to which extent the capacity of a social norm to foster cooperation depends on the community size. We consider a population of individuals who randomly interact in pairs through a donation game, where one player is a potential provider of help (donor) to the other (recipient). The donor may cooperate and help the recipient at a cost c to herself/himself, conferring a benefit b to the recipient (with $b > c$); otherwise no one pays any costs nor distributes any benefits. Reputations are public and attributed by a bystander who witnesses a pairwise interaction. We adopt a world of binary reputations, Good (G) and Bad (B), which in our case are mere labels with no a-priori meaning. Their significance emerges in association with individual behavior in connection with the donation game. This binary reputation scheme, despite its formal simplicity, allows to consider a plethora of moral rules with variable complexity and it is specially amenable to a systematic mathematical treatment, in the framework of population dynamics. To perform an evaluation, the bystander uses a social norm, that is, a rule that converts the combined information stemming from the action of the donor and the reputation of the recipient into a new reputation for the donor. Social norms encoding this type of information are classified as 2nd-order norms (Ohtsuki and Iwasa, 2006). Four of these social norms have been given special attention (see matrices on Fig. 1): Stern-judging (SJ, also known as Kandori), which assigns a good reputation to a donor that helps a good recipient or refuses help to a bad one, assigning a bad reputation in the other cases (Pacheco et al., 2006); Simple-Standing (SS), similar to SJ, but more "benevolent" by assigning a good reputation to any donor that cooperates; Shunning (SH), similar to SJ but less "benevolent", by assigning a bad reputation to any donor that defects; and Image Score (IS, actually a first order norm) where all that matters is the action of the donor, who acquires a good reputation if playing C and a bad reputation if playing D (Nowak and Sigmund, 2005). In the space of 2nd-order norms that we consider, a duple p suffices to unambiguously define a strategy, by specifying the action directed at a G or B recipient. This leads to the following 4 possible strategies: unconditional Defection (AllD, $p = (D, D)$), unconditional Cooperation (AllC, $p = (C, C)$), Discriminator strategy (Disc, $p = (C, D)$), that is, cooperate with those in good reputation, and defect otherwise), and paradoxical Discriminator strategy (pDisc,

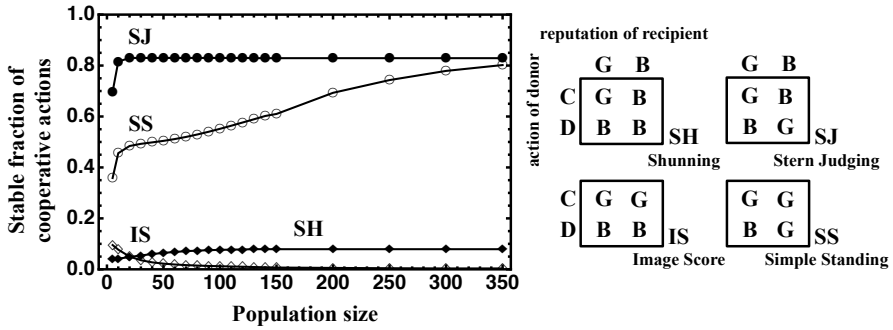

Figure 1: Stern-judging (SJ) is able to foster the highest rates of cooperation, independently of the (finite and small) population size; Differently, the efficiency of SS in fostering cooperation strongly depends on the population size and on the error rate committed by individuals (Santos et al., 2016). SH harms cooperation by being too strict compared to SJ due to the abusive widespread assignment of bad labels. See main text and Santos et al. (2016) for details on the strategic dynamics induced by each social norm. The matrices illustrate the 4 dominant social norms in terms of the new reputation (B/G, inside each square) attributed to a donor given its action (C/D, rows) and the reputation (B/G, columns) of the recipient.

$p = (D, C)$, the opposite of Disc). Unlike previous studies, in Santos et al. (2016) we investigate the evolutionary dynamics of these 4 strategies within finite populations by means of a stochastic birth-death process, both analytically and through large-scale computer simulations (for details, please see Santos et al. (2016)). As detailed in Fig. 1, we show that SJ clearly stands out for small population sizes, dominating with SS for large population sizes. Indeed, it can be shown that only SJ and SS are able to combine a high prevalence of an ALL-Disc configuration with the incidence of Good reputations in this configuration, efficiently fostering high levels of cooperation (Fig. 1). Yet, in small-scale societies, SJ significantly promotes more cooperation than SS, as the latter fails to prevent the invasion of unconditional defectors (AllD) in small populations. On the other hand, SJ fosters an ideal coordination between strategy and prevailing reputations, assuring the stability of configurations where individuals cooperate in the donation game. Indeed, the high degree of symmetry of SJ allows the promotion of cooperation, irrespectively of the emerging meaning of G and B labels, also allowing "paradoxical" strategies to prevail and promote cooperation. These results remain valid for a wide interval of reputation assignment time-scales, errors of execution and reputation-assignment inaccuracies.

To conclude, a single social norm (SJ) emerges as the leading norm in small-scale societies. That simple norm dictates that only whoever cooperates with good individuals, and defects against bad ones, deserves a good reputation. Remarkably this pattern is consistent with recent empirical results (Hamlin et al., 2011) showing that toddlers positively evaluate i) those who treat others prosocially, ii) those who behave negatively towards those who have acted antisocially, and iii) puppets that harm antisocial puppets. This said, behavioral experiments in this context remain a vast open territory and very active area of research. Finally, our modeling framework has the advantage of being naturally extendable to social norms of higher order, enlarging the complexity of the norms studied to date. Work along these lines is in progress, with promising preliminary results of interest within the area of evolution of biological complexity, and the ALife community in general.

Acknowledgements

This research was supported by Fundação para a Ciência e Tecnologia (FCT) through grants SFRH/BD/94736/2013, PTDC/EEI-SII/5081/2014, PTDC/MAT/STA/3358/2014 and by multi-annual funding of CBMA and INESC-ID (under the projects UID/BIA/04050/2013 and UID/CEC/50021/2013 provided by FCT).

References

Hamlin, J. K., Wynn, K., Bloom, P., and Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proc Natl Acad Sci USA*, 108(50):19931–19936.

Imhof, L. A., Fudenberg, D., and Nowak, M. A. (2005). Evolutionary cycles of cooperation and defection. *Proc Natl Acad Sci USA*, 102(31):10797–10800.

Nowak, M. A. and Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437(7063):1291–1298.

Ohtsuki, H. and Iwasa, Y. (2006). The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J Theor Biol*, 239(4):435–444.

Pacheco, J. M., Santos, F. C., and Chalub, F. (2006). Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Comput. Biol*, 2(12):e178.

Santos, F. P., Santos, F. C., and Pacheco, J. M. (2016). Social norms of cooperation in small-scale societies. *PLoS Comput. Biol*, 12(1):e1004709.