The effects of imperfect information on Stern-judging in networks

Group 13
João Carranca - 102600 || Javier Santa Maria - 99240
Masters in Computer Science and Engineering and Masters in Engineering and Data Science
Network Science
Instituto Superior Técnico

Abstract

Indirect reciprocity is one of the most fundamental concepts in the study of population dynamics and cooperation dynamics. We can find the effects and marks of this mechanism all over the planet in a variety of species, including humans. Stern-judging has proven to be one of the most robust assessment rules in indirect reciprocity, generally outperforming others and providing stable rates of cooperation in populations of various sizes when all individuals have access to the reputations of all other members of the population and no information errors occur. In this work, we analyze the effect of assessment error and imperfect information on the performance of stern-judging on its own and relative to full cooperation and full defection represented by agents that fully cooperator or fully defect every time, first in a well-mixed population and then in the context of a static random network (maybe scale free? might be too resource consuming). We show that in both of these population scenarios, the norm´s effectiveness collapses.

I. Introduction

Reciprocity can be used to describe several behavioural patterns found in nature. The idea of "You scratch my back, I scratch yours" has been present in several species for hundreds of thousands of years. Standard reciprocity as a societal model faces a logical problem [1]. A society as a whole does benefit more from being made up of cooperators than defectors, however, on an individual level, each defector can extract more benefits than each cooperator since helping someone incurs a cost. This thought process would lead to societies full of defectors and no cooperators. This doesn´t, however, happen in real life. A variety of biological systems actively use cooperation in their organization [3].

Social norms allow us to craft more accurate models of interactions in populations, especially human ones, that address this collapse in cooperation [2].These social norms are derived from the idea of indirect reciprocity, which has proven to maintain mutual cooperation in populations [4] [5]. *Indirect reciprocity* makes use of the concept of reputations in individuals which adds a higher level of cognitive capacity to interactions [6]. An individual can, for example, choose to cooperate or not cooperate, based on the reputation of the individual he is interacting with. The way he might use this information on reputation will constitute a specific social norm. The degree of a norm depends on the amount of information used to decide on whether to cooperate or not cooperate, with each level of information constituting a "layer" [7].

Stern-Judging in particular is a second degree social norm that, given an interaction between two individuals, one donor and one recipient, will take into account the reputations of both the donor and recipient as well as the donor´s action to calculate the donor´s new reputation [7]. Stern-Judging has proven to be the most successful social norm when it comes to maintaining high levels of cooperation across many population sizes [8], but how does it behave when faced with imperfect information in a way similar to what we would expect to see in real life interactions amongst individuals in a population? If we create a population mix with different types of agents like defectors cooperators and discriminators, what will be the distribution of these agents at the sate of equilibrium? Here we extend on previous work related to imperfect information in social norm models, in particular [9], first in a Well-Mixed population model and then in a random network.

II. Methods

I Evolution Model

Our evolution model, which will be used to evaluate which agents dominate given certain simulation parameters, is grounded in *evolutionary game theory*. This evolution dynamic will be applied to the three types of agents we will be mixing in our populations: cooperators, defectors and discriminators, where discriminators will follow the Stern-Judging social norm. A cooperator can, for example, turn into a defector and vice-versa.

In evolutionary game theory, the success of a strategy is tied to its *fitness*—a measure of how well an individual with that strategy performs relative to others in the population. The higher the fitness, the more likely a strategy is to proliferate. The evolution of strategies in a population follows the *replicator equation*, which describes how the frequency of a given strategy changes over time based on its fitness.

The general form of the *replicator equation* is given by:

$$\dot{x}_i = x_i \left(f_i(\mathbf{x}) - \phi(\mathbf{x}) \right)$$

Where $\dot{x}_i$ is the rate of change in the frequency of strategy i, x_i is the current frequency of strategy i, $f_i(\mathbf{x})$ is the fitness of individuals using strategy i, and $\phi(\mathbf{x})$ is the average fitness of the population. This equation implies that strategies whose fitness exceeds the population's average fitness will increase in frequency, while strategies with below-average fitness will decrease.

In addition to the natural selection process described by the replicator equation, our model also includes *mutation* based on fitness comparisons where each individual A has a certain probability of imitating individual B based on the difference in their fitness. That probability is given by the following equation:

$$p = \left[1 + e^{-\beta(f_B - f_A)} \right]^{-1}$$

Where p is the probability, f_B is the fitness of individual B and f_A is the fitness of individual A and β will have a value of 0.05.

II Well-Mixed population model, Stern-Judging and imperfect information

A large, well-mixed population of players is considered. From time to time, two players are chosen at random from the population and they engage in a one-shot prisoner´s dilemma with the following payoff matrix:

	C	D
C	R, R	S, T
D	T, S	P, P

	C	D
C	$0.4, 0.4$	$-1, 0.6$
D	$0.6, -1$	$-0.1, -0.1$

Each individual has an assessment rule by which to judge the action of the donor. We assume a binary judgment: either the label 'good' or 'bad', represented by 1 and 0, is assigned to the donor. The social norm called stern-judging is considered for discriminator agents while cooperators and defectors always either cooperate or defect.

Stern-judging views those as good who, in their previous game,gave help to a good recipient or refused help to a bad recipient [7]. Discriminators using Stern-Judging will make use a reputation matrix to keep track of which individuals they consider 'good´ or 'bad´ and update said Matrix based on observations of actions. We assume that an assessment error occurs with a small probability.

When it comes to implementing the idea of imperfect information in this model, we have followed studies on the *Three Degrees of Influence* [10], where information has a certain probability of reaching individuals depending on their distance to the action. We have used the value of 61%, used in [10], as the experimental value of "influence" for first-degree neighbours. Only this value is needed since in a well mixed population there are only first-degree neighbours. In the case of networks, the percentages for second and third-degree neighbours will be obtained from the same source. It should be noted that only discriminators make use of reputation to guide their decisions.

III Reputation matrix for Well-Mixed Population

Let us the notation r_{ij} to represent the reputation of player j in the eyes of player i. The values $r_{ij} = 1$ and $r_{ij} = 0$ correspond to the situations where i thinks that j is good and bad, respectively. The matrix (r_{ij}) is called the *reputation matrix* .

The update rule is described as follows: let u, v, and W denote two random individuals, and the set of observers, which will be, on average 61% of the whole population. The number of observers is qN, where N is the total number of individuals in the population. For the analytical calculation, the population is assumed to be infinite.

After the game, the updated reputation matrix for discriminators is given by:

$$b'_{ij} = \begin{cases} b_{ij} & \text{if } i \notin W \text{ or } j \neq u \\ f(i)(a_u, b_{iv}) & \text{if } i \in W \text{ and } j = u \end{cases}$$

where a_u is the action of u toward v in the game.

Using the abbreviations C for "help" and D for "refuse", the function $f(i)(a, b) \in \{0, 1\}$ with $a \in \{C, D\}$ and $b \in \{0, 1\}$ is the assessment, from the viewpoint of i, of the action of random node, a towards the recipient. For *stern-judging*, the function $f(i)$ is defined as:

$$f(i)(C, 1) = f(i)(D, 0) = 1$$

$$f(i)(C, 0) = f(i)(D, 1) = 0$$

and

$$f(i)(C, 1) = f(i)(D, 0) = 0$$

$$f(i)(C, 0) = f(i)(D, 1) = 1$$

with probability λ (i.e., i made a mistake).

For any given updating step, only the u-column of the image matrix changes its values. Since the action a_u is probabilistically determined by b_{uv}, the new assessment b'_{wu} in the eyes of observer $w \in W$ probabilistically depends on the assessments of recipient v both in the eyes of u (b_{uv}) and w (b_{wv}) before the game.

IV Network Model

V Evaluation Criteria

For the purposes of evaluating our simulation results, we shall be using four criteria: Average fitness of the population ate the end of a simulation, the percentage of cooperative interactions per time-step in the simulation, the overall population mix in comparison the starting mix, overall imperfection of information at the end of the simulation, taking a sample of 5% of the population for that calculation.

III. Results and Discussion

IV. Conclusion

V. Future work

VI. Acknowledgements

We would like to express our gratitude to Professor Inês Pereira, Professor Alexandre Francisco and Professor Francisco Santos for their invaluable help and availability.

References

[1] Dawes, R. M. (1980). Social Dilemmas. Annual Review of Psychology, 31(1), 169–193

[2] Fehr, E., and Fischbacher, U. 2004. Social norms and human cooperation. Trends Cogn Sci 8(4):185–190.

[3] Smith JM, Szathmáry E (1995) The major transitions in evolution. Oxford: Freeman

[4] Trivers R. The evolution of reciprocal altruism. Quart Rev Biol 1971;46:35–57.

[5] Rand DG, Nowak MA. Human cooperation trends in cognitive sciences. Trends Cogn Sci 2013;17:413–25.

[6] Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437: 1291–1298.

[7] Pacheco, J. M., Santos, F. C., Chalub, F. A. C. C. (2006). Stern-Judging: A Simple, Successful Norm Which Promotes Cooperation under Indirect Reciprocity. PLoS Computational Biology, 2(12), e178.

[8] Santos, F., Pacheco, J. Santos, F. Evolution of cooperation under indirect reciprocity and arbitrary exploration rates. Sci Rep 6, 37517 (2016).

[9] Uchida, S., Sasaki, T. (2013). Effect of assessment error and private information on stern-judging in indirect reciprocity. Chaos, Solitons Fractals, 56, 175–180.

[10] Christakis, N. A., Fowler, J. H. (2008). The Collective Dynamics of Smoking in a Large Social Network. New England Journal of Medicine, 358(21), 2249–2258.

A Appendix