Institutional Incentives for promoting Evolution of Commitment Compliance and Cooperation in Spatial Public Goods Games

Lucas Flores[1] and The Anh Han[2]
[1] Instituto de Física, Universidade Federal do Rio Grande do Sul, CP 15051, CEP 91501-970 Porto Alegre - RS, Brazil
[2] School of Computing, Engineering and Digital Technologies, United Kingdom
E-Mail: lucassflores42@gmail.com , T.Han@tees.ac.uk

Introduction

Prior to embarking on a collective project, individuals involved may solicit commitment from group members and estimate how interested they are in contributing to the group's efforts. This assessment helps them determine whether it is worthwhile to initiate the endeavour and/or if it would be beneficial to join. Commitment mechanisms for enhancing cooperation are widespread in nature, which exist in various forms and contexts, including legal contracts and pledges (Nesse, 2001), marriage (Swensen and Trahaug, 1985), deposit-refund schemes (Cherry and McEvoy, 2013), emotion-based (Frank, 2001) or reputation-based commitment (Frank, 1988; Krellner and Han, 2023). Both empirical and theoretical studies demonstrated that high levels of cooperation can be achieved through reliable commitments (Ostrom, 2005; Dannenberg, 2016; Han et al., 2022; Chen and Komorita, 1994; Ogbo and Han, 2024). They enable individuals to reach mutual cooperation even when there is little knowledge about others' past behaviours (Han, 2022; Sasaki et al., 2015), as it requires them to reveal their preferences or intentions (Han et al., 2015; Tomasello et al., 2005).

However, prior models of commitment have mainly focused on well-mixed population settings (Han, 2022; Sasaki et al., 2015), potentially overlooking the significant influence of commitment dynamics within a population's actual network structure. This structure dictates who may form commitments with whom (Szabó and Fáth, 2007; Barabasi, 2014; Sayama, 2015), ultimately determining the worthiness of arranging conditional commitments. Network reciprocity plays a crucial role in shaping human social interactions, fostering the creation of cooperative clusters and thereby influencing cooperation dynamics (Szabó and Fáth, 2007; Perc et al., 2013; Szolnoki et al., 2014; Cimpeanu et al., 2023).

Herein, we highlight recent findings from an agent-based simulation analysis of the impact of commitment formation and institutional incentives for promoting the evolution of cooperation in spatial populations (Flores and Han, 2024). The study is structured by studying commitment strategies in the absence of incentives, and later introducing punishment and reward in the framework.

Model

Flores and Han (2024) study the dynamics of commitment formation and cooperation in the Public Goods Game (PGG) in a square lattice. Neighboring players compose a group of size N and can choose to contribute or not to a common pool, where the total contributions are summed and multiplied by a factor r ($1 < r < N$). In the end, the pool is divided equally among all players.

For a player to decide to contribute to the common pool, a commitment threshold τ was introduced that allows them to conditionally consider their choice based on the level of commitment in the group. Each player must commit or not to cooperate prior to the actual interaction. Now, players can decide whether to cooperate if the commitment threshold is reached by the group members, as well as whether to cooperate otherwise. The notation ijk is used to denote each strategy, where i is the decision to accept or not to commit ($i = A$ or N); j the decision to cooperate or not if the commitment is formed ($j = C$ or D); and k the decision to cooperate or not if the commitment is not formed ($k = C$ or D). For one group X, a focal individual with strategy ijk has the following payoff

$$\Pi_{ijk} = \frac{r}{N} \sum_{x \in X} c_x - c_{ijk} + incentive, \qquad (1)$$

where c_x is the contribution from group member x (including the focal player ijk) and c_{ijk} is the contribution from the focal player. The $incentive$ is used for rewarding commitment-compliant players (ACC and ACD) or punish commitment non-compliant players (ADC and ADD).

A player (ijk) and a random neighbor ($i'j'k'$) are selected randomly, and the former adopts the later strategy according to a probability given by the Fermi update rule,

$$W_{ijk \to i'j'k'} = \frac{1}{1 + e^{-(\Pi_{i'j'k'} - \Pi_{ijk})/K}} \ , \qquad (2)$$

This evolutionary step is repeated Z times where Z is the population size, characterising one Monte Carlo step (MCS). The total simulation time is set to $t = 10^5$ MCS, $K = 0.1$, and $Z = 100^2$.

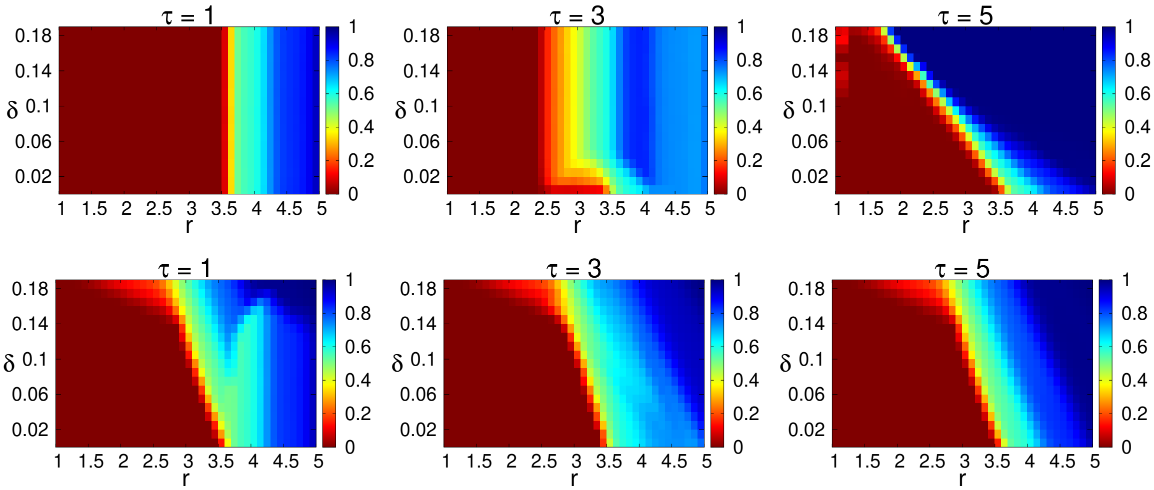

Figure 1: Phase diagram $r \times \delta$ for the density of cooperative behaviour for different commitment thresholds, $\tau = 1, 3$ and 5, in the case of **institutional punishment (top row)** and **institutional reward (bottom row)**. Here r is the multiplicative factor and τ is the commitment threshold. The strength of punishment and reward are given respectively by $\frac{N}{n_{non-com}}\delta$ and $\frac{N}{n_{com}}\delta$ where $n_{non-com}$ is the number of commitment non-compliant players in the group and n_{com} commitment-compliant ones
.

Results and Discussion

An important remark is that spatial reciprocity plays a crucial role in games in structured populations (Santos et al., 2008; Perc et al., 2017; Nowak, 2006). In such games, cooperators can form clusters to survive, by avoiding defecting neighbours. Noticeably, only some commitment-based strategies (namely, NDC, NCC, ACC, and ACD) have spatial reciprocity, meaning that when they cluster together they cooperate with each other. Flores and Han (2024) found that, low thresholds select for the non-committing strategies ($i = N$), while high thresholds for the committing ones ($i = A$). Thus, strategies will always cooperate or defect unconditionally, recovering the classical PGG.

Interestingly, for $\tau = 3$, it leads to a different scenario than the classical PGG, even without punishment or reward. It was shown that non-committing strategies (NDD and NDC) coexist with committing ones (ADD and ACD).

This creates a scenario where strategies act conditionally. Therefore cooperative strategies can avoid exploitation sometimes by defecting giving an advantage over unconditional cooperators. Nevertheless, for high r values, the benefit from cooperating increases benefiting unconditional cooperators. In this case, clustering does not happen since strategies invade one another..

Thus, in the presence of commitment and absence of an incentive mechanism, conditional behaviour based on commitment alone can enhance cooperation for low r values, especially when considering a specific commitment threshold value (Flores and Han, 2024).

In the presence of punishment, the work shows that the survival of cooperation is most likely at intermediate commitment thresholds, see Figure 1 (top row). Notably, cooperation is maximised at high commitment thresholds, when punishment occurs more frequently. Moreover, even when cooperation rarely survives, a cyclic behaviour emerges, facilitating the persistence of cooperation. In the context of the institutional reward, see Figure 1 (bottom row), it is found that cooperation is highly frequent regardless of the commitment threshold adopted.

These findings have useful implications for institutional mechanisms' design for promoting pro-social behaviour in complex networks, especially when communication is allowed to establish commitments (Nesse, 2001; Han, 2016).

In short, we have summarised here an analysis of different forms of institutional incentive for promoting cooperation and commitment compliance in interactions with a prior commitment formation in a spatial population (Flores and Han, 2024). The results highlight the comparative efficiency of the reward and punishment mechanisms in relation to the commitment threshold required for commitment to be formed and the multiplication factor (r) of the PGG. That said, in order to use institutional incentives efficiently for ensuring commitment compliance on networks, it is crucial to consider the contextual information of the interactions.

It is noteworthy that institutional approaches have been widely adopted to study biological and artificial life systems (Andras et al., 2018; Powers et al., 2018; Duong and Han, 2021; Ostrom, 1990; Han et al., 2022; Perc et al., 2017; Cimpeanu et al., 2019; Powers et al., 2023; Han, 2013). The summarised analysis provides new, fundamental insights into a cost-efficient design of commitment- and institution-based solutions for promoting pro-social behaviours and behavioural compliance in biological, social as well as artificial systems (Singh, 2013; Nesse, 2001; Ostrom, 1990).

References

Andras, P., Esterle, L., Guckert, M., Han, T. A., Lewis, P. R., Milanovic, K., Payne, T., Perret, C., Pitt, J., Powers, S. T., et al. (2018). Trusting intelligent machines: Deepening trust within socio-technical systems. *IEEE Technology and Society Magazine*, 37(4):76–83.

Barabasi, A.-L. (2014). *Linked-how Everything is Connected to Everything Else and what it Means F*. Perseus Books Group.

Chen, X.-P. and Komorita, S. S. (1994). The effects of communication and commitment in a public goods social dilemma. *Organizational Behavior and Human Decision Processes*, 60(3):367–386.

Cherry, T. L. and McEvoy, D. M. (2013). Enforcing compliance with environmental agreements in the absence of strong institutions: An experimental analysis. *Environmental and Resource Economics*, 54:63–77.

Cimpeanu, T., Han, T. A., and Santos, F. C. (2019). Exogenous rewards for promoting cooperation in scale-free networks. In *Artificial Life Conference Proceedings*, pages 316–323. MIT Press.

Cimpeanu, T., Santos, F. C., and Han, T. A. (2023). Does spending more always ensure higher cooperation? an analysis of institutional incentives on heterogeneous networks. *Dynamic Games and Applications*, 13(4):1236–1255.

Dannenberg, A. (2016). Non-binding agreements in public goods experiments. *Oxford Economic Papers*, 68(1):279–300.

Duong, M. H. and Han, T. A. (2021). Cost efficiency of institutional incentives for promoting cooperation in finite populations. *Proceedings of the Royal Society A*, 477(2254):20210568.

Flores, L. S. and Han, T. A. (2024). Evolution of commitment in the spatial public goods game through institutional incentives. *Applied Mathematics and Computation*, 473:128646.

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. Norton and Company.

Frank, R. H. (2001). Cooperation through Emotional Commitment. In Nesse, R. M., editor, *Evolution and the capacity for commitment*, pages 55–76. New York: Russell Sage.

Han, T. A. (2013). *Intention Recognition, Commitments and Their Roles in the Evolution of Cooperation: From Artificial Intelligence Techniques to Evolutionary Game Theory Models*, volume 9. Springer SAPERE series.

Han, T. A. (2016). Emergence of social punishment and cooperation through prior commitments. In *Proceedings of the thirtieth AAAI conference on artificial intelligence*, pages 2494–2500.

Han, T. A. (2022). Institutional incentives for the evolution of committed cooperation: Ensuring participation is as important as enhancing compliance. *Journal of the Royal Society Interface*, 19(188):20220036.

Han, T. A., Lenaerts, T., Santos, F. C., and Pereira, L. M. (2022). Voluntary safety commitments provide an escape from over-regulation in ai development. *Technology in Society*, 68:101843.

Han, T. A., Santos, F. C., Lenaerts, T., and Pereira, L. M. (2015). Synergy between intention recognition and commitments in cooperation dilemmas. *Scientific reports*, 5(9312).

Krellner, M. and Han, T. A. (2023). The importance of commitment for stable cooperation. *Physics of Life Reviews*, 46:255–257.

Nesse, R. (2001). *Evolution and the capacity for commitment*. Russell Sage Foundation.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *science*, 314(5805):1560–1563.

Ogbo, N. B. and Han, T. A. (2024). Coordination dynamics in technology adoption: Lessons from an evolutionary game theoretical analysis. In *Multisector Insights in Healthcare, Social Sciences, Society, and Technology*, pages 295–326. IGI Global.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Ostrom, E. (2005). *Understanding institutional diversity*. Princeton university press.

Perc, M., Gómez-Gardenes, J., Szolnoki, A., Floría, L. M., and Moreno, Y. (2013). Evolutionary dynamics of group interactions on structured populations: a review. *Journal of the royal society interface*, 10(80):20120997.

Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., and Szolnoki, A. (2017). Statistical physics of human cooperation. *Physics Reports*, 687:1–51.

Powers, S. T., Ekárt, A., and Lewis, P. R. (2018). Modelling enduring institutions: The complementarity of evolutionary and agent-based approaches. *Cognitive Systems Research*, 52:67–81.

Powers, S. T., Linnyk, O., Guckert, M., Hannig, J., Pitt, J., Urquhart, N., Ekárt, A., Gumpfer, N., Han, T. A., Lewis, P. R., et al. (2023). The stuff we swim in: Regulation alone will not lead to justifiable trust in ai. *IEEE Technology and Society Magazine*, 42(4):95–106.

Santos, F. C., Santos, M. D., and Pacheco, J. M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216.

Sasaki, T., Okada, I., Uchida, S., and Chen, X. (2015). Commitment to cooperation and peer punishment: Its evolution. *Games*, 6(4):574–587.

Sayama, H. (2015). *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks [Imprint].

Singh, M. P. (2013). Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):21.

Swensen, C. H. and Trahaug, G. (1985). Commitment and the long-term marriage relationship. *Journal of Marriage and the Family*, pages 939–945.

Szabó, G. and Fáth, G. (2007). Evolutionary games on graphs. *Phys Rep*, 97-216(4-6).

Szolnoki, A., Mobilia, M., Jiang, L.-L., Szczesny, B., Rucklidge, A. M., and Perc, M. (2014). Cyclic dominance in evolutionary games: a review. *Journal of the Royal Society Interface*, 11(100):20140735.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and brain sciences*, 28(05):675–691.