Applied Mathematics and Computation 473 (2024) 128646

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Full Length Article

Evolution of commitment in the spatial public goods game through institutional incentives

Lucas S. Flores [a], The Anh Han [b,*]

[a] Instituto de Física, Universidade Federal do Rio Grande do Sul, CP 15051, CEP 91501-970 Porto Alegre - RS, Brazil
[b] School of Computing, Engineering and Digital Technologies, United Kingdom

A B S T R A C T

Studying social dilemmas prompts the question of how cooperation can emerge in situations where individuals are expected to act selfishly. Here, in the framework of the one-shot Public Goods Game (PGG), we introduce the concept that individuals can adjust their behaviour based on the cooperative commitments made by other players in the group prior to the actual PGG interaction. To this end, we establish a commitment threshold that group members must meet for a commitment to be formed. We explore the effects of punishing commitment non-compliant players (those who commit and defect if the commitment is formed) and rewarding commitment-compliant players (those who commit and cooperate if the commitment is formed). In the presence of commitment and absence of an incentive mechanism, we observe that conditional behaviour based on commitment alone can enhance cooperation, especially when considering a specific commitment threshold value. In the presence of punishment, our results suggest that the survival of cooperation is most likely at intermediate commitment thresholds. 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. For the reward case, we found that cooperation is highly frequent regardless of the commitment threshold adopted.

1. 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 [1], marriage [2], deposit-refund schemes [3], emotion-based [4] or reputation-based commitment [5,6]. Both empirical and theoretical studies demonstrated that high levels of cooperation can be achieved through reliable commitments [7–11]. They enable individuals to reach mutual cooperation even when there is little knowledge about others' past behaviours [12–15], as it requires them to reveal their preferences or intentions [16–18].

However, prior models of commitment have mainly focused on well-mixed population settings [12,14,19], potentially overlooking the significant influence of commitment dynamics within a population's actual network structure. This structure dictates who may form commitments with whom [20,21], 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 [20–24]. Here, our study reveals that this critical aspect also significantly influences how commitments contribute to the emergence of cooperation.

* Corresponding author.
E-mail address: T.Han@tees.ac.uk (T.A. Han).

https://doi.org/10.1016/j.amc.2024.128646
Received 8 February 2024; Received in revised form 29 February 2024; Accepted 1 March 2024

Available online 12 March 2024

Moreover, the initiation of many collective projects hinges on the majority of participants making a commitment to contribute towards a common good. For instance, for a cooperative hunting endeavour to take place, it typically requires a sufficient number of participants ready and willing to participate [25,26]. While some international agreements require ratification by all parties before entering into force, most (especially global treaties) require a minimum of less than the total number of negotiating countries [3,27]. In general, it appears that the necessary level of commitment is contingent on the specific nature of the problem at hand. However, this issue has been under-explored in theoretical modelling, especially in the context of spatial group-interaction settings.

As motivated, herein we investigate the potential of a conditionally applied commitment strategy, contingent on the required commitment level from the group, to promote the evolution of cooperation in a structured population. Our analysis is carried out in the context of the one-shot Public Goods Game (PGG) [28–32]. In this game, a group of G players has the choice to invest (cooperate, paying a cost c) or not (defect, paying nothing) in a common pool of the group. All their contributions are then multiplied by a factor r ($1 < r < G$) and in the end the result is divided equally among all players, regardless of their initial choice. Before engaging in a PGG game, players can choose whether or not to join a commitment and cooperate in the game. The commitment is formed if a threshold τ ($0 \leq \tau \leq G$) regarding the number of committed players, is met. Subsequently, players make decisions in the game based on whether the commitment is formed or not.

In addition, for understanding optimal incentive mechanisms that enable commitment compliance [12], we assess the comparative effectiveness of institutional reward and punishment in promoting cooperation, given a commitment threshold. Our findings indicate that both institutional punishment and reward mechanisms can positively impact cooperation. Furthermore, even in the absence of these mechanisms, the potential for cooperative players to switch to defection can still contribute to cooperation, especially with a specific commitment threshold. Notably, in the presence of punishment, we observe that high punishment and high commitment thresholds are most effective, while intermediate thresholds can better sustain cooperation for difficult PGGs (i.e. those with small values of the multiplication factor r). In the case of reward, high levels of cooperation can be achieved regardless of the commitment threshold.

2. Model

Players interact following the one-shot PGG of size G. For a player to decide to contribute to the common pool, we introduce a commitment threshold τ that allows them to reconsider their choice. 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. We use the notation ijk 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 example, the NCD strategy does not commit ($i = N$), cooperates if the commitment is formed ($j = C$), and defects if it is not formed ($k = D$). In total, there are eight possible strategies, summarised in Table I.

For one group X, an individual with strategy ijk has the following payoff

$$\Pi_{ijk} = \frac{r}{G} \sum_{x \in X} c_x - c_{ijk} + incentive, \tag{1}$$

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. A player contributes to the common pool, i.e. $c_{ijk} = c$, if they cooperate when the commitment is formed ($j = C$) and the commitment was actually formed, or if they cooperate when the commitment is not formed ($k = C$) and the commitment was actually not formed. Otherwise, the player does not contribute. In that case, $c_{ijk} = 0$. Without loss of generality, we set the cost of contribution $c = 1$.

The *incentive* is provided by an institution, which can be used for rewarding commitment-compliant players (ACC and ACD) or punish commitment non-compliant players (ADC and ADD). The total budget allocated for providing incentives, whether as reward or punishment, is represented by $G\delta$ per group, with δ denoting the per capita budget. We use a weight ω to define which incentive is being applied, with the possibility of a mixed punishment and reward. The reward part is equally divided among all commitment-compliant players n_{com}, resulting in an increase of $\omega G\delta / n_{com}$ in their payoff. The punishment part is equally applied to all $n_{non-com}$ commitment non-compliant players, resulting in a decrease of $(1-\omega)G\delta / n_{non-com}$ in their payoff. For clarity, we present a summary of all incentives in Table I.

In an evolutionary step, first a random player (ijk) is selected from the population. Its payoff is calculated according to Equation (1) for all groups that they participate. Then, a random neighbour ($i'j'k'$) of (ijk) is selected, and we repeat the same calculation for its payoff. Player (ijk) will adopt ($i'j'k'$) strategy according to a probability given by the Fermi update rule,

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

where K is a noise related to irrationality. This evolutionary step is repeated N times where N is the population size, characterising one Monte Carlo step (MCS). We set the total simulation time as $t = 10^5$ MCS which is enough for the equilibrium to be reached. In line with previous works [20,22,33,34], we perform our simulations on a square lattice with von Neumann neighbourhood, with periodic boundary conditions, size $N = 100^2$ and $K = 0.1$. We initialise the population using a uniform distribution of the eight strategies. Since each player can cooperate or not according to their neighbourhood, we define the cooperation frequency as the number of groups in which an interacting player cooperates divided by the total number of groups G.

Table I
Eight strategies with commitment formation and their incentives. It is important to notice that incentives are only applied if a prior commitment is formed. If the commitment is not formed, no incentive is provided to any strategy.

Strategies	Accept commitment?	Cooperate in presence of commitment?	Cooperate in absence of commitment?	incentives
ACC	Yes	Yes	Yes	$\omega G\delta/n_{com}$
ACD	Yes	Yes	No	$\omega G\delta/n_{com}$
ADC	Yes	No	Yes	$-(1-\omega)G\delta/n_{non-com}$
ADD	Yes	No	No	$-(1-\omega)G\delta/n_{non-com}$
NCC	No	Yes	Yes	0
NCD	No	Yes	No	0
NDC	No	No	Yes	0
NDD	No	No	No	0

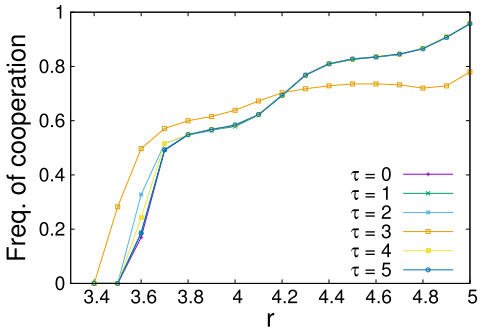

Fig. 1. Frequency of cooperation as a function of the public goods multiplication factor, r. For all $\tau \neq 3$, a similar outcome is observed to that of the classical PGG game ($\tau = 0$). Despite that, we observe a notable increase in cooperation for $\tau = 3$ for small r and a decrease in cooperation for higher r. When $\tau \neq 3$, we observe that committing and non-committing players cannot coexist. This results in unconditional behaviour since commitment is either always met or always not met. For $\tau = 3$, coexistence of committers and non-committers is possible. This brings an advantage to cooperation where some strategies (ACD and NDC) can avoid exploitation by defecting sometimes. For high r values, this advantage is lost since now unconditional cooperation becomes more viable.

3. Results

The structure of our results presentation unfolds as follows. As the baseline, we begin by exploring the effect of commitment on the evolution of cooperation in the spatial PGG in the absence of any incentive mechanism (Section 3.1). We then study the impact of punishment of commitment non-compliant behaviour (Section 3.2) and rewarding of commitment-compliant behaviour (Section 3.3).

3.1. Commitment without incentives ($\delta = 0$)

We show in Fig. 1 the cooperative density for varying r for all threshold values. For $\tau = 0$, we reproduce the classical PGG (i.e., without commitment), where every player has a fixed choice in the PGG since the commitment is always formed. For all other thresholds, except $\tau = 3$, we observe a similar cooperation outcome to the classical case. An important remark is that spatial reciprocity plays a crucial role in games in structured populations [20–24,35]. In such games, cooperators can form clusters to survive, by avoiding defecting neighbours. We notice that only some commitment-based strategies (Table I) have spatial reciprocity, meaning that when they cluster together they cooperate with each other. They are NDC, NCC, ACC, and ACD. Nevertheless, if the commitment threshold is low and thus easily formed, NDC players would have an advantage over the committing strategies above, since they can cluster and cooperate with each other while at the same time defecting against other strategies. Moreover, they are not exploitable by defectors that commit, because they would defect in such players' presence. Note that, unlike NDC, NCC possesses neither of those two advantages. Thus, defectors that do not commit have a benefit and unconditional cooperators that commit are exploitable. On the other hand, if the commitment threshold is high, it would be harder to be formed, and therefore ACD would have spatial reciprocity while at the same time can exploit the other spatial reciprocity capable strategies. They are not exploited by defectors who do not commit since they would defect in their presence. Thus, defectors that commit have a benefit and cooperators that do not commit or behave unconditionally are exploitable.

In summary, low thresholds select for the non-committing strategies ($i = N$), while high thresholds for the committing ones ($i = A$). If we have only committers or only non-committers in the population, players will always behave in the same way. For example, if the strategy ACD always interacts with committers, a commitment is always formed and therefore they always cooperate. Thus, strategies will always cooperate or defect unconditionally, recovering the classical PGG.

Interestingly, for $\tau = 3$, we have a different scenario from the classical PGG, even without punishment or reward. We observe that non-committing strategies (NDD and NDC) coexist with committing ones (ADD and ACD). The intermediate threshold $\tau = 3$ is

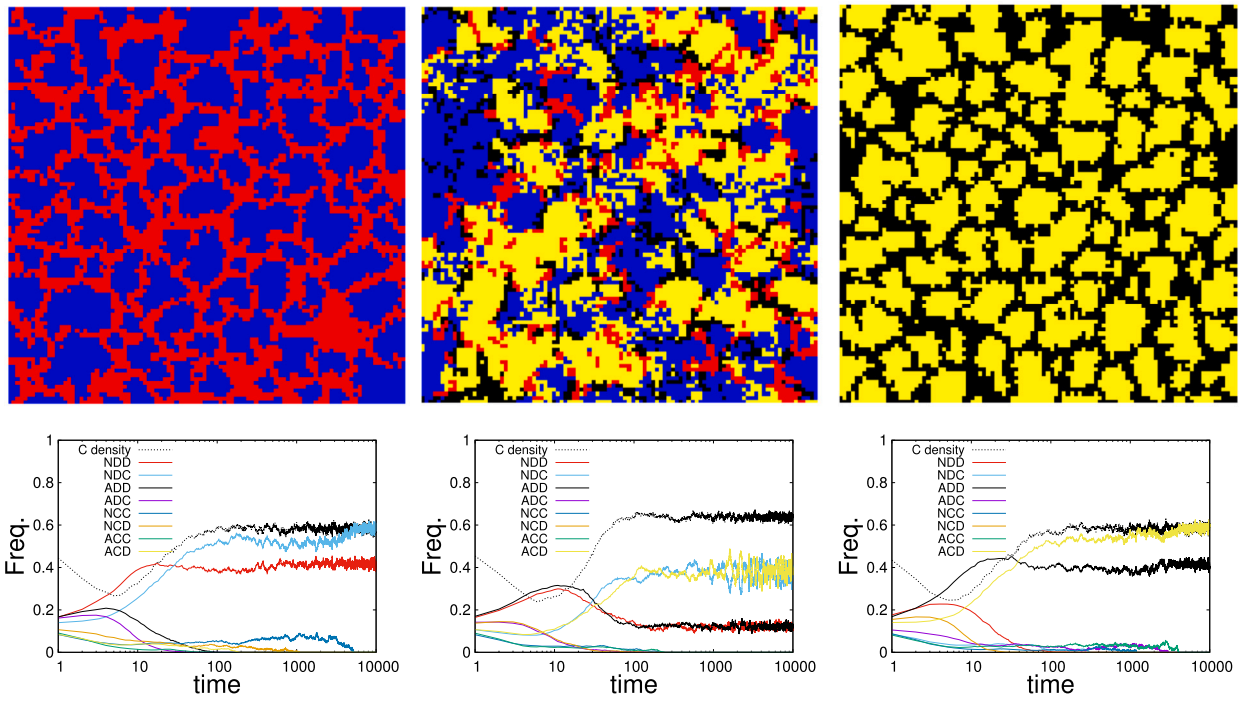

Fig. 2. Snapshots of the population at equilibrium and the time evolution of strategies for $r = 4$ and different threshold values: $\tau = 2$ (left column), 3 (middle column) and 4 (right column). For $\tau = 2$, only non-committers are present, where NDC players (blue) cluster around NDD (red). For $\tau = 4$, only committers survive, where ACD players (yellow) cluster in a sea of ADD (black). Both situations reassemble the classical PGG game where cooperation survives by clustering in a sea of defectors. For $\tau = 3$, there are no longer any clear clusters: we observe that ACD and NDC continue to invade each other in the presence of defectors (NDD and ADD).

high enough for ACD to exploit NDC (capable of invading a cluster without reaching the commitment thresholds) and low enough for the NDC to exploit the ACD (able to invade a cluster and still meet the commitment). In the end, we observe that their densities always converge to the same value when interacting alone, independently of r. As such, we end up with committing and non-committing strategies being able to coexist. If we set only two strategies at a time in the population, we observe that ACD performs better against NDD than NDC. The same is true for ADD, where NDC performs better than ACD. This can be understood by the fact that, when we have an interaction between NDD and NDC, we have the classical scenario, where they never change strategy. When NDD interacts with ACD, a cooperator can sometimes defect and avoid exploitation. Therefore, acting conditionally to avoid exploitation is better than unconditionally cooperating, for low r values. For high r values, unconditional cooperation becomes viable and therefore can outperform defective strategies.

We illustrate the observations above with snapshots of the population when the equilibrium is reached, and the time evolution of the strategies, see Fig. 2. We observe that for low threshold values ($\tau < 3$), only non-committers survive by clustering (NDC and NCC) in a sea of NDD and NCD players, whereas for high threshold values ($\tau > 3$) only committers survive by clustering (ACD and ACC) in a sea of ADC and ADD players. For $\tau = 3$ there are no longer clearly formed clusters as in the previous cases. This is because ACD and NDC can invade one another.

3.2. *Punishment ($\omega = 0$, for varying δ)*

Here we explore the effect of punishment, where players who committed to cooperate yet defect after the formation of a commitment are punished and thus have a decrease in their payoff. Indeed, Fig. 3 shows the frequency of cooperation for different thresholds, in the presence of punishment. We observe that, for a sufficiently high threshold, namely, $\tau \geq 2$, punishment leads to improved cooperation compared to the classical scenario (note that for $\tau = 0$ the model is equivalent to the classical PGG). For $\tau = 1$, there is no improvement compared to the classical scenario. This is because in this case a commitment is never formed, therefore no punishment is applied. This results in the classical scenario since cooperators and defectors who do not commit are possible strategies. We observe significant improvements of cooperation for high thresholds (namely, $\tau = 4$ and 5), which is because commitments are formed and therefore punishment is applied.

We also observe that for all threshold values eventually increasing punishment strength stops to affect cooperation. This observation is in line with previous models of commitments in well-mixed populations [13,15,19]. The lower the threshold, the sooner this behaviour occurs, resulting in cooperation enhancement in high thresholds and low r values for stronger punishments. Therefore, high threshold values are most conducive to the emergence and dominance of cooperation. Interestingly, cooperation is not dominant in the population for $\tau = 3$ in our parameters range. This is due to the coexistence between ACD and NDC players. For the lowest

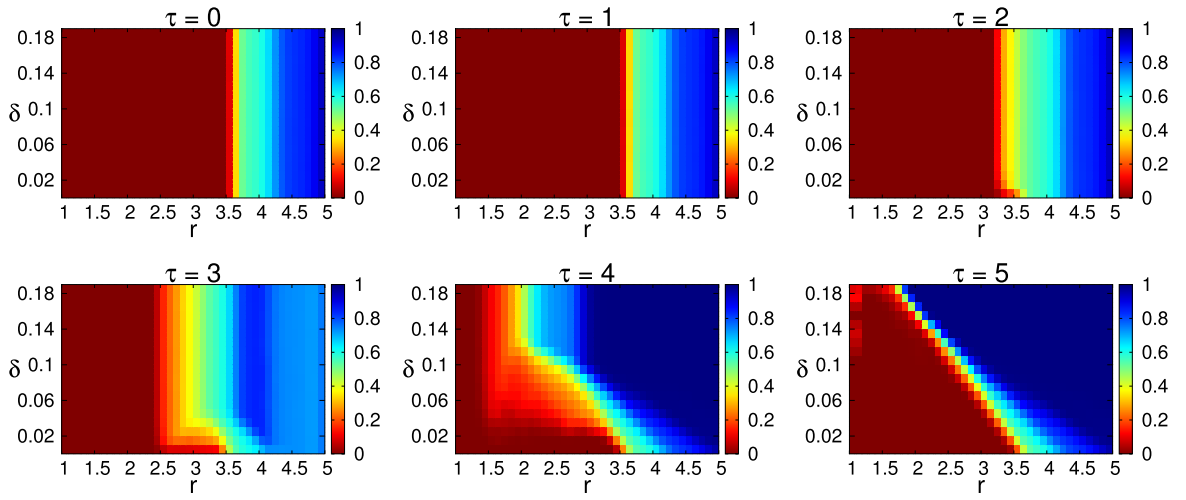

Fig. 3. Phase diagram $r \times \delta$ for the density of cooperative behaviour for all threshold values, in the presence of punishment. We observe that for $\tau \geq 2$, cooperation benefits from the presence of punishment, in the sense of a reduced critical r for cooperation to prevail. For most thresholds we observe a stagnation in the punishment effect. The effect of stronger punishments stagnates slower for high thresholds, allowing cooperation to survive for lower r values. Despite that, intermediate thresholds (i.e. $\tau = 3,\ 4$) can sustain cooperation for the same r value with a smaller punishment value. Another interesting observation is that for $\tau = 5$ cooperation can even survive for very hard PGG (i.e. small r, $1 < r \leq 1.1$), due to a cyclic behaviour.

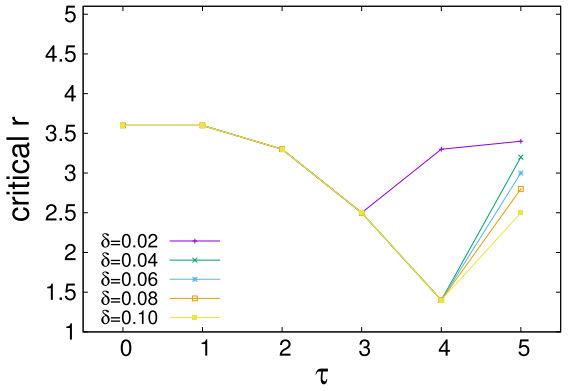

Fig. 4. We show the critical value of r that sustains cooperation, as a function of the threshold τ, for different costs of punishment δ. We observe no effect of punishment for small thresholds. This is due the fact that punishment is only applied if the commitment is met. For higher thresholds, we observe a positive effect of punishment. We observe that weaker punishments are more effective for intermediate thresholds to sustain cooperation.

r values where cooperation survives, we observe that only ACD players coexist with NDD ones. Increasing r benefits cooperative strategies and therefore, increases cooperative density. But at the same time, the increase of r allows NDC players to survive in the population. Their presence is detrimental for the overall cooperative density since they are less effective against non-committing defectors, as discussed in the previous section. Despite that, intermediate thresholds can sustain cooperation for smaller r values if the punishment is low enough. We illustrate this in Fig. 4, where we plot the critical values of r for the survival of cooperation (for low punishment), as a function of the threshold τ. In general, cooperation is enhanced in the sense of survival (minimal r that cooperation survives) for intermediate thresholds. Moreover, cooperation is dominant (minimal r that cooperation dominates) when the threshold is high and punishment is strong.

Now, we examine which strategies contribute to the presence of cooperation in Fig. 3. Fig. 5 shows the frequency of each behaviour as a function of the threshold for $r = 3.6$ and $\delta = 0.19$. The same trend is observed from the one-shot PGG when changing thresholds. For low thresholds, only non-committers survive, while for high thresholds, only committers do. But now, there is a more continuous transition since we observe the coexistence of committers and non-committers for more thresholds, such as 2, 3, and 4 (depending on the r value). It can also be seen that for all thresholds, the cooperative density matches the density of those strategies that would change their choice in the PGG if the commitment is not formed. The case $\tau = 0$ is an exception because the commitment is always formed.

Recalling that for the case of commitment without incentives (Section 3.1), we observed that high thresholds selected for defectors that commit and low thresholds for defectors that do not commit. As such, a commitment is likely being formed for high thresholds, thus, defectors are always be punished. This is why high punishments under high thresholds are highly effective (for promoting cooperation). For lower thresholds, a commitment is less likely to be formed, and thus defectors are punished less often. Therefore

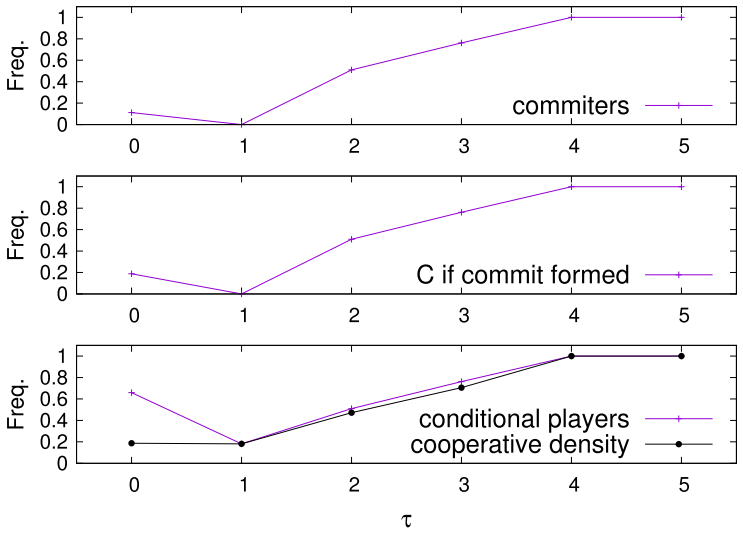

Fig. 5. Density of strategies for varying the threshold τ, for $r = 3.6$ and $\delta = 0.19$. Increasing τ selects for ACD strategy. We observe that for low thresholds cooperation evolves due to the conditional non-committers, while for high thresholds, it evolves due to the conditional committers. Note that the density of cooperation is slightly smaller than the density of conditional players. If committers and non-committers coexist, conditional cooperative players can sometimes defect, decreasing the cooperative density.

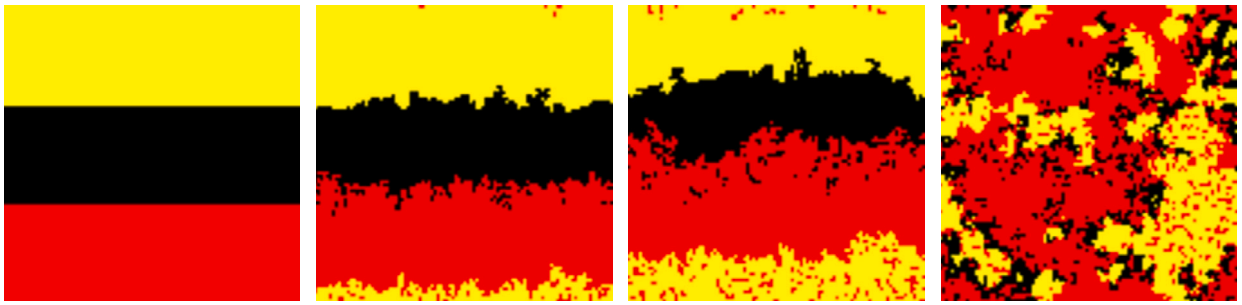

Fig. 6. Snapshots of a prepared initial condition of NDD players (red), ADD (black) and ACD (yellow) for $r = 2.5$, $\delta = 0.06$ and $\tau = 4$ for $MCS = 1,\ 20,\ 50,\ 500$. We observe the cyclic invasions of $NDD > ADD > ACD > NDD$, where NDD invades ADD due to the latter being punished. ADD exploits and invades ACD due to the low r value. And ACD invades NDD since the former can reciprocate with each other. The same dynamics can happen for different τ, r and δ values, but mostly with a negligible presence of cooperative behaviour.

even high punishments are not highly effective. Another interesting observation is that for $\tau = 5$, cooperation can even survive for very low values of r (corresponding to PGG games where it is very hard for cooperation to survive, e.g. $r = 1.1$), due to a cyclic behaviour that will be explored in the next section.

Cyclic behaviour

For very low values of r and δ, we observe that a cyclic behaviour is possible, where $ADD > ACD > NDD > ADD$ (here $X > Y$ or $Y < X$ means X invades Y). One interesting observation is that all strategies involved defect if the commitment is not formed. This reinforces the idea that cooperating when a commitment is not formed is disadvantageous since it indicates the players' intention to cooperate is unlikely. Each step of the cycle $ADD > ACD > NDD > ADD$ can be explained as follows. For low enough r, ADD invades ACD while for high values of r, the opposite occurs. It is because the commitment is always formed and thus the former defects in the interactions. Next, ACD players invade NDD ones. Both defect if the commitment is not being formed. Despite that, ACD has spatial reciprocity while NDD doesn't. Thus, even if they defect when interacting with each other the ACD players have higher payoffs as they contribute in the groups where there are only ACD players. Now, NDD players invade ADD ones, due to the fact that the latter are always being punished when a commitment is formed, while the former are not.

The cycle also occurs when replacing NDD with NCD, where the same explanation applies. Since the cycle has a ADD strategy, for a strong enough punishment, they become extinct, and only ACD survive, breaking the cycle. However, if the punishment is sufficiently low, the invasion $NDD > ADD$ is slow, which is detrimental for ACD players' success and results in the dominance of defection. The cycles can happen for a variety of r values but with a negligible frequency of cooperation. The most relevant regions are for extremely low r values if $\tau = 5$ and for weak punishments for $\tau = 3$ and 4 (below the plateau). In Fig. 6, we show snapshots of a prepared initial condition to illustrate the cyclic dynamic for $r = 2.5$, $\delta = 0.06$ and $\tau = 4$.

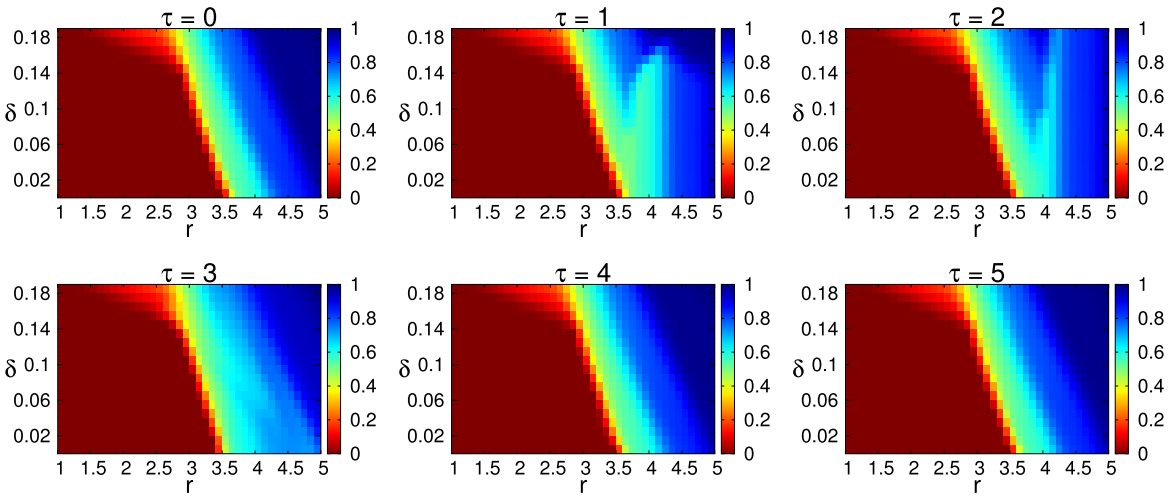

Fig. 7. Phase diagram $r \times \delta$ for the density of cooperative behaviour for all commitment thresholds τ, in the case of institutional reward. We observe that for all values of τ, the critical values of r for the survival of cooperation are similar. Reward is provided to commitment-compliant strategies, giving and advantage for any threshold if δ is high enough. Despite that, some thresholds differ for high r values. For small thresholds, we see cooperation can be harmed for a small region when increasing r.

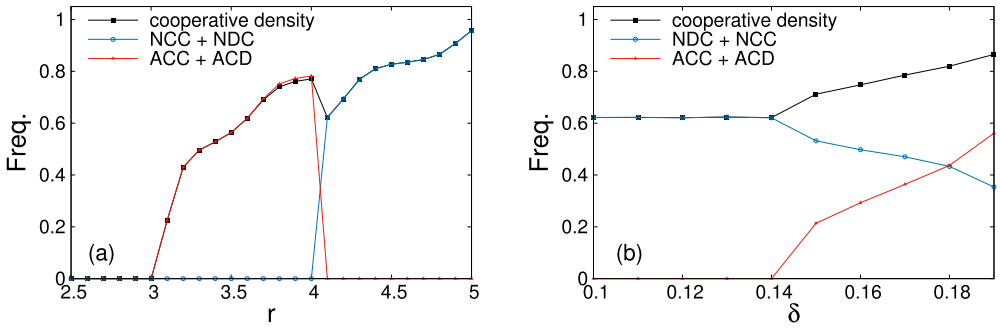

Fig. 8. Frequency of cooperation, cooperative committers ($ACC + ACD$) and cooperative non-committers ($NCC + NDC$) for $\tau = 2$, for (a) $\delta = 0.1$ varying r and (b) $r = 4.1$ varying δ. In the former, similar to the behaviour discussed in the punishment case for $\tau = 3$, we observe that for a high enough r, non-committers can survive, thereby jeopardising cooperation. In (b), we observe that for a sufficiently strong reward, committers are favoured where non-committers have an advantage.

3.3. Reward ($\omega = 1$, for varying δ)

We now consider the situation where commitment-compliant players, i.e. those who commit to cooperate and actually cooperate when the commitment is formed, are rewarded. The rewarded players gain an increase in their payoff according to Table I. In Fig. 7, we show the phase diagrams $r \times \delta$ of cooperative intensity for varying the threshold τ. We observe that the effect of reward is almost invariant among the threshold values. Recall that in the absence of incentives (Section 3.1), low thresholds selected for non-committers due to NDC clustering and exploiting the other clustering strategies (namely, NCC, ACC, and ACD). Now, in the presence of reward, ACD and ACC are rewarded for committing and cooperating, thus even when being exploited by NDC strategy, they still prevail.

We show, in Fig. 8(a), the density of cooperative committing ($ACC + ACD$) and non-committing ($NCC + NDC$) strategies for varying r (fixing $\tau = 2$ and $\delta = 0.1$). An interesting dynamics takes place now, where for $r < 4$ only committers survive, including the defective ones; while for $r > 4$, the population transitions into the dominance of non-committing strategies only. This happens due to the fact that for low thresholds, NDC can exploit and invade ACD. But if r is low enough, non-committing cooperators cannot survive in the sea of defectors, since we would recover the classic PGG. Therefore, for low incentives there occurs a similar phase diagram to the punishment case for $\tau = 1$ and 2, while increasing the reward results in a new region of survival of cooperation due to committers. This transition between committing and non-committing strategies ends up harming cooperation momentarily since cooperators were being rewarded. This however ceases to happen for high r.

Now, in Fig. 8(b), we show the cooperative densities as functions of δ (fixing $r = 4.1$ and $\tau = 2$). We observe that for low rewards there is no coexistence between committers and non-committers, as was the case in Fig. 8(a). But for sufficiently high reward values, the committing cooperators are selected for and outperform the non-committing ones while they coexist. A sufficiently high reward benefits committers and therefore cooperation for low τ. For higher thresholds, committers are selected for even without a reward, as previously explained.

Overall, given the analyses above, we arrive, first of all, at the conclusion that reward is preferred to punishment for promoting cooperation when the commitment threshold is low. When it is high, punishment and reward have a similar effect on the evolutionary outcome of cooperation. Nevertheless, there exists clear difference between punishment and reward, which is most evident for $\tau = 4$, where punishment ensures a higher advantage to cooperation. Another notable finding from our analyses is that conditional strategies were essential for the maintenance of cooperation under punishment. This is due to the fact that punishment does not necessarily mean a benefit to cooperators since non-committing defectors could still exist. For the reward case, unconditional cooperators become more viable if the reward is high enough. Finally, another noticeable finding is that ADC and NCD strategies are the least favourable in general. They can be deemed irrational or contradictory since there is no clear benefit to commit and only cooperate if the commitment is not met. Despite that, NCD could still survive in the punishment case due to possible cyclic dominance that involves the strategy.

4. Discussion

In this paper, we have explored evolutionary dynamics in the spatial Public Goods Game (PGG) with the introduction of commitments. In this setting, behavioural strategies can become conditional, choosing cooperation or defection based on the level of commitment from the group members. Interestingly, with only this assumption, we found an increase in cooperation for a specific commitment threshold. This happens because cooperative players that can change to defection in some situations outperform unconditional cooperators. In this case, we also found that having low commitment thresholds selected for only non-committing strategies, while having high commitment thresholds selected for only committing ones. We have also studied the effect of punishing commitment non-compliant players, showing that intermediate commitment thresholds required the least severe punishment to sustain cooperative behaviour. Nevertheless, to achieve highest levels of cooperation, it is necessary to strictly impose high commitment thresholds and strong punishment.

When considering the effect of rewarding commitment-compliant players, we found that all commitment thresholds led to similar cooperation outcomes. This is because, for any threshold, a high enough reward could sustain committing cooperators. This was not the case with punishment since defectors could still avoid punishment by not committing. This observation is in line with previous results from a well-mixed population analysis of institutional incentives for commitment compliance in the one-shot Prisoner's Dilemma game [12], showing the advantage of reward over punishment in commitment-based interactions. However, as this previous work focuses on two-player game, there was no notion of commitment threshold, which as we show in our analysis, can lead to some advantage for punishment when the threshold is high. The insight gleaned from this finding provides valuable implications for designing institutional mechanisms that foster pro-social behaviour, particularly in situations where pre-communication is permitted to establish mutual cooperative agreements [1,3,10,11,15].

It is noteworthy that institutional incentives have been studied as an important pathway for promoting the emergence of cooperation in social dilemma situations, both in well-mixed and spatial settings [31,36–49]. However, these works have not explored the repercussions of introducing a commitment before interactions in networks, a common occurrence in real-world personal and business settings [1,3,5,27,50,51]. Moreover, an issue with pro-social incentives aimed at promoting cooperation is the potential for antisocial reward and punishment dynamics. In this scenario, defectors may choose to punish cooperators or reward fellow defectors, thereby impeding the evolutionary progress of cooperation [33,52–54]. This concern diminishes when a prior commitment is established, as it clarifies the expected behaviour of all parties involved in the interaction. As such, only those who commit to cooperation can face repercussions for defection or receive rewards for cooperation [12].

When deploying institutional incentives for advocating cooperation, a key consideration is how the incentive-providing institution is set up and maintained. The question of which individuals and how many of them contribute to the incentive pool is a social dilemma itself. Various solutions have been created to address this issue, including pool incentives with second order punishments [31,46,55,56] and hybrid incentives [40,43,44,57]. In this work, we assume that an institution exists to manage and enforce commitment compliance. Previous works on commitment modelling in a pairwise interaction setting, consider a cost assigned to those who agree to join a prior commitment, to sustain the institution [12,19,54]. In our current model, we assume this cost is negligible, in order to focus on unravelling the impact of network reciprocity and commitment threshold on the more complex, multi-player games. Nevertheless, it would be interesting to explore other non-institutional mechanisms that might underline commitment compliance on networks, such as reputation-based mechanisms (that is, lower reputation scores are given to those who do not comply, compared to those who simply defect without making a prior commitment) [58–61], emotional incentives (e.g. individuals feel more guilty for violating a commitment compared to those who simply defect without making a prior commitment) [1,62,63], and how these mechanisms might be integrated with the institutional approach for better promoting prosocial behaviours [1].

Furthermore, many previous evolutionary game models have shown that different heterogeneous aspects, including but not limited to network structures, update mechanisms and incentives, play an important role for the emergence of cooperation [34,35,64–68]. As a future work, it would be interesting to incorporate those heterogeneous factors into our model. This would specifically involve examining the influence of individualised commitment thresholds for different players, rather than treating them as global parameters; as well as that of different heterogeneous population structures.

CRediT authorship contribution statement

L. S. Flores contributed with Software and all authors contributed equally in Conceptualization, Formal analysis and Writing.

Data availability

Data will be made available on request.

Acknowledgements

L.S.Flores thanks the Brazilian funding agency CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the Ph.D. scholarship. The simulations were performed on the IF-UFRGS computing cluster infrastructure. TAH acknowledges generous support from the Future of Life Institute.

References

[1] Randolph Nesse, Evolution and the Capacity for Commitment, Russell Sage Foundation, 2001.
[2] Clifford H. Swensen, Geir Trahaug, Commitment and the long-term marriage relationship, J. Marriage Fam. (1985) 939–945.
[3] Todd L. Cherry, David M. McEvoy, Enforcing compliance with environmental agreements in the absence of strong institutions: an experimental analysis, Environ. Resour. Econ. 54 (2013) 63–77.
[4] Robert H. Frank, Cooperation through emotional commitment, in: Randolf M. Nesse (Ed.), Evolution and the Capacity for Commitment, Russell Sage, New York, 2001, pp. 55–76.
[5] Robert H. Frank, Passions Within Reason: The Strategic Role of the Emotions, Norton and Company, 1988.
[6] Marcus Krellner, The Anh Han, The importance of commitment for stable cooperation, Phys. Life Rev. 46 (2023) 255–257.
[7] Elinor Ostrom, Understanding Institutional Diversity, Princeton University Press, 2005.
[8] Astrid Dannenberg, Non-binding agreements in public goods experiments, Oxf. Econ. Pap. 68 (1) (2016) 279–300.
[9] Zumbansen Peer, The law of society: governance through contract, Indiana J. Glob. Legal Stud. 14 (2) (2007) 191–233.
[10] The Anh Han, Tom Lenaerts, Francisco C. Santos, Luís Moniz Pereira, Voluntary safety commitments provide an escape from over-regulation in ai development, Technol. Soc. 68 (2022) 101843.
[11] Xiao-Ping Chen, Samuel S. Komorita, The effects of communication and commitment in a public goods social dilemma, Organ. Behav. Hum. Decis. Process. 60 (3) (1994) 367–386.
[12] The Anh Han, Institutional incentives for the evolution of committed cooperation: ensuring participation is as important as enhancing compliance, J. R. Soc. Interface 19 (188) (2022) 20220036.
[13] The Anh Han, Luis Moniz Pereira, Tom Lenaerts, Evolution of commitment and level of participation in public goods games, Auton. Agents Multi-Agent Syst. 31 (3) (2017) 561–583.
[14] Tatsuya Sasaki, Isamu Okada, Satoshi Uchida, Xiaojie Chen, Commitment to cooperation and peer punishment: its evolution, Games 6 (4) (2015) 574–587.
[15] Ndidi Bianca Ogbo, Aiman Elragig, The Anh Han, Evolution of coordination in pairwise and multi-player interactions via prior commitments, Adapt. Behav. 30 (3) (2022) 257–277.
[16] T.A. Han, F.C. Santos, T. Lenaerts, L.M. Pereira, Synergy between intention recognition and commitments in cooperation dilemmas, Sci. Rep. 5 (9312) (2015).
[17] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, Henrike Moll, Understanding and sharing intentions: the origins of cultural cognition, Behav. Brain Sci. 28 (05) (2005) 675–691.
[18] J.H. Silk, Girneys, and good intentions: the origins of strategic commitment in nonhuman primates, in: Randolf M. Nesse (Ed.), Evolution and the Capacity for Commitment, Russell Sage, New York, 2001, pp. 138–158.
[19] The Anh Han, Luis Moniz Pereira, Francisco C. Santos, Tom Lenaerts, Good agreements make good friends, Sci. Rep. 3 (1) (2013) 2695.
[20] G. Szabó, G. Fáth, Evolutionary games on graphs, Phys. Rep. 97-216 (4-6) (2007).
[21] Albert-Laszlo Barabasi, Linked-How Everything Is Connected to Everything Else and What It Means F, Perseus Books Group, 2014.
[22] Matjaž Perc, Jesús Gómez-Gardenes, Attila Szolnoki, Luis M. Floría, Yamir Moreno, Evolutionary dynamics of group interactions on structured populations: a review, J. R. Soc. Interface 10 (80) (2013) 20120997.
[23] Attila Szolnoki, György Szabó, Matjaž Perc, Phase diagrams for the spatial public goods game with pool punishment, Phys. Rev. E 83 (3) (2011) 036101.
[24] Attila Szolnoki, Mauro Mobilia, Luo-Luo Jiang, Bartosz Szczesny, Alastair M. Rucklidge, Matjaž Perc, Cyclic dominance in evolutionary games: a review, J. R. Soc. Interface 11 (100) (2014) 20140735.
[25] Michael S. Alvard, David A. Nolin, Rousseau's whale hunt? Coordination among big-game hunters, Curr. Anthropol. 43 (4) (2002) 533–559.
[26] Philip E. Stander, Cooperative hunting in lions: the role of the individual, Behav. Ecol. Sociobiol. 29 (1992) 445–454.
[27] Scott Barrett, Environment and Statecraft: The Strategy of Environmental Treaty-Making: The Strategy of Environmental Treaty-Making, OUP, Oxford, 2003.
[28] Josef Hofbauer, Karl Sigmund, Evolutionary Games and Population Dynamics, Cambridge University Press, 1998.
[29] C. Hauert, A. Traulsen, H. Brandt, M.A. Nowak, K. Sigmund, Via freedom to coercion: the emergence of costly punishment, Science 316 (2007) 1905–1907.
[30] Vincent Ostrom, Elinor Ostrom, Public goods and public choices, in: Alternatives for Delivering Public Services, Routledge, 2019, pp. 7–49.
[31] Karl Sigmund, Hannelore De Silva, Arne Traulsen, Christoph Hauert, Social learning promotes institutions for governing the commons, Nature 466 (7308) (2010) 861–863.
[32] Anna Dreber, David G. Rand, Drew Fudenberg, Martin A. Nowak, Winners don't punish, Nature 452 (7185) (2008) 348–351.
[33] David G. Rand, Martin A. Nowak, The evolution of antisocial punishment in optional public goods games, Nat. Commun. 2 (2011) 434.
[34] Matjaž Perc, Jillian J. Jordan, David G. Rand, Zhen Wang, Stefano Boccaletti, Attila Szolnoki, Statistical physics of human cooperation, Phys. Rep. 687 (2017) 1–51.
[35] Francisco C. Santos, Marta D. Santos, Jorge M. Pacheco, Social diversity promotes the emergence of cooperation in public goods games, Nature 454 (7201) (2008) 213–216.
[36] Yuji Zhang, Ziyan Zeng, Bin Pi, Minyu Feng, An evolutionary game with revengers and sufferers on complex networks, Appl. Math. Comput. 457 (2023) 128168.
[37] Dirk Helbing, Attila Szolnoki, Matjaž Perc, György Szabó, Punish, but not too hard: how costly punishment spreads in the spatial public goods game, New J. Phys. 12 (8) (2010) 083005.
[38] Attila Szolnoki, Matjaz Perc, Reward and cooperation in the spatial public goods game, Europhys. Lett. 92 (3) (2010) 38003.
[39] Tatsuya Sasaki, Åke Brännström, Ulf Dieckmann, Karl Sigmund, The take-it-or-leave-it option allows small penalties to overcome social dilemmas, Proc. Natl. Acad. Sci. 109 (4) (2012) 1165–1169.
[40] Xiaojie Chen, Tatsuya Sasaki, Åke Brännström, Ulf Dieckmann, First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation, J. R. Soc. Interface 12 (102) (2015) 20140935.
[41] Manh Hong Duong, The Anh Han, Cost efficiency of institutional incentives for promoting cooperation in finite populations, Proc. R. Soc. A 477 (2254) (2021) 20210568.

[42] Theodor Cimpeanu, Cedric Perret, The Anh Han, Cost-efficient interventions for promoting fairness in the ultimatum game, Knowl.-Based Syst. 233 (2021) 107545.
[43] António R. Góis, Fernando P. Santos, Jorge M. Pacheco, Francisco C. Santos, Reward and punishment in climate change dilemmas, Sci. Rep. 9 (1) (2019) 1–9.
[44] Weiwei Sun, Linjie Liu, Xiaojie Chen, Attila Szolnoki, Vítor V. Vasconcelos, Combination of institutional incentives for cooperative governance of risky commons, iScience 24 (8) (2021).
[45] Robert Boyd, Herbert Gintis, Samuel Bowles, Coordinated punishment of defectors sustains cooperation and can proliferate when rare, Science 328 (5978) (2010) 617–620.
[46] Christian Hilbe, Arne Traulsen, Torsten Röhl, Manfred Milinski, Democratic decisions establish stable authorities that overcome the paradox of second-order punishment, Proc. Natl. Acad. Sci. USA 111 (2) (2014) 752–756.
[47] Lucas S. Flores, Heitor C.M. Fernandes, Marco A. Amaral, Mendeli H. Vainstein, Symbiotic behaviour in the public goods game with altruistic punishment, J. Theor. Biol. 524 (2021) 110737.
[48] Linjie Liu, Xiaojie Chen, Effects of interconnections among corruption, institutional punishment, and economic factors on the evolution of cooperation, Appl. Math. Comput. 425 (2022) 127069.
[49] Zhen Wang, Zhao Song, Chen Shen, Shuyue Hu, Emergence of punishment in social dilemma with environmental feedback, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, 2023, pp. 11708–11716.
[50] The Anh Han, Intention Recognition, Commitments and Their Roles in the Evolution of Cooperation: From Artificial Intelligence Techniques to Evolutionary Game Theory Models, vol. 9, Springer SAPERE Series, 2013.
[51] Daniel Balliet, Communication and cooperation in social dilemmas: a meta-analytic review, J. Confl. Resolut. 54 (1) (2010) 39–57.
[52] Benedikt Herrmann, Christian Thöni, Simon Gächter, Antisocial punishment across societies, Science 319 (5868) (March 2008) 1362–1367.
[53] Miguel dos Santos, The evolution of anti-social rewarding and its countermeasures in public goods games, Proc. R. Soc. B, Biol. Sci. 282 (1798) (2015) 20141994.
[54] The Anh Han, Emergence of social punishment and cooperation through prior commitments, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 2494–2500.
[55] Attila Szolnoki, Matjaž Perc, Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment, Phys. Rev. X 7 (4) (2017) 041027.
[56] Matjaž Perc, Sustainable institutionalized punishment requires elimination of second-order free-riders, Sci. Rep. 2 (1) (2012) 344.
[57] M.H. Duong, C.M. Durbac, T.A. Han, Cost optimisation of hybrid institutional incentives for promoting cooperation in finite populations, J. Math. Biol. 87 (2023).
[58] Martin A. Nowak, Karl Sigmund, Evolution of indirect reciprocity, Nature 437 (7063) (2005) 1291–1298.
[59] Cedric Perret, Marcus Krellner, The Anh Han, The evolution of moral rules in a model of indirect reciprocity with private assessment, Sci. Rep. 11 (1) (2021) 1–10.
[60] Isamu Okada, A review of theoretical studies on indirect reciprocity, Games 11 (3) (2020) 27.
[61] Minyu Feng, Bin Pi, Liang-Jian Deng, Jürgen Kurths, An evolutionary game with the game transitions based on the Markov process, IEEE Trans. Syst. Man Cybern. Syst. (2023).
[62] Luís Moniz Pereira, Tom Lenaerts, Luis A. Martinez-Vaquero, The Anh Han, Social manifestation of guilt leads to stable cooperation in multi-agent systems, in: AAMAS, 2017, pp. 1422–1430.
[63] Christoph Vanberg, Why do people keep their promises? An experimental test of two explanations 1, Econometrica 76 (6) (2008) 1467–1480.
[64] Matjaž Perc, Attila Szolnoki, A double-edged sword: benefits and pitfalls of heterogeneous punishment in evolutionary inspection games, Sci. Rep. 5 (2015) 11027.
[65] Chaoqian Wang, Attila Szolnoki, A reversed form of public goods game: equivalence and difference, New J. Phys. 24 (12) (2022) 123030.
[66] Marco Antonio Amaral, Marco Alberto Javarone, Heterogeneous update mechanisms in evolutionary games: mixing innovative and imitative dynamics, Phys. Rev. E 97 (Apr. 2018) 042305.
[67] Xian-Bin Cao, Wen-Bo Du, Zhi-Hai Rong, The evolutionary public goods game on scale-free networks with heterogeneous investment, Physica A 389 (6) (2010) 1273–1280.
[68] Lucas S. Flores, Mendeli H. Vainstein, Heitor C.M. Fernandes, Marco A. Amaral, Heterogeneous contributions can jeopardize cooperation in the public goods game, Phys. Rev. E 108 (2) (2023) 024111.