The efficient interaction of costly punishment and commitment

The Anh Han
Tesside University
t.han@tees.ac.uk

Tom Lenaerts
Université Libre de Bruxelles and
Vrije Universiteit Brussel
Tom.Lenaerts@ulb.ac.be

ABSTRACT

To ensure cooperation in the Prisoner's Dilemma, agents may require prior commitments from others, subject to compensations when defecting after agreeing to commit. Alternatively, agents may prefer to behave reactively, without arranging prior commitments, by simply punishing those who misbehave. These two mechanisms have been shown to promote the emergence of cooperation, yet are complementary in the way they aim to instigate cooperation. In this work, using Evolutionary Game Theory, we describe a computational model showing that there is a wide range of parameters where the combined strategy is better than either strategy by itself, leading to a significantly higher level of cooperation. Interestingly, the improvement is most significant when the cost of arranging commitments is sufficiently high and the penalty reaches a certain threshold, thereby overcoming the weaknesses of both strategies.

1. INTRODUCTION

The problem of explaining the evolution of cooperation have been actively addressed in different fields of research, ranging from Economics, Political Science, Biology to Artificial Intelligence and Multi-agent Systems [9, 4, 6, 3, 8]. Various mechanisms, such as direct and indirect reciprocity, kin and group selection, and networked reciprocity have been proposed to explain the evolution of cooperation (see a survey in [9]). Among these mechanisms, costly punishment and commitments are those that can lead to the evolution of cooperation in a one-shot interaction, and both of them have been studied widely in the literature [4, 1, 8, 5, 7]. However, a synergy of these two apparently complementary mechanisms has never been studied before.

Commitment proposers force participants in a game to reveal their intentions. Yet, even when co-players accept the commitment and behave appropriately, they can still decide not to initiate such agreements themselves as this is costly, and defect when no agreement is established. Especially when the commitment cost is high, this kind of free-riders, which benefit directly from the efforts of commitment proposing strategies, can dominate [2, 5, 7], leading to destruction of cooperation and social welfare.

Punishing strategies do not experience this problem. They can effectively deal with these different types of free-riding players, always punishing them even when no agreement was established. Yet, costly punishment requires high efficiency or an excessive effect-to-cost ratio to thrive [3, 5], which is not the case for commitment proposing strategies: Arranging a prior commitment regarding the posterior compensation reduces the cost-to-impact ratio, as was shown in [5].

As a consequence of these observations, we might expect that a weighted combination of these two mutually complementary strategies can lead to a better solution to deal with free-riding behaviors. In the sequel we will show whether and when that is the case, within the context of the one-shot Prisoner's Dilemma and using methods of Evolutionary Game Theory, which are described in the next section.

2. MODELS AND METHODS

In a Prisoner's Dilemma (PD) [9], a player can choose either to cooperate (C) or defect (D). A player who chooses to cooperate with someone who defects receives the sucker's payoff S, whereas the defecting player gains the temptation to defect, T. Mutual cooperation (resp., defection) yields the reward R (resp., punishment P) for both players. The PD is characterized by the ordering, $T > R > P > S$, where in each interaction defection is the rational choice but cooperation is the desired outcome.

Now, before playing the PD, a commitment strategy (denoted by COMP), proposes to her co-player to commit to the game and cooperate. To make the deal reliable, the proposer has to pay an arrangement cost ϵ_1. If the co-player agrees with the deal, then COMP assumes that the opponent will cooperate, yet there is no guarantee that this will actually be the case. When the opponent accepts the commitment though later does not cooperate, she has to compensate the non-defaulting player at a personal cost δ_1 [5].

Next to the traditional unconditional cooperators (C, who always commit when being proposed a commitment deal, cooperate whenever the PD is played, but do not propose commitment themselves) and unconditional cooperators (D, who do not accept commitment, defect when the PD takes place, and do not propose commitment), we consider two commitment free-riding strategies, which have been shown to become dominant under certain conditions in the one-shot PD situation [5]: (ii) Fake committers (FAKE), who accept a commitment proposal yet do not cooperate whenever the PD takes place. These players assume that they can exploit the commitment proposing players without suffering the consequences; (ii) Commitment free-riders (FREE), who defect unless being proposed a commitment, which they then accept and cooperate subsequently in the PD. In other words, these players are willing to cooperate when a commitment is proposed but are not prepared to pay the cost of setting it up.

We consider a well-mixed, finite population of a constant size N, composed of those five strategies, i.e. COMP, C, FREE, D, and FAKE. In each round, two random players are chosen from the population for an interaction. For the row player, the (average)

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

payoff matrix reads

$$M_1 = \begin{array}{c} \\ \text{COMP} \\ C \\ D \\ \text{FAKE} \\ \text{FREE} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{COMP} & \text{C} & \text{D} & \text{FAKE} & \text{FREE} \end{array} \\ \begin{pmatrix} R-\frac{\epsilon_1}{2} & R-\epsilon_1 & 0 & S+\delta_1-\epsilon_1 & R-\epsilon_1 \\ R & R & S & S & S \\ 0 & T & P & P & P \\ T-\delta_1 & T & P & P & P \\ R & T & P & P & P \end{pmatrix} \end{array}.$$

We consider now the costly punishment strategy, denoted by CP [1, 8]. It behaves as a standard C player in the PD. But differently from a C player, the CP player punishes the co-player, at a personal cost ϵ_2, if she defects. That punishment consists in reducing the defector's payoff by δ_2. Replacing COMP with CP in the previous payoff matrix, we obtain a new payoff matrix, denoted by M_2.

We now introduce a new strategy, denoted by CPP, that combines COMP and CP in the following way. With probability q, CPP plays the strategy COMP, and CP otherwise (i.e. with probability $1-q$). Except for the payoff when CPP player encounters another CPP player, the payoff matrix in case of CPP reads

$$M_{CPP} = q \times M_1 + (1-q) \times M_2. \quad (1)$$

The average payoff of a CPP player when playing with another CPP player is:

$$(1-q)^2 R + (1-q)qR + q(1-q)(R-\epsilon_1) + q^2(R-\epsilon_1/2).$$

Evolutionary dynamics.

We adopt a standard approach to implementing social learning or imitation [8, 9]. Namely, at each time-step, one individual i with a fitness f_i is randomly chosen for behavioural revision. i will adopt the strategy of a randomly chosen individual j with fitness f_j with a probability given by the Fermi function $\left(1+e^{-\beta(f_j-f_i)}\right)^{-1}$, where the quantity β controls the intensity of selection. Furthermore, we adopt the small mutation approach, i.e. a single mutant in a monomorphic population will fixate or will become extinct long before the occurrence of another mutation. This allows one to describe the evolutionary dynamics of our population in terms of a reduced Markov Chain of a size equal to the number of different strategies (i.e. five in our model). The stationary distribution of the Markov Chain characterises the average time the population spends in each of these monomorphic states and can be computed analytically. Due to lack of space, we refer to the Method session in [5] for a full description of the stationary distribution computation.

3. RESULTS AND CONCLUSIONS

We compute the frequencies of the five strategies CPP, C, D, FAKE and FREE as a function of q (see Figure 1a). We observe that FAKE players can be restrained better as q increases – that is, when commitment is used more frequent. On the other hand, the FREE players are better coped with for smaller q, i.e. when punishment is used more frequent. This means that a balance between arranging prior commitments and using reactive costly punishment may provide a strategy that performs better than either strategy by itself. Indeed, we observe that there is a large range of q where CPP is better than both COMP and CP. Additional analysis shows this observation is robust for other parameter values, including ϵ and δ.

We now check what is the actual improvement, in terms of the improved level of cooperation, that one may obtain with the combined strategy. To that extent, we search for the optimal value of q, at which CPP has the highest frequency for varying both ϵ and δ (Figure 1b). We observe a significant improvement compared to the highest frequency of COMP and CP. The improvement is

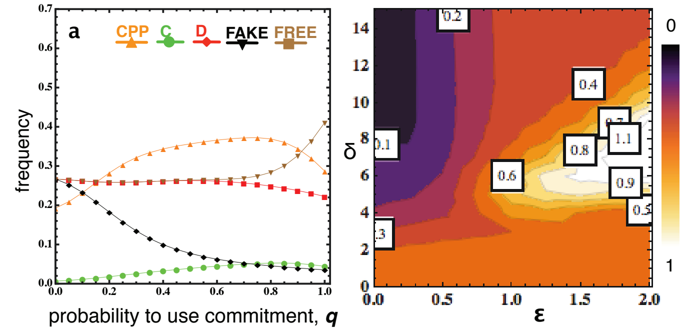

Figure 1: (a) Frequency of each strategy as a function of q, in a population of five strategies CPP, C, D, FAKE and FREE. For a large range of q, CPP has a higher frequency than both the commitment proposing strategy COMP (i.e. CPP with $q = 1$) and the costly punishment strategy CP (i.e. CPP with $q = 0$). (b) Improvement in percentage of CPP in comparison to maximum of CPP and CP. The payoffs being used are, $T = 2,\ R = 1,\ P = 0,\ S = -1$; $\beta = 0.1$; $N = 100$; $\epsilon_1 = \epsilon_2 = \epsilon$; $\delta_1 = \delta_2 = \delta$ where $\epsilon = 0.75$ and $\delta = 5$ in panel a.

most significant (even more than 100%) when δ reaches a certain threshold. And interestingly, it occurs when the cost ϵ is sufficiently large, because in that case, the performance of COMP is severely demolished. This is a notable observation since the performance of the commitment strategy is demolished in the former case and the performance of costly punishment is reduced in the latter one. As such, our results have clearly shown that the combined strategy can overcome the weaknesses of both strategy when using separately.

To conclude, we have shown that, although both commitment and costly punishment might promote the evolution of cooperation in the one-shot interaction setting, they can actually complement each other to lead to a better combined solution that ensures a more favorable outcome for cooperative behavior. As such, our work provides important insights into the design of multi-agent systems, whether they are self-organized or distributed, that make use of commitments or costly punishment for regulation purposes.

4. REFERENCES

[1] Robert Boyd, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100(6):3531–3535, March 2003.

[2] Todd L Cherry and David M McEvoy. Enforcing compliance with environmental agreements in the absence of strong institutions: An experimental analysis. *Environmental and Resource Economics*, 54(1):63–77, 2013.

[3] Martijn Egas and Arno Riedl. The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637):871–878, 2008.

[4] T. A. Han, L. M. Pereira, and F. C. Santos. The emergence of commitments and cooperation. In *AAMAS'2012*, pages 559–566, 2012.

[5] T. A. Han, L. M. Pereira, F. C. Santos, and T. Lenaerts. Good agreements make good friends. *Scientific reports*, 3(2695), 2013.

[6] T. A. Han, L. M. Pereira, F. C. Santos, and T. Lenaerts. Why Is It So Hard to Say Sorry: The Evolution of Apology with Commitments in the Iterated Prisoner's Dilemma. In *IJCAI'2013*, 2013.

[7] The Anh Han, Luís Moniz Pereira, and Tom Lenaerts. Avoiding or Restricting Defectors in Public Goods Games? *Journal of the Royal Society Interface*, 12(103):20141203, 2014.

[8] C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund. Via freedom to coercion: The emergence of costly punishment. *Science*, 316:1905–1907, 2007.

[9] Karl Sigmund. *The Calculus of Selfishness*. Princeton University Press, 2010.