The Cost of Interference in Evolving Systems*

The Anh Han[1], Long Tran-Thanh[2] and Nicholas R. Jennings[2]

[1] Vrije Universiteit Brussel
h.anh@ai.vub.ac.be
[2] University of Southampton
{ltt08r,nrj}@ecs.soton.ac.uk

Abstract. We study the situation of a decision-maker who aims to encourage the players of an evolutionary game theoretic system to follow certain desired behaviours. To do so, she can interfere in the system to reward her preferred behaviour patterns. However, this action requires certain cost (e.g., resource consumption). Given this, her main goal is to maintain an efficient trade-off between achieving the desired system status and minimising the total cost spent. Our initial numerical results reveal interesting observations, which in fact imply that further investigations in the future are required.

1 Introduction

In this paper we consider the following problem. Given a system with a finite number of players, who interact with each other either repeatedly or in a one-shot manner. A decision-maker, who is not part of the system, aims to force the players to maintain certain strategy profiles. However, the decision-maker does not fully control all the behaviours and actions of the players, due to some (physical) limitations. Instead, she can interfere in the system at any particular time step, (partially) modifying the system dynamics. By doing so, she has to consume a certain amount of her (typically limited) resources, which is an increasing function of the degree of interference. Put differently, in order to make more impact on the change of the system dynamics, she has to consume more from her resources. Thus, her actions and the total impact is restricted by her resource constraints. Given this, the research challenge is to identify a sequence of actions that can balance between achieving the decision-maker's objective (i.e., to maintain a desired state) and minimising the resource consumption. This model is motivated by many real-world applications, such as the peace-keeping process of the United Nations [1,2], or population control in habitat management [3,4].

Although the (sequential) decision-making literature provides a number of techniques to tackle similar resource constraint optimisation problems [5,6,7,8,9], these approaches typically ignore the fact that the players, with whom the decision-maker has to interact, also have their own strategic behaviours that together drive the dynamics of the system. Given this, we argue that such solutions will not be able to exploit the system characteristics, and thus, will fail in providing efficient performance in achieving the desired goals. On the other hand, game theoretic literature typically focusses on the extremes. In particular, researchers either assume that the system is fully closed (i.e., there is no outsider decision-makers), or the decision-maker has full control on the behaviour of the players. Typical models for the former are classical (both non-cooperative and coalitional) game theoretical models. The latter includes models from mechanism design, where the decision-maker is the system designer, and can define some set of norms and penalties such that the players are not incentivised to deviate from the norms.

Against this background, this paper aims to start filling the gap by addressing the problem as follows. We combine the decision-making process design with an evolutionary game theoretic perspective (described in Section 2). While the former aims to capture the behaviour of the decision-maker, the latter can be used to formalise the dynamics of the system of players. In particular, we consider a population where the players interact through the Prisoner's Dilemma [10,11], a two-player game model. Suppose that as an outsider decision-maker, we aim to promote a certain strategy profile (more generally with given composition). We also have a budget that can be used to interfere by rewarding particular strategists/individuals in the population at concrete moments (e.g. depending on the current composition of the population). In particular, at each time step, we can reward the players who follow the desired strategy. Hence, the

* The Anh Han was supported by the F.W.O. Belgium (postdoctoral fellowship 05_05 TR 7142). Long Tran-Thanh was supported by the ORCHID project (EP/I011587/1).

research question here is to identify when and how much we want to pay the players, in order to achieve our goals. Within this paper, we demonstrate that the answer to this question is not trivial, and we point out some interesting observations through experimental evaluations.

The remainder of the paper is as follows. We first introduce our formal model. We then continue with the discussion of the numerical results. The last section concludes.

2 Model and Methods

In the current work we focus on a two-player game model, where the one-shot Prisoner's Dilemma (PD) is used as the interaction model of agents in a population. The PD game is a well-known framework to study the problem of the evolution of cooperation [12,11], where without any supporting mechanisms such as kin selection, reciprocities, structured population [12,11], punishment and reward [13,14] and commitments [15], cooperation is rare and cannot evolve. Here, differently from all previous work, we study what are the appropriate interference strategies (by rewarding cooperative acts) leading to high level of cooperation while minimising the investment budget.

2.1 Two-player model

We consider a well-mixed population of fixed size N. The players's interactions are modelled using the one-shot PD, as defined by the payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R, R & S, T \\ T, S & P, P \end{pmatrix}. \end{array}$$

A player who chooses to cooperate (C) with someone who defects (D) 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 [10,11]. Hence, if we consider a population of two strategies, one always cooperates (denoted by C), and another always defects (denoted by D), the latter will spread under evolutionary dynamics (see next section).

Now, suppose we have a budget that can be used to interfere by rewarding particular strategists, which, in the current case, the cooperators (C-players). If the budget is unlimited, it is easy to promote cooperation, but with a limited budget the question could also be, to what extent cooperation can be maintained with it? Put differently, how can we interfere to reach a certain level of cooperation while minimising the cost of interference?

In the well-mixed population of N individuals, with k C-players and $(N-k)$ D-players, the average payoff a C- and a D-player can be written as follows, respectively,

$$\begin{aligned} \Pi_C(k) &= \frac{1}{N-1}\sum_{j=1}^{m}[(k-1)R + (N-k)S] \\ \Pi_D(k) &= \frac{1}{N-1}\sum_{j=1}^{m}[(N-k-1)T + kP] \end{aligned} \tag{1}$$

2.2 Evolutionary Dynamics

The accumulated payoff from all interactions (defined in Eq. (1)) emulates the individual *fitness* or social *success* and the most successful individuals will tend to be imitated by others, implementing a simple form of social learning [11]. A strategy update event is defined in the following way, corresponding to the so-called pairwise comparison [16]. 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 (from statistical physics)

$$p(f_i, f_j) = \left(1 + e^{-\beta(f_j - f_i)}\right)^{-1}$$

where the quantity β, which in physics corresponds to an inverse temperature, controls the intensity of selection. When $\beta = 0$ we obtain the limit of neutral drift, and with the increasing of β one strengthens the role played by the game payoff in the individual fitness, and behavioural evolution [16].

In the absence of mutations, the end states of evolution are inevitably monomorphic, as a result of the stochastic nature of the evolutionary dynamics and update rule. As we are interested in a global analysis of the population dynamics with multiple strategies, we further assume that with a small probability μ individuals switch to a randomly chosen strategy, freely exploring the space of possible behaviours. By introducing a small probability of mutation or exploration, the eventual appearance of a single mutant in a monomorphic population, this mutant will fixate or will become extinct long before the occurrence of another mutation and, for this reason, the population will spend all of its time with a maximum of two strategies present simultaneously [17,11]. 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, where each state represents a possible monomorphic end-state of the population associated with a given strategy, and the transitions between states are defined by the fixation probabilities of a single mutant of one strategy in a population of individuals who adopt another strategy. The resulting stationary distribution characterises the average time the population spends in each of these monomorphic states, and can be computed analytically [17,11] (see below).

In the presence of two strategies the payoffs of each are given by Eq. (1), whereas the probability to change the number k of individuals with a strategy **A** (by ± one in each time step) in a population of $(N-k)$ **B**-strategists is

$$T^{\pm}(k) = \frac{N-k}{N}\frac{k}{N}\left[1 + e^{\mp\beta[\Pi_A(k)-\Pi_B(k)]}\right]^{-1}.$$

The fixation probability of a single mutant with a strategy **A** in a population of $(N-1)$ **B**s is given by [16]

$$\rho_{B,A} = \left(\sum_{i=0}^{N-1}\prod_{j=1}^{i}\lambda_j\right)^{-1} \tag{2}$$

where $\lambda_j = T^-(j)/T^+(j)$.

In the limit of neutral selection (i.e., $\beta = 0$), $\lambda_j = 1$. Thus, $\rho_{B,A} = 1/N$. Considering a set $\{1, ..., n_S\}$ of different strategies, these fixation probabilities determine a transition matrix $[T_{ij}]_{i,j=1,...,n_S}$, with $T_{ii} = 1 - \sum_{k=1,k\neq i}^{n_S}\rho_{k,i}/(n_S-1)$ and $T_{ij,j\neq i} = \rho_{ji}/(n_S-1)$, of a Markov Chain. The normalized eigenvector associated with the eigenvalue 1 of the transposed of M provides the stationary distribution described above [11], describing the relative time the population spends adopting each of the strategies.

2.3 Optimization problem

Let us consider the two-strategy model (C and D), where with a limited budget we would like to check what would be the optimal rewarding/investment strategy, leading to the highest possible frequency of cooperation. We consider that the investment strategy solely depends on the current state of the population. Namely, whenever there are i C-players (i.e. $N-i$ D-players) in the population, an (per-generation) investment, θ_i, is made. That is, each C-player gets an increase of θ_i/i in the average payoff. We denote by $\Theta = \{\theta_1, ..., \theta_{N-1}\}$ the overall investment strategy (scheme).

In order to compute the expected total amount of investment we need to compute the expected number of times the population contains i C-players, $1 \leq i \leq N-1$. For that, we consider an absorbing Markov chain of $(N+1)$ states, $\{S_0, ..., S_N\}$, where S_i represents a population with i C-players. S_0 and S_N are absorbing states. Let $U = \{u_{ij}\}_{i,j=1}^{N-1}$ denote the transition matrix between the $N-1$ transient states, $\{S_1, ..., S_{N-1}\}$. The transition probabilities can be defined as follows. For $1 \leq i \leq N-1$,

$$\begin{aligned} u_{i,i\pm j} &= 0 \qquad \text{for all } j \geq 2 \\ u_{i,i\pm 1} &= \frac{N-i}{N}\frac{i}{N}\left(1 + e^{\mp\beta[\Pi_C(i)-\Pi_D(i)+\theta_i/i]}\right)^{-1} \\ u_{i,i} &= 1 - u_{i,i+1} - u_{i,i-1} \end{aligned} \tag{3}$$

The entries n_{ij} of the so-called fundamental matrix $N = (I-U)^{-1}$ of the absorbing Markov chain gives the expected number of times the population is in the state S_j if it is stated in the transient state S_i [18, Chapter 3].

As a mutant can randomly occur either at S_0 or S_N, the expected number of visits at state S_i is: $\frac{1}{2}(n_{1i} + n_{N-1,i})$. Hence, the expected total investment is

$$Q = \frac{1}{2}\sum_{i=1}^{N-1}(n_{1i} + n_{N-1,i})\theta_i \tag{4}$$

In a population with two strategies C and D, the fixation probabilities of a C (respectively, D) player in a population of D (respectively, C) players when the investment strategy is carried out are, respectively,

$$\begin{aligned} \rho_{D,C} &= \left(1 + \sum_{i=1}^{N-1}\prod_{k=1}^{i}\frac{1+e^{\beta(\Pi_k(C)-\Pi_k(D)+\theta_k/k)}}{1+e^{-\beta(\Pi_k(C)-\Pi_k(D)+\theta_k/k)}}\right)^{-1} \\ \rho_{C,D} &= \left(1 + \sum_{i=1}^{N-1}\prod_{k=1}^{i}\frac{1+e^{\beta(\Pi_k(D)-\Pi_k(C)-\theta_k/k)}}{1+e^{-\beta(\Pi_k(D)-\Pi_k(C)-\theta_k/k)}}\right)^{-1} \end{aligned} \tag{5}$$

Computing the stationary distribution using these fixation probabilities, we can show that the frequency of cooperation can be maximised by maximizing

$$\max_{\Theta}(\rho_{D,C}/\rho_{C,D}) \tag{6}$$

In short, the goal is to find an investment strategy $\Theta = \{\theta_1, ..., \theta_{N-1}\}$ that maximizes the cooperation level (or guarantees a certain level of cooperation) while minimising the expected total investment as defined in Eq. 4. In the next section we use numerical simulations to analyse some (reasonable) investment strategies.

3 Numerical Evaluation

First, suppose in each generation we have a fixed amount of resource, θ, for rewarding cooperative acts, i.e. $\theta_i = \theta\ \forall i$. We ask, should one focus the effort to reward only a few C players rather than spreading the effort to reward all C players but that may not be sufficient for them to survive? For that, we consider investment strategies that invest only when the number of C-players does not exceed a given threshold t, $1 \leq t \leq N-1$.

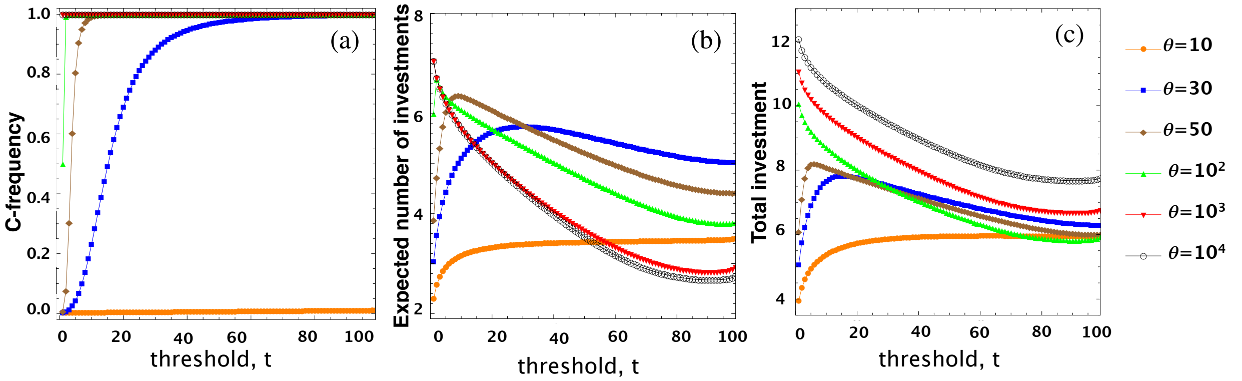

Fig. 1: Level of cooperation (a), expected number of investments (b), and expected total investment (c), as functions of the investment threshold t and for different per-generation investment values θ. In panel (b) and (c), the plot is on a log(10)-scale. Parameters: $R = 1,\ T = 2, P = 0,\ S = -1$; population size $N = 100$; imitation strength $\beta = 0.1$.

We compute the stationary distribution in a population consisting of the two strategies, C and D (see Methods in Section 2). In Figure 1a, we plot the frequency (level) of cooperation varying the investment threshold t, and for different per-generation investment θ. It is not surprising that the larger threshold t, i.e. the more spreading the investment, and the larger the per-generation investment (θ), the higher level of cooperation is obtained. For a too small θ, defection is prevalent even when the investment is always made (see $\theta = 10$). For a sufficiently large θ, a rather spreading investment strategy can lead to a high level of cooperation.

But does a more spreading investment scheme necessarily mean a larger amount of total investment, let alone the higher level of cooperation it leads to? If the stochastic and dynamic aspects of the system are not taken into account, the answer is clearly the positive one. However, as one can see from Figure 1c where the expected total investment is shown for varying t, above a certain threshold of t, a more spreading investment strategy mostly leads to a lower total investment expected to be made (for $\theta \geq 30$). Moreover, the larger θ is, the lower that threshold and the more significant the decreasing are. But it is important to note that this decreasing tendency stops when t reaches a certain threshold (then it slowly increases) (e.g. for $\theta = 100$, the optimal is $t = 91$, and for $\theta = 50$, the optimal is $t = 97$).

The explanation for this observation can be seen from Figure 1b, where the expected number of times of investment is depicted for varying t. We can see that it is still important to make investments when there is a rather large fraction of C-players (i.e. high enough value of t) in the population, because otherwise defection can still fight back and becomes more frequent, leading to further investments latter (hence, wasting the earlier investment efforts). The investment can be ceased only when the fraction of C-players in the population is sufficiently large (around 90%) to be able to maintain their abundance themselves. That is, once we decide to interfere (to help with sustaining high cooperation), we should interfere until the cooperators can survive and fight defection on their own.

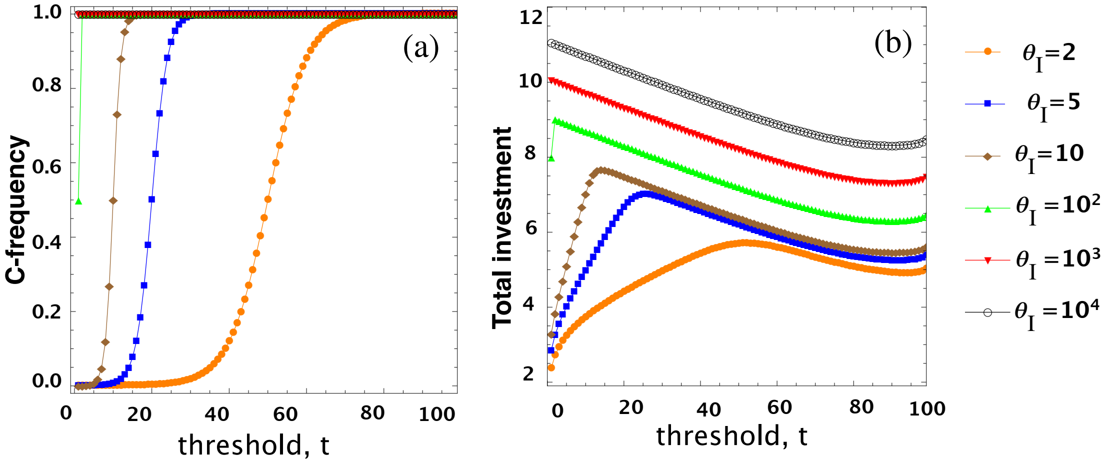

Fig. 2: Level of cooperation (a) and total investment (b), as functions of the investment threshold t and for different individual investment values θ_I. The plots are made on a log(10)-scale. Parameters: $R = 1,\ T = 2, P = 0,\ S = -1$; population size $N = 100$; imitation strength $\beta = 0.1$.

We now consider a different investment strategy where, instead of using a fixed per-generation amount, we employ a fixed, per-individual, investment amount, θ_I; that is, $\theta_i = i \times \theta_I$. We have a similar effect as seen in the previous case (see Figure 2). That is, the greater θ_I and the more spreading the investment strategy are, the higher level of cooperation we obtain. Assuming we would like to maintain a high level of cooperation (say, at least 90%), for any θ_I, the optimal investment scheme is with a sufficiently high t : when reaching a certain threshold of t, the expected total investment increases while not leading to any improvement of the cooperation level. For instance, for $\theta_I = 2$ and 5, the optimal t are 94 and 91, respectively.

Furthermore, to some degree the second investment strategy is more efficient than the first one (comparing the minimal total investment of the two strategies, $\theta_I = 2$ for the second and $\theta = 100$ for the first). That is, if the per-generation investment can be adapted according to the number of players need to be invested in the current state (cooperators), it can lead to a better expected total investment.

4 Conclusions and Future Work

In this paper, we seek an answer to the question of how to interfere in a close system of agents in order to achieve desired system states. In particular, the cost of interference is measured in the consumption of certain (monetary) resources, and the higher impact we want to make, the higher cost we have to pay. To tackle this problem, we combine a decision-making concept with a simple evolutionary game theoretic model. Our initial results reveal non-trivial observations, which need further investigations. In particular, we observed the following from the obtained results:

- Interference must be carried out thoroughly and sufficiently, until the desired behaviour is prevalent enough to sustain itself; and then it is not necessary to invest further.
- In many cases, although a less spreading investment strategy leads to high cooperation, interestingly, more spreading ones lead to much lower total investments (let alone that they clearly lead to at least the same level of cooperation). An inadequately spreading investment scheme gives the undesired behaviour a chance to fight back hence wasting earlier interference efforts.

Our possible future work can be described as follows. In the current work, interference was carried out through rewarding desired behaviour, but it also can be done via punishing non-desired behaviours. We aim to compare these two interference strategies and perhaps find a way to combine them hopefully leading to a better combined interference scheme. Also, as both punishment and rewarding are studied, jointly and separately, in the Evolutionary Game Theory modeling literature (see, e.g. [14]) we aim to reflect on which of two approaches, the external interference and internal strategic motive (of reward and punishment), can provide a more efficient way to reach a certain desired behaviour. Furthermore, in reality interacting agents are usually distributed in a structured network instead of in a well-mixed one as in the current work. We envisage that in such cases the cost of interference can be reduced by identifying the (strongly) influential agents in the network [19], and then focusing investment efforts to promote their survival.

References

1. Roberts, A., Kingsbury, B.: United Nations, Divided World: The UN's Roles in International Relations. Oxford University Press (1993)
2. Lowe, V., Roberts, A., Welsh, J., , Zaum, D.: The United Nations Security Council and War: The Evolution of Thought and Practice since 1945. Oxford University Press (2010)
3. Levin, S.A.: Multiple scales and the maintenance of biodiversity. Ecosystems **3**(6) (2000) 498–506
4. Ausden, M.: Habitat Management for Conservation: A Handbook of Techniques 5. Techniques in Ecology & Conservation. Oxford University Press (2007)
5. Madani, O., Lizotte, D.J., Greiner, R.: The budgeted multi–armed bandit problem. In Proceedings of the Seventeenth Annual Conference on Learning Theory (2004) 643–645
6. Guha, S., Munagala, K.: Approximation algorithms for budgeted learning problems. In Proceedings of the Thirty-Ninth Annual ACM symposium on Theory of Computing (2007) 104–113
7. Bachrach, Y., Elkind, E., Meir, R., Pasechnik, D., Zuckerman, M., Rothe, J., Rosenschein, J.: The cost of stability in coalitional games. In: Algorithmic Game Theory. Volume 5814 of LNCS. (2009) 122–134
8. Tran-Thanh, L., Chapman, A., Rogers, A., Jennings, N.R.: Knapsack based optimal policies for budget–limited multi–armed bandits. In Proceedings of the Twenty-Sixth National Conference on Artificial Intelligence (AAAI) (2012) 1134–1140
9. Ding, W., Qin, T., Zhang, X.D., Liu, T.Y.: Multi-armed bandit with budget constraint and variable costs. In: In Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI). (2013) 232–238
10. Coombs, C.H.: A reparameterization of the prisoner's dilemma game. Behavioral Science **18**(6) (1973) 424–428
11. Sigmund, K.: The Calculus of Selfishness. Princeton University Press (2010)
12. Nowak, M.A.: Five rules for the evolution of cooperation. Science **314**(5805) (2006) 1560 DOI: 10.1126/science.1133755.
13. Fehr, E., Gachter, S.: Altruistic punishment in humans. Nature **415** (2002) 137–140
14. Sigmund, K., Hauert, C., Nowak, M.: Reward and punishment. P Natl Acad Sci USA **98**(19) (2001) 10757–10762
15. Han, T., Pereira, L., Santos, F., Lenaerts, T.: Good agreements make good friends. Scientific reports **3**(2695) (2013)
16. Traulsen, A., Nowak, M.A., Pacheco, J.M.: Stochastic dynamics of invasion and fixation. Phys. Rev. E **74** (2006) 11909
17. Fudenberg, D., Imhof, L.A.: Imitation processes with small mutations. Journal of Economic Theory **131** (2005) 251–262
18. Kemeny, J., Snell, J.: Finite Markov Chains. Undergraduate Texts in Mathematics. Springer (1976)
19. Franks, H., Griffiths, N., Jhumka, A.: Manipulating convention emergence using influencer agents. Autonomous Agents and Multi-Agent Systems **26**(3) (2013) 315–353