

A synergy of costly punishment and commitment in cooperation dilemmas

Journal Title
XX(X):1–11
© The Author(s) 2016
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/

The Anh Han¹ and Tom Lenaerts^{2,3}

Abstract

To ensure cooperation in the Prisoner's Dilemma, individuals may require prior commitments from others, subject to compensations when agreements to cooperate are violated. Alternatively, individuals 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 promote cooperation. Although both mechanisms have their specific limitations, either one of them can overcome the problems of the other. On one hand, costly punishment requires an excessive effect-to-cost ratio to be successful, and this ratio can be significantly reduced by arranging a prior commitment with a more limited compensation. On the other hand, commitment-proposing strategies can be suppressed by free-riding strategies that commit only when someone else is paying the cost to arrange the deal, whom in turn can be dealt with more effectively by reactive punishers. Using methods from Evolutionary Game Theory, we present here an analytical model showing that there is a wide range of settings for which the combined strategy outperforms either strategy by itself, leading to significantly higher levels 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 mechanisms.

Keywords

Commitment, Costly Punishment, Cooperation, Prisoner's Dilemma, Evolutionary Game Theory

Introduction

The problem of explaining the evolution of cooperative behaviour has been actively investigated in many fields, from Evolutionary Biology, Ecology, Computer Science to Economics and Social Science. Several mechanisms responsible for the evolution of cooperation have been proposed, from kin and group selection to direct and indirect reciprocity, to structured populations, and to punishment (Nowak 2006b,a; West et al. 2007; Perc et al. 2013; Sigmund 2010). An extensive body of theoretical and experimental evidence have shown that arranging prior commitments with posterior compensations promotes the evolution of cooperation (Ostrom 1990; Nesse 2001; Gerber and Wichardt 2009; Cherry and McEvoy 2013; Miettinen 2013; Han et al. 2013a, 2015b,c), even when the interaction is one-shot, i.e. not repeated. Arranging prior commitments, such as through enforceable contracts or pledges (Chen and Komorita 1994), deposit-refund schemes (Gerber and Wichardt 2009; Cherry and McEvoy 2013; Sasaki et al. 2015) or even emotional or reputation-based commitment devices (Frank 1988; Nesse 2001), enforces others to cooperate, as it requires them to reveal their preferences or intentions (Chen and Komorita 1994;

Sterelny 2012; Han et al. 2011, 2012a; Han 2013). This behaviour is ubiquitous in various human activities, ranging from personal to group and international relationships (Frank 1988; Chen and Komorita 1994; Nesse 2001; Cherry and McEvoy 2013). For instance, contracts are a popular kind of commitment, playing a key role in enforcing cooperation in modern societies (Skyrms 1996; Nesse 2001). Commitments are also widely studied and utilised in multi-agent and autonomous agent systems, in order to ensure high levels of cooperation among agents (Schelling 1990; Wooldridge and Jennings 1999; Castelfranchi and Falcone 2010; Winikoff 2007).

Another important mechanism that promotes cooperation in one-shot interactions is costly punishment, where a punisher pays a cost to punish another player who

¹School of Computing, Teesside University

²AI lab, Computer Science Department, Vrije Universiteit Brussel

³MLG, Département d'Informatique, Université Libre de Bruxelles

Corresponding author:

The Anh Han, School of Computing, Teesside University, Borough Road, Middlesbrough, TS1 3BA, UK

Email: T.Han@tees.ac.uk

misbehaves (Fehr and Gächter 2002; Fehr and Gächter 2000; Boyd et al. 2003; Henrich et al. 2006; Hauert et al. 2007; Egas and Riedl 2008; Wu et al. 2009; Guala 2012; Traulsen et al. 2012). Unlike commitment, this strategy does not request a prior agreement from the co-player before the interaction. Instead, players will reactively punish those that misbehave (if he or she can be identified) once the interaction has taken place. Several theoretical and experimental studies have shown that this kind of costly punishment can evolve in well-mixed populations only if it is cost-effective, i.e. when the punished agent suffers a sufficiently higher cost than the punisher (Sigmund et al. 2001; Boyd et al. 2003; Carpenter 2007; Anderson and Putterman 2006; Nikiforakis and Normann 2008; Egas and Riedl 2008; Wu et al. 2009) *. This issue has also been shown to be facilitated in structured populations as a result of strategies segregation (Brandt et al. 2003; Nakamaru and Iwasa 2005; Perc and Szolnoki 2012; Szolnoki and Perc 2013): spatial interactions may make it possible for costly punishers to avoid being exploited by second-order free-riders and fight against defectors more effectively (Helbing et al. 2010a,b).

These two strategies, i.e. commitments and costly punishment, compel others to cooperate in a complementary manner. Commitment proposers force participants in a game to reveal their intentions or preferences (Chen and Komorita 1994; Han et al. 2015c,a). 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 (Cherry and McEvoy 2013; Han et al. 2013a), leading to destruction of cooperation and social welfare. Punishing strategies do not experience this problem. They can effectively deal with different types of players free-riding on the investments of commitment proposers, especially since they defect when no agreement was established. Yet, within the context of well-mixed populations as was mentioned earlier, costly punishment only thrives when there is an excessive effect-to-cost ratio (Carpenter 2007; Anderson and Putterman 2006; Nikiforakis and Normann 2008; Egas and Riedl 2008; Wu et al. 2009; Han et al. 2013a), which is not the case for commitment proposing strategies: Arranging a prior agreement regarding the posterior compensation does not require that compensation to scale with the cost of setting up the agreement, as was shown in (Han et al. 2013a, 2015b), reducing significantly the punishment fine necessary to induce an effect on the level of cooperation. Given this complementarity, we hypothesise that a weighted

combination of these two mutually complementary strategies should lead to a better solution in coping with free-riding behaviours and as a consequence with the evolution of cooperation.

Resorting to Evolutionary Game Theory (EGT) methods (Maynard-Smith 1982; Hofbauer and Sigmund 1998; Sigmund 2010), we study the conditions for when a weighted combination of these two strategies may lead to a more favourable outcome for cooperative behaviour in the one-shot Prisoner's Dilemma (PD) (Trivers 1971; Sigmund 2010). In this game, rational choice determines that it is better for each player not to cooperate, even though both would be better off cooperating (see detailed description in the next section). Consequently, evolutionary game dynamics predicts that under those conditions cooperation disappears (Hardin 1968; Nowak 2006b,a; Sigmund 2010). Here, the synergy of the two strategies, arranging prior commitments and costly punishment, will be characterised by a single parameter, which describes the probability that either strategy is used in the PD. We study both analytically and through numerical simulations, the range of this value that will allow the combined strategy to overcome the weaknesses of both strategies, thereby leading to a higher level of cooperation than the one achieved by either strategy by itself. Our results show that there is always a wide range of values for this parameter where the synergistic strategy performs better than both strategies independently.

The remainder of this paper is structured as follows. The next section describes the model combining commitment and punishment strategies in the one-shot Prisoner's Dilemma, and the EGT methods used to analyse the model. Then, in Results section, analytical and numerical results are presented. The paper ends with a discussion and some directions for future work. We also provide a separate Supporting Information (SI) text that contains additional results and supporting analysis, which is referred to in the paper where relevant.

Models and Methods

We first recall the definition of the Prisoner's Dilemma game and its extension with strategies that allow for the creation of commitment strategies (Han et al. 2012b, 2013a) or strategies using posterior costly punishment (Sigmund

*Note that this argument appears to be valid in well-mixed populations only when antisocial punishment, which has defectors punishing cooperators, is not allowed or unavailable (Herrmann et al. 2008; Rand et al. 2010; Rand and Nowak 2011; Hilbe and Traulsen 2012). Institutionalising punishment (a.k.a. pool punishment) wherein the punishment duty is shared or coordinated among several players rather than being carried by a single player, can overcome the cost efficiency issue (Boyd et al. 2010; Sigmund et al. 2010; Sasaki et al. 2012; Szolnoki et al. 2011b; Hilbe et al. 2014; Schoenmakers et al. 2014).

et al. 2001; Boyd et al. 2003). We then describe our model for the synergy of prior commitments and posterior punishment.

One-shot Prisoner's Dilemma (PD)

As usual, for the one-shot Prisoner's Dilemma (PD) game the four possible outcomes resulting from the action choices of the two players can be written down as a symmetric payoff matrix (Trivers 1971; Sigmund 2010)

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

with the entries of this matrix satisfying the ordering $T > R > P > S$ (Coombs 1973). The interpretation of these entries is explained as follows: Once the interaction is established and both players have decided to play C (D), both players receive the same reward R (penalty P) for mutual cooperation (mutual defection). Unilateral cooperation provides the sucker's payoff S for the cooperative player and the temptation to defect T for the defective one. Changing the ordering of the matrix entries will result in different kinds of social dilemmas with their specific Nash Equilibria. For the sake of mathematical simplicity, the Donor game (Sigmund 2010), a special case of the PD, is sometimes used: $T = b$, $R = b - c$, $P = 0$, $S = -c$, where b and c correspond to the benefit and cost of cooperation, respectively.

Commitment and costly punishment in the Prisoner's Dilemma

Before playing the PD, a commitment strategy (denoted by COMP), proposes to her co-player to commit to the game and cooperate. As arranging agreements or exposing others' intentions may be costly, 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 accepted the commitment and later does not cooperate, she has to compensate the non-defaulting player at a personal cost δ_1 .

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 pair-wise PD situation (Han et al. 2013a):

1. 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;

2. 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.

In the following sections, we consider well-mixed, finite populations of a constant size N , composed of those five strategies, i.e. COMP, C, FREE, D, and FAKE. We have shown, both analytically and numerically, that COMP dominates when the cost of arranging commitment ϵ_1 is justified with respect to the cost of cooperation c and the compensation δ_1 is sufficiently high, leading to a substantial level of cooperation (Han et al. 2013a) (see also SI, Figure S1). If these conditions are not satisfied, then either the FREE or FAKE players will dominate the population making it also possible for D players to dominate commitment proposers at higher initiation costs. Important to understand is that the cost of setting up the agreement (ϵ_1) is the essential factor, since when the compensation (δ_1) reaches a certain threshold, increasing it does not lead to any noticeable improvement for the level of cooperation (Han et al. 2013a).

Adding now all five strategies, the following payoff matrix for the pairwise PD is obtained (Han et al. 2013a):

$$M_1 = \begin{array}{c} \text{COMP} \\ C \\ D \\ \text{FAKE} \\ \text{FREE} \end{array} \begin{array}{ccccc} \text{COMP} & C & D & \text{FAKE} & \text{FREE} \\ \left(\begin{array}{ccccc} R - \frac{\epsilon_1}{2} & R - \epsilon_1 & 0 & S + \delta_1 - \epsilon_1 & R - \epsilon_1 \\ R & R & S & S & S \\ 0 & T & P & P & P \\ T - \delta_1 & T & P & P & P \\ R & T & P & P & P \end{array} \right) \end{array} \quad (1)$$

Let us in turn determine the dominance conditions for the costly punishment strategy (CP) (Sigmund et al. 2001; Boyd et al. 2003; Hauert et al. 2007). It behaves as a standard C player in the PD, but unlike a C player, it punishes its co-player if she defects in the game. That punishment consists in paying a personal cost ϵ_2 to reduce the defector's payoff by δ_2 . Replacing COMP with CP in the previous payoff matrix, provides the following new payoff matrix

$$M_2 = \begin{array}{c} \text{CP} \\ C \\ D \\ \text{FAKE} \\ \text{FREE} \end{array} \begin{array}{ccccc} \text{CP} & C & D & \text{FAKE} & \text{FREE} \\ \left(\begin{array}{ccccc} R & R & S - \epsilon_2 & S - \epsilon_2 & S - \epsilon_2 \\ R & R & S & S & S \\ T - \delta_2 & T & P & P & P \\ T - \delta_2 & T & P & P & P \\ T - \delta_2 & T & P & P & P \end{array} \right) \end{array} \quad (2)$$

Similar to COMP, the lower the cost ϵ_2 , the higher frequency of CP. But in contrast to COMP, the frequency of CP increases with δ_2 when assuming a specific value of ϵ_2 (see SI, Figure S1). However, when the cost of commitment is sufficiently small, to reach the same level of cooperation as in the commitment model a much more severe punishment is required for an equivalent cost, especially when the PD is less harsh. As such, commitment effectively reduces the cost-to-impact ratio in this case, an inefficient situation typically observed for CP in well-mixed populations (Anderson and Putterman 2006; Egas and Riedl 2008; Han et al. 2013a).

The synergy of commitment and punishment strategies.

We now introduce a new strategy, denoted by CPP, that combines COMP and CP in the following manner: With probability q , CPP uses 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

$$M_{CPP} = q \times M_1 + (1 - q) \times M_2. \quad (3)$$

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

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

In this expression, the four terms correspond to the payoff they get for playing both C taking into account the probabilities that i) both players do not propose the commitment, ii) the focal player proposes and the other does not, iii) the focal player does not propose and the other does, and iv) both players propose.

Evolutionary dynamics in finite populations

Both the analytical and numerical results obtained here use Evolutionary Game Theory (EGT) methods for finite populations (Nowak et al. 2004; Imhof et al. 2005; Nowak 2006a; Sigmund 2010). In such a setting, agents' payoff represents their *fitness* or social *success*, and evolutionary dynamics is shaped by social learning (Hofbauer and Sigmund 1998; Sigmund 2010), whereby the most successful agents will tend to be imitated more often by the other agents. In the current work, social learning is modeled using the so-called pairwise comparison rule (Traulsen et al. 2006), assuming that an agent A with fitness f_A adopts the strategy of another agent B with fitness f_B with probability p given by the Fermi function (Szabó and Tóke 1998; Traulsen et al. 2006; Sigmund 2010; Perc and

Szolnoki 2010),

$$p_{A,B} = \left(1 + e^{-\beta(f_B - f_A)}\right)^{-1}.$$

The parameter β represents the 'imitation strength' or 'intensity of selection', i.e., how strongly the agents base their decision to imitate on fitness difference between themselves and the opponents. For $\beta = 0$, we obtain the limit of neutral drift – the imitation decision is random. For large β , imitation becomes increasingly deterministic.

In the absence of mutations or exploration, the end states of evolution are inevitably monomorphic: once such a state is reached, it cannot be escaped through imitation. We thus further assume that, with a certain mutation probability, an agent switches randomly to a different strategy without imitating another agent. In the limit of small mutation rates, the dynamics will proceed with, at most, two strategies in the population, such that the behavioural dynamics can be conveniently described by a Markov Chain, where each state represents a monomorphic population, whereas the transition probabilities are given by the fixation probability of a single mutant (Fudenberg and Imhof 2005; Imhof et al. 2005; Nowak et al. 2004). The resulting Markov Chain has a stationary distribution, which characterises the average time the population spends in each of these monomorphic end states.

Let N be the size of the population. Suppose there are at most two strategies in the population, say, k agents using strategy A ($0 \leq k \leq N$) and $(N - k)$ agents using strategy B. Thus, the (average) payoff of the agent that uses A and B can be written as follows, respectively,

$$\begin{aligned} \Pi_A(k) &= \frac{(k-1)\pi_{A,A} + (N-k)\pi_{A,B}}{N-1}, \\ \Pi_B(k) &= \frac{k\pi_{B,A} + (N-k-1)\pi_{B,B}}{N-1}. \end{aligned} \quad (4)$$

Now, the probability to change the number k of agents using strategy A by \pm one in each time step can be written as (Traulsen et al. 2006)

$$T^\pm(k) = \frac{N-k}{N} \frac{k}{N} \left[1 + e^{\mp\beta[\Pi_A(k) - \Pi_B(k)]}\right]^{-1}. \quad (5)$$

The fixation probability of a single mutant with a strategy A in a population of $(N - 1)$ agents using B is given by (Traulsen et al. 2006; Karlin and Taylor 1975; Nowak et al. 2004)

$$\rho_{B,A} = \left(1 + \sum_{i=1}^{N-1} \prod_{j=1}^i \frac{T^-(j)}{T^+(j)}\right)^{-1}. \quad (6)$$

In the limit of neutral selection (i.e. $\beta = 0$), $\rho_{B,A}$ equals the inverse of population size, $1/N$.

Considering a set $\{1, \dots, s\}$ of different strategies, these fixation probabilities determine a transition matrix

$M = \{T_{ij}\}_{i,j=1}^s$, with $T_{ij,j \neq i} = \rho_{ji}/(s-1)$ and $T_{ii} = 1 - \sum_{j=1, j \neq i}^s T_{ij}$, of a Markov Chain. The normalized eigenvector associated with the eigenvalue 1 of the transposed of M provides the stationary distribution described above (Karlin and Taylor 1975; Fudenberg and Imhof 2005; Imhof et al. 2005), describing the relative time the population spends adopting each of the strategies.

Risk-dominance An important measure to compare the two strategies A and B is which direction the transition is stronger or more probable, an A mutant fixating in a population of agents using B, $\rho_{B,A}$, or a B mutant fixating in the population of agents using A, $\rho_{A,B}$. It can be shown that the former is stronger, in the limit of large N , if (Kandori et al. 1993; Sigmund 2010)

$$\pi_{A,A} + \pi_{A,B} - \pi_{B,A} - \pi_{B,B} > 0. \quad (7)$$

Results

Conditions for the viability of CPP

We derive analytical conditions for which CPP can be a viable strategy for the emergence of cooperation in the Donation game (see Methods). In other words, we wish to determine when CPP is successful against the defective and free-riding strategies, and this is relative to q . Using the inequality in Equation (7), we obtain the conditions under which CPP is risk-dominant against the three strategies D and FAKE and FREE, respectively

$$\delta_2(1-q) + q(b+c-\epsilon_1+\epsilon_2) + \frac{q^2\epsilon_1}{2} - 2c - \epsilon_2 > 0, \quad (8)$$

$$\delta_2(1-q) + 2q\delta_1 + q(\epsilon_2 - 2\epsilon_1) + \frac{q^2\epsilon_1}{2} - 2c - \epsilon_2 > 0, \quad (9)$$

$$\delta_2(1-q) + q(b+c) + q(\epsilon_2 - 2\epsilon_1) + \frac{q^2\epsilon_1}{2} - 2c - \epsilon_2 > 0. \quad (10)$$

We observe that the left hand side of Equation (10) is smaller than or equal to the left hand side of Equation (8). Thus, satisfying the inequality in Equation (10) implies satisfying the inequality in Equation (8). Hence, in order for CPP to be risk-dominant against the three defective strategies, we only need to guarantee that the two inequalities in Equation (9) and (10) hold. By solving the system of these two inequalities, we can analytically derive the range of q for which CPP is risk-dominant against all free-riders (see SI for details), which are corroborated by numerical simulation results in the next section.

We now derive some properties of these inequalities for varying q , and for some special cases. First, we observe that whenever punishment is carried out, at least to some degree (i.e. $q < 1$), the two conditions are satisfied whenever δ_2 is sufficiently large, regardless of values of the other costs

(see a proof in SI). This is in contrast to COMP where if the cost of arranging commitment, ϵ_1 , exceeds a certain limit, COMP is not risk-dominant against the commitment free-riders FREE, however large δ_1 is. This can be seen explicitly by simplifying Equation (10), substituting for $q = 1$, which is equivalent to (Han et al. 2013a)

$$b - c - \frac{3\epsilon_1}{2} > 0. \quad (11)$$

In order to obtain further meaningful comparison of CPP to COMP and CP, we consider that prior commitment and costly punishment have the same cost of setting up as well as the same effect of compensation/punishment when the co-player misbehaves (i.e. defects in the PD after having agreed to cooperate in the former case, and defects in the PD in the latter). That is, we assume $\epsilon_1 = \epsilon_2 = \epsilon$ and $\delta_1 = \delta_2 = \delta$. Then, Equation (9) can be simplified to

$$\delta > \epsilon + \frac{2c}{1+q} - \frac{q^2\epsilon}{2(1+q)}. \quad (12)$$

The right hand side of Equation (12) is a decreasing function of q (see SI), implying that the larger q is, the easier this condition can be satisfied. In other words, the larger q the better FAKE players can be restrained by CPP players (as can already be inferred from the numerical results shown in Figure 1).

When $q = 0$, CPP is reduced to CP. Hence, CP is risk-dominant against the three defective strategies when

$$\delta > \epsilon + 2c. \quad (13)$$

Similarly, when $q = 1$, CPP is reduced to COMP, which is risk-dominant against all the defective strategies when

$$\begin{aligned} \epsilon &< \frac{2(b-c)}{3}, \\ x\delta &> \frac{3\epsilon}{4} + c. \end{aligned} \quad (14)$$

Comparing conditions specified in Equation (13) and Equation (14), it follows that when ϵ is sufficiently small (namely, bounded up by $2(b-c)/3$), then a smaller δ is required for the risk-dominance of COMP against all defective strategies than CP (see also SI, Figure S1).

Now, Equation (10) can be simplified to

$$\delta - 2c - \epsilon + q \left(b + c - \delta - \epsilon \left(1 - \frac{q}{2} \right) \right) > 0. \quad (15)$$

When δ is sufficiently large compared to the benefit and cost of cooperation, the left hand side is a decreasing function of q (see SI). Hence, it is more difficult for CPP to be risk-dominant against FREE players as q increases (see also Figure 1).

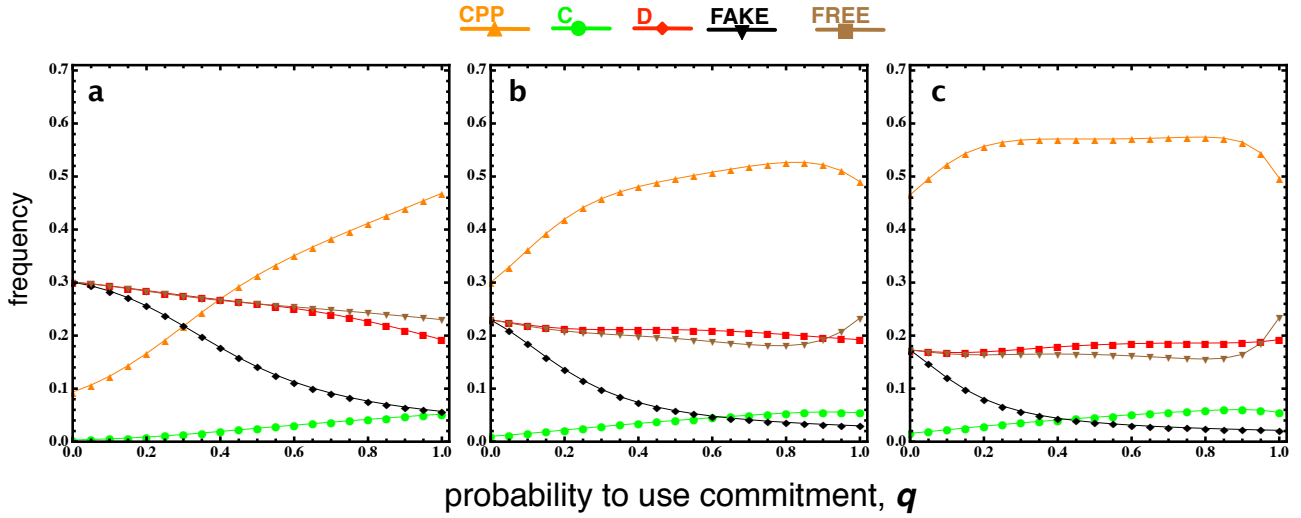


Figure 1. Frequency of each strategy as a function of q , in a population of five strategies CPP, C, D, FAKE and FREE. For a large range of q , CPP has a higher frequency than both the commitment proposing strategy COMP (i.e. CPP with $q = 1$) and the costly punishment strategy CP (i.e. CPP with $q = 0$). The payoffs being used are, $T = 2$, $R = 1$, $P = 0$, $S = -1$; imitation strength, $\beta = 0.1$; population size, $N = 100$; $\epsilon_1 = \epsilon_2 = \epsilon = 0.75$; $\delta_1 = \delta_2 = \delta$ where $\delta = 5$ in panel (a), $\delta = 10$ in panel (b) and $\delta = 15$ in panel (c).

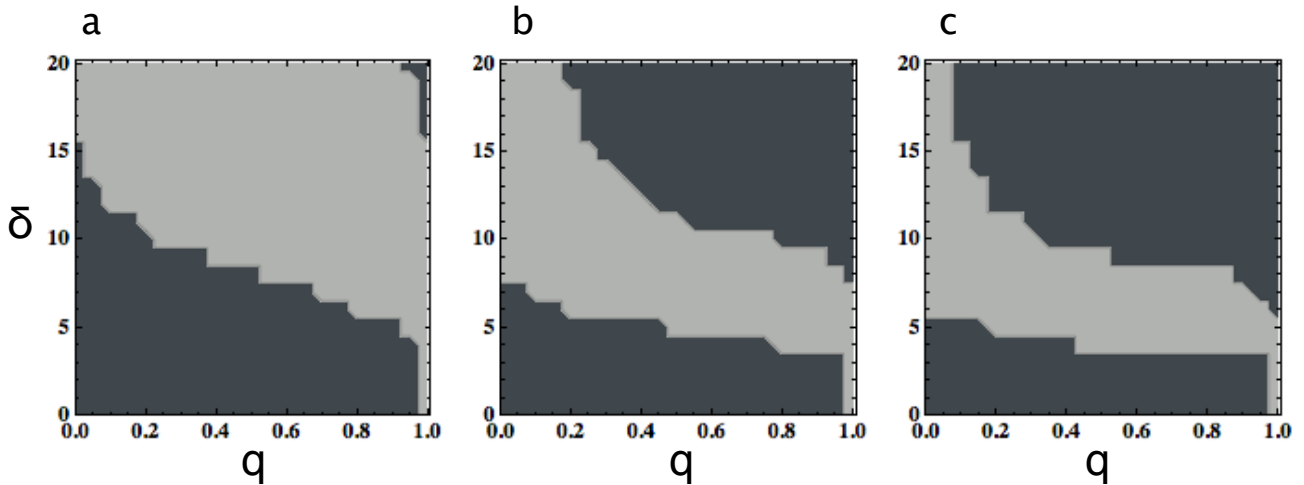


Figure 2. The range of q in which CPP is more frequent than both COMP and CP (the light grey area), as a function of δ , for (a) $\epsilon = 0.1$, (b) $\epsilon = 0.5$ and (c) $\epsilon = 1$. In general, CPP is more frequent for a wide range of q , which is larger for smaller ϵ . The range is largest when δ is sufficiently high (but not too high). Both the lower and upper bounds of the range decrease (not strictly) with δ . The payoffs being used are, $T = 2$, $R = 1$, $P = 0$, $S = -1$; imitation strength, $\beta = 0.1$; population size, $N = 100$.

Varying usage of prior commitments and punishment can cope better with various types of free-riders

The analytical observations above are clarified by looking at Figure 1, where we plot the frequencies of the five strategies CPP, C, D, FAKE and FREE (see Methods) as a function of q , for different values of δ and identical ϵ . For a large enough δ that sufficiently restrains FAKE players, increasing it further does not enhance the frequency or

performance of COMP (i.e. CPP with $q = 1$). But that is not the case for CP, since the frequency of CP (i.e. CPP with $q = 0$) against D, FAKE and FREE always increases with δ (see also Figure S1 in SI). In general, we observe that FAKE players can be restrained better as q increases, i.e. when commitment is used more often. On the other hand, the FREE players are better coped with for smaller q , i.e. when punishment is used more often, especially for larger δ (comparing the three panels). This means that a balance between arranging prior commitments and using reactive

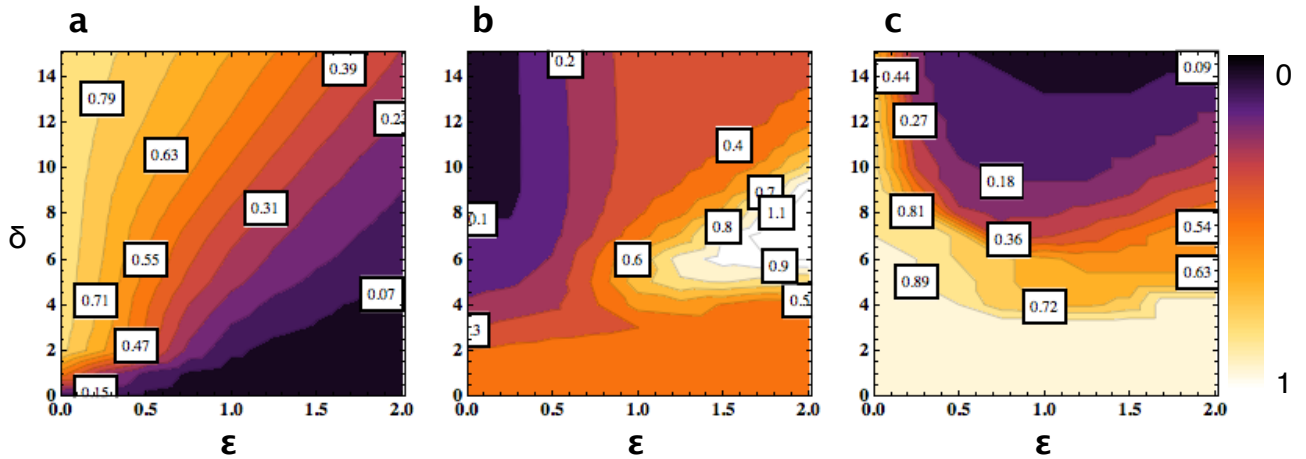


Figure 3. (a) Frequency of CPP at optimal value of q , (b) the improvement in percentage of CPP in comparison to maximum of CPP and CP, and (c) optimal value of $q \in [0, 1]$ where CPP has the highest frequency, as a function of ϵ and δ (where $\epsilon_1 = \epsilon_2 = \epsilon$ and $\delta_1 = \delta_2 = \delta$). The frequency of CPP increases with δ , and exhibits a significant improvement in comparison to the average of CP and COMP. Furthermore, the larger δ , the lower optimal value of q . The payoffs being used are, $T = 2$, $R = 1$, $P = 0$, $S = -1$; imitation strength, $\beta = 0.1$; population size, $N = 100$.

costly punishment may provide a strategy that performs better than either strategy by itself.

Indeed, we observe that there is a wide range of q where CPP is better than both COMP and CP, in all the panels of Figure 1. The ranges are in close accordance with the analytical results (see SI, Section 3). Furthermore, in Figure 2 we plot the range of q in which CPP is better than both COMP and CP, for varying δ as well as for different values of ϵ . In general, CPP is more frequent for a wide range of q , and this range is larger for smaller ϵ . The range is largest when δ reaches a certain threshold, as whereafter commitment does not provide further advantage for the combined strategy. However, when δ is small, COMP is crucially important to reduce the impact-to-cost ratio, especially for a smaller ϵ , as can be seen from the fact that the upper bound of the range is close to 1 (i.e. 100% use of commitment). Furthermore, we can observe that both the lower and upper bounds of this range decrease when δ increases, because punishment is more beneficial then.

As we have seen, there is always a wide range of parameters' values in which CPP is better than both COMP and CP individually. But what is the actual improvement, e.g. 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 (Figures 3a and 3c), for varying both ϵ and δ . We can see that, CPP retains the important property of CP, i.e. its frequency increases with δ (Figure 3a). Furthermore, we observe a significant improvement compared to the highest frequency of COMP and CP (Figure 3b). As expected, the improvement is most significant (even more than 100%) when δ reaches

a certain threshold. But interestingly, it occurs when the cost ϵ is sufficiently large, because in that case, the performance of COMP is severely demolished. These interesting observations are robust for varying the benefit-to-cost ratio (see SI, Figure S2).

Moreover, to gain further understanding about when CPP performs best, in Figure 3c we plot as a function of ϵ and δ the optimal value of q where CPP reaches its highest frequency. We observe that the optimal value of q is a decreasing function of δ , which reaches its minimum for intermediate values of ϵ . That is, it is more advantageous to use more punishment when the effect-to-cost is sufficiently high. This observation is robust for various configurations of the PD (see SI, Figure S3). We also observe that the less harsh the PD becomes, using more commitments is more beneficial to reach high levels of cooperation.

Discussion

Both costly punishment and arranging prior commitment have been shown to provide important pathways for the evolution of cooperation (Boyd et al. 2003; Hauert et al. 2007; Hilbe and Traulsen 2012; Chen and Komorita 1994; Han et al. 2013a, 2015b,c; Han 2016), and both are widely used in diverse human activities to enforce cooperation and regulate collective behaviour (Fehr and Gächter 2002; Fehr and Gächter 2000; Nesse 2001; Henrich et al. 2006; Sterelny 2012). Interestingly, we have shown in this paper that a simple synergy of the two mechanisms can lead to a even better one that promotes a higher level of cooperation than either strategy by itself, for a wide range of parameters' values. Each strategy has its own weakness, which the

combined strategy can overcome. On one hand, arranging prior commitment reduces the effect-to-cost ratio required by costly punishment to perform efficiently, particularly when the cost of arrangement is sufficiently low. On the other hand, costly punishment can enable one to deal with commitment free-riders, who can escape sanctioning when interacting with the commitment strategy. In addition, one important feature of costly punishment is that its efficiency always increases with the effect-to-cost ratio of punishment, which is not possessed by the commitment strategy, even when the cost of arrangement is low. Our results show that the combined strategy retains this important property of costly punishment. Furthermore, the improvement that can be achieved through the combined strategy, in terms of frequency (compared to the best of the two strategies separately), is most significant when the cost is sufficiently large and the impact of punishment reaches a threshold. 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 shown that the combined strategy can overcome the weaknesses of both strategies. Hence, as an implication, our results provide novel insights for the design of self-organised or distributed multi-agent and autonomous agent systems (Bonabeau et al. 1999) that require cooperation among agents in a competitive environment.

In addition, our results suggest that free-riders of various types might be better dealt with by simply varying the use of two complementary mechanisms that can efficiently deal with the same kind of strategic situations (i.e. the one-shot PD herein). It would be optimal to be able to identify what kind of free-riders one is dealing with and use the most appropriate mechanism, but it might be difficult to do so when the information available about the co-players is insufficient for decision making, which is particularly the case in the non-repeated interactions. By varying the use of the available mechanisms (in our case, costly punishment and commitment), we can suppress more types of free-riders. Related to this, we envisage that our combined strategy can be improved through using additional cognitive skills, e.g. intention recognition (Han et al. 2011, 2015c), to recognise the type of free-riders one is dealing with and then use the best mechanism to deal with that type.

Our model is closely related to the models in (Sigmund et al. 2010; Szolnoki et al. 2011a) where peer and pool punishments are studied when present as two separate strategies at the same time in either a well-mixed or structured (network) population. These models were designed to investigate which of the two sanctioning mechanisms is preferred when both of them are available to a society, a question which has also been studied experimentally in (Putterman et al. 2011; Traulsen et al.

2012). Although pool punishment and prior commitment are similar in the sense that both require prior arrangements for the sanctioning of defectors at a later stage, our work differs in that, commitment and peer punishment are combined into one single strategy, aiming to show that these two mechanisms can support each other in dealing with various types of free-riders. Nonetheless, as shown in SI (Section 6, Figure S4) where we analyse various types of CPP strategies in co-presence in the population (including CP and COMP as the extreme types of CPP, with $q = 0$ and 1, respectively), a similar observation regarding the emergence of a weighted combination of the two mechanisms is obtained. Namely, there is a wide range of parameters where CPP with an intermediate value of q is the most abundant in the population: neither CP nor COMP can be dominant when using separately. That said, it would be interesting to see in the context of the models in (Sigmund et al. 2010; Szolnoki et al. 2011a) whether a weighted combination of pool and peer punishment can outperform either of them when using separately.

Similarities can also be found with the model that analyses the probabilistic sharing of punishment duty in the context of the Public Goods Game (PGG) (Chen et al. 2014). In a PGG group interaction, punishers share the duty of punishing defectors. That is, when facing defectors, a punisher only punishes with a certain probability, and cooperates otherwise. A similar approach is conditional punishment, wherein punishment duty is proportional to the number of punishers in the group (Szolnoki and Perc 2013). Both approaches have been shown to be more efficient for promoting cooperation in structured populations than mere punishment as they enable costly punishers to defer defectors more efficiently, especially when punishment is expensive (i.e. having a low effect-to-cost ratio). We go beyond those results by showing that the synergy between costly punishment and commitment provides a more efficient solution than mere punishment even in well-mixed populations, wherein cooperation is harder to emerge (Santos et al. 2006). Moreover, the efficiency of the model in (Chen et al. 2014) is based on the fact that punishers can share punishment duty, which would clearly become less efficient, if not inapplicable, for a two-player game as in our setting. To the contrary, our model can be readily extended to group interactions typically found in the PGG (Han et al. 2015b).

Various extensions to the current model can be described. First, in this work we have not taken into account the fact that punishment might be antisocial, in which defectors can also punish cooperators. Antisocial punishment is widespread in nature (Herrmann et al. 2008), which has been shown to be detrimental for the emergence of cooperation (Rand and Nowak 2011; Hilbe and Traulsen 2012; Han 2016). To the contrary, arranging a prior

commitment ensures that this kind of antisocial behaviour does not occur, because only those who agreed to commit can be punished for misbehaving (Han et al. 2013a). Hence, it would be interesting to see whether our combined strategy can overcome this weakness of costly punishment when antisocial punishment is possible. Another setting where the combination of commitment and costly punishment might be of interest is that of the repeated interaction setting, where both commitment and costly punishment have been used in support of apology (Okamoto and Matsumura 2000; Han et al. 2013b; Martinez-Vaquero et al. 2015). Hence, it would be interesting to see whether and how the combined strategy can support apology more efficiently, leading to a better outcome for cooperation. Last but not least, as discussed above, the cost efficiency issue of costly punishment can be efficiently dealt with in structured populations (Brandt et al. 2003; Nakamaru and Iwasa 2005; Helbing et al. 2010a,b). Thus, it would be interesting to study how the synergistic effect of commitment and costly punishment would change in such a setting.

In short, our results 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 assemble a better combined solution that ensures a more favourable outcome for cooperative behaviour. By varying the use of the available mechanisms one can suppress more various types of free-riders, even without looking at which mechanism is best for the situation at hand.

Acknowledgements

T.A.H. acknowledges the support provided by Teesside URF funding (11200174). T.L. acknowledges the support provided by the F.R.S. - F.N.R.S Belgium (grant number 2.4614.12) and the F.W.O. Belgium (grant number G.0391.13N).

References

- Anderson, C. M. and Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? the demand for punishment in the voluntary contribution mechanism. *Games and Economic Behavior*, 54(1):1–24.
- Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, USA.
- Boyd, R., Gintis, H., and Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978):617–620.
- Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100(6):3531–3535.
- Brandt, H., Hauert, C., and Sigmund, K. (2003). Punishment and reputation in spatial public goods games. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1519):1099–1104.
- Carpenter, J. P. (2007). The demand for punishment. *Journal of Economic Behavior & Organization*, 62(4):522–542.
- Castelfranchi, C. and Falcone, R. (2010). *Trust Theory: A Socio-Cognitive and Computational Model (Wiley Series in Agent Technology)*. Wiley.
- Chen, X., Szolnoki, A., and Perc, M. (2014). Probabilistic sharing solves the problem of costly punishment. *New Journal of Physics*, 16(8):083016.
- 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(1):63–77.
- Coombs, C. H. (1973). A reparameterization of the prisoner's dilemma game. *Behavioral Science*, 18(6):424–428.
- Egas, M. and Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637):871–878.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.
- Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415:137–140.
- Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. W. W. Norton and Company.
- Fudenberg, D. and Imhof, L. A. (2005). Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–262.
- Gerber, A. and Wichardt, P. C. (2009). Providing public goods in the absence of strong institutions. *Journal of Public Economics*, 93(3):429–439.
- Guala, F. (2012). Reciprocity: Weak or strong? what punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(1):1.
- 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 Conference of the American Association of Artificial Intelligence (AAAI'2016)*, pages 2494–2500, Phoenix, Arizona, USA.
- Han, T. A., Lenaerts, T., Santos, F. C., and Pereira, L. M. (2015a). Emergence of cooperation via intention recognition, commitment and apology—a research summary. *AI*

Communications, pages 1–7.

- Han, T. A., Moniz Pereira, L., and Lenaerts, T. (2015b). Avoiding or Restricting Defectors in Public Goods Games? *Journal of the Royal Society Interface*, 12(103):20141203.
- Han, T. A., Pereira, L. M., and Santos, F. C. (2011). Intention recognition promotes the emergence of cooperation. *Adaptive Behavior*, 19(3):264–279.
- Han, T. A., Pereira, L. M., and Santos, F. C. (2012a). Corpus-based intention recognition in cooperation dilemmas. *Artificial Life journal*, 18(4):365–383.
- Han, T. A., Pereira, L. M., and Santos, F. C. (2012b). The emergence of commitments and cooperation. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'2012)*, pages 559–566. ACM.
- Han, T. A., Pereira, L. M., Santos, F. C., and Lenaerts, T. (2013a). Good agreements make good friends. *Scientific reports*, 3(2695).
- Han, T. A., Pereira, L. M., Santos, F. C., and Lenaerts, T. (2013b). Why Is It So Hard to Say Sorry: The Evolution of Apology with Commitments in the Iterated Prisoner's Dilemma. In *Proceedings of the 23rd international joint conference on Artificial intelligence (IJCAI'2013)*. AAAI Press.
- Han, T. A., Santos, F. C., Lenaerts, T., and Pereira, L. M. (2015c). Synergy between intention recognition and commitments in cooperation dilemmas. *Scientific reports*, 5(9312).
- Hardin, G. (1968). The tragedy of the commons. *Science*, 162:1243–1248.
- Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A., and Sigmund, K. (2007). Via freedom to coercion: The emergence of costly punishment. *Science*, 316:1905–1907.
- Helbing, D., Szolnoki, A., Perc, M., and Szabó, G. (2010a). Evolutionary establishment of moral and double moral standards through spatial interactions. *PLoS Comput Biol*, 6(4):e1000758.
- Helbing, D., Szolnoki, A., Perc, M., and Szabó, G. (2010b). Punish, but not too hard: how costly punishment spreads in the spatial public goods game. *New Journal of Physics*, 12(8):083005.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., et al. (2006). Costly punishment across human societies. *Science*, 312(5781):1767–1770.
- Herrmann, B., Thöni, C., and Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, 319(5868):1362–1367.
- Hilbe, C. and Traulsen, A. (2012). Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Scientific reports*, 2.
- Hilbe, C., Traulsen, A., Röhl, T., and Milinski, M. (2014). Democratic decisions establish stable authorities that overcome the paradox of second-order punishment. *Proceedings of the National Academy of Sciences*, 111(2):752–756.
- Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.
- Imhof, L. A., Fudenberg, D., and Nowak, M. A. (2005). Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences of the United States of America*, 102:10797–10800.
- Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61:29–56.
- Karlin, S. and Taylor, H. E. (1975). *A First Course in Stochastic Processes*. Academic Press, New York.
- Martinez-Vaquero, L. A., Han, T. A., Pereira, L. M., and Lenaerts, T. (2015). Apology and forgiveness evolve to resolve failures in cooperative agreements. *Scientific reports*, 5(10639).
- Maynard-Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.
- Miettinen, T. (2013). Promises and conventions—an approach to pre-play agreements. *Games and Economic Behavior*, 80:68–84.
- Nakamaru, M. and Iwasa, Y. (2005). The evolution of altruism by costly punishment in lattice-structured populations: score-dependent viability versus score-dependent fertility. *Evolutionary ecology research*, 7(6):853–870.
- Nesse, R. M. (2001). *Evolution and the capacity for commitment*. Russell Sage Foundation series on trust. Russell Sage.
- Nikiforakis, N. and Normann, H.-T. (2008). A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11(4):358–369.
- Nowak, M. A. (2006a). *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, Cambridge, MA.
- Nowak, M. A. (2006b). Five rules for the evolution of cooperation. *Science*, 314(5805):1560. DOI: 10.1126/science.1133755.
- Nowak, M. A., Sasaki, A., Taylor, C., and Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650.
- Okamoto, K. and Matsumura, S. (2000). The evolution of punishment and apology: an iterated prisoner's dilemma model. *Evolutionary Ecology*, 14(8):703–720.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.
- Perc, M., Gómez-Gardeñes, 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. and Szolnoki, A. (2010). Coevolutionary games? a mini review. *BioSystems*, 99(2):109–125.
- Perc, M. and Szolnoki, A. (2012). Self-organization of punishment in structured populations. *New Journal of Physics*, 14(4):043013.
- Putterman, L., Tyran, J.-R., and Kamei, K. (2011). Public goods and voting on formal sanction schemes. *Journal of Public*

- Economics*, 95(9):1213–1222.
- Rand, D. G., Armao IV, J. J., Nakamaru, M., and Ohtsuki, H. (2010). Anti-social punishment can prevent the co-evolution of punishment and cooperation. *Journal of Theoretical Biology*, 265(4):624 – 632.
- Rand, D. G. and Nowak, M. A. (2011). The evolution of antisocial punishment in optional public goods games. *Nature Communications*, 2:434.
- Santos, F. C., Pacheco, J. M., and Lenaerts, T. (2006). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences of the United States of America*, 103:3490–3494.
- Sasaki, T., Brännström, Å., Dieckmann, U., and Sigmund, K. (2012). The take-it-or-leave-it option allows small penalties to overcome social dilemmas. *Proceedings of the National Academy of Sciences*, 109(4):1165–1169.
- Sasaki, T., Okada, I., Uchida, S., and Chen, X. (2015). Commitment to cooperation and peer punishment: Its evolution. *Games*, 6(4):574–587.
- Schelling, T. C. (1990). *The strategy of conflict*. London: Oxford University Press.
- Schoenmakers, S., Hilbe, C., Blasius, B., and Traulsen, A. (2014). Sanctions as honest signals the evolution of pool punishment by public sanctioning institutions. *Journal of Theoretical Biology*, 356:36 – 46.
- Sigmund, K. (2010). *The Calculus of Selfishness*. Princeton University Press.
- Sigmund, K., Hauert, C., and Nowak, M. (2001). Reward and punishment. *P Natl Acad Sci USA*, 98(19):10757–10762.
- Sigmund, K., Silva, H. D., Traulsen, A., and Hauert, C. (2010). Social learning promotes institutions for governing the commons. *Nature*, 466:7308.
- Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge University Press.
- Sterelny, K. (2012). *The evolved apprentice*. MIT Press.
- Szabó, G. and Töke, C. (1998). Evolutionary prisoner’s dilemma game on a square lattice. *Physical Review E*, 58(1):69.
- Szolnoki, A. and Perc, M. (2013). Effectiveness of conditional punishment for the evolution of public cooperation. *Journal of theoretical biology*, 325:34–41.
- Szolnoki, A., Szabó, G., and Czakó, L. (2011a). Competition of individual and institutional punishments in spatial public goods games. *Physical Review E*, 84(4):046106.
- Szolnoki, A., Szabó, G., and Perc, M. (2011b). Phase diagrams for the spatial public goods game with pool punishment. *Physical Review E*, 83(3):036101.
- Traulsen, A., Nowak, M. A., and Pacheco, J. M. (2006). Stochastic dynamics of invasion and fixation. *Phys. Rev. E*, 74:11909.
- Traulsen, A., Röhl, T., and Milinski, M. (2012). An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B: Biological Sciences*, 279(1743):3716–3721.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46:35–57.
- West, S., Griffin, A., and Gardner, A. (2007). Evolutionary explanations for cooperation. *Current Biology*, 17:R661–R672.
- Winikoff, M. (2007). Implementing commitment-based interactions. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, AAMAS ’07*, pages 868–875, New York, NY, USA. ACM.
- Wooldridge, M. and Jennings, N. R. (1999). The cooperative problem-solving process. In *Journal of Logic and Computation*, pages 403–417.
- Wu, J. J., Zhang, B. Y., Zhou, Z. X., He, Q. Q., Zheng, X. D., Cressman, R., and Tao, Y. (2009). Costly punishment does not always increase cooperation. *Proc Natl Acad Sci U S A*, 106(41):17448–51.