Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Emergence of Social Punishment and Cooperation through Prior Commitments

The Anh Han
School of Computing and Digital Futures Institute
Teesside University, UK

Abstract

Social punishment, whereby cooperators punish defectors, has been suggested as an important mechanism that promotes the emergence of cooperation or maintenance of social norms in the context of the one-shot (i.e. non-repeated) interaction. However, whenever antisocial punishment, whereby defectors punish cooperators, is available, this antisocial behavior outperforms social punishment, leading to the destruction of cooperation. In this paper, we use evolutionary game theory to show that this antisocial behavior can be efficiently restrained by relying on prior commitments, wherein agents can arrange, prior to an interaction, agreements regarding posterior compensation by those who dishonor the agreements. We show that, although the commitment mechanism by itself can guarantee a notable level of cooperation, a significantly higher level is achieved when both mechanisms, those of proposing prior commitments and of punishment, are available in co-presence. Interestingly, social punishment prevails and dominates in this system as it can take advantage of the commitment mechanism to cope with antisocial behaviors. That is, establishment of a commitment system helps to pave the way for the evolution of social punishment and abundant cooperation, even in the presence of antisocial punishment.

Introduction

Punishment has been suggested as one of the most important instruments or mechanisms for enforcing cooperation and compliance of social norms in various societies (Fehr and Gachter 2002; Herrmann, Thöni, and Gächter 2008; Powers, Taylor, and Bryson 2012; Airiau, Sen, and Villatoro 2014; Sigmund, Hauert, and Nowak 2001; Boyd et al. 2003; Sigmund et al. 2010). Numerous empirical studies show that human subjects are eager to incur a cost to punish unjust behaviors and violation of social norms (Fehr and Gachter 2002; Herrmann, Thöni, and Gächter 2008; De Quervain et al. 2004). While sanctioning systems have been widely implemented in modern societies for law enforcement, many social norms are still enforced through private sanctions (Fehr and Gachter 2002). Punishment has also been implemented in several computerized systems such as virtual online agent societies (Savarimuthu et al. 2009) and e-marketplace (Michalak et al. 2009), in order to enhance cooperative behavior and norms compliance, e.g. by both customers and sellers (Michalak et al. 2009).

Given the abundance of the social punishment behavior (i.e. punishment targeting at wrongdoers or defectors), the puzzle is, how this behavior could evolve as it is costly to punish others, hence unlikely to evolve unless it results in direct or indirect benefits. Various studies have argued that this kind of punishment can evolve if it is cost-effective, i.e. the punished agent suffers a sufficiently higher cost compared to that of the punisher (Sigmund, Hauert, and Nowak 2001; Boyd et al. 2003; Sigmund et al. 2010). However, it appears that this argument is only valid if antisocial punishment (i.e. defectors punish cooperators) is not allowed or unavailable, as shown in recent work (Rand et al. 2010; Rand and Nowak 2011). In fact, as defined, an antisocial punisher can take advantage of the cost-efficiency effect of punishment even after its wrongdoing, by punishing any cooperative co-player. As such, antisocial punishment reduces even further the chance that cooperation can survive. Moreover, empirical studies show that antisocial punishment does exist across cultures, see e.g. (Herrmann, Thöni, and Gächter 2008). Thus, a more subtle question arises: whether and how social punishment could evolve in different societies under the presence of antisocial behaviors?

A major approach to addressing this query has been based on reputation (for a review, see (Raihani and Bshary 2015)). The key idea is, if agents' actions are assumed or made publicly available, social punishers can gain indirect benefits through maintaining a good reputation of punishing wrongdoers and unjust behavior (Hilbe and Traulsen 2012) (more discussion on this in Related Work).

In this paper, we propose and analyze a novel approach to resolving this antisocial behavior issue, without relying on reputation and the assumption that agents' actions are non-anonymous, which is not realistic in many contexts or application domains (Nowak 2006), e.g. in populations or systems with a large number of agents, or when activities observation is not easy. We use methods from evolutionary game theory (Hofbauer and Sigmund 1998; Sigmund 2010) (see Models and Methods section) to show that antisocial punishment can be significantly restrained by relying on prior commitments (Nesse 2001; Han et al. 2013; Han, Pereira, and Lenaerts 2015), wherein agents can arrange, prior to an interaction, agreements regarding posterior compensation by those who dishonor the agreements. This commitment proposing mechanism has been studied widely in Artificial Intelligence (AI) and Multi-agent System (MAS) literature (see e.g. (Castelfranchi and Falcone 2010; Winikoff 2007; Hasan and Raja 2013)), and has been shown to promote the evolution of cooperation (Han, Pereira, and Santos 2012; Han 2013; Han et al. 2013; Martinez-Vaquero et al. 2015; Han et al. 2015). By proposing a commitment deal regarding posterior punishment (prior to an interaction) one can prevent antisocial behavior as only those who agree with the deal (and its terms) then default on it can be punished (detailed description in Models and Methods). Most interestingly, as we shall show in Results section, the presence of the commitment mechanism helps to pave the way for the prevalence of social punishment and cooperation, rather than the commitment proposing agents themselves.

As the underlying agent interaction framework, we adopt here the one-shot Prisoner's Dilemma (PD) (Sigmund 2010), one of the most popular frameworks to study the evolution of cooperation amongst self-regarding agents. This game also stands for the most difficult pairwise social dilemma for cooperation to emerge (Hofbauer and Sigmund 1998). In this game, two players simultaneously decide to either cooperate (C) or defect (D). If both play C, they get more than if both play D, but if one defects and the other cooperates, the defector gets the highest payoff and the cooperator gets the lowest (further details on the PD in Models and Methods section). That is, rational choice determines that it is better for each player to defect, even though both would be better off cooperating. Consequently, evolutionary game dynamics predicts that under those conditions cooperation disappears (Nowak 2006; Sigmund 2010).

Related Work

Punishment has been a major explanation for the evolution of cooperation in the context of the one-shot interaction (Fehr and Gachter 2002; De Quervain et al. 2004; Sigmund et al. 2010; Powers, Taylor, and Bryson 2012) (for other explanations, see a survey in (Nowak 2006)). Several theoretical models (Sigmund, Hauert, and Nowak 2001; Boyd et al. 2003; Hauert et al. 2007) have been analyzed, showing that social punishment can promote the emergence of social punishment and cooperation. However, these early works do not integrate the possibility of antisocial punishment in their models, assuming that only cooperators can punish defectors. Only when experimental evidence, from numerous countries, e.g. in (Herrmann, Thöni, and Gächter 2008), pointed out that antisocial punishment does exist, this aspect has been paid more attention. In the models of (Rand et al. 2010; Rand and Nowak 2011) wherein antisocial punishment is integrated, the authors show that the evolution of social punishment, hence also that of cooperation, is prohibited. Yet these works do not show how to overcome the antisocial punishment issue, i.e. how social punishment could evolve in the presence of antisocial punishers. Going beyond these previous works, this paper presents a solution to this issue, showing that when accompanied by the possibility to arrange prior commitments, social punishment and cooperation can thrive even in the presence of antisocial punishment.

Furthermore, it is noteworthy that commitment, by itself, provides a pathway for the evolution of cooperation in the one-shot PD (Han, Pereira, and Santos 2012; Nesse 2001; Hasan and Raja 2013; Han et al. 2013; 2015; Sasaki et al. 2015). This mechanism, however, has never been studied in co-presence with antisocial punishment, even when commitment and social punishment have been analyzed together (Han and Lenaerts 2015). Additionally, this latter work aims to analyze how to integrate the two mechanisms into a single one in order to better deal with commitment free-riders and defectors; while the aim of the current work is to find a new solution for the antisocial punishment issue, thereby providing an explanation for the conundrum of the evolution of social punishment. Nevertheless, it would be interesting to study whether the integrated strategy described in (Han and Lenaerts 2015) can help deal even better with the antisocial punishment issue. We intend to study this in future work.

A major approach to addressing the antisocial punishment issue is to rely on reputation (Raihani and Bshary 2015), wherein agents' actions are assumed to be non-anonymous to others in a population. As such, social punishers can gain indirect benefits through maintaining a good reputation of punishing wrongdoers and unjust behavior. It is because being identified as a social punisher reduces the risk of being cheated badly (Hilbe and Traulsen 2012). Differently from this reputation-based approach, our commitment based solution does not generally require the assumption of non-anonymity, which is not reasonable in many contexts; for instance, it is difficult to track others' activities in a large MAS or society of agents. In fact, commitments can be enforced through several different means, such as legal contracts, pledges, emotion and also the reputation device itself (Nesse 2001; Han, Pereira, and Santos 2012; Han, Pereira, and Lenaerts 2015; Han et al. 2015). As such, it appears that our approach is applicable to more diverse situations, and is more easily facilitated and deployed, than the reputation-based one—especially given that contractual commitments are generally enforceable in modern societies.

Finally, it is important to note that both punishment and commitment have been studied extensively in AI and MAS literature (Airiau, Sen, and Villatoro 2014; Wooldridge and Jennings 1999; Castelfranchi and Falcone 2010; Winikoff 2007; Harrenstein, Brandt, and Fischer 2007; Savarimuthu et al. 2009; Chopra and Singh 2009; Savarimuthu and Cranefield 2011). Differently from our work, these studies aim at using the cooperation enforcing power of the two mechanisms for the purpose of regulating individual and collective behaviors, formalizing different relevant aspects of these mechanisms (such as norms and conventions) in a MAS. Moreover, to our knowledge, no work exists in this literature that studies the two mechanisms in co-presence (for such regulating purposes). As we shall show later, the co-presence of the two mechanisms in a system leads to a significantly higher level of cooperation than what can be achieved with either mechanism solely. Therefore, our results and approach provide insights into the design of computerized and MAS systems that can, on the one hand, prevent antisocial behaviors, and on the other hand, maximize the benefit of an appropriate sanctioning system; for instance, together with sanctioning, one should also facilitate arrangement and enforcement of prior commitments.

Models and Methods

We first recall the Prisoner's Dilemma game and its extensions integrating the option of using costly punishment or arranging prior commitments. We then describe the main model where both strategic options are included.

Prisoner's Dilemma (PD)

To begin with, the (one-shot) PD can be described with the following payoff matrix:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R,R & S,T \\ T,S & P,P \end{pmatrix} \end{array} \tag{1}$$

Once the interaction is established and both players have decided to play C or 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 defecting one. The payoff matrix corresponds to the preferences associated with the PD when the parameters satisfy the ordering, $T > R > P > S$ (Sigmund 2010). For the sake of a simple representation, we sometimes use the Donor game (Sigmund 2010), a special case of the PD, with $T = b,\ R = b - c,\ P = 0,\ S = -c$, where b and c stand for the benefit and cost of cooperation, respectively.

PD with Punishment

We extend the PD with the option of costly punishment. After the PD has taken place, a player can choose to punish her opponent, which consists in paying a cost ϵ_1 to make the opponent incur a cost δ_1. As usual (Rand et al. 2010; Boyd et al. 2003), we assume that $\epsilon_1 < \delta_1$.

We now define the social and antisocial punishment strategies, denoted by CP and AP, respectively. CP cooperates in the PD, and punishes a co-player who defected in the game. In contrast to CP, AP defects in the PD, and punishes a co-player who cooperated in the game. These two strategies, together with the two traditional unconditional strategies, pure cooperator (C) and pure defector (D) (i.e. they do not use the punishment option), forms a minimal model that allows us to analyze evolutionary dynamics and viability of punishment strategies and the induced cooperation level (Rand et al. 2010). The 4×4 payoff matrix (for row player) is given by

$$\begin{array}{cc} & \begin{array}{cccc} \text{CP} & \text{C} & \text{D} & \text{AP} \end{array} \\ \begin{array}{c} \text{CP} \\ C \\ D \\ \text{AP} \end{array} & \begin{pmatrix} R & R & S-\epsilon_1 & S-\delta_1-\epsilon_1 \\ R & R & S & S-\delta_1 \\ T-\delta_1 & T & P & P \\ T-\delta_1-\epsilon_1 & T-\epsilon_1 & P & P \end{pmatrix} \end{array} \tag{2}$$

PD with Commitment

We now recall the commitment variant of the PD game, as described in (Han et al. 2013). Before playing the PD, a commitment strategy (denoted by COM), proposes to her co-player to commit to the game and cooperate. To make the commitment deal reliable, the proposer pays an arrangement cost ϵ_2. If the co-player agrees with the deal, then COM 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 co-player at a personal cost δ_2. The PD interaction does not occur if the co-player does not accept the commitment deal. As in (Han et al. 2013), we consider a minimal model with five (basic) strategies in this commitment variant of the PD game:

(i) Commitment proposers (COM), who proposes commitment before the PD, cooperates if the game occurs;

(ii) Unconditional cooperators (C), who always commit when being proposed a commitment deal, cooperate whenever the PD is played, but do not propose commitment themselves;

(iii) Unconditional defectors (D), who do not accept commitment, defect when the PD takes place, and do not propose commitment;

(iv) 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;

(v) 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.

The rationale for considering this restricted set of strategies is that the other possible strategies are dominated by at least one of these five strategies in any configuration of the game: they can be omitted without changing the outcome of the analysis (Han et al. 2013).

The 5×5 payoff matrix (for row player) is given by

$$\begin{array}{cc} & \begin{array}{ccccc} \text{COM} & \text{C} & \text{D} & \text{FAKE} & \text{FREE} \end{array} \\ \begin{array}{c} \text{COM} \\ C \\ D \\ \text{FAKE} \\ \text{FREE} \end{array} & \begin{pmatrix} R-\frac{\epsilon_2}{2} & R-\epsilon_2 & 0 & S+\delta_2-\epsilon_2 & R-\epsilon_2 \\ R & R & S & S & S \\ 0 & T & P & P & P \\ T-\delta_2 & T & P & P & P \\ R & T & P & P & P \end{pmatrix} \end{array} \tag{3}$$

PD with Punishment and Commitment

We now extend the commitment variant to integrate the costly punishment strategic behavior. First, when a commitment deal is set up, the terms of the commitment regarding posterior compensation is called upon, if necessary, after the PD occurs. That is, punishment is not utilized in this case [1].

Next, if no commitment deal is set up before the PD interaction—for instance when neither of the players in the

[1] Punishment could occur, e.g. after the terms of commitment have been carried out as conflict might occur in this process. This is indeed an interesting issue but it is beyond the scope of this work.

PD is a commitment proposer (COM), or when a commitment proposal is rejected—costly punishment can be used after the PD has occurred. By definition of the punishment strategies, a C player will adopt social punishment (i.e. CP), while D, FREE and FAKE will adopt antisocial punishment (i.e. AP). Abusing notation, they are denoted by CP, AP, FREE-AP and FAKE-AP, respectively.

For a clear and fair comparison of the punishment and commitment strategies as well as the induced cooperation levels in different models, we consider, next to the five strategies obtained via the extension (i.e. COM, CP, AP, FREE-AP and FAKE-AP), also the unconditional cooperator (C) and unconditional defector (D) strategies (which do not use commitment and punishment). That is, for the current PD variant with both strategic options, we will analyze a minimal model of seven strategies: COM, C, D, CP, AP, FREE-AP, and FAKE-AP. Their payoff matrix can be directly derived from matrices (2) and (3), substituting for each pair of strategies the corresponding payoff outcomes.

Note that some strategies can be (and are) omitted in this model, including FREE and FAKE that do not use antisocial punishment, because they are dominated by FREE-AP and FAKE-AP, respectively (since $\epsilon_1 < \delta_1$).

Population Setup and Evolutionary Dynamics

All the analysis and numerical results in this paper (see next section) are obtained using evolutionary game theory (EGT) methods for finite populations (Nowak et al. 2004; Imhof, Fudenberg, and Nowak 2005). 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, Nowak, and Pacheco 2006), a standard approach in EGT, 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, $p_{A,B} = \left(1 + e^{-\beta(f_B - f_A)}\right)^{-1}$. 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 behavioral 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 (Imhof, Fudenberg, and Nowak 2005; Nowak et al. 2004). The resulting Markov Chain has a stationary distribution, which characterizes the average time the population spends in each of these monomorphic end states (for illustration, see already examples in Figure 1).

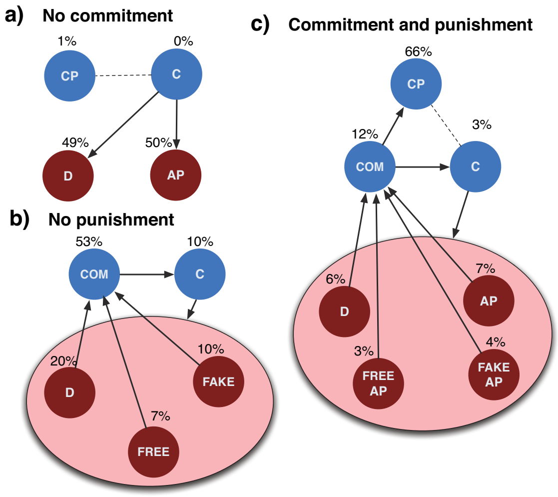

Figure 1: **Transitions and stationary distributions;** a) Antisocial punishment and defection dominate when proposing commitment is not an option; b) Commitment dominates when no punishment option is available; c) When both punishment and commitment are present, CP dominates and a higher level of cooperation is achieved. COM catalyzes for the dominance of CP and cooperation. For clarity, only transitions that are stronger than neural are shown (arrow). Dashed lines stand for neutral transitions. Blue and red circles represent cooperative and defective strategies, respectively. Parameters: $T = 4, R = 3, P = 0, S = 1$; $\epsilon_1 = \epsilon_2 = 1$, $\delta_1 = \delta_2 = 3$; $\beta = 0.1$; population size $N = 100$.

Let N be the size of the population. Denote $\pi_{X,Y}$ the payoff a strategist X obtains in a pairwise interaction with strategist Y (defined in the payoff matrices). Suppose there are at most two strategies in the population, say, k agents using strategy A ($0 \le k \le N$) and $(N-k)$ agents using strategies B. Thus, the (average) payoff of the agent that uses A and B can be written as follows, respectively,

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

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, Nowak, and Pacheco 2006)

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

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

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

Considering a set $\{1, ..., q\}$ of different strategies, these fixation probabilities determine a transition matrix $M = \{T_{ij}\}_{i,j=1}^{q}$, with $T_{ij,j\neq i} = \rho_{ji}/(q-1)$ and $T_{ii} = 1 - \sum_{j=1,j\neq i}^{q} T_{ij}$, of a Markov Chain. The normalized eigenvector associated with the eigenvalue 1 of the transposed of M provides the stationary distribution described above (Imhof, Fudenberg, and Nowak 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 (Nowak et al. 2004; Sigmund 2010)

$$\pi_{A,A} + \pi_{A,B} > \pi_{B,A} + \pi_{B,B}. \tag{7}$$

Results

We study the performance of punishment and commitment strategies and the induced cooperation levels in three aforementioned scenarios: no commitment (CP, C, D, AP), no punishment (COM, C, D, FREE, FAKE) and when both are available (COM, C, D, CP, AP, FREE-AP, FAKE-AP).

Stationary Distributions

To begin with, we compute stationary distributions and transition probabilities in each of the three scenarios. First, in the system where proposing commitment is not an option (Figure 1a), antisocial punishment (AP) and unconditional defector (D) dominate the population. Although there is no above-neutral transition from CP to those strategies, its neutral transitions to C, which is in turn dominated by D and AP, leads to the dominance of these defective strategies. Note that, using the inequality (7) and the property of the PD ($T > R > P > S$) we can show that AP is risk-dominant against both C and CP [2], and C is dominated by D, showing that the transitions in Figure 1a hold not only for the specific game configuration being considered.

Next, in the system where punishment is not an option (Figure 1b), defective strategies (i.e. D, FREE and FAKE) are dominated by COM, which is taken away by C, then back to defective strategies, creating a cycle among cooperative and defective strategies . This cycle has been shown to hold in general for sufficiently strong commitment deals (i.e. when δ_2 is roughly larger than the cost of arranging commitment ϵ_2 plus the cost of cooperation c) (Han et al. 2013). This cycle results in a high fraction of COM but defectors are not vanished (the total fraction of the three defective strategies in Figure 1b is 37%).

We now consider the system where both strategic behaviors are available (Figure 1c). Interestingly, the cycle observed in Figure 1b (i.e. from defective strategies D, AP, FREE-AP, FAKE-AP to COM, to C, and back to the defective strategies) is broken, when the system reaches CP state. On the one hand, as long as arranging a prior commitment is costly, the regime of COM is taken down by CP mutants that can establish cooperation among themselves without having to pay any cost. Note that when COM interacts with CP, COM pays the cost of arranging commitment (ϵ_2). Using the inequality (7), we can easily show that CP is risk-dominant against COM as long as $\epsilon_2 > 0$, confirming that the transition from COM to CP is not particular to the game configuration being considered. On the other hand, as COM requires a prior agreement regarding posterior punishment/compensation, antisocial punishment can be prevented (see transitions from D, AP, FREE-AP, FAKE-AP to COM). As such, the level of cooperation is significantly increased (the total fraction of cooperation is 81%). More interestingly, COM is not the most frequent strategy as in the second scenario (Figure 1b): instead, it catalyzes the emergence and dominance of CP, hence cooperation.

[2]They are equivalent to, $T + P > R + S - \delta_1$ and $T + P > R + S$, respectively. Both clearly hold.

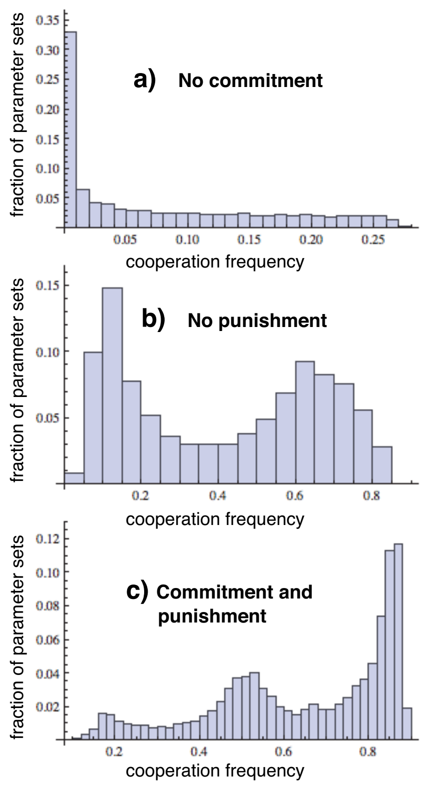

Figure 2: **Robustness of the results across game configurations and parameters of punishment and commitment**. Shown are the results of 10000 numerical computations using $T = b, R = b - 1, P = 0, S = -1$ (Donor game, with $c = 1$) and randomly sampling from uniform distributions on the intervals: $2 \leq b \leq 6$; $0 \leq \epsilon_1, \epsilon_2 \leq 3$; $\epsilon_1 \leq \delta_1 \leq 10\epsilon_1$; $\epsilon_2 \leq \delta_2 \leq 10\epsilon_2$. Other parameters: $N = 100, \beta = 0.1$.

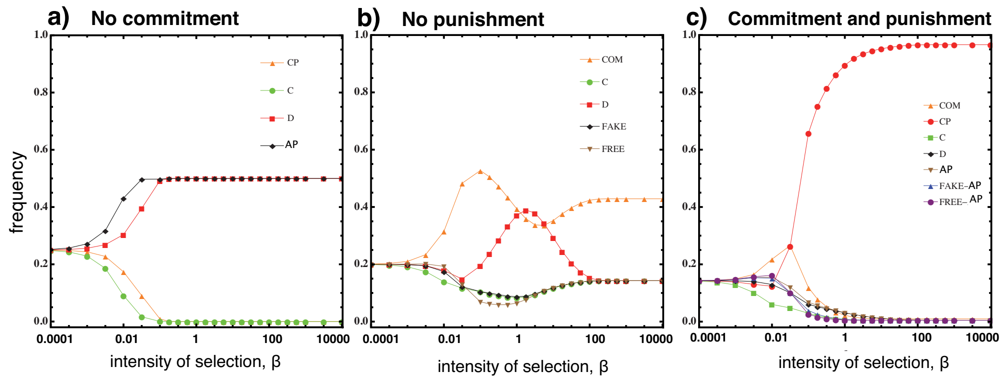

Figure 3: **Stationary distributions for varying intensity of selection;** a) Antisocial punishment and defection are dominant when no commitment is present; b) Although commitment performs best, all strategies are present in population; c) Costly punishment and cooperation are prevalent for increasing intensity of selection when both strategic options are available. Parameters: $T = 4, R = 3, P = 0, S = 1$; $\epsilon_1 = \epsilon_2 = 1$, $\delta_1 = \delta_2 = 3$; $N = 100$.

Robustness to Parameters Change

These interesting observations and results are not particular to the parameter setting in Figure 1. We compute stationary distributions for 10000 randomly drawn parameter sets (see Figure 2). The average result is very similar to what was observed in Figure 1. When commitment is not an option (Figure 2a), the cooperation frequency is very low (8% on average), with no sample having more than 50% of cooperation. When punishment is not available (Figure 2b), cooperation is more frequent (41% on average), but defection is also prevalent (59% on average). In 45% of the samples there is more than 50% of cooperation. Finally, when both options are present (Figure 2c), a significantly higher level of cooperation is achieved (65% on average), with 75% of the samples having more than 50% of cooperation.

Varying the Intensity of Selection

In the sampling described above, we fixed the intensity of selection β since unlike the sampled parameters which characterize the strategic nature of the game, it is an external, environment-wide parameter. As this parameter is an important factor in determining population dynamics and stationary distributions (Nowak et al. 2004; Imhof, Fudenberg, and Nowak 2005), we provide a numerical analysis for varying β (Figure 3).

First, when commitment is not an option, the antisocial punishment and pure defector strategies are dominant for sufficiently large β (Figure 3a). When punishment is not an option, COM is most frequent most of the time, but the other defective strategies are still present in the population (Figure 3b). However, when both commitment and punishment are available, CP is dominant especially for large β (Figure 3c). Although COM is not significant in this scenario, it has paved the way for the emergence of social punishment CP.

In short, our results show that, on the one hand, social punishment cannot survive by itself when antisocial punishment is allowed. On the other hand, the prior commitment proposing mechanism can prevent antisocial punishing behaviors but the unavoidable cost needed to sustain the system leads to a notable presence of defectors. Happily, if both mechanisms are available, in co-presence, social punishment thrives leading to significantly enhanced levels of cooperation.

Concluding Remarks

This paper presents a novel approach to coping with the antisocial punishment issue, providing an explanation for the conundrum of the evolution of social punishment and cooperation. We showed, in the context of the one-shot PD, that if in addition to using punishment, agents in a population can propose cooperation agreements to their co-players prior to an interaction, social punishment and cooperation can thrive even in the presence of antisocial punishment. First of all, antisocial punishers can be significantly restrained by commitment proposing agents since only those who dishonor a commitment deal can be enforced to pay compensation. Moreover, as arranging a commitment deal is costly, its regime can be replaced by social punishers who do not have to pay this cost while still can maintain cooperation among themselves. Indeed, our results showed that when both strategic options are available, social punishment dominates its population (which contains antisocial punishment players). As a consequence, a significantly higher level of cooperation is achieved compared to the case when either of the strategic options is absent. This is a rather surprising outcome since arranging prior commitments, by itself, is already a strong mechanism that can enforce a notable level of cooperation. But by sacrificing for a strategy that is vulnerable to antisocial behaviors and defection, it results in significant improvements in terms of cooperation. That is, the commitment mechanism catalyzes for the emergence of social punishment and cooperation.

These observations and results suggest new insights into the design of computerized and MAS systems that can prevent antisocial behaviors while at the same time, maximize the benefit of deploying an appropriate sanctioning system.

References

Airiau, S.; Sen, S.; and Villatoro, D. 2014. Emergence of conventions through social learning. *Autonomous Agents and Multi-Agent Systems* 28(5):779–804.

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.

Castelfranchi, C., and Falcone, R. 2010. *Trust Theory: A Socio-Cognitive and Computational Model (Wiley Series in Agent Technology)*. Wiley.

Chopra, A. K., and Singh, M. P. 2009. Multiagent commitment alignment. In *AAMAS'2009*, 937–944.

De Quervain, D. J.; Fischbacher, U.; Treyer, V.; Schellhammer, M.; et al. 2004. The neural basis of altruistic punishment. *Science* 305(5688):1254.

Fehr, E., and Gachter, S. 2002. Altruistic punishment in humans. *Nature* 415:137–140.

Han, T. A., and Lenaerts, T. 2015. The efficient interaction of costly punishment and commitment. In *AAMAS'2015*, 1657–1658.

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

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

Han, T. A.; Pereira, L. M.; and Lenaerts, T. 2015. 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. 2012. The emergence of commitments and cooperation. In *AAMAS'2012*, 559–566.

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.

Harrenstein, P.; Brandt, F.; and Fischer, F. 2007. Commitment and extortion. In *AAMAS '07*, 1–8.

Hasan, M. R., and Raja, A. 2013. Emergence of cooperation using commitments and complex network dynamics. In *IEEE/WIC/ACM Intl Joint Conf. on Web Intelligence and Intelligent Agent Technologies*, volume 2, 345–352.

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.

Herrmann, B.; Thöni, C.; and Gächter, S. 2008. Antisocial Punishment Across Societies. *Science* 319:1362–1367.

Hilbe, C., and Traulsen, A. 2012. Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Scientific reports* 2(458).

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* 102:10797–10800.

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).

Michalak, T.; Tyrowicz, J.; McBurney, P.; and Wooldridge, M. 2009. Exogenous coalition formation in the e-marketplace based on geographical proximity. *Electronic Commerce Research and Applications* 8(4):203–223.

Nesse, R. M. 2001. *Evolution and the capacity for commitment*. Series on trust. Russell Sage.

Nowak, M. A.; Sasaki, A.; Taylor, C.; and Fudenberg, D. 2004. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428:646–650.

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

Powers, S. T.; Taylor, D. J.; and Bryson, J. J. 2012. Punishment can promote defection in group-structured populations. *Journal of theoretical biology* 311:107–116.

Raihani, N. J., and Bshary, R. 2015. The reputation of punishers. *Trends in ecology & evolution* 30(2):98–103.

Rand, D. G., and Nowak, M. A. 2011. The evolution of antisocial punishment in optional public goods games. *Nature Communications* 2:434.

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.

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

Savarimuthu, B. T. R., and Cranefield, S. 2011. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems* 7(1):21–54.

Savarimuthu, B. T. R.; Purvis, M.; Purvis, M.; and Cranefield, S. 2009. Social norm emergence in virtual agent societies. In *Declarative Agent Languages and Technologies VI*. Springer. 18–28.

Sigmund, K.; Silva, H. D.; Traulsen, A.; and Hauert, C. 2010. Social learning promotes institutions for governing the commons. *Nature* 466:7308.

Sigmund, K.; Hauert, C.; and Nowak, M. 2001. Reward and punishment. *Proceedings of the National Academy of Sciences* 98(19):10757–10762.

Sigmund, K. 2010. *The Calculus of Selfishness*. Princeton University Press.

Traulsen, A.; Nowak, M. A.; and Pacheco, J. M. 2006. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* 74:11909.

Winikoff, M. 2007. Implementing commitment-based interactions. In *AAMAS '07*, 868–875.

Wooldridge, M., and Jennings, N. R. 1999. The cooperative problem-solving process. In *Journal of Logic and Computation*, 403–417.