First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation

Xiaojie Chen1,2,†, Tatsuya Sasaki1,3,†, Åke Brännström1,4 and Ulf Dieckmann1

1Evolution and Ecology Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg 2361, Austria
2School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, People’s Republic of China
3Faculty of Mathematics, University of Vienna, Vienna 1090, Austria
4Department of Mathematics and Mathematical Statistics, Umeå University, Umeå 90187, Sweden

TS, 0000-0002-4635-1389; UD, 0000-0001-7089-0393

Social institutions often use rewards and penalties to promote cooperation. Providing incentives tends to be costly, so it is important to find effective and efficient policies for the combined use of rewards and penalties. Most studies of cooperation, however, have addressed rewarding and punishing in isolation and have focused on peer-to-peer sanctioning as opposed to institutional sanctioning. Here, we demonstrate that an institutional sanctioning policy we call ‘first carrot, then stick’ is unexpectedly successful in promoting cooperation. The policy switches the incentive from rewarding to punishing when the frequency of cooperators exceeds a threshold. We find that this policy establishes and recovers full cooperation at lower cost and under a wider range of conditions than either rewards or penalties alone, in both well-mixed and spatial populations. In particular, the spatial dynamics of cooperation make it evident how punishment acts as a ‘booster stage’ that capitalizes on and amplifies the pro-social effects of rewarding. Together, our results show that the adaptive hybridization of incentives offers the ‘best of both worlds’ by combining the effectiveness of rewarding in establishing cooperation with the effectiveness of punishing in recovering it, thereby providing a surprisingly inexpensive and widely applicable method of promoting cooperation.

1. Introduction

Cooperation is desirable whenever groups of cooperating individuals can reap higher benefits than groups of individuals acting out of individual self-interest. Promoting cooperation can be difficult, however, because a single non-cooperating individual (‘defector’) in a group of cooperators often achieves a higher net benefit by free-riding on the others’ contributions. An efficient policy for promoting cooperation needs to overcome two fundamental challenges: to ensure that cooperators can gain a foothold in a population of defectors and to protect a population of cooperators from exploitation by defectors once cooperation has been established.

Incentives can help overcome these challenges [1–3]. The promise of reward or the threat of punishment can induce cooperation among self-interested individuals who would otherwise prefer actions that undermine the public good. At first glance, there might seem to be little difference between a reward and a penalty: after all, cooperation is induced whenever the size of the incentive exceeds the pay-off difference between a cooperator and a defector, regardless of whether the incentive is positive or negative [4]. This equivalence ceases to hold, however, when one considers the costs of implementing an incentive...
scheme. Rewarding a large number of cooperators or penalizing a large number of defectors is either very costly or becomes ineffective when a limited budget for incentives is stretched out too far. Pamela Oliver exemplifies this with the problem of fundraising: ‘If only 5% of the population needs to contribute to an Arts Fund for it to be successful, they can be rewarded by having their names printed in a program: it would be silly and wasteful to try to punish the 95% who did not contribute’ [5, p. 125]. While the challenges of implementing positive and negative incentives are separately well known [2,3] and weight has traditionally been given to peer-to-peer punishment [6–8], no study to date has established how such incentives should best be combined at an institutional level to promote cooperation.

Here, we demonstrate how an institution implementing incentives can effectively establish and recover cooperation at low cost. Institutional sanctioning is widespread [1,4,9–22]; however, surprisingly few theoretical studies have thus far considered the effects of institutionalized incentives on the evolution of cooperation, and the few studies that exist have either considered rewarding and punishing in isolation [4,9,19] or did not consider how optimal incentives change with the frequency of cooperators [10–12]. Indeed, sanctioning agents, such as officers and managers, often alter rewards and penalties as events unfold. We address this question in an established game-theoretical framework for studying the evolution of cooperation under institutionalized incentives [4,19]. By considering the strengths of positive and negative incentives as independent variables, we can encompass a range of hybrid incentive policies. In particular, by allowing the relative allocation of incentives to rewarding and punishing to vary with the frequency of cooperators, our framework includes hybrid incentive policies controlled by adaptive feedback from the population’s current state.

2. Model and methods

2.1. Institutional incentives

We aim to determine the best way to allocate a budget available to an institution for promoting cooperation through positive and negative incentives. As criteria for assessing the performance of alternative sanctioning policies, we consider their effectiveness and efficiency in promoting cooperation. For measuring effectiveness, we assess the diversity of conditions for which full cooperation can be established or recovered with certainty, and for measuring efficiency, we determine the cumulative cost and total time required to convert a population of defectors to full cooperation or to recover full cooperation after the invasion of a single defector.

2.2. Public good games with dynamic incentives

Our model is based on the public good game for cooperation (C) and defection (D), widely recognized as the most suitable mathematical metaphor for studying cooperation in large groups [23–28]. We posit well-mixed populations of interacting individuals. From time to time, individuals randomly selected from the population form an n-player group with \( n \geq 2 \). A cooperator invests a fixed amount \( c > 0 \) into a common pool, whereas a defector invests nothing. The total contribution to the pool is then multiplied by a public-benefit factor \( r > 1 \) and distributed equally among all \( n \) group members. The infamous ‘tragedy of the commons’ [29] arises when \( r < n \) and no incentives are applied, because single individuals can then improve their pay-offs by withholding their contributions.

The total budget for providing incentives is given by \( n\delta \) per group, where \( \delta > 0 \) is the average per capita incentive. This budget \( n\delta \) is then divided into two parts based on a relative weight \( w \) with \( 0 \leq w \leq 1 \). The part \( w\delta \) is equally shared among the \( n_C \) cooperators in the group (see Chen et al. [17] for a similar application to the n-person volunteer’s dilemma), who thus each obtain a reward \( \frac{w\delta}{n_C} \), whereas the remainder \( (1 - w)\delta \) is used for equally punishing the \( n - n_C \) defectors, who thus each have their pay-offs reduced by \( b(1 - w)\delta n_C \). The factors \( a, b > 0 \) are the respective leverages of rewarding and punishing, and/or the factors by which a recipient’s pay-off is increased or decreased relative to the cost of implementing the incentive. ‘Antisocial’ incentives, rewarding defectors or punishing cooperators [30], could in principle be considered, but as such incentives only reduce cooperation and promote defection, they are not studied here.

We account for feedback from the population’s state by allowing the weight \( w \) to depend on the frequency of cooperators. Pure rewarding and pure punishing correspond to \( w = 1 \) or \( w = 0 \), respectively. Therefore, a cooperator and a defector obtain the pay-offs

\[
\frac{r n_C}{n} - c + \frac{awn \delta}{nc} \quad \text{and} \quad \frac{r n_C}{n} - b(1 - wn \delta) \left( n - n_C \right),
\]

respectively.

2.3. Replicator dynamics

We assume replicator dynamics [31], which describe how the frequencies of different strategies change in infinitely large, well-mixed populations. Replicator dynamics are governed by a system of differential equations, \( x_i = x_i(P_i - P) \), in which \( x_i, P_i \) and \( P \) denote, respectively, the frequency of strategy \( i \), the average pay-off for strategy \( i \), and the average pay-off in the whole population (\( P = \sum x_i P_i \), with \( \sum x_i = 1 \)). In the public good game studied here, we consider cooperation and defection with respective frequencies \( x \) and \( 1 - x \). The replicator dynamics are therefore given by a single differential equation,

\[
\dot{x} = x(P_C - P). \quad \text{With} \quad P = xP_C + (1 - x)P_D, \quad \text{we obtain} \quad \dot{x} = -x(1 - x)(P_D - P_C).
\]

This differential equation has at least two equilibria: \( x = 0 \), at which all individuals defect, and \( x = 1 \), at which all individuals cooperate.

Our model extends the traditional public good game [23–28] by incorporating incentives. Specifically, letting \( k \) denote the number of cooperators among the \( n - 1 \) co-players in a group, the expected pay-offs for a defector and a cooperater are given by

\[
P_D = \sum_{k=0}^{n-1} \frac{(n - 1 - k)}{k} x^k (1 - x)^{n-1-k} \left( \frac{r k}{n} - \frac{b(1 - w)n \delta}{n - k} \right)
\]

and

\[
P_C = \sum_{k=0}^{n-1} \frac{(n - 1 - k)}{k} x^k (1 - x)^{n-1-k} \left( \frac{r k + 1}{n} - c + \frac{awn \delta}{k + 1} \right)
\]

Without incentives, \( \delta = 0 \), we have \( P_D - P_C = c(1 - r / n) = F \), which is the defector’s advantage in the public good game.
The replicator dynamics in equation (2.2) thus lead to full defection, $x = 0$, when $r < n$ and to full cooperation, $x = 1$, when $r > n$. According to equations (2.3), incentives $\delta > 0$ modify the defector’s advantage for $0 < x < 1$ as follows:

$$P_D - P_C = F - n\delta \sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-1-k} \left( \frac{aw}{k+1} + \frac{b(1-w)}{n-k} \right).$$

(2.4a)

Using $\binom{n-1}{k} = \binom{n}{k+1}$ and $\binom{n}{k} = \binom{n}{n-k}$, this yields

$$P_D - P_C = F - \delta \left[ \sum_{k=0}^{n-1} \binom{n}{k+1} x^k (1-x)^{n-1-k} \right] + b(1-w) \sum_{k=0}^{n-1} \binom{n}{k} x^k (1-x)^{n-1-k}.$$

(2.4b)

When rewarding or punishing are applied in isolation ($w = 1$ or $w = 0$), the defector’s advantage simplifies to $F - \delta \sum_{k=0}^{n-1} (1-x)^k$ and $F - b\delta \sum_{k=0}^{n-1} x^k$, respectively [4,19]. Thus, the defector’s advantage is strictly increasing with the frequency $x$ of cooperators for pure rewarding and strictly decreasing with the frequency $x$ for pure punishing. In both cases, there exists a unique interior equilibrium of the replicator dynamics if and only if the per capita incentive $\delta$ lies within an intermediate range, $\delta_- < \delta < \delta_+$, with

$$\delta_- = \frac{F}{aw} \quad \text{and} \quad \delta_+ = \frac{F}{aw}.$$

(2.5)

Here the value of $\alpha$ depends on the type of incentive being applied: $\alpha = a$ for rewarding and $\alpha = b$ for punishing. The unique interior equilibrium is globally asymptotically stable for pure rewarding and unstable for pure punishing. Therefore, when incentives are intermediate, $\delta_- < \delta < \delta_+$, the replicator dynamics for pure rewarding lead to a mixture of defectors and cooperators, while for pure punishing, they lead to bistability between full defection and full cooperation. By contrast, if incentives are very small, $\delta \leq \delta_-$, or very large, $\delta \geq \delta_+$, the replicator dynamics for pure rewarding and for pure punishing lead to full defection or full cooperation, respectively.

3. Results

We first demonstrate, in §3.1, that an institutional sanctioning policy we call ‘first carrot, then stick’, which switches from rewarding to punishing when the frequency of cooperators exceeds a threshold, minimizes the defector’s advantage. Its effectiveness and efficiency are investigated in §3.2 and compared with those of pure rewarding and pure punishing. We extend our results to spatial populations in §3.3 and conclude by verifying, in §3.4, that our results are robust to other parameter combinations and other model variants.

3.1. ‘First carrot, then stick’ as an optimal sanctioning policy

By allowing the fraction $w$ of the incentive budget that is allocated to rewarding, rather than to punishing, to change with the frequency of cooperators, $w = w(x)$, we can represent a broad range of institutional sanctioning policies. Below we show that the ‘first carrot, then stick’ sanctioning policy is optimal in that it minimizes the defector’s advantage $P_D - P_C$; it thus maximizes the selection gradient $\frac{df}{dx}$ at any frequency $x$ of cooperators. This means that the ‘first carrot, then stick’ sanctioning policy results in the highest level of cooperation for each parameter combination and that it consequently is the most effective institutional sanctioning policy.

To see that the ‘first carrot, then stick’ sanctioning policy minimizes the defector’s advantage, we first write equation (2.4b) as

$$P_D - P_C = F - \delta \left[ w \sum_{k=0}^{n-1} (a(1-x)^k - bx^k) + b \sum_{k=0}^{n-1} x^k \right].$$

(3.1)

As this equation is linear with respect to the weight $w$, a value of $w$ of either 0 or 1 is optimal depending on whether the sum $\sum_{k=0}^{n-1} (a(1-x)^k - bx^k)$ is positive or negative, respectively. (In the degenerate case when this sum equals zero, any value of $w$ will be optimal.) One can show that this sum is a decreasing function of $x$ with exactly one root $x = \hat{x}$ satisfying $0 < \hat{x} < 1$, about which the sum changes sign from positive to negative as $x$ increases. Thus, $P_D - P_C$ is minimized for the following on-off control

$$w(x) = \begin{cases} 1 & \text{if } 0 < x < \hat{x}, \\ 0 & \text{if } x < \hat{x} \leq 1, \end{cases}$$

(3.2)

with $\sum_{k=0}^{n-1} (a(1-x)^k - bx^k) = 0$. This means that rewarding is optimal when the fraction of cooperators is below $\hat{x}$; otherwise, punishing is optimal. For obvious reasons, we call this institutional sanctioning policy ‘first carrot, then stick’.

3.2. Effectiveness and efficiency of ‘first carrot, then stick’

Figure 1 shows how the replicator dynamics are affected by per capita incentives $\delta$ under pure rewarding, pure punishing and the optimal policy in equation (3.2), with the assumption that rewarding is equally efficient ($a = b$) or less efficient ($a < b$) than punishing [32] (see also figure 2a-f). In particular, the effects of the optimal policy, which are illustrated in figure 1ef can be understood analytically. At the boundaries $x = 0$ and $x = 1$, $P_D - P_C$ in equation (3.1) takes the values $F - an\delta$ and $F - b\delta$, respectively. With the weight $w(x)$ from equation (3.2), $P_D - P_C$ takes its maximum value, $F - \delta \sum_{k=0}^{n-1} (1-x)^k$ (or equivalently, $F - b\delta \sum_{k=0}^{n-1} x^k$), at $x = \hat{x}$; $P_D - P_C$ is strictly increasing for $0 \leq x < \hat{x}$ and strictly decreasing for $\hat{x} < x \leq 1$. Hence, $P_D - P_C = 0$ can have at most two interior roots $x$. Indeed, for $\delta < \min\{F/(an),F/(bn)\}$, the replicator dynamics converge to $x = 0$. For $\delta > F/(an)$, a stable equilibrium enters the interior state space $0 < x < 1$ at $x = 0$, and the full-defection equilibrium, $x = 0$, becomes unstable. For $\delta > F/(bn)$, an
unstable equilibrium enters the interior state space \(0 < x < 1\) at \(x = 1\), and the full-cooperation equilibrium, \(x = 1\), becomes stable. Thus, the interior state space \(0 < x < 1\) has both stable and unstable equilibria if

\[
\max\left\{ \frac{F}{bn}, \frac{F}{am} \right\} \leq \delta < \frac{F}{a \sum_{k=0}^{1-x}(1-x)^k} = \frac{F}{b \sum_{k=0}^{x}(x)^k}. \tag{3.3}
\]

As \(\delta\) increases within this interval, the two interior equilibria approach each other and eventually merge and vanish at the upper bound of this interval. For values of \(\delta\) beyond this interval’s upper bound, the replicator dynamics converge to \(x = 1\). In the special case when rewarding and punishing are equally efficient, \(a = b\), the stable and unstable equilibria enter the interior state space simultaneously (figure 1c). Conversely, in the extreme case when rewarding is much less efficient than punishing, \(a < b\), punshing alone is sufficient from the very beginning, and hybridization with relatively expensive rewarding is irrelevant.

For well-mixed populations, we have proved in §3.1 that the hybridization of positive and negative incentives according to the ‘first carrot, then stick’ sanctioning policy described by equation (3.2) minimizes the defector’s advantage, which ensures that this sanctioning policy is most effective for converting a population of
defectors to cooperation (figure 2a–f). By combining the differential advantages of rewarding and punishing, this sanctioning policy is far more effective than pure punishing in establishing cooperation (figure 2a,c) and far more effective than pure rewarding in recovering cooperation (figure 2b,f). Offering the ‘best of both worlds’, the ‘first carrot, then stick’ sanctioning policy for combining rewarding with punishing is therefore hereafter also called the ‘adaptive hybrid’ sanctioning policy.

Although it is natural to expect that the threshold $\hat{x}$ at which the adaptive hybrid policy switches from rewarding to punishing might depend on other parameters, this is not the case: the threshold remains the same independent of the per capita incentive $\delta$ and the public-benefit factor $r$. What is more, when there is no difference in leverage between positive and negative incentives ($a = b$), this threshold is situated at a 50% frequency of cooperators, $\hat{x} = 0.5$. In practice, punishing is often more effective than rewarding ($a < b$) [32], in which case the switching point for optimal hybridization is situated at a frequency of cooperators of less than 50%, $\hat{x} < 0.5$ (figure 1f).

The adaptive hybrid policy is not only more effective, but also more efficient, for establishing and recovering cooperation than either rewarding or punishing alone (figure 3a–f).
Once a state of full cooperation has been reached, punishing is cheaper as a means of recovering cooperation, as it needs to be used only occasionally. As the adaptive hybrid policy stipulates punishment once the frequency of cooperators surpasses the threshold \( x \), it is similar to pure punishment in this respect. The two policies differ markedly, however, in the cost of converting a population of defectors to a population of cooperators. The adaptive hybrid policy has the lowest cumulative cost of all three sanctioning policies, and hence requires both the lowest establishment cost and the lowest recovery cost for full cooperation. With respect to conversion speed, it generically takes a similar (finite) time for all three policies to establish and recover cooperation (electronic supplementary material, figure S1).
Also in spatial populations, the adaptive hybrid policy turns out to be superior (figure 2g–i). Unexpectedly, it gives rise to spatial patterns of cooperation and defection that cannot easily be predicted from the patterns arising under either rewarding or punishing alone. For small and large incentives, the patterns emerging from the adaptive hybrid policy when considering a single cooperator in a population of defectors resemble the patterns observed under pure rewarding and punishing, respectively. Cooperators thrive under a policy of pure rewarding (figure 4a), forming fragmented islands in which they are locally interspersed with defectors, but ultimately fail to establish full cooperation for the incentive strength considered. Under a policy of pure punishing (figure 4b), spatio-temporal dynamics are different: an invasion that begins with a single cooperator in a population of defectors always results in a contiguous cluster of cooperators that grows and eventually displaces all defectors. The adaptive hybrid policy, by contrast, for intermediate incentive strengths exhibits an intriguing transition between these two distinct patterns: fragmented islands of cooperators, initially supported by rewarding, create circumstances under which punishing can act as a ‘booster stage’ that capitalizes on and amplifies the pro-social effects of rewarding, promoting the rapid growth of contiguous clusters of cooperators that eventually displace all defectors (figure 4c).

All three policies are capable of recovering cooperation in much the same way as for well-mixed populations. The only qualitative difference is that the successful spread of defectors originating from an initially single defector can occasionally split contiguous clusters of cooperators. This phenomenon, however, which has previously been described for the spatial extension of the well-studied Prisoner’s Dilemma [35], occurs in our model only for vanishing or very small incentives.

### 3.4. Robustness

In the electronic supplementary material, we demonstrate the robustness of our results with respect to the following model variants. (i) First, we establish that in spatial populations the adaptive hybrid policy with either local or global feedback establishes and recovers full cooperation at lower cost and under a wider range of conditions than an alternative hybridization of positive and negative incentives in which the weight $w$ is proportional to the frequency of cooperators, $w(x) = x$ (electronic supplementary material, figure S2). We also show that information about the local frequency of cooperation allows a sanctioning institution that implements the adaptive hybrid policy to establish full cooperation more readily than information about the global (i.e. population-wide) frequency of cooperation [20]. This result is in line with expectations, as tailoring a strategy to local conditions should generally achieve better results than a strategy that depends on conditions averaged across large spatial scales.

We also explore (ii) a variant of the public good game in which a cooperators does not benefit from its own investment [4,19] (electronic supplementary material, figure S3) and (iii) a variant of the incentive scheme in which we relax the assumption that the received incentive is inversely proportional to the number of cooperators or defectors in an interacting group [4,19] (electronic supplementary material, figure S4). Furthermore, we test variants of our spatial model with (iv) interactions encompassing the eight nearest neighbours [33–35] (chess-king move, $n = 9$, electronic supplementary material, figure S5), (v) smaller population size (electronic supplementary material, figure S6), (vi) asynchronous updating
of experimental studies that have explored the combined use of positive and negative incentives in peer-to-peer sanctioning [38–41] or in sanctioning by an assigned team leader [42]. Although these studies differ significantly in their experimental design, they share two common characteristics. First, punishment is typically more effective than rewarding at promoting high contributions to the public good. Second, players initially have a propensity for rewarding cooperation, which is soon superseded by a propensity for punishing defectors [38–40]. While the latter trend might superficially be interpreted as corroborative evidence for the effectiveness of the institutional sanctioning policy developed here, the rationale for shifting from positive to negative incentives is strikingly different. In the experimental studies, this shift typically coincides with declining average contributions and can thus be interpreted as a response to the emergence of defectors [42]. In particular, a study on team leadership concludes that ‘leaders who experience frequent complete free-riding and high variance in contributions in their teams are more likely to switch from positive to negative incentives’ [42], while other studies find that punishing is more effective than rewarding at staying off complete free-riding [38–40]. By contrast, we have demonstrated the advantage of shifting from positive to negative incentives as contributions increase, and we predict that rewarding is more effective than punishing in staying off complete free-riding [43].

We have determined the optimal sanctioning policy for a social institution charged with overseeing rational agents. Two complementary studies on peer-to-peer sanctioning that account, respectively, for reputation effects and for the potential of group selection have similarly highlighted the role of positive incentives in promoting incipient cooperation among defectors [36,44]. These theoretical predictions, derived under the assumption of rational behaviour, clearly question the wisdom of the human behaviour observed in the aforementioned experimental studies. Understanding whether punishment in the face of rampant defection is a human fallacy or a rational choice under circumstances other than the ones analysed here is a key challenge for future research.

Acknowledgements. We thank Karl Sigmund and Mitsuhiro Nakamura for their valuable comments and suggestions.

Funding statement. This study was enabled by financial support by the Austrian Science Fund (FWF) to U.D. (TECT I-106 G11), through a grant for the research project The Adaptive Evolution of Mutualistic Interactions as part of the multi-national collaborative research project Mutualisms, Contracts, Space, and Dispersal (BIOCONTRACT) selected by the European Science Foundation as part of the European Collaborative Research (EUROCORES) Programme The Evolution of Cooperation and Trading (TEC). U.D. gratefully acknowledges additional support by the European Commission, the European Science Foundation, the Austrian Ministry of Science and Research, and the Vienna Science and Technology Fund. T.S. acknowledges additional support by the Foundational Questions in Evolutionary Biology Fund (RFT-12-21) and the Austrian Science Fund (FWF): P27018-G11.
References


