Research

Author for correspondence:
Manh Hong Duong
e-mail: h.duong@bham.ac.uk

THE ROYAL SOCIETY
PUBLISHING

Cost efficiency of institutional incentives for promoting cooperation in finite populations

Manh Hong Duong[1] and The Anh Han[2]

[1]School of Mathematics, University of Birmingham, Birmingham B15 2TT, UK
[2]School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough TS1 3BX, UK

MHD, 0000-0002-4361-0795; TAH, 0000-0002-3095-7714

Institutions can provide incentives to enhance cooperation in a population where this behaviour is infrequent. This process is costly, and it is thus important to optimize the overall spending. This problem can be mathematically formulated as a multi-objective optimization problem where one wishes to minimize the cost of providing incentives while ensuring a minimum level of cooperation, sustained over time. Prior works that consider this question usually omit the stochastic effects that drive population dynamics. In this paper, we provide a rigorous analysis of this optimization problem, in a finite population and stochastic setting, studying both pairwise and multi-player cooperation dilemmas. We prove the regularity of the cost functions for providing incentives over time, characterize their asymptotic limits (infinite population size, weak selection and large selection) and show exactly when reward or punishment is more cost efficient. We show that these cost functions exhibit a phase transition phenomenon when the intensity of selection varies. By determining the critical threshold of this phase transition, we provide exact calculations for the optimal cost of the incentive, for any given intensity of selection. Numerical simulations are also provided to demonstrate analytical observations. Overall, our analysis provides for the first time a selection-dependent calculation of the optimal cost of institutional incentives (for both reward and punishment) that guarantees a minimum level of

cooperation over time. It is of crucial importance for real-world applications of institutional incentives since the intensity of selection is often found to be non-extreme and specific for a given population.

1. Introduction

The problem of promoting the evolution of cooperative behaviour within populations of self-regarding individuals has been intensively investigated across diverse fields of behavioural, social and computational sciences [1–5]. Various mechanisms responsible for promoting the emergence and stability of cooperative behaviours among such individuals have been proposed. They include kin and group selection [6,7], direct and indirect reciprocities [8–12], spatial networks [13–16], reward and punishment [17–22] and pre-commitments [23–27]. Institutional incentives, namely rewards for cooperation and punishment for wrongdoing, are among the most important ones [22,28–36]. Different from other mechanisms, in order to carry out institutional incentives, it is assumed that there exists an *external* decision maker (e.g. institutions such as the United Nations and the European Union) that has a budget to interfere in the population to achieve a desirable outcome. Institutional enforcement mechanisms are crucial for enabling large-scale cooperation. Most modern societies implement certain forms of institutions for governing and promoting collective behaviours, including cooperation, coordination and technology innovation [37–42].

Providing incentives is costly and it is therefore important to minimize the cost while ensuring a sustained level of cooperation over time [28,31,41]. Despite its paramount importance, so far there have been only a few works exploring this question. In particular, Wang *et al*. [35] use optimal control theory to provide an analytical solution for cost optimization of institutional incentives assuming deterministic evolution and infinite population sizes (modelled using replicator dynamics). This work therefore does not take into account various stochastic effects of evolutionary dynamics such as mutation and non-deterministic behavioural update [4,43,44]. In a deterministic system consisting of cooperators and defectors, once the latter disappear (for instance through strong institutional punishment), there is no further change to the system and thus no further interference in it is required. When mutation is present, this behaviour can however recur and become abundant over time, requiring institutions to spend more of their budget on providing further incentives. Moreover, a key factor of behavioural update, the intensity of selection [4]—which determines how strongly an individual bases their decision to copy another individual's strategy on their fitness difference—might strongly impact an institutional incentives strategy and its cost efficiency. Its value is usually found to be specific for a given population [45–48] and thus should be taken into account when designing suitable cost-efficient incentives. For instance, when selection is weak such that behavioural update is close to a random process (i.e. an imitation decision is independent of how large the fitness difference is), providing incentives would make little difference to cause behavioural change, however strong it is. When selection is strong, incentives that ensure a minimum fitness advantage to cooperators would ensure a positive behavioural change.

In a stochastic, finite-population context, so far this problem has been investigated primarily using agent-based and numerical simulations [28,31,49–52]. Results demonstrate several interesting phenomena, such as the significant influence of the intensity of selection on incentive strategies and optimal costs. However, there is no satisfactory rigorous analysis available at present that allows one to determine the optimal way of providing incentives. This is a challenging problem because of the large but finite population size and the complexity of stochastic processes governing the population dynamics.

In this paper, we provide exactly such a rigorous analysis. We study cooperation dilemmas in both pairwise (the Donation game (DG)) and multi-player (the Public Goods game (PGG)) settings [4]. They are among the most well-studied models for investigating the evolution of

cooperative behaviour where individual defection is always preferred over cooperation while mutual cooperation is the preferred collective outcome for the population as a whole. Adopting a popular stochastic evolutionary game approach for analysing well-mixed finite populations [53–55], we derive the total expected costs of providing institutional reward or punishment, characterize their asymptotic limits (namely, for an infinite population, weak selection and strong selection) and show the existence of a phase transition phenomenon in the optimization problem when the intensity of selection varies. We calculate the critical threshold of phase transitions and study the minimization problem when the selection is less than and greater than the critical value. We furthermore provide numerical simulations to demonstrate the analytical results.

The rest of the paper is organized as follows. In §2, we introduce the models and methods, deriving mathematical optimization problems that will be studied. The main results of the paper are presented in §3. In §4, we discuss possible extensions for future work. Finally, detailed computations, technical lemmas and proofs of the main results are provided in the electronic supplementary material.

2. Models and methods

(a) Cooperation dilemmas

We consider a well-mixed, finite population of N self-regarding individuals or players, who interact with each other using one of the following one-shot (i.e. non-repeated) cooperation dilemmas: the DG or its multi-player version, the PGG. In these games, a player can choose either to cooperate (i.e. a cooperator or C player) or to defect (i.e. a defector, or D player).

Let $\Pi_C(i)$ and $\Pi_D(i)$ be the average pay-offs of a C player and a D player in a population with i C players and $N - i$ D players, respectively (see also §2.3 for more details). We show below that the difference $\delta = \Pi_C(i) - \Pi_D(i)$ does not depend on i. For cooperation dilemmas, it is always the case that $\delta < 0$.

(i) Donation game

The pay-off matrix of the DG (for a row player) is given as follows:

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} C \quad\quad D \\ \begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix}, \end{array}$$

where c and b represent the cost and benefit of cooperation, where $b > c$. DG is a special version of the Prisoner's Dilemma (PD) game.

Denoting $\pi_{X,Y}$ as the pay-off of a strategist X when playing with strategist Y from the pay-off matrix above, we obtain

$$\Pi_C(i) = \frac{(i - 1)\pi_{C,C} + (N - i)\pi_{C,D}}{N - 1} = \frac{(i - 1)(b - c) + (N - i)(-c)}{N - 1}$$

and

$$\Pi_D(i) = \frac{i\pi_{D,C} + (N - i - 1)\pi_{D,D}}{N - 1} = \frac{ib}{N - 1}.$$

Thus,

$$\delta = \Pi_C(i) - \Pi_D(i) = -\left(c + \frac{b}{N - 1}\right).$$

(ii) Public Goods game

In a PGG, players interact in a group of size n, where they decide to cooperate, contributing an amount $c > 0$ to a common pool, or to defect, contributing nothing to the pool. The total contribution in a group will be multiplied by a factor r, where $1 < r < n$ (for the PGG to be a

social dilemma), which is then shared equally among all members of the group, regardless of their strategy.

We obtain [56]

$$\Pi_C(i) = \sum_{j=0}^{n-1} \frac{\binom{i-1}{j}\binom{N-i}{n-1-j}}{\binom{N-1}{n-1}} \left(\frac{(j+1)rc}{n} - c\right) = \frac{rc}{n}\left(1 + (i-1)\frac{n-1}{N-1}\right) - c$$

and

$$\Pi_D(i) = \sum_{j=0}^{n-1} \frac{\binom{i}{j}\binom{N-1-i}{n-1-j}}{\binom{N-1}{n-1}} \frac{jrc}{n} = \frac{rc(n-1)}{n(N-1)}i.$$

Thus,

$$\delta = \Pi_C(i) - \Pi_D(i) = -c\left(1 - \frac{r(N-n)}{n(N-1)}\right).$$

(b) Cost of institutional reward and punishment

To reward a cooperator (respectively, punish a defector), the institution has to pay an amount θ/a (resp., θ/b) so that the cooperator's (defector's) pay-off increases (decreases) by θ, where $a, b > 0$ are constants representing the efficiency ratios of providing the corresponding incentive. As we study reward and punishment separately, without losing generality, we set $a = b = 1$ [22,28]. Thus, the key question here is: *What is the optimal value of the individual incentive cost θ that ensures a sufficient desired level of cooperation in the population (in the long run) while minimizing the total cost spent by the institution?*

(i) Deriving the expected cost of providing institutional incentives

We adopt here the finite population dynamics with the Fermi strategy update rule [44], stating that a player A with fitness f_A adopts the strategy of another player B with fitness f_B with a probability given by $P_{A,B} = (1 + e^{-\beta(f_B - f_A)})^{-1}$, where β represents the intensity of selection (see details in §2c). We compute the expected number of times the population contains i C players, $1 \leq i \leq N-1$. For that, we consider an absorbing Markov chain of $(N+1)$ states, $\{S_0, \ldots, S_N\}$, where S_i represents a population with i C players. S_0 and S_N are absorbing states. Let $U = \{u_{ij}\}_{i,j=1}^{N-1}$ denote the transition matrix between the $N-1$ transient states, $\{S_1, \ldots, S_{N-1}\}$. The transition probabilities can be defined as follows, for $1 \leq i \leq N-1$:

$$\left.\begin{aligned}
u_{i,i\pm j} &= 0 \quad \text{for all } j \geq 2, \\
u_{i,i\pm 1} &= \frac{N-i}{N}\frac{i}{N}(1 + e^{\mp\beta[\Pi_C(i) - \Pi_D(i) + \theta]})^{-1} \\
u_{i,i} &= 1 - u_{i,i+1} - u_{i,i-1}.
\end{aligned}\right\} \tag{2.1}$$

and

The entries n_{ij} of the so-called fundamental matrix $\mathcal{N} = (n_{ij})_{i,j=1}^{N-1} = (I-U)^{-1}$ of the absorbing Markov chain give the expected number of times the population is in the state S_j if it starts in the transient state S_i [57]. As a mutant can randomly occur at either S_0 or S_N, the expected number of visits at state S_i is, thus, $\frac{1}{2}(n_{1i} + n_{N-1,i})$.

The total cost per generation is

$$\theta_i = \begin{cases} i \times \theta & \text{in the case of institutional reward,} \\ (N-i) \times \theta & \text{in the case of institutional punishment.} \end{cases}$$

Hence, the expected total costs of interference for institutional reward and institutional punishment are, respectively,

$$E_r(\theta) = \frac{\theta}{2} \sum_{i=1}^{N-1} (n_{1i} + n_{N-1,i})i \quad \text{and} \quad E_p(\theta) = \frac{\theta}{2} \sum_{i=1}^{N-1} (n_{1i} + n_{N-1,i}) (N-i). \tag{2.2}$$

(ii) Cooperation frequency

Since the population consists of only two strategies, the fixation probabilities of a C (D) player in a homogeneous population of D (C) players when the interference scheme is carried out are, respectively,

$$\rho_{D,C} = \left(1 + \sum_{i=1}^{N-1} \prod_{k=1}^{i} \frac{1 + e^{\beta(\Pi_C(k) - \Pi_D(k) + \theta)}}{1 + e^{-\beta(\Pi_C(k) - \Pi_D(k) + \theta)}}\right)^{-1}$$

and

$$\rho_{C,D} = \left(1 + \sum_{i=1}^{N-1} \prod_{k=1}^{i} \frac{1 + e^{\beta(\Pi_D(k) - \Pi_C(k) - \theta)}}{1 + e^{-\beta(\Pi_D(k) - \Pi_C(k) - \theta)}}\right)^{-1}.$$

Computing the stationary distribution using these fixation probabilities, we obtain the frequency of cooperation (see §2.3),

$$\frac{\rho_{D,C}}{\rho_{D,C} + \rho_{C,D}}.$$

Hence, this frequency of cooperation can be maximized by maximizing

$$\max_{\theta}(\rho_{D,C}/\rho_{C,D}). \tag{2.3}$$

The fraction in equation (2.3) can be simplified as follows [54]:

$$\frac{\rho_{D,C}}{\rho_{C,D}} = \prod_{k=1}^{N-1} \frac{T^-(k)}{T^+(k)} = \prod_{k=1}^{N-1} \frac{1 + e^{\beta[\Pi_C(k) - \Pi_D(k) + \theta]}}{1 + e^{-\beta[\Pi_C(k) - \Pi_D(k) + \theta]}}$$

$$= e^{\beta \sum_{k=1}^{N-1}(\Pi_C(k) - \Pi_D(k) + \theta)}$$

$$= e^{\beta(N-1)(\delta+\theta)}. \tag{2.4}$$

In the above transformation, $T^-(k)$ and $T^+(k)$ are the probabilities of decreasing or increasing the number of C players (i.e. k) by one in each time step, respectively.

We consider non-neutral selection, i.e. $\beta > 0$ (under neutral selection, there is no need to use incentives). Assuming that we desire to obtain at least an $\omega \in [0, 1]$ fraction of cooperation, i.e. $\rho_{D,C}/(\rho_{D,C} + \rho_{C,D}) \geq \omega$, it therefore follows from equation (2.4) that

$$\theta \geq \theta_0(\omega) = \frac{1}{(N-1)\beta} \log\left(\frac{\omega}{1-\omega}\right) - \delta. \tag{2.5}$$

Therefore, it is guaranteed that, if $\theta \geq \theta_0(\omega)$, at least an ω fraction of cooperation can be expected. This condition implies that the lower bound of θ monotonically depends on β. Namely, when $\omega \geq 0.5$ it increases with β, while when $\omega < 0.5$ it decreases with β.

(iii) Optimization problems

Bringing all these factors together, we obtain the following cost-optimization problems of institutional incentives in stochastic finite populations:

$$\min_{\theta \geq \theta_0(\omega)} E(\theta), \tag{2.6}$$

where E is either E_r or E_p, defined in (2.2), which respectively corresponds to institutional reward or punishment. We show in the electronic supplementary material that $\theta \mapsto E(\theta)$ is a smooth function on $\mathbb{R}$.

(c) Methods: evolutionary dynamics in finite populations

We adopt in our analysis the evolutionary game theory (EGT) methods for finite populations [53–55]. Herein, individuals' pay-offs represent their *fitness* or social *success*, and evolutionary dynamics is shaped by social learning [4,43], whereby the most successful players will tend to be imitated more often by the other players. Here, social learning is modelled using the pairwise comparison rule [44], that is, a player A with fitness f_A adopts the strategy of another player B with fitness f_B with probability given by the Fermi function,

$$P_{A,B} = (1 + e^{-\beta(f_B - f_A)})^{-1},$$

where β conveniently describes the selection intensity ($\beta = 0$ represents neutral drift while $\beta \to \infty$ represents increasingly deterministic selection).

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 social learning. We assume that, with a certain mutation probability, an individual switches randomly to a different strategy without imitating another individual. In addition, we assume here the small mutation limit [53,55,58]. Thus, at most two strategies are present in the population at a time. The evolutionary dynamics can be described by a Markov chain, where each state represents a homogeneous population and the transition probabilities between any two states are given by the fixation probability of a single mutant [53,55,58]. The resulting Markov chain has a stationary distribution, which describes the average time the population spends in an end state. The small mutation limit allows us to obtain an analytical form of the frequency of cooperation (see below). It is noteworthy that, although we focus here on the small mutation limit, this approach has been shown to be widely applicable to scenarios which go well beyond the strict limit of very small mutation rates [45,46,48,59].

The fixation probability of a single mutant A taking over a whole population with $(N - 1)$ B players is as follows (see [44,55,60] for details)

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

where $T^{\pm}(k) = ((N - k)/N)(k/N)[1 + e^{\mp\beta[\Pi_A(k) - \Pi_B(k)]}]^{-1}$ describes the probability of changing the number of A players by $\pm$ one in a time step. Specifically, when $\beta = 0$, $\rho_{B,A} = 1/N$, representing the transition probability at the neutral limit.

Considering the set of two strategies C and D (see [53,58] for the calculation for any number of strategies). Their stationary distribution is given by the normalized eigenvector associated with the eigenvalue 1 of the transpose of a matrix [53,58]

$$M = \begin{pmatrix} 1 - \rho_{C,D} & \rho_{C,D} \\ \rho_{D,C} & 1 - \rho_{D,C} \end{pmatrix},$$

which is $\{\rho_{D,C}/(\rho_{D,C} + \rho_{C,D}), \rho_{C,D}/(\rho_{D,C} + \rho_{C,D})\}$. The first term is the frequency of cooperation and the second one is that of defection.

3. Main results

The present paper provides a rigorous analysis of the expected total cost of providing an institutional incentive (2.2) and the associated optimization problem (2.6). In this section, we state our main analytical results, theorems 3.1–3.4, and provide numerical simulations to illustrate the analytical results. The proofs of these results, which require a delicate analysis of the cost functions, are presented in the electronic supplementary material.

In the following theorems, E denotes the cost function for either institutional reward, E_r, or institutional punishment, E_p, as obtained in (2.2). Also, H_N denotes the well-known harmonic number

$$H_N := \sum_{j=1}^{N-1} \frac{1}{j}. \tag{3.1}$$

Our first main result provides qualitative properties and asymptotic limits of E.

Theorem 3.1 (qualitative properties and asymptotic limits of total cost functions).

(I) *(finite population estimates) The expected total cost of providing an incentive satisfies the following estimates for all finite populations of size N:*

$$\frac{N^2\theta}{2}\left(H_N + \frac{1}{N-1}\right) \leq E(\theta) \leq N(N-1)\theta(H_N + 1). \tag{3.2}$$

(II) *(infinite population limit) The expected total cost of providing an incentive satisfies the following asymptotic behaviour when the population size N tends to $+\infty$:*

$$\lim_{N\to+\infty} \frac{E(\theta)}{\frac{N^2\theta}{2}(\ln N + \gamma)} = \begin{cases} 1 + e^{-\beta|\theta - c|} & \text{for DG,} \\ 1 + e^{-\beta|\theta - c|}e^{\beta c\frac{r}{n}} & \text{for PGG,} \end{cases} \tag{3.3}$$

where $\gamma = 0.5772\cdots$ is the Euler–Mascheroni constant.

(III) *(weak selection limit) The expected total cost of providing an incentive satisfies the following asymptotic limit when the selection strength β tends to 0:*

$$\lim_{\beta\to 0} E(\theta) = N^2\theta H_N. \tag{3.4}$$

(IV) *(strong selection limit) The expected total cost of providing an incentive satisfies the following asymptotic limit when the selection strength β tends to $+\infty$:*

$$\lim_{\beta\to+\infty} E_r(\theta) = \begin{cases} \dfrac{N^2}{2}\theta\left(\dfrac{1}{N-1} + H_N\right) & \text{for } \theta < -\delta, \\[2mm] N^2\theta H_N & \text{for } \theta = -\delta, \\[2mm] \dfrac{N^2}{2}\theta(1 + H_N) & \text{for } \theta > -\delta \end{cases} \tag{3.5}$$

and

$$\lim_{\beta\to+\infty} E_p(\theta) = \begin{cases} \dfrac{N^2\theta}{2}(1 + H_N) & \text{for } \theta < -\delta, \\[2mm] N^2\theta H_N & \text{for } \theta = -\delta, \\[2mm] \dfrac{N^2\theta}{2}\left(H_N + \dfrac{1}{N-1}\right) & \text{for } \theta > -\delta. \end{cases} \tag{3.6}$$

The lower and upper bounds obtained in part (I) of the theorem suggest that the total expected cost function E for both reward and punishment behaves asymptotically in the order of $(N^2 H_N) \times \theta$ for sufficiently large N. This is confirmed in part (II), noting that $H_N \sim \ln N$. We also show that the leading asymptotic coefficient of E depends on the game (i.e. DG or PGG) and its parameters. Hence, it is important to adopt a precise optimal value of θ (e.g. obtained by solving the optimization problem (2.6)), as a small increase in this individual incentive cost can lead to a significant increase in E, especially when the population size is large. Figure 1 numerically demonstrates this asymptotic limit.

Parts (III) and (IV) of the theorem provide theoretical estimations of E under the weak ($\beta \to 0$) and strong ($\beta \to +\infty$) selection limits. For the weak selection limit, the expected total costs are the same for reward and punishment, i.e. $E_r(\theta) = E_p(\theta)$. For the strong selection limit, E_r is smaller
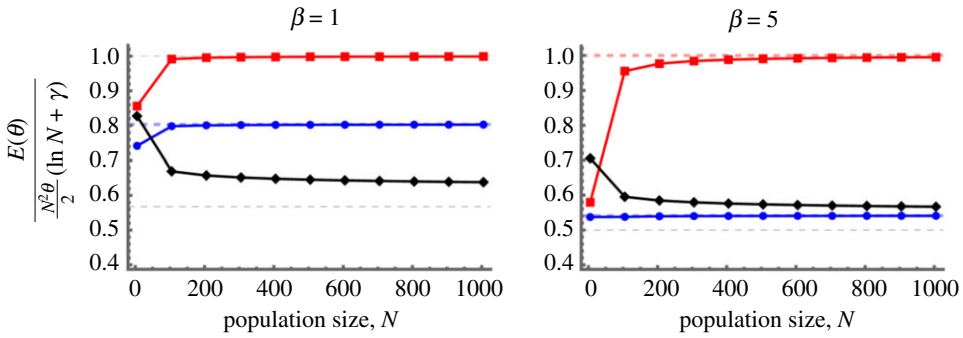
Figure 1. Large population size limit. We calculate numerically the expected total cost of incentive E for reward and punishment, varying population size N, for different values of θ and β. The dashed lines represent the corresponding theoretical limiting values obtained in theorem 3.1 for the large population size limit, $N \to +\infty$. We observe that numerical results are in close accordance with those obtained theoretically. Results are obtained for DG with $b = 2, c = 1$. (Online version in colour.)
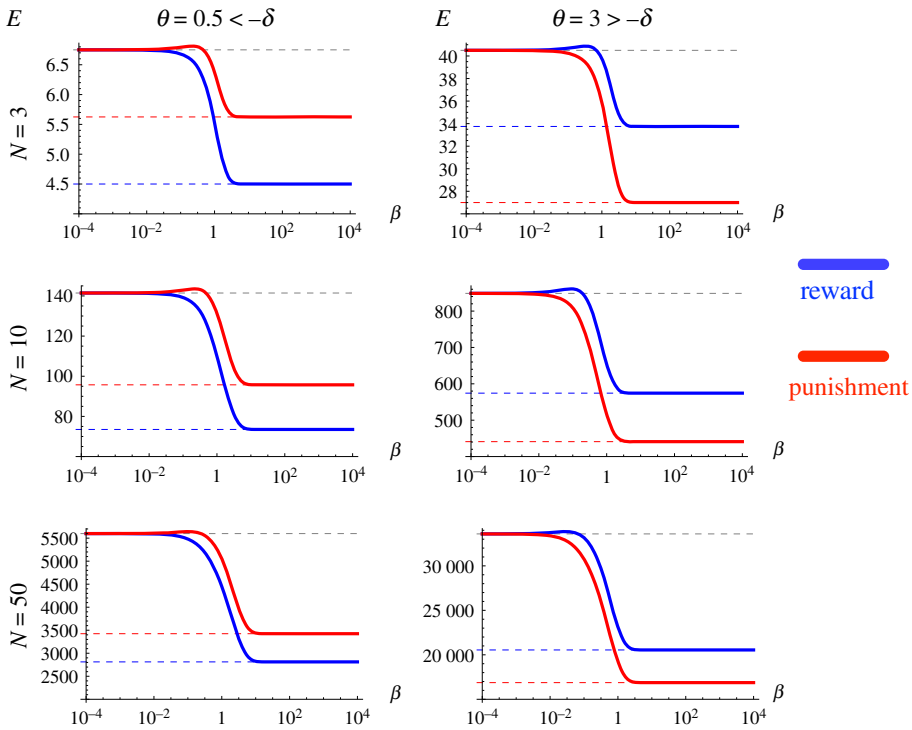
Figure 2. Weak and strong selection limits. We calculate numerically the total expected cost of incentive E for reward and punishment, by varying the intensity of selection, for different values of N and β. The dashed lines represent the corresponding theoretical limiting values obtained in theorem 3.1 for weak and strong selection limits. We observe that numerical results are in close accordance with those obtained theoretically. Results are obtained for DG with $b = 2, c = 1$. (Online version in colour.)

than, equal to or greater than E_p, depending on whether θ is smaller than, equal to or greater than $-\delta$. Figure 2 provides numerical validation of the theoretical weak and strong selection asymptotic behaviours of E, for different population sizes N. We can observe that, for a given individual incentive cost θ, the range of E increases significantly for larger N.

Our second main result concerns the optimization problem (2.6). We show that the cost function E exhibits a phase transition when the selection intensity β varies.

Theorem 3.2 (optimization problems and phase transition phenomenon).

(I) *(phase transition phenomena and behaviour under the threshold) Define*

$$F^* = \begin{cases} \min\{F(u) : P(u) > 0\} & \text{in the reward case,} \\ \min\{\hat{F}(u) : \hat{P}(u) > 0\} & \text{in the punishment case,} \end{cases}$$

where $P(u)$ and $F(u)$ as well as $\hat{P}$ and $\hat{F}$ are defined in the electronic supplementary material (see §§1 and 2 there, respectively). There exists a threshold value β^* given by

$$\beta^* = -\frac{F^*}{\delta} > 0,$$

such that $\theta \mapsto E(\theta)$ is non-decreasing for all $\beta \leq \beta^*$ and is non-monotonic when $\beta > \beta^*$. As a consequence, for $\beta \leq \beta^*$

$$\min_{\theta \geq \theta_0} E(\theta) = E(\theta_0). \tag{3.7}$$

(II) *(behaviour above the threshold value) For $\beta > \beta^*$, the number of changes of the sign of $E'(\theta)$ is at least two for all N and there exists an N_0 such that the number of changes is exactly two for $N \leq N_0$. As a consequence, for $N \leq N_0$, there exist $\theta_1 < \theta_2$ such that, for $\beta > \beta^*$, $E(\theta)$ is increasing when $\theta < \theta_1$, decreasing when $\theta_1 < \theta < \theta_2$ and increasing when $\theta > \theta_2$. Thus, for $N \leq N_0$,*

$$\min_{\theta \geq \theta_0} E(\theta) = \min\{E(\theta_0), E(\theta_2)\}.$$

The proofs of theorems 3.1 and 3.2 for the cases of reward and punishment are given in §§1 and 2 in the electronic supplementary material, respectively. We also provide explicit computations for $N = 3$ and $N = 4$ to illustrate these theorems in §3 in the electronic supplementary material. Based on numerical simulations, we conjecture that the requirement $N \leq N_0$ could be removed and theorem 3.2 is true for all finite N. In electronic supplementary material, figure S2, using numerical calculation we have shown that $N_0 = 100$ satisfies the conjecture, ensuring the validity of the numerical examples below. Theorem 3.2 gives rise to the following algorithm to determine the optimal value θ^* for $N \leq N_0$.

Algorithm 3.3 (finding optimal cost of incentive $\theta^\star$). *Inputs: (i) $N \leq N_0$: population size, (ii) β: intensity of selection, (iii) game and parameters: PD (c and b) or PGG (c, r and n), and (iv) ω: minimum desired cooperation level.*

(1) *Compute δ {in PD: $\delta = -(c + (b/(N-1)))$; in PGG: $\delta = -c(1 - ((r(N-n))/(n(N-1))))$}.*
(2) *Compute $\theta_0 = (1/(N-1)\beta) \log(\omega/(1-\omega)) - \delta$.*
(3) *Compute*

$$F^* = \begin{cases} \min\{F(u) : P(u) > 0\} & \text{in the reward case,} \\ \min\{\hat{F}(u) : \hat{P}(u) > 0\} & \text{in the punishment case,} \end{cases}$$
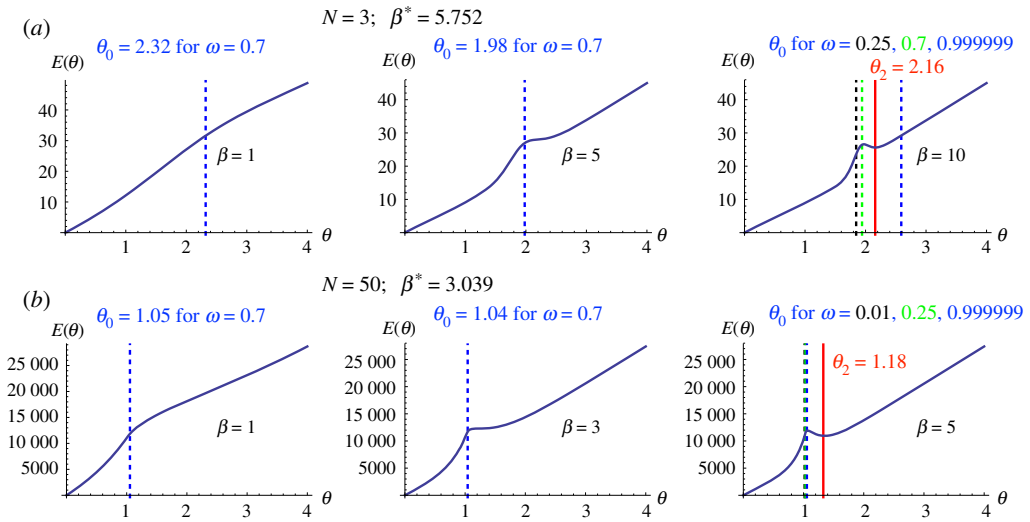
where $P(u)$ and $F(u)$, as well as $\hat{P}$ and $\hat{F}$, are defined in the electronic supplementary material.
(4) *Compute $\beta^* = -(F^*/\delta)$.*
(5) *If $\beta \leq \beta^*$:*

$$\theta^* = \theta_0, \quad \min E(\theta) = E(\theta_0).$$

(6) *Otherwise (i.e. if $\beta > \beta^*$)*

(a) *Compute u_2 that is the largest root of the equation $F(u) + \beta\delta = 0$ for the reward case or that of $\hat{F}(u) + \beta\delta = 0$ for the punishment case.*
(b) *Compute $\theta_2 = ((\log u_2)/\beta) - \delta$:*

— *if $\theta_2 \leq \theta_0$: $\theta^* = \theta_0$, $\quad \min E(\theta) = E(\theta_0)$.*

Figure 3. We use algorithm 3.3 to find optimal θ that minimizes $E(\theta)$ (for institutional reward) while ensuring a minimum level of cooperation ω. We also use as examples a small population size ($N = 3(a)$) and a larger one ($N = 50(b)$), for DG ($b = 1.8$, $c = 1$). (Online version in colour.)

— *Otherwise (if $\theta_2 > \theta_0$):*
— if $E(\theta_0) \le E(\theta_2)$: $\theta^* = \theta_0$, $\min E(\theta) = E(\theta_0)$;
— if $E(\theta_2) < E(\theta_0)$: $\theta^* = \theta_2$, $\min E(\theta) = E(\theta_2)$.

Output: θ^* and $E(\theta^*)$.

To illustrate theorem 3.2 and algorithm 3.3, we focus on the case of reward. Figure 3 shows the cost function E_r as a function of θ, for different values of N, β and ω to illustrate the phase transition when varying β, in a DG. We can see that, in all cases, these numerical observations are in close accordance with theoretical results. For example, with $N = 3$ (figure 3a), we found $\beta^* = f^*/\delta = 10.9291/1.9 = 5.752$. For $\beta < \beta^*$, $E(\theta)$ are increasing functions of θ. Thus, the optimal cost of incentive $\theta^* = \theta_0$, for a given required minimum level of cooperation ω. For example, with $N = 3$, for $\beta = 1$ to ensure at least 70% of cooperation ($\omega = 0.7$), then $\theta^* = \theta_0 = 2.32$. When $\beta \ge \beta^*$ one needs to compare $E(\theta_0)$ and $E(\theta_2)$. For example, with $N = 3$, $\beta = 10$: for $\omega = 0.25$ (black dashed line), then $E(\theta_0) = 23.602 < 25.6124 = EC(\theta_2)$, so $\theta^* = \theta_0 = 1.845$; for $\omega = 0.7$ (green dashed line), then $E(\theta_0) = 26.446 > 25.6124 = EC(\theta_2)$, so $\theta^* = \theta_2 = 2.16$ (red solid line); for $\omega = 0.999999$ (blue dashed line), since $\theta_2 < \theta_0$, $\theta^* = \theta_0 = 2.59078$.

Similarly, with a larger population size ($N = 50$; see figure S1 in the electronic supplementary material, bottom row), we obtained $\beta^* = 3.15/1.03673 = 3.039$. In general, similar observations are obtained as in the case of a small population size $N = 3$. Except that, when N is large, the values of θ_0 for different non-extreme values of minimum required cooperation ω (say, $\omega \in (0.01, 0.99)$) are very small (given the log scale of $\omega/(1 - \omega)$ in the formula of ω_0). This value is also smaller than θ_0, with a cost $E(\theta_0) > E(\theta_2)$, making θ_2 the optimal cost of incentive. Similar results are obtained for PGG (figure 4). When ω is extremely high (i.e. greater than $1 - 10^{-k}$, for a large k) (we do not look at extremely low values since we would like to ensure at least a sufficient level of cooperation), then we can also see other scenarios where the optimal cost is θ_0 (see figure S1 in the electronic supplementary material, bottom row). We thus can observe that for $\omega \in (0.01, 0.99)$, for sufficiently large population size N and large enough β ($\beta > \beta^* +$ a bit more), then the optimal value of ω is always θ_2. Otherwise, θ_0 is the optimal cost.

Our last result provides a comparison of the expected total costs for providing institutional reward and punishment, for different individual incentive costs θ.
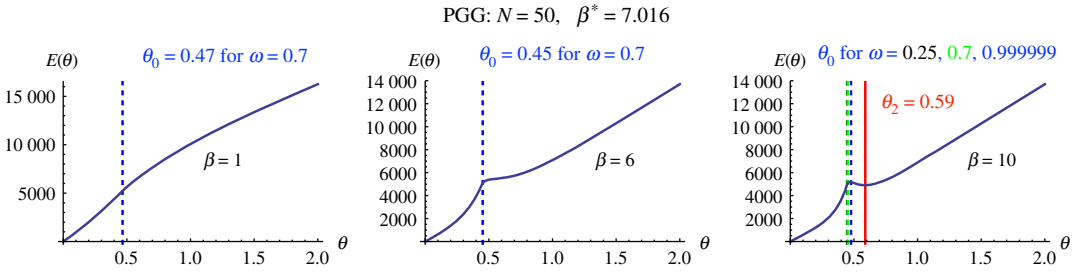
Figure 4. We use algorithm 3.3 to find optimal θ that minimizes $E(\theta)$ while ensuring a minimum level of cooperation ω, for PGG ($r = 3, n = 5, c = 1$) with $N = 50$. Similar observations to those for DG are obtained. (Online version in colour.)
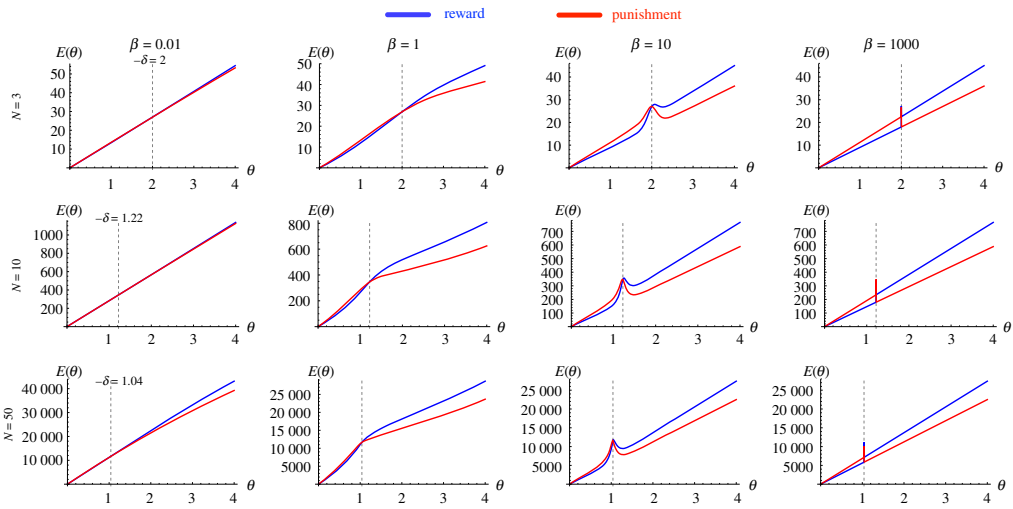
Figure 5. Comparison of the total costs E for reward and punishment as a function of θ, for different values of N and β. Reward is less costly than punishment ($E_r < E_p$) for small θ, and vice versa. The threshold of θ for this change was obtained analytically (see theorem 3.1), which is exactly equal to $-\delta$. Results are obtained for DG with $b = 2, c = 1$. (Online version in colour.)

Theorem 3.4 (reward versus punishment costs). *The difference between the expected total costs of reward and punishment is given by*

$$(E_r - E_p)(\theta) = \begin{cases} < 0, & \text{for } \theta < -\delta, \\ = 0, & \text{for } \theta = -\delta, \\ > 0, & \text{for } \theta > -\delta. \end{cases} \tag{3.8}$$
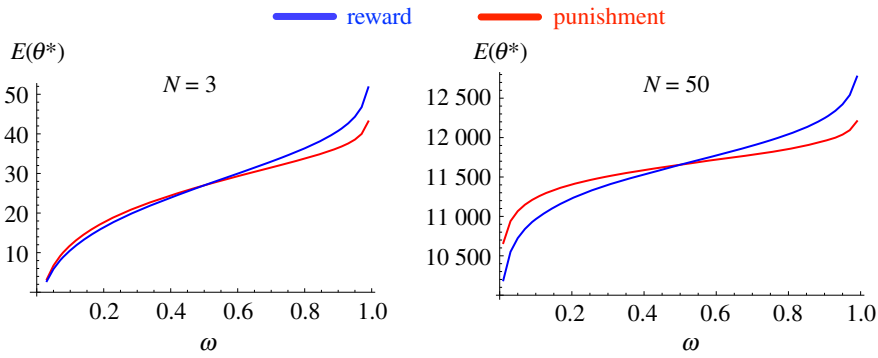
As a consequence, when $\beta \leq \min\{\beta_r^*, \beta_p^*\}$ we have

$$E_r^* = E_r(\theta_0) \quad \text{and} \quad E_p^* = E_p(\theta_0).$$

In this case,

$$(E_r^* - E_p^*) = E_r(\theta_0) - E_p(\theta_0) = \begin{cases} < 0 & \text{for } \omega < 0.5, \\ = 0 & \text{for } \omega = 0.5, \\ > 0 & \text{for } \omega > 0.5. \end{cases} \tag{3.9}$$

The proof of theorem 3.4 is given in §3 in the electronic supplementary material. Numerical calculation in figure 5 shows the expected total costs for reward and punishment (DG), for

Figure 6. Compare the total costs E for reward and punishment at the optimal value $\theta^\star$ (obtained using algorithm 3.3), by varying the minimum required level of cooperation, ω. Reward is more cost efficient for small ω, while punishment is more cost efficient when ω is larger. In both cases, the threshold is around $\omega = 0.5$. Other parameters: $\beta = 1$, DG with $b = 2, c = 1$. (Online version in colour.)

varying θ. We observe that reward is less costly than punishment ($E_r < E_p$) for $\theta < -\delta$ and vice versa when $\theta > -\delta$. It is exactly as shown analytically in theorem 3.4. This analytical result is confirmed here for different population size N and intensity of selection β. Figure 6 also confirms the second part of the theorem, where for small β, if one can choose the type of incentive to use, either reward or punishment, then the former can provide a lower cost when requiring less than 50% cooperation at minimum and the latter otherwise. This is in line with previous work showing that reward mechanisms work very well to promote cooperation in environments in which it is rare, while punishment mechanisms are better at maintaining high levels of cooperation (e.g. [28,35,52]).

4. Discussion

Institutional incentives such as punishment and reward provide an effective tool for promoting the evolution of cooperation in social dilemmas. Both theoretical and experimental analysis has been carried out [29,36,37,52,61–63]. However, past research usually ignores the question of how institutions' overall spending, i.e. the total cost of providing these incentives, can be minimized, while at the same time guaranteeing a minimum desired level of cooperation over time. Answering this question allows one to estimate exactly how incentives should be provided, that is, how much to reward a cooperator and how severely to punish a wrongdoer. Existing works that consider this question usually omit the stochastic effects that drive population dynamics, namely when the intensity of selection varies.

Resorting to a stochastic evolutionary game approach for finite, well-mixed populations, we have provided theoretical results for the optimal cost of incentives that ensure a desired level of cooperation while minimizing the total budget, for a given intensity of selection, β. We show that this cost strongly depends on the value of β, owing to the existence of a phase transition in the cost functions when β varies. This behaviour is missing in works that consider a deterministic evolutionary approach [35]. The intensity of selection plays an important role in evolutionary processes. Its value differs depending on the pay-off structure (i.e. scaling game pay-off matrix by a factor is equivalent to dividing β by that factor) and is usually found to be specific for a given population, which can be estimated through behavioural experiments [45–48]. Thus, our analysis provides a way to calculate the optimal incentive cost for a given population and game pay-off matrix at hand.

With regard to theoretical importance, we characterized asymptotic behaviours of the total cost functions for both reward and punishment (namely, in the limits of a large population, weak

selection and strong selection) and compared these functions for the two types of incentive. We showed that punishment is always more costly for a small (individual) incentive cost (θ) but less so when this cost is above a certain threshold. We provided an exact formula for this threshold. This result provides insights into the choice of which type of incentives to use.

In the context of institutional incentives modelling, a crucial issue is the question of how to maintain the budget for providing incentives [59,64]. The problem of who pays or contributes to the budget is a social dilemma in itself, and how to escape this dilemma is a critical research question. In this work, we focus on the question of how to optimize the budget used for the provided incentives.

There are several simplifications made for the theoretical analysis to be possible. First, in order to derive the analytical formula for the frequency of cooperation, we assumed the small mutation limit. Despite the simplified assumption, this small mutation limit approach has been shown to be widely applicable to scenarios which go well beyond the strict limit of very small mutation rates [46,48,59]. Relaxing this assumption would make the derivation of a close form for the frequency of cooperation intractable.

Second, we focused in this paper on two important cooperation dilemmas, the DG and the PGG. They have in common a useful property that the difference in (average) pay-off between a cooperator and a defector, $\delta = \Pi_C(i) - \Pi_D(i)$, does not depend on i, the number of cooperators in the population. This property allows us to simplify the fundamental matrix to a tridiagonal form and apply the techniques of matrix analysis to obtain a close form of its inverse matrix (see electronic supplementary material). In games with more complex pay-off matrices such as the PD in its general form and the collective risk game [65], the difference δ depends on i and the technique in this paper cannot be directly applied. We might consider other approaches to approximate the inverse matrix, exploiting its block structure.

References

1. Han TA. 2013 *Intention recognition, commitments and their roles in the evolution of cooperation: from artificial intelligence techniques to evolutionary game theory models*, vol. 9. Springer SAPERE Series. Berlin, Germany: Springer Verlag.

2. Nowak MA. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560. (doi:10.1126/science.1133755)

3. Perc M, Jordan JJ, Rand DG, Boccaletti S, Szolnoki A. 2017 Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51. (doi:10.1016/j.physrep.2017.05.004)

4. Sigmund K. 2010 *The calculus of selfishness*. Princeton, NJ: Princeton University Press.

5. West SA, Griffin AA, Gardner A. 2007 Evolutionary explanations for cooperation. *Curr. Biol.* **17**, R661–R672. (doi:10.1016/j.cub.2007.06.004)

6. Hamilton WD. 1964 The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16. (doi:10.1016/0022-5193(64)90038-4)

7. Traulsen A, Nowak MA. 2006 Evolution of cooperation by multilevel selection. *Proc. Natl Acad. Sci. USA* **103**, 10952. (doi:10.1073/pnas.0602530103)

8. Han TA, Pereira LM, Santos FC. 2012 Corpus-based intention recognition in cooperation dilemmas. *Artif. Life* **18**, 365–383. (doi:10.1162/ARTL_a_00072)

9. Krellner M, Han TA. 2020 Putting oneself in everybody's shoes—pleasing enables indirect reciprocity under private assessments. In *Proc. ALIFE 2020: The 2020 Conf. on Artificial Life, Virtual, 13–18 July 2020*, pp. 402–410. Cambridge, MA: MIT Press.

10. Nowak MA, Sigmund K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131)

11. Ohtsuki H, Iwasa Y. 2006 The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **239**, 435–444. (doi:10.1016/j.jtbi.2005.08.008)
12. Okada I. 2020 A review of theoretical studies on indirect reciprocity. *Games* **11**, 27. (doi:10.3390/g11030027)
13. Antonioni A, Cardillo A. 2017 Coevolution of synchronization and cooperation in costly networked interactions. *Phys. Rev. Lett.* **118**, 238301. (doi:10.1103/PhysRevLett.118.238301)
14. Peña J, Wu B, Arranz J, Traulsen A. 2016 Evolutionary games of multiplayer cooperation on graphs. *PLoS Comput. Biol.* **12**, e1005059. (doi:10.1371/journal.pcbi.1005059)
15. Perc M, Gómez-Gardeñes J, Szolnoki A, Floría LM, Moreno Y. 2013 Evolutionary dynamics of group interactions on structured populations: a review. *J. R. Soc. Interface* **10**, 20120997. (doi:10.1098/rsif.2012.0997)
16. Santos FC, Pacheco JM, Lenaerts T. 2006 Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc. Natl Acad. Sci. USA* **103**, 3490–3494. (doi:10.1073/pnas.0508201103)
17. Boyd R, Gintis H, Bowles S, Richerson PJ. 2003 The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535. (doi:10.1073/pnas.0630443100)
18. Boyd R, Gintis H, Bowles S. 2010 Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **328**, 617–620. (doi:10.1126/science.1183665)
19. Fehr E, Gachter S. 2000 Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994. (doi:10.1257/aer.90.4.980)
20. Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K. 2007 Via freedom to coercion: the emergence of costly punishment. *Science* **316**, 1905–1907. (doi:10.1126/science.1141588)
21. Herrmann B, Thöni C, Gächter S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–1367. (doi:10.1126/science.1153808)
22. Sigmund K, Hauert C, Nowak M. 2001 Reward and punishment. *Proc. Natl Acad. Sci. USA* **98**, 10 757–10 762. (doi:10.1073/pnas.161155698)
23. Han TA, Pereira LM, Santos FC, Lenaerts T. 2013 Good agreements make good friends. *Sci. Rep.* **3**, 2695. (doi:10.1038/srep02695)
24. Han TA, Pereira LM, Lenaerts T. 2016 Evolution of commitment and level of participation in public goods games. *Auton. Agents Multi-Agent Syst.* **31**, 561–583. (doi:10.1007/s10458-016-9338-4)
25. Martinez-Vaquero LA, Han TA, Pereira LM, Lenaerts T. 2017 When agreement-accepting free-riders are a necessary evil for the evolution of cooperation. *Sci. Rep.* **7**, 1–9. (doi:10.1038/s41598-016-0028-x)
26. Nesse RM. 2001 *Evolution and the capacity for commitment*. Foundation Series on Trust. New York, NY: Russell Sage.
27. Sasaki T, Okada I, Uchida S, Chen X. 2015 Commitment to cooperation and peer punishment: its evolution. *Games* **6**, 574–587. (doi:10.3390/g6040574)
28. Chen X, Sasaki T, Brännström Å, Dieckmann U. 2015 First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation. *J. R. Soc. Interface* **12**, 20140935. (doi:10.1098/rsif.2014.0935)
29. García J, Traulsen A. 2019 Evolution of coordinated punishment to enforce cooperation from an unbiased strategy space. *J. R. Soc. Interface* **16**, 20190127. (doi:10.1098/rsif.2019.0127)
30. Góis AR, Santos FP, Pacheco JM, Santos FC. 2019 Reward and punishment in climate change dilemmas. *Sci. Rep.* **9**, 1–9. (doi:10.1038/s41598-019-52524-8)
31. Han TA, Tran-Thanh L. 2018 Cost-effective external interference for promoting the evolution of cooperation. *Sci. Rep.* **8**, 1–9. (doi:10.1038/s41598-018-34435-2)
32. Powers ST, Ekárt A, Lewis PR. 2018 Modelling enduring institutions: the complementarity of evolutionary and agent-based approaches. *Cogn. Syst. Res.* **52**, 67–81. (doi:10.1016/j.cogsys.2018.04.012)
33. Sigmund K, De Silva H, Traulsen A, Hauert C. 2010 Social learning promotes institutions for governing the commons. *Nature* **466**, 7308. (doi:10.1038/nature09203)
34. Vasconcelos VV, Santos FC, Pacheco JM. 2013 A bottom-up institutional approach to cooperative governance of risky commons. *Nat. Clim. Change* **3**, 797. (doi:10.1038/nclimate1927)
35. Wang S, Chen X, Szolnoki A. 2019 Exploring optimal institutional incentives for public cooperation. *Commun. Nonlinear Sci. Numer. Simul.* **79**, 104914. (doi:10.1016/j.cnsns.2019.104914)

36. Wu J-J, Li C, Zhang B-Y, Cressman R, Tao Yi. 2014 The role of institutional incentives and the exemplar in promoting cooperation. *Sci. Rep.* **4**, 6421. (doi:10.1038/srep06421)
37. Bardhan P. 2005 Institutions matter, but which ones? *Econ. Transit.* **13**, 499–532. (doi:10.1111/ecot.2005.13.issue-3)
38. Bowles S. 2009 *Microeconomics: behavior, institutions, and evolution*. Princeton, NJ: Princeton University Press.
39. Bowles S, Gintis H. 2002 Social capital and community governance. *Econ. J.* **112**, F419–F436. (doi:10.1111/1468-0297.00077)
40. Han TA, Pereira LM, Lenaerts T, Santos FC. 2021 Mediating artificial intelligence developments through negative and positive incentives. *PLoS ONE* **16**, e0244592. (doi:10.1371/journal.pone.0244592)
41. Ostrom E. 1990 *Governing the commons: the evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.
42. Scotchmer S. 2004 *Innovation and incentives*. Cambridge, MA: MIT Press.
43. Hofbauer J, Sigmund K. 1998 *Evolutionary games and population dynamics*. Cambridge, UK: Cambridge University Press.
44. Traulsen A, Nowak MA, Pacheco JM. 2006 Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 11909. (doi:10.1103/PhysRevE.74.011909)
45. Domingos EF, Grujić J, Burguillo JC, Kirchsteiger G, Santos FC, Lenaerts T. 2020 Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *Iscience* **23**, 101752. (doi:10.1016/j.isci.2020.101752)
46. Rand DG, Tarnita CE, Ohtsuki H, Nowak MA. 2013 Evolution of fairness in the one-shot anonymous ultimatum game. *Proc. Natl Acad. Sci. USA* **110**, 2581–2586. (doi:10.1073/pnas.1214167110)
47. Traulsen A, Semmann D, Sommerfeld RD, Krambeck H-J, Milinski M. 2010 Human strategy updating in evolutionary games. *Proc. Natl Acad. Sci. USA* **107**, 2962–2966. (doi:10.1073/pnas.0912515107)
48. Zisis I, Di Guida S, Han TA, Kirchsteiger G, Lenaerts T. 2015 Generosity motivated by acceptance—evolutionary analysis of an anticipation games. *Sci. Rep.* **5**, 18076. (doi:10.1038/srep18076)
49. Cimpeanu T, Han TA, Santos FC. 2019 Exogenous rewards for promoting cooperation in scale-free networks. In *Proc. of the 2018 Conf. on Artificial Life: A Hybrid of the European Conf. on Artificial Life (ECAL) and the Int. Conf. on the Synthesis and Simulation of Living Systems (ALIFE)*, pp. 316–323. Cambridge, MA: MIT Press.
50. Cimpeanu T, Perret C, Han TA. 2021 Promoting fair proposers, fair responders or both? Cost-efficient interference in the spatial ultimatum game. In *Proc. of the 20th Int. Conf. on Autonomous Agents and MultiAgent Systems, London (Virtual Event), 3–7 May 2021*, pp. 1480–1482. International Foundation for Autonomous Agents and Multiagent Systems.
51. Han TA, Lynch S, Tran-Thanh L, Santos FC. 2018 Fostering cooperation in structured populations through local and global interference strategies. In *Proc. IJCAI-ECAI'2018, Stockholm, Sweden, 13–19 July 2018*, pp. 289–295. International Joint Conferences on Artificial Intelligence.
52. Sasaki T, Brännström Å, Dieckmann U, Sigmund K. 2012 The take-it-or-leave-it option allows small penalties to overcome social dilemmas. *Proc. Natl Acad. Sci. USA* **109**, 1165–1169. (doi:10.1073/pnas.1115219109)
53. Imhof LA, Fudenberg D, Nowak MA. 2005 Evolutionary cycles of cooperation and defection. *Proc. Natl Acad. Sci. USA* **102**, 10 797–10 800. (doi:10.1073/pnas.0502589102)
54. Nowak MA. 2006 *Evolutionary dynamics: exploring the equations of life*. Cambridge, MA: Harvard University Press.
55. Nowak MA, Sasaki A, Taylor C, Fudenberg D. 2004 Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650. (doi:10.1038/nature02414)
56. Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K. 2007 Via freedom to coercion: the emergence of costly punishment. *Science* **316**, 1905–1907. (doi:10.1126/science.1141588)
57. Kemeny J, Snell J. 1976 *Finite Markov chains*. Undergraduate Texts in Mathematics. Berlin, Germany: Springer.
58. Fudenberg D, Imhof LA. 2005 Imitation processes with small mutations. *J. Econ. Theory* **131**, 251–262. (doi:10.1016/j.jet.2005.04.006)
59. Sigmund K, De Silva H, Traulsen A, Hauert C. 2010 Social learning promotes institutions for governing the commons. *Nature* **466**, 861–863. (doi:10.1038/nature09203)

60. Karlin S, Taylor HE. 1975 *A first course in stochastic processes*. New York, NY: Academic Press.
61. Baldassarri D, Grossman G. 2011 Centralized sanctioning and legitimate authority promote cooperation in humans. *Proc. Natl Acad. Sci. USA* **108**, 11 023–11 027. (doi:10.1073/pnas.1105456108)
62. Dong Y, Sasaki T, Zhang B. 2019 The competitive advantage of institutional reward. *Proc. R. Soc. B* **286**, 20190001. (doi:10.1098/rspb.2019.0001)
63. Gürerk Ö, Irlenbusch B, Rockenbach B. 2006 The competitive advantage of sanctioning institutions. *Science* **312**, 108–111. (doi:10.1126/science.1123633)
64. Hilbe C, Traulsen A, Röhl T, Milinski M. 2014 Democratic decisions establish stable authorities that overcome the paradox of second-order punishment. *Proc. Natl Acad. Sci. USA* **111**, 752–756. (doi:10.1073/pnas.1315273111)
65. Santos FC, Pacheco JM. 2011 Risk of collective failure provides an escape from the tragedy of the commons. *Proc. Natl Acad. Sci. USA* **108**, 10 421–10 425. (doi:10.1073/pnas.1015648108)