Supplementary Materials

Commitment to Cooperation and Peer Punishment: Its Evolution

Tatsuya Sasaki, Isamu Okada, Satoshi Uchida and Xiaojie Chen

This file includes:
Supplementary Text, Sections S1 and S2
Supplementary Figures, Figure S1

S1. Replicator Dynamics in Donation Games with Deposit-Based Commitment

We analyze the replicator dynamics in the case of deposit-based commitment to costly peer punishment. Let x_S and P_S be, respectively, the frequency and expected payoff value for strategy S with S = ALLC (cooperator), ALLD (defector) and COM (faithful committer). The replicator dynamics are described as, $\dot{x}_S = x_S(P_S - \overline{P})$, where $\overline{P} = \sum_S x_S P_S$ is the average payoff over the population. Indeed,

$$\begin{aligned} \dot{x}_{ALLC} &= x_{ALLC}(P_{ALLC} - \overline{P}), \\ \dot{x}_{ALLD} &= x_{ALLD}(P_{ALLD} - \overline{P}), \\ \dot{x}_{COM} &= x_{COM}(P_{COM} - \overline{P}). \end{aligned} \tag{S1}$$

From Table 2 (in the main text),

$$\begin{aligned} P_{ALLC} &= -f_0 x_{COM} - c, \\ P_{ALLD} &= -f_0 x_{COM} - f_1 x_{COM}, \\ P_{COM} &= -g_0(x_{ALLC} + x_{ALLD}) - g_1 x_{ALLD} - c, \end{aligned} \tag{S2}$$

in which, for simplicity, we removed a common term $b(x_{ALLC} + x_{COM})$ for the expected benefit (the replicator dynamics are not affected by such a transformation). Thus,

$$\begin{aligned} \overline{P} &= x_{ALLC}P_{ALLC} + x_{ALLD}P_{ALLD} + x_{COM}P_{COM} \\ &= -(f_0 + g_0)x_{COM}(1 - x_{COM}) - (f_1 + g_1)x_{ALLD}x_{COM} - c(1 - x_{ALLD}). \end{aligned} \tag{S3}$$

From Equations (S2) and (S3), it follows that when $x_{ALLD} = 0$ (that is, on the COM-ALLC edge),

$$\begin{aligned} \dot{x}_{COM} &= x_{COM}(P_{COM} - \overline{P}) \\ &= x_{COM}(1 - x_{COM})(P_{COM} - P_{ALLC}) \\ &= x_{COM}(1 - x_{COM})((f_0 + g_0)x_{COM} - g_0). \end{aligned} \tag{S4}$$

$$\therefore P_{COM} = P_{ALLC} \;\Leftrightarrow\; x_{COM} = \frac{g_0}{f_0 + g_0} \;\; (=: x^*_{COM}). \tag{S5}$$

Let R_2 be the boundary point on the edge with $x_{ALLD} = 0$ and $x_{COM} = x^*_{COM}$. Following the replicator dynamics in Equation (S4), thus, the point R_2 divides the edge into the basins of attraction for the COM and ALLC nodes.

Similarly, when $x_{ALLC} = 0$ (that is, on the COM-ALLD edge),

$$\begin{aligned}\dot{x}_{COM} &= x_{COM}(P_{COM} - \bar{P}) \\ &= x_{COM}(1 - x_{COM})(P_{COM} - P_{ALLD}) \\ &= x_{COM}(1 - x_{COM})((f_0 + f_1 + g_0 + g_1)x_{COM} - g_0 - g_1 - c).\end{aligned} \tag{S6}$$

$$\therefore P_{COM} = P_{ALLD} \Leftrightarrow x_{COM} = \frac{g_0 + g_1 + c}{f_0 + f_1 + g_0 + g_1} \; (=: x^{**}_{COM}). \tag{S7}$$

Let R1 be the boundary point on the edge with $x_{ALLC} = 0$ and $x_{COM} = x^{**}_{COM}$. Following Equation (S6), thus, the point R1 divides the edge into basins of attraction for the COM and ALLD nodes.

Then, Equation (2) also yields,

$$\therefore P_{ALLC} = P_{ALLD} \Leftrightarrow x_{COM} = \frac{c}{f_1}, \tag{S8}$$

and:

$$\therefore P_{COM} = P_{ALLC} \Leftrightarrow (f_0 + g_0)x_{COM} = g_1 x_{ALLD} + g_0. \tag{S9}$$

Therefore, solving Equations (S8) and (S9) yields a unique interior equilibrium $\mathrm{Q} = (x_{ALLC}, x_{ALLD}, x_{COM})$ such that:

$$x_{COM} = \frac{c}{f_1} \text{ and } x_{ALLD} = \frac{1}{g_1}\left(\frac{c(f_0 + g_0)}{f_1} - g\right). \tag{S10}$$

The interior equilibrium Q is a saddle point. To check this, we shall show that the Jacobian matrix at Q, J_{Q}, has a negative determinant, $\det J_{\mathrm{Q}} < 0$. We introduce a new coordinate system $(w, z) = \left(\frac{x_{ALLC}}{x_{ALLC} + x_{ALLD}}, x_{COM}\right)$. This yields, for (w, z) on the open square $]0,1[^2$,

$$\begin{aligned}\dot{w} &= w(1 - w)(-f_1 z + c), \\ \dot{z} &= z(P_{COM} - \bar{P}).\end{aligned} \tag{S11}$$

In Equation (S11),

$$\begin{aligned}P_{COM} - \bar{P} &= (1 - z)((f_0 + g_0)z - g_0) + \underbrace{(1 - w)(1 - z)}_{= x_{ALLD}}((f_1 + g_1)z - g_1 - c) \\ &= (1 - z)(A(z) + wB(z)),\end{aligned} \tag{S12}$$

where:

$$\begin{aligned}A(z) &:= (f_0 + g_0)z - g_0, \\ B(z) &:= -(f_1 + g_1)z + g_1 + c.\end{aligned} \tag{S13}$$

From Equation (10), at $\mathrm{Q} = (w_{\mathrm{Q}}, z_{\mathrm{Q}})$,

$$B(z_{\mathrm{Q}}) = -(f_1 + g_1)z_{\mathrm{Q}} + g_1 + c = g_1\left(1 - \frac{c}{f_1}\right) < 0. \tag{S14}$$

and thus,

$$J_{Q=(w_Q,z_Q)} = \begin{pmatrix} 0 & f_1 w_Q(1-w_Q) \\ z_Q(1-z_Q)B(z_Q) & z_Q(1-z_Q)(A'(z_Q)+w_Q B'(z_Q)) \end{pmatrix} = \begin{pmatrix} 0 & + \\ + & (*) \end{pmatrix}. \tag{S15}$$

Whatever the value of the lower diagonal component, therefore, we obtain that $\det J_Q < 0$.

S2. Replicator Dynamics for ALLC, ALLD, PEER and COM

Competition in peer punishment (PEER) and deposit-based commitment (COM) (Figure S1): We analyze the replicator dynamics for ALLC, ALLD, PEER and COM. Since playing ALLC is clearly better off than playing PEER by saving the punishing fee, whatever others do, all interior orbits in the state space will converge to the boundaries of the state space (see Figure S1A). Similarly, ALLD dominates ALLC. Dynamics for other pairs of strategies, PEER and ALLD, COM and PEER, COM and ALLC, COM and ALLD, each can be bistable. This leads to separating the state space into three basins of attractions for attractors: COM node, ALLD node and PEER-adjacent segment of the PEER-ALLC edge (a continuum of fixed points). We note that there is no interior fixed point in the three-dimensional replicator system. Restricted to the ALLD-PEER-COM and ALLC-PEER-COM surfaces, yet, interior fixed points (on the surfaces) are possible, because punishment of ALLC, ALLD and PEER by COM can make the payoffs for the three strategies equal one another.

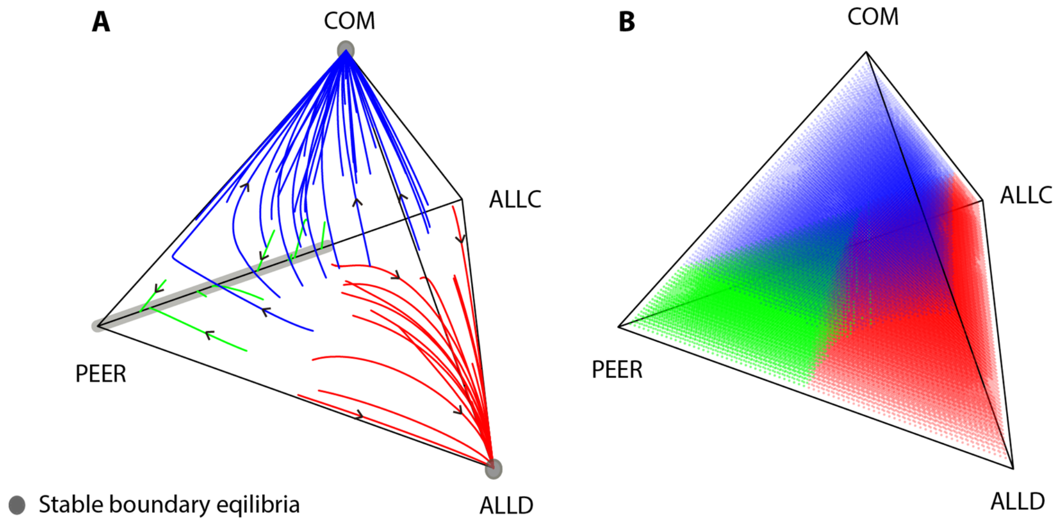

Figure S1. Competition of peer punishment (PEER) and deposit-based commitment (COM). (**A**) Interior trajectories are attracted either to a homogeneous state of COM (COM-node) or a homogeneous state of ALLD (ALLD node), a mixed state of PEER and ALLC (segment of the PEER-ALLC edge, adjacent to the PEER node); (**B**) colored regions correspond to three different basins of attraction, respectively, for the COM-node (blue, relative size: 40%), ALLD-node (red, 43%) and PEER-ALLC edge (green, 17%), in the parameter specific settings. Parameters are as in Figure 2.