Supporting Information

Wu et al. 10.1073/pnas.0905918106

SI Text

As reported in the main text, the primary objective of our experiments on the pairwise repeated Prisoner's Dilemma game is to determine the effect that the option of costly punishment has on the level of cooperation. Because our experimental design is the same as Dreber et al. (1), it is also interesting to see whether our data match their conclusion that participants who gain higher payoffs tend not to use punishment in the two treatments. This conclusion is shown through Figs. S1–S3. First, Fig. S1 divides the participants in each treatment (24 for T1 and 22 for T2) into six groups according to their payoff (with lower ranked groups corresponding to higher payoffs). Similar to Fig. 3 in Dreber et al. (1), there is a clear trend; players with lower rank (higher payoffs) punish less than players with higher rank (lower payoffs).

Second, Fig. S2*a* shows the frequency of cooperation in punishers ($P > 0$), and in nonpunishers ($P = 0$) whereas Fig. S2*b* shows that the average payoffs of punishers ($P > 0$) is less than that of nonpunishers ($P = 0$). These results also match those of Dreber et al. (1) (see their supplementary figure 4). Finally, Fig. S3 shows that average payoffs decrease in those interactions that have a higher use of costly punishment immediately after the opponent defects, matching the result shown in figure 4 in Dreber et al. (1).

From Figs. S1–S3, there is a strong negative correlation between average payoff and the use of costly punishment. We conclude that, although cultural differences in attitude to costly punishment may affect the level of cooperation, these differences do not change the outcome that winners don't punish.

Another common result in many experiments (e.g., refs 1–2), is that the rate of defection increases in later rounds of the repeated PD game. In our two control experiments, the prevalence of defection increases as expected over the first six rounds of a single interaction as seen in Fig. S4*a* and confirmed by Pearson χ^2 tests ($P = 0.001$ in C1 and $P < 0.001$ in C2). The regression lines for the frequency D(n) of defection in round n for C1 and C2 are D(n) = 69.3%+(2.4%)n and D(n) = 71.0%+(2.6%)n respectively with standard deviations 4.68% and 5.85% respectively. (Although some interactions lasted nine rounds, there is limited data for rounds seven to nine because the average number of rounds in our experiments is only 3.6.) Moreover, in T2, there is a significant increase in defection over the first six rounds (Pearson χ^2 test has $P = 0.043$), although not as significant as in the two controls. Only in T1 was there no significant increase (Pearson χ^2 test has $P = 0.643$) as is clear from Fig. S4*b* where the defection level is basically constant.

Finally, we examine whether the higher frequency of first P use during round 1 in our experiments compared with those of Dreber et al. (1) is a stable phenomenon over the course of a session or a transient behavior. From Fig. S5, the regression lines for the frequencies $P(n)$, $C(n)$ and $D(n)$ of costly defection, cooperation and defection respectively in round 1 of interaction n for T1 are $P(n) = 7.4\% - (0.2\%)n$, $C(n) = 25.6\% - (0.6\%)n$, $D(n) = 67\% + (0.8\%)n$ with standard deviations 2%, 3.2%, and 4% respectively. For T2, these are $P(n) = 5.7\% - (0.2\%)n$, $C(n) = 36.9\% - (0.5\%)n$, $D(n) = 57.4\% + (0.7\%)n$ with standard deviations 1.5%, 3.6%, 3.6% respectively. Because the regression lines are nearly constant, there are no trends in T1 and T2. In particular, the frequency of first P use during round 1 is maintained throughout the entire session in both T1 and T2. Furthermore, cross tabulation (a type of multidimensional Pearson χ^2 test) tests the significance between the individual choices [which is a 3-dim variable ($C\%$, $D\%$, $P\%$)] and the interactions. $P = 0.589$ in T1 and $P = 0.665$ in T2 and so there are no significant relations between individual choices and interactions in the first round of both experiments.

1. Dreber A, Rand DG, Fudenberg D, Nowak MA (2008) Winners don't punish. *Nature* 452:348–351.

2. Selten R, Stoecker R (1986) End behavior in sequences of finite prisoner's dilemma supergames. *J Econ Behav Org* 7:47–70.

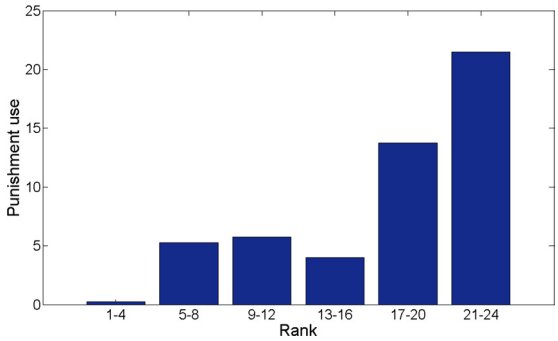

Fig. S1. Average punishment use according to payoff rank. In both T1 (Fig. S1*a*) and T2 (Fig. S1*b*), participants with lower rank (higher payoff) punish less than participants with high rank.

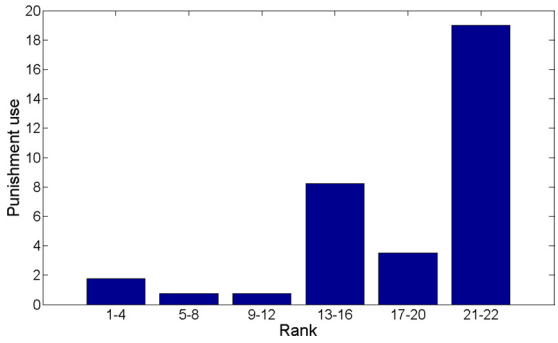

Fig. S1 (continued).

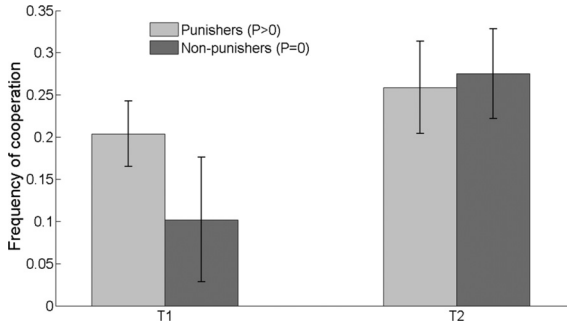

Fig. S2. Average frequency of cooperation and average payoff for punishers and nonpunishers. In *a*, the average frequency of cooperation in punishers ($P > 0$) is bigger than in nonpunishers in T1. However, there is no significant difference [$P = 0.0945$ and $z = -1.6722$ (Mann–Whitney test)]. In T2, the frequency of cooperation in punishers ($P > 0$) is slightly less than in nonpunishers but no significant change with $P = 0.5618$ and $z = 0.5801$. In *b*, the average payoff of punishers ($P > 0$) in T1 is negative but the average payoff of nonpunishers ($P = 0$) is positive. Here, the Mann–Whitney test has $P = 0.0229$ and $z = 2.2746$. In T2, the average payoff of punishers ($P > 0$) is less than the average payoff of nonpunishers ($P = 0$) with $P = 0.0222$ and $z = 2.2865$.

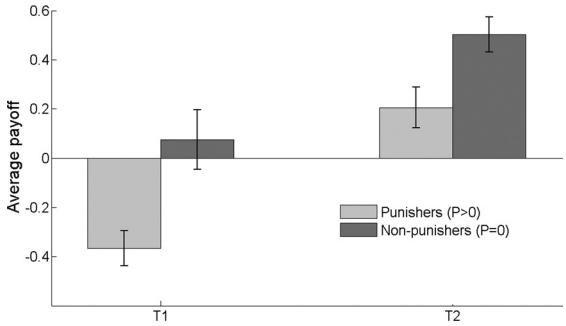

Fig. S2 (continued).

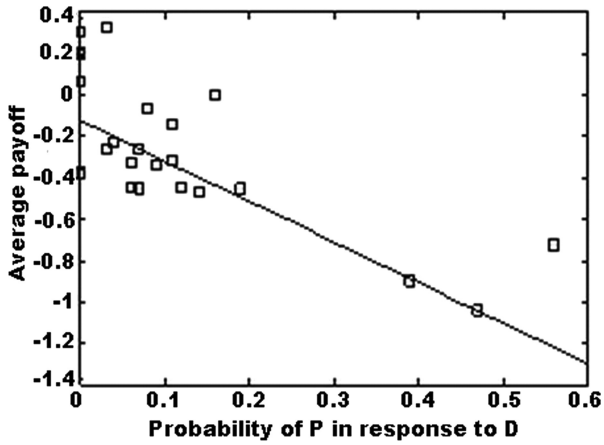

Fig. S3. Correlation between average payoff and probability of P in response to D in the treatments. The average payoff vs. probability P in response to D in T1 (panel a) has slope $= -0.01058$ with $P < 0.0001$. The average payoff vs. probability P in response to D in T2 (panel b) has slope $= -0.02428$ with $P < 0.0001$.

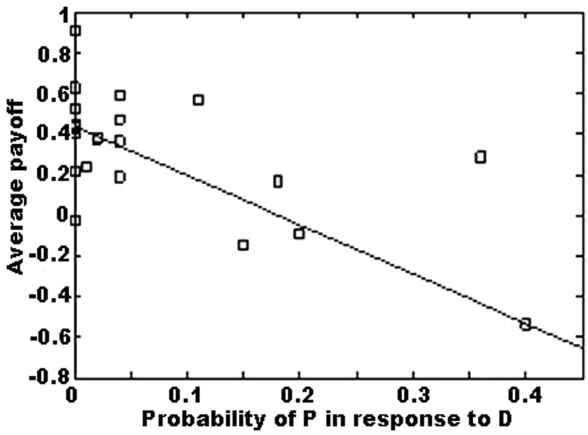

Fig. S3 (continued).

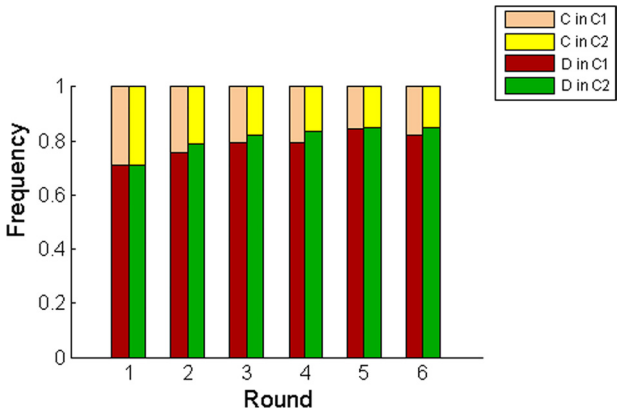

Fig. S4. Average strategy frequencies from rounds 1 to 6 in our four experiments. *a* shows the frequencies of D and C in C1 and C2. *b* shows the frequencies of D, P and C in T1 and T2.

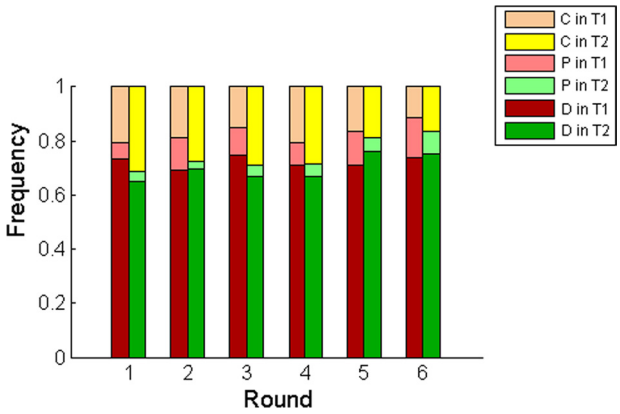

Fig. S4 (continued).

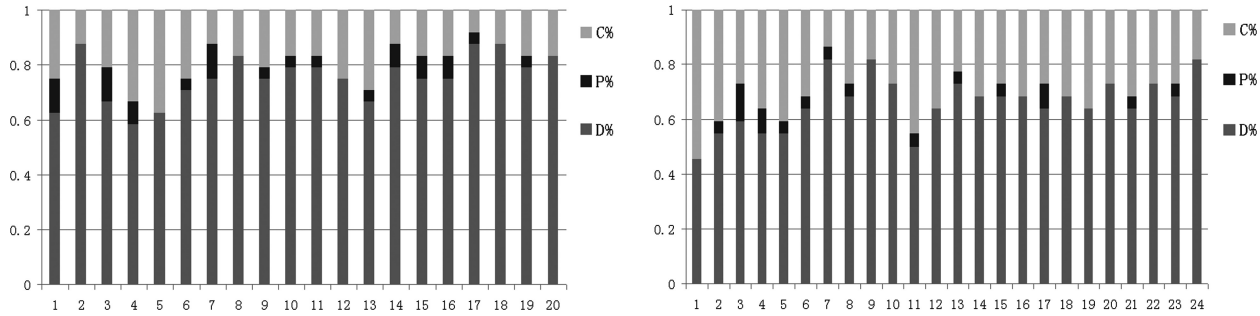

Fig. S5. Frequencies of D, P and C in round 1 of each interaction in T1 (*a*) and T2 (*b*). There are 20 interactions for each subject in T1 and 24 in T2. The average combined frequency of P and D in the first round is 79.9% in T1 and 69.1% in T2.