Appendix

Belonging to

Do descriptive social norms drive peer punishment?

Conditional punishment strategies and their impact on cooperation

By

Xueheng Li, Lucas Molleman and Dennie van Dolder

Contents

1. Supplementary Figures

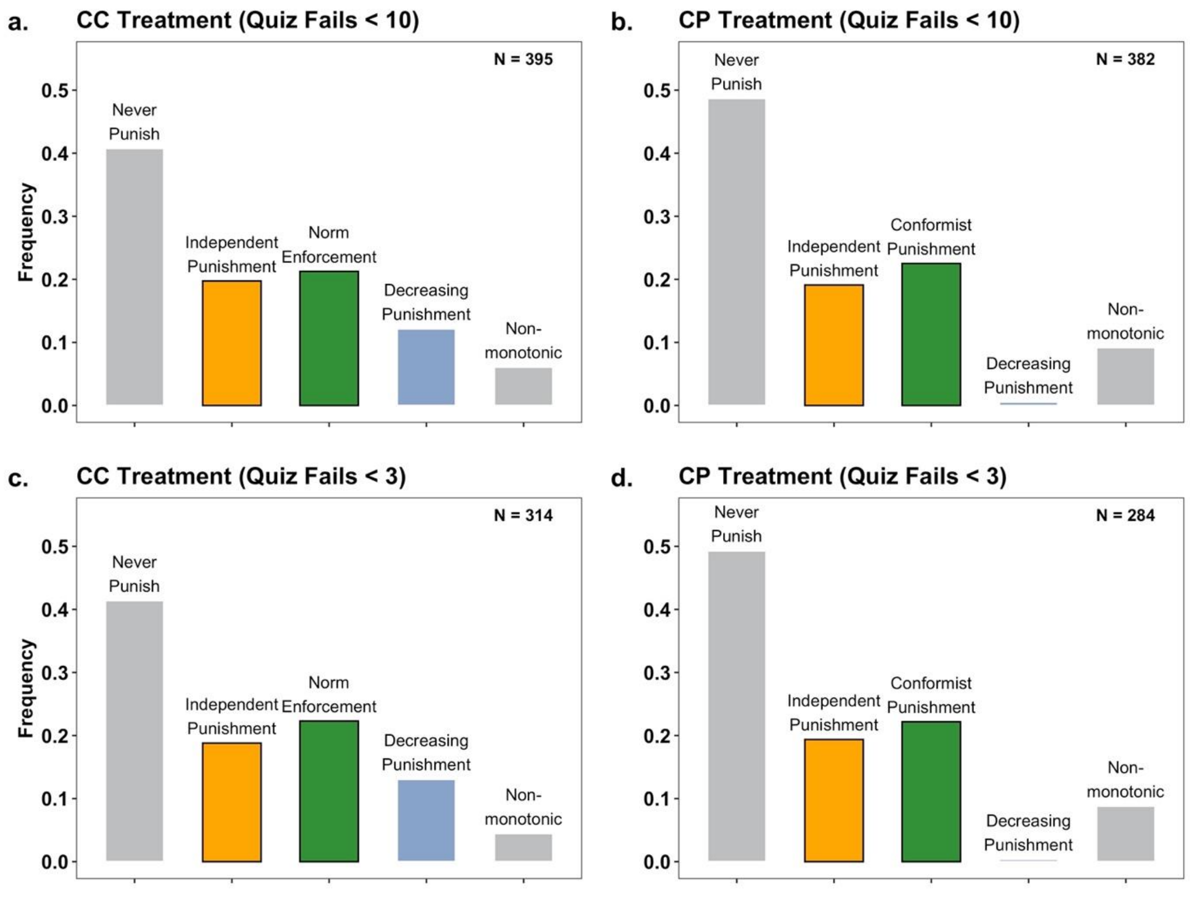


Fig. A1. Distributions of punishment strategies in each experiment, excluding participants with more than 10 Quiz Fails (**a** and **b**) and those more than 3 Quiz Fails (**c** and **d**), respectively. These plots suggest that the distribution of strategies remains stable when we consider different subsets of participants based on their number of attempts on the control questions.

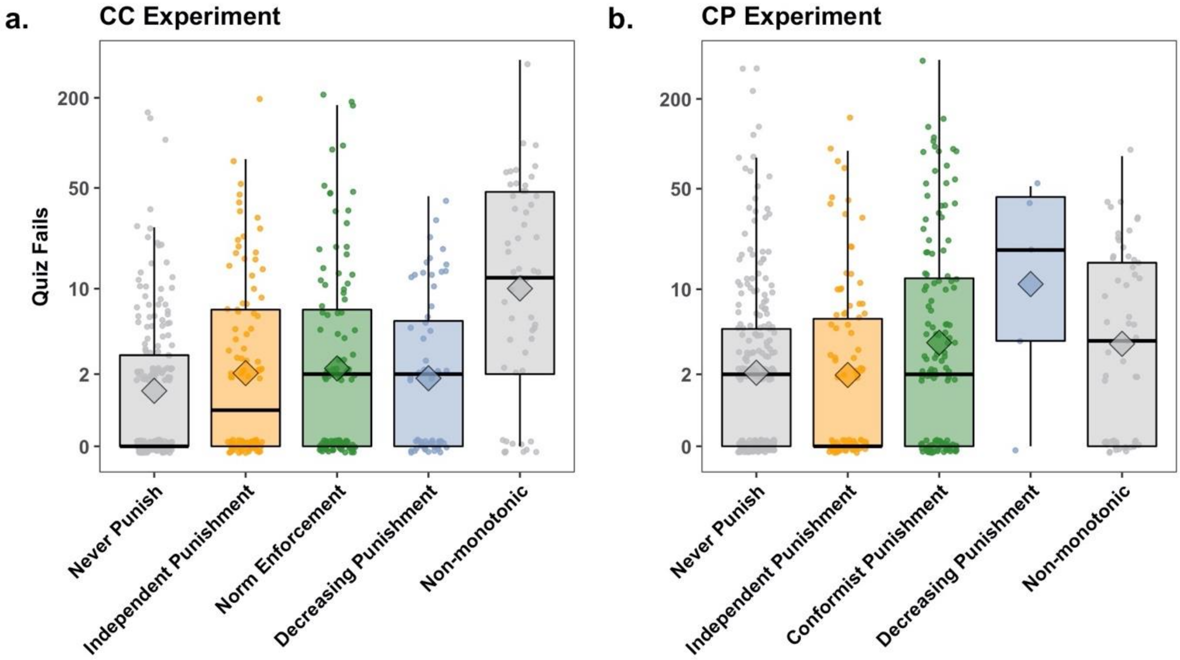


Fig. A2. Distributions of failed attempts at the compulsory control questions, broken down by punishment strategy in each experiment. Each jittered data point shows a participant. The y-axis is displayed in log scale to account for outliers with many attempts before passing the 9 control questions (7 of which are open questions about game payoffs). Diamond symbols show the means.

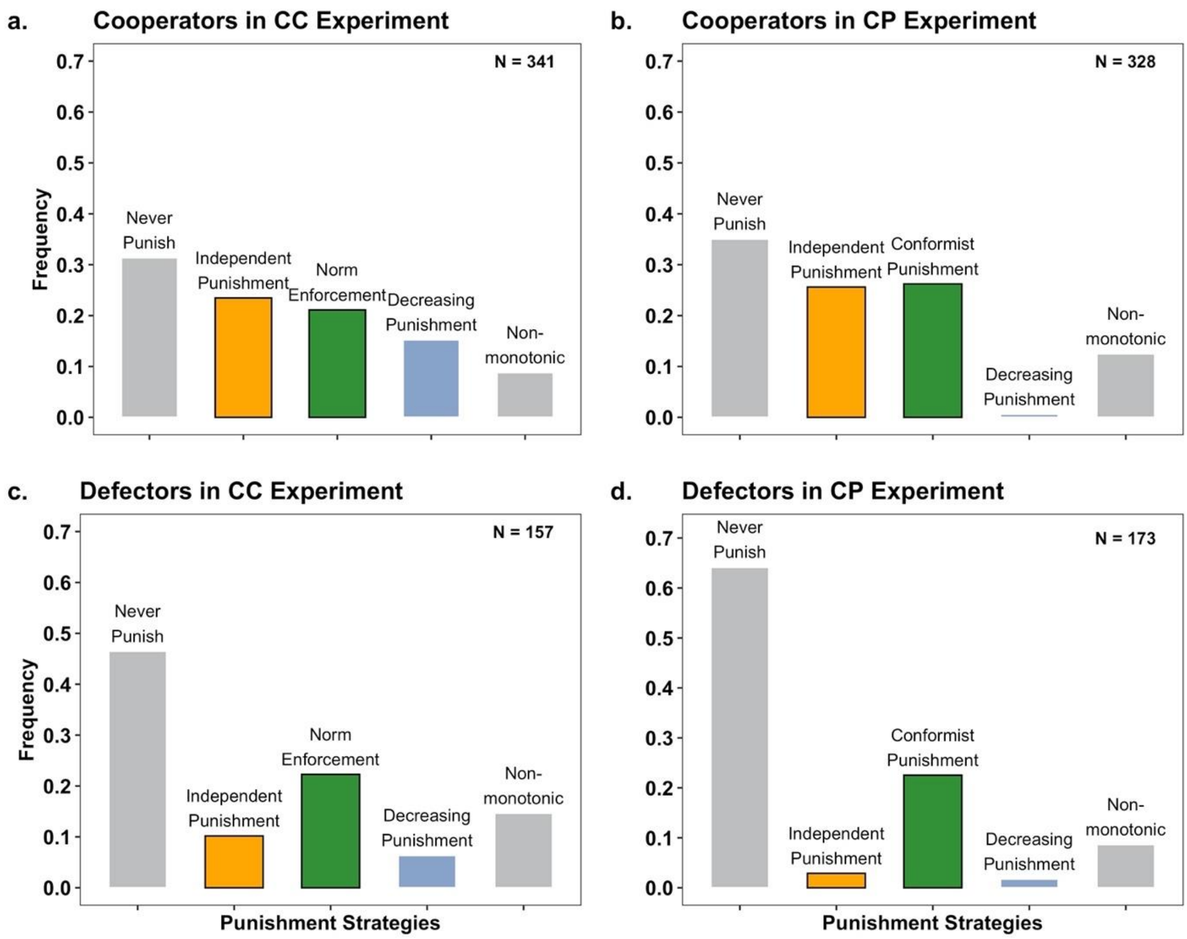


Fig. A3. Distributions of punishment strategies in each experiment, broken down by cooperation decision in Stage 1.

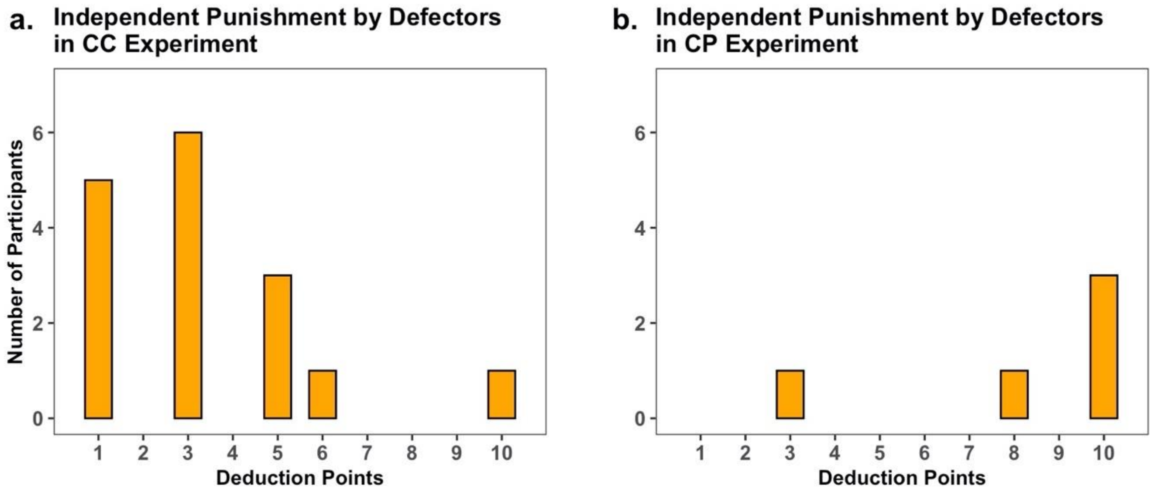


Fig. A4. Distribution of deduction points for independent punishers who defected in Stage 1 in each experiment. As in the equivalent figure showing behavior of independent punishers who cooperated in Stage 1 (main text Fig. 3a,b), vertical axes show counts. Note that, in contrast to cooperators, defectors do not equalize payoffs between themselves and their partners by assigning 8 deduction points (or rather, only does so if they believe that their partner assigns 8 deduction points to them as well), potentially explaining why this response was much more frequent among cooperators than among defectors.

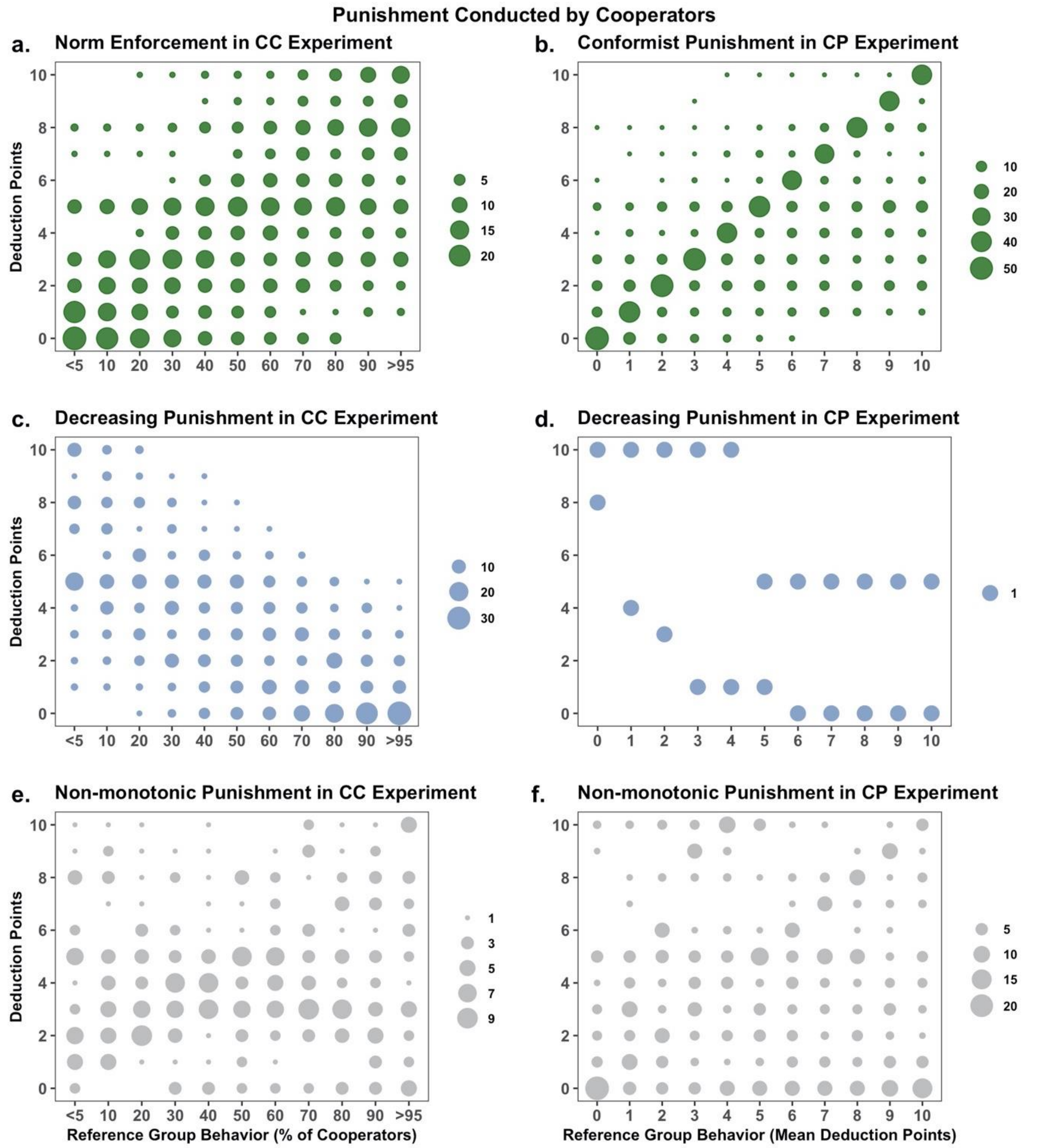


Fig. A5. Deduction points assigned by cooperators, broken down by punishment strategy in each experiment. Sizes of dots indicate numbers of observations.

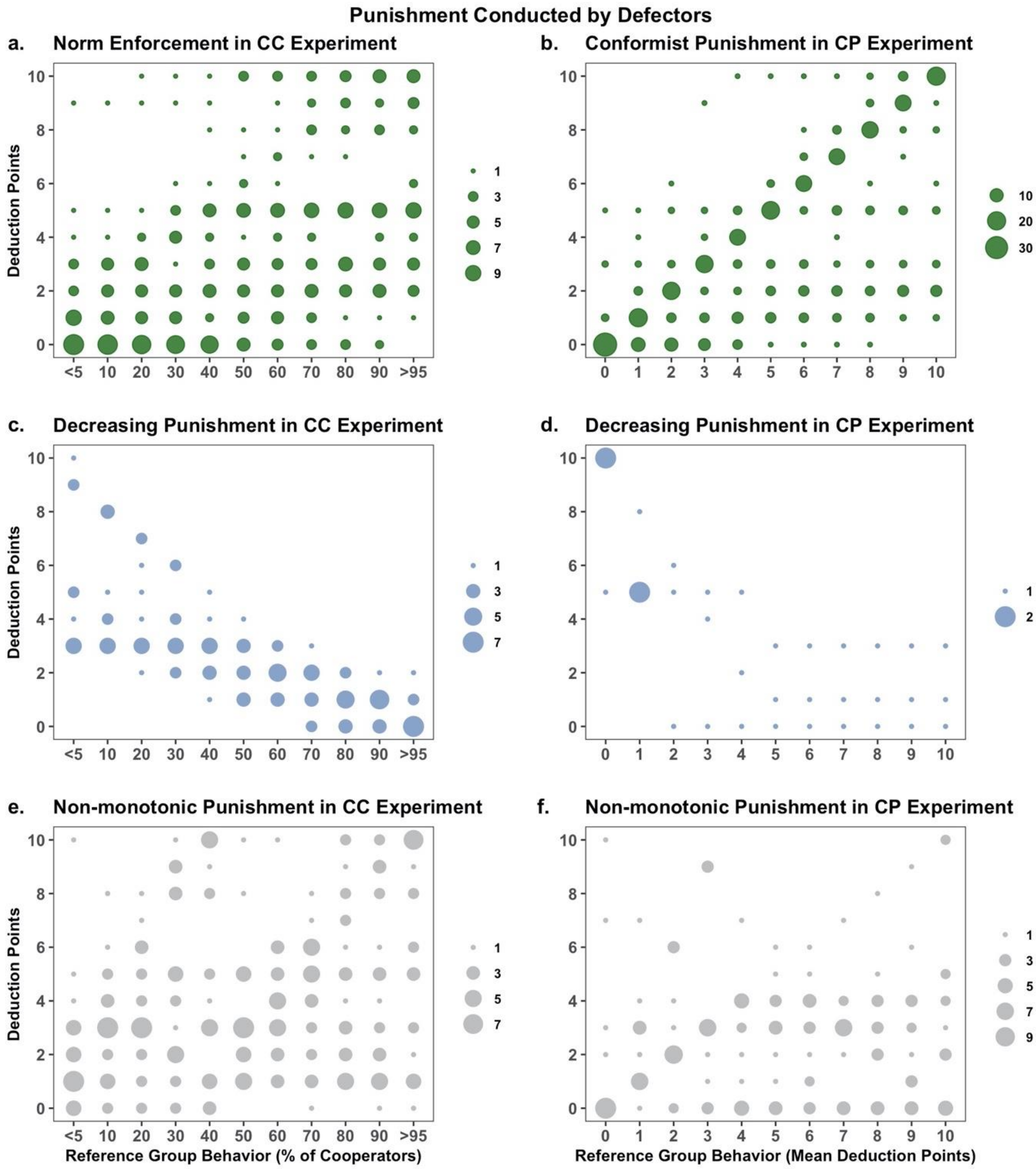


Fig. A6. Deduction points assigned by defectors, broken down by punishment strategy in each experiment. Sizes of dots indicate numbers of observations.

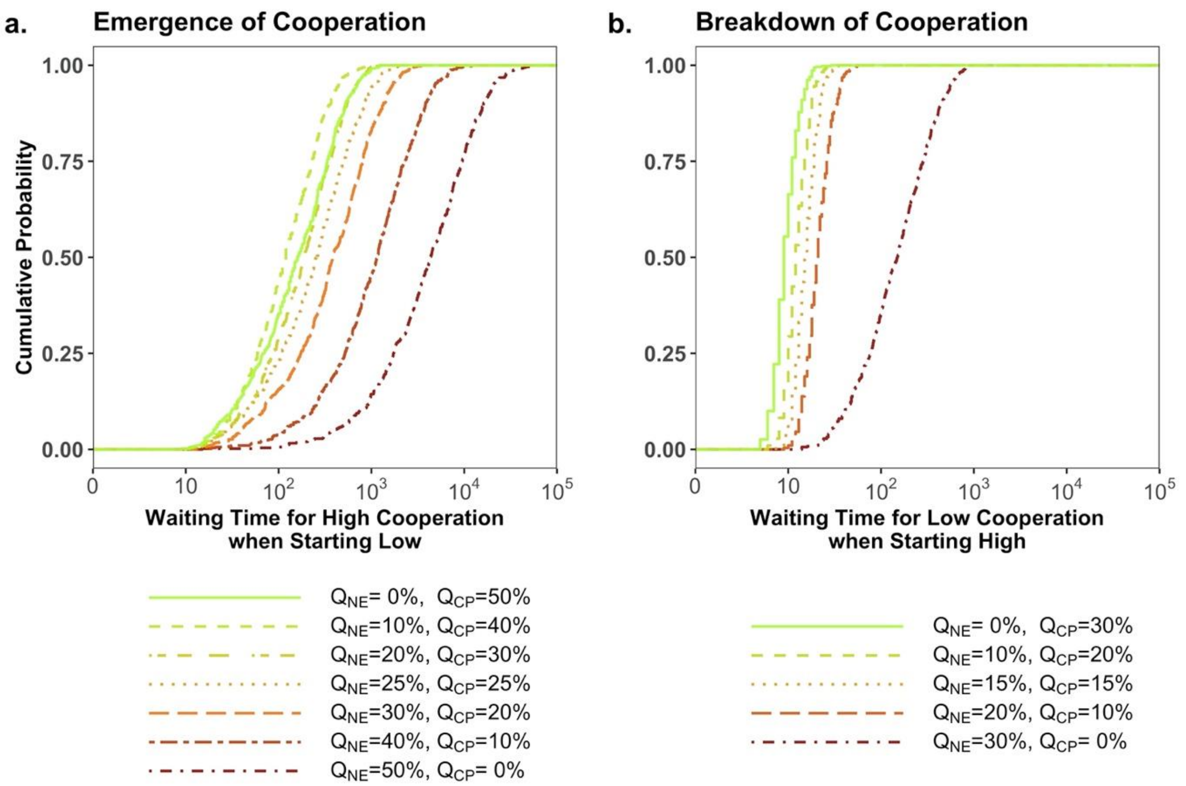


Fig. A7. Effects of conditional punishment strategies on the emergence and breakdown of cooperation. Lines show the cumulative probability of cooperation to rise above 75 percent (**a**) or fall below 25 percent (**b**), as a function of time. Time is shown on a logarithmic scale, and each line represents 500 simulation runs. Across both panels, we hold fixed the frequency of independent punishers at 30%. Q_{NE} is the frequency of norm enforcement, and Q_{CP} is the frequency of conformist punishment. Frequencies of these conditional strategies were chosen such that—according to our analytical results—cooperation would emerge (Panel **a**) or break down (panel **b**) in the long run. Initial beliefs regarding cooperation and punishment levels start low in Panel **a** ($b_c = b_p = 0.25$), and high in Panel **b** ($b_c = b_p = 0.75$; see Methods for details). Each simulation runs for 100,000 (10^5) periods. Further simulation settings: *n* = 100, *m* = 10, $u = 0.5$, ε=0.05.

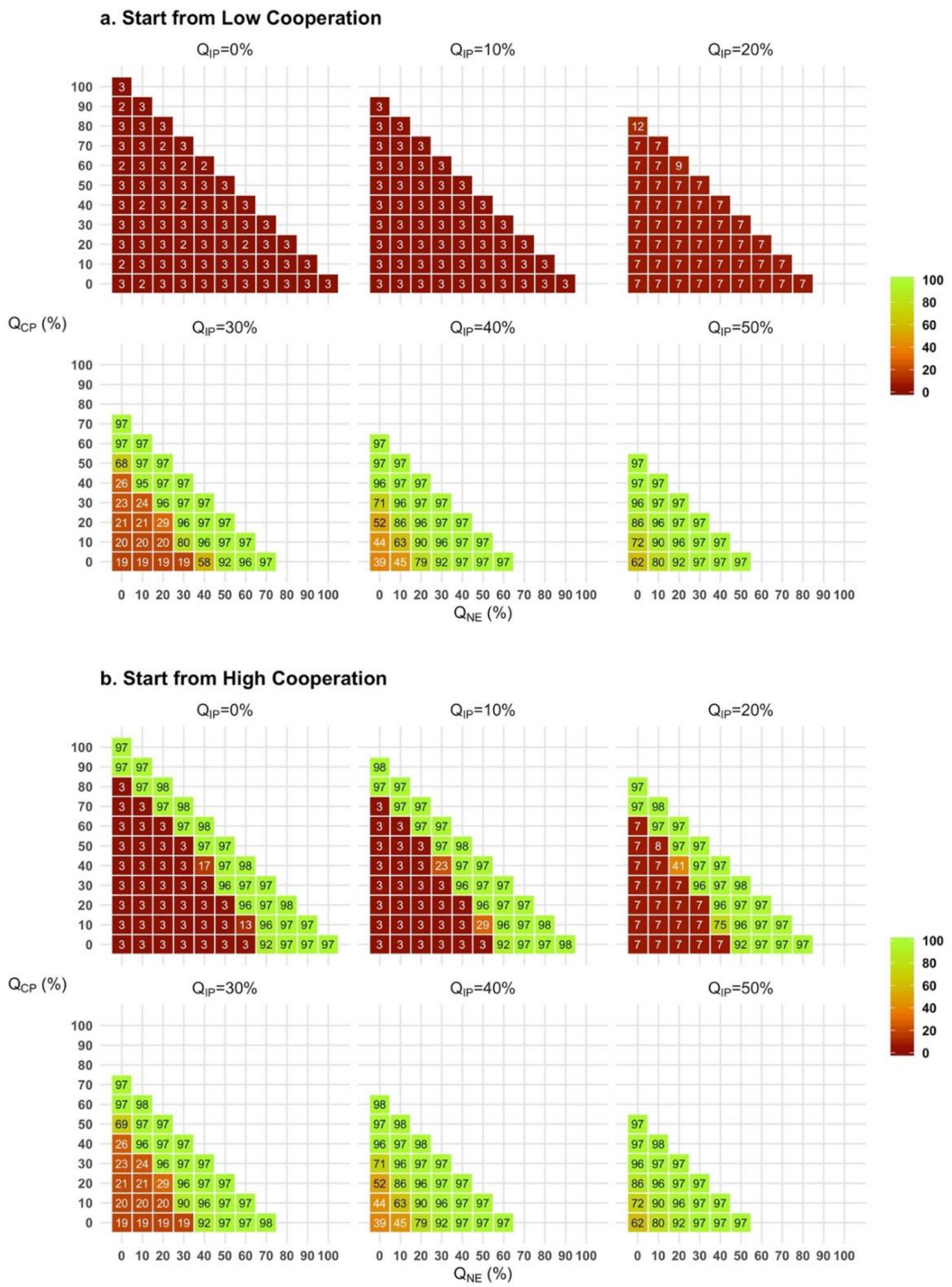


Fig. A8. Frequency of cooperation for different population compositions and different initial conditions. Colors (and number in the cells) show average percentages of cooperation across all periods of the simulation (see right hand side for color key). Each cell represents 30 simulation runs. Initial beliefs regarding cooperation and punishment start low (Panel **a**; $b_c = b_p = 0.25$), or high (Panel **b**; $b_c = b_p = 0.75$). Q_{IP} is the frequency of independent punishment, Q_{NE} is the frequency of norm enforcement, and Q_{CP} is the frequency of conformist punishment. Each simulation runs for 100,000 (10^5) periods. Further simulation settings: n = 100, m = 10, $u = 0.5$, ε=0.05. These analyses give an overall impression of how cooperation levels in our model depend on the relative frequencies of independent punishment (Q_{IP}), norm enforcement (Q_{NE}) and conformist punishment (Q_{CP}). The results suggest that conditional punishment strategies can play a particularly important role in the emergence and persistence of cooperation when independent punishment occurs at intermediate frequencies. That is, when independent punishment occurs at low frequencies, cooperation is unlikely to arise regardless of conditional punishment. Conversely, when independent punishment occurs at high frequencies, conditional punishment is not needed for cooperation to arise and persist. Overall, these simulations confirm the analytical model results presented in sections 3.3. and 3.4 of this Appendix. For the effects of the relative frequencies of punishment strategies on the speed of the dynamics (emergence or collapse of cooperation), see Figure A9.

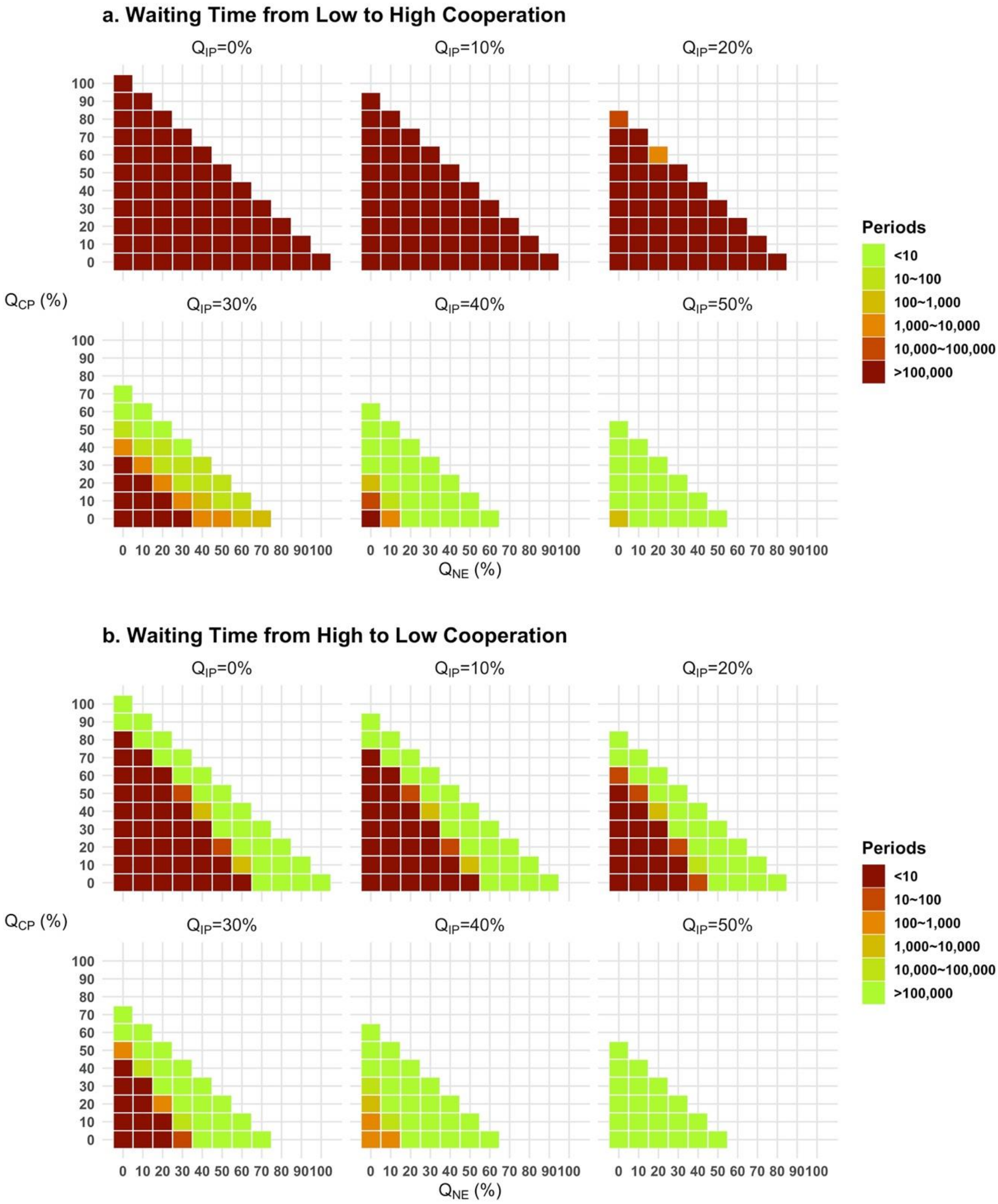


Fig A9. Waiting time for transitions between low and high cooperation for different population compositions. The color of each cell shows the median number of waiting periods across 30 simulation runs for cooperation to rise above 75% (Panel **a**) or fall below 25% (Panel **b**). Initial beliefs regarding cooperation and punishment start low in Panel **a** ($b_c = b_p = 0.25$), and high in Panel **b** ($b_c = b_p = 0.75$). Q_{IP} is the frequency of independent punishment, Q_{NE} is the frequency of norm enforcement, and Q_{CP} is the frequency of conformist punishment. The plots are based on the same simulation data used for Fig. A8.

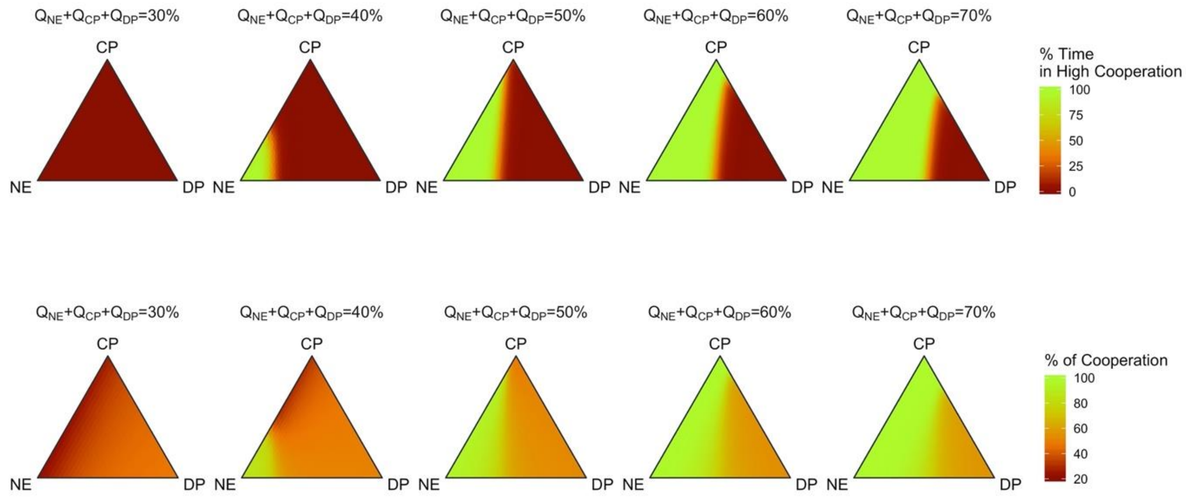


Fig A10. Effects of 'Decreasing Punishment' (DP) (cf. Fig 1a, blue bar) on cooperation dynamics. Agents with this strategy punish free riders if they believe cooperation rates are lower than 50% (b_c < 0.5). Triangles show outcomes of simulations that vary the relative proportion of norm enforcement, conformist punishment, and decreasing punishment, with independent punishment fixed at 30%. The top row shows the percentage of periods for which cooperation was higher than 75% for each combination of strategies, whereas the bottom row shows the frequency of cooperation over all periods. Q_{DP} is the frequency of decreasing punishment, Q_{NE} is the frequency of norm enforcement, and $Q_{CP}$is the frequency of conformist punishment. Results are the average outcome of simulations where cooperation either started high ($b_c = b_p = 0.75$) or low ($b_c = b_p = 0.25$). In particular, for each possible combination of Q_{NE}, Q_{CP}, and Q_{DP} averages are based on 10 simulations (5 with high and 5 with low initial beliefs). Each simulation runs for 10,000 (10^4) periods. Further simulation settings: *n* = 100, *m* = 10, $u = 0.5$, *ε*=0.05.

2. Supplementary Tables

Table A1. Dropouts during the experiment. The table show the number of participants completing each of the different phases of our experiment. About one in three individuals who clicked the study link on Amazon Mechanical Turk (MTurk) did not pass the control questions. Many of these individuals clicked the link and never moved beyond the first page of the experiment (a welcome screen), while others took multiple attempts at solving the control questions but did not succeed in completing them (see Section 4 in this Appendix for screenshots of the control questions). Dropouts were rare among those participants who did pass the control questions. The results reported in our paper are based on the 999 participants who completed the punishment stage of our experiment (Stage 2).

	Entered the experimental pages	Passed control questions for Stage 1	Completed Stage 1	Passed control questions for Stage 2	Completed Stage 2	Completed the game and the questionnaire
Number of participants (percentage)	1,624 (100%)	1,061 (65.3%)	1,058 (65.1%)	1,008 (62.1%)	999 (61.5%)	992 (61.1%)

Table A2. Determinants of punishment. The table displays results from ordinary least squares (OLS) regressions of the number of deduction points assigned to defecting partners. The predictor 'Cooperation in the reference group' ranges from 0 to 10, corresponding to the 11 situations presented in Stage 2 of the CC experiment. Similarly, the predictor 'Punishment in the reference group' ranges from 0 to 10, reflecting the 11 situations in the CP treatment. Numbers in parentheses are robust standard errors (SEs) corrected for clustering at the participant level. *P*-values are presented below the SEs. Model (i) shows that in the CC experiment, an increase of 10% of cooperators in the reference group leads to an average increase of 0.051 deduction points assigned to a defecting interaction partner (F-test: $P<0.01$). Model (ii) shows that in the CP experiment, an average increase of one deduction point assigned in the reference group leads to an average increase of about 0.156 deduction points (F-test: $P<0.001$).

	CC experiment (i)	CP experiment (ii)
Cooperation in the reference group (1 unit = 10% increase)	0.051 (0.016) *P*=0.002	
Punishment in the reference group (mean number of deduction points assigned in the reference group)		0.156 (0.016) *P*<0.001
Cooperator (1 if the punisher cooperated in stage 1; 0 otherwise)	1.239 (0.249) *P*<0.001	1.776 (0.241) *P*<0.001
Male	0.199 (0.257) *P*=0.438	0.212 (0.260) *P*= 0.416
Age	-0.012 (0.013) *P*=0.340	0.019 (0.012) *P*=0.100
Constant	1.907 (0.495) *P*<0.001	-0.271 (0.473) *P*=0.567
Observations	5,434	5,478
Number of participants	494	498
R-squared	0.038	0.095

3. Supplementary Analysis

Contents

Here we use analytical methods to evaluate our model, addressing how the experimentally identified punishment strategies interact to shape the dynamics of cooperation. Section 1 describes and formalizes the interaction setting. Section 2 describes the strategies we consider. Sections 3 and 4 analyze the effects of conditional punishment strategies on cooperation in the short run and in the long run, respectively.

3.1. Setting

We consider the following decision setting, which is similar to the task used in the experiment. Two agents, A and B, are randomly drawn from a large population to play a two-stage game. In Stage 1, they can either *cooperate* or *defect*. Table A3 shows how the Stage 1 material payoffs for both agents depend on their choices.

Table A3. Payoffs from stage 1.

		Agent B	
		Cooperate	*Defect*
Agent A	*Cooperate*	a, a	d, e
	Defect	e, d	c, c

Note: we consider a prisoner's dilemma, which is characterized by $e > a > c > d$. In the experiment, the values used were: $e = 25, a = 18, c = 16, d = 9$.

In Stage 2, each agent can punish their partner if their partner defected in Stage 1. We depart from the experiment by considering binary punishment decisions (rather than choosing integers on a 0-10 scale). Punishment incurs a cost $k > 0$ to the punisher and a loss $l > 0$ to the defector. As consistent with our experiment, our model only considers punishment of defectors and ignores antisocial punishment of cooperators. The final payoffs from the game are the payoffs from Stage 1 minus the costs of conducting punishment and losses from being punished in Stage 2.

We focus on binary punishment decisions for the sake of exposition and tractability. Compared with the task in our experiment, focusing on binary punishment decisions in our model is not without loss of generality. Binary punishment excludes the possibility that an individual's punishment is not weakly monotonic—i.e., that it is neither independent, nor weakly increasing or weakly decreasing—in response to increasing cooperation rate or punishment rate in the population. Our experimental results, however, suggest that non-monotonic punishment behavior is much less common than independent punishment, norm enforcement, and conformist punishment (Fig. 1 in the main text). Furthermore, the group of participants who show non-monotonic punishment behavior becomes very small (less than 10 percent) if we exclude participants who had difficulty answering the nine compulsory control questions (Fig. A7), suggesting that such non-monotonic behavior is likely to be the result of inattentive choice behavior, rather than a real preference.

3.2. Strategies

3.2.1. Cooperation

We assume that an agent's choice to cooperate or defect depends on which choice generates the highest expected material payoffs. Let $b_c \in [0,1]$ denote an agent's belief about the cooperation rate in the population, and $b_p \in [0,1]$ the punishment rate. From Table A3 we can see that the expected payoff from choosing *cooperate* is

(1) $$b_c a + (1 - b_c)d.$$

The expected payoff from choosing *defect* is

(2) $$b_c e + (1 - b_c)c - b_p l.$$

An agent cooperates if and only if (1) ≥ (2) (assuming they cooperate if expected payoffs are the same). Rearranging the terms leads to the condition

(3) $$b_p \geq \theta_C \equiv \frac{1}{l}[b_c(e - a) + (1 - b_c)(c - d)].$$

This shows that an agent cooperates if and only if their beliefs of being punished if they defect (b_p) exceeds a threshold. This threshold is linearly increasing in the temptation to defect $b_c(e - a) + (1 - b_c)(c - d)$, and decreasing in the loss from being punished l. In the analysis presented in the main text, we assume $\theta_C = 0.5$. Here, we consider the general case of arbitrary threshold values.

3.2.2. Punishment

Our implementation of punishment strategies is informed by our experimental results. We consider four distinct 'types' of agents: *i*) *independent punishers* who punish independently of b_c and b_p, *ii*) *norm enforcers* who punish if and only if b_c is high enough, and *iii*) *conformist punishers* who punish if and only if b_p is high enough; and *iv*) *non-punishers*, who never punish. For simplicity and ease of illustration, we assume that the four strategies above are mutually exclusive and stable: each individual has a unique strategy that doesn't change over time.

The frequencies of punishment types in the population—independent punishers (Q_{IP}), norm enforcers (Q_{NE}), conformist punishers (Q_{CP}) and non-punishers (Q_0)—sum up to 1. Agents do not know the punishment strategy of their interaction partners. Norm enforcers punish if and only if $b_c \geq \theta_{NE}$. Conformist punishers punish if and only if $b_p \geq \theta_{CP}$. In the main text, we assume $\theta_{NE} = \theta_{CP} = 0.5$. Here we will consider the case of arbitrary threshold values. Table A4 summarizes all model parameters.

Table A4. Collection of model parameters.

Agent-level parameters	
$b_c \in [0,1]$	An agent's belief about the cooperation rate in the population
$b_p \in [0,1]$	An agent's belief about the punishment rate in the population
$\theta_C \geq 0$	If and only if $b_p \geq \theta_C$ then the agent will cooperate; derived from $\frac{1}{l}[b_c(e-a)+(1-b_c)(c-d)]$
$\theta_{NE} \geq 0$	If and only if $b_c \geq \theta_{NE}$ then norm enforcers will punish
$\theta_{CP} \geq 0$	If and only if $b_p \geq \theta_{CP}$ then conformist punishers will punish
Population-level parameters	
n	Number of agents
$Q_0 \in [0,1]$	Frequencies of non-punishers
$Q_{IP} \in [0,1]$	Frequencies of independent punishers
$Q_{NE} \in [0,1]$	Frequencies of norm enforcers
$Q_{CP} \in [0,1]$	Frequencies of conformist punishers
T	Number of periods
m	Number of sampling agents
S	Set of strategy profiles in the population (the state space of the dynamic)
$u \in (0,1)$	Probability of having an opportunity to update
$\varepsilon \in (0,1)$	Probability of making mistakes when updating
$P^{\varepsilon} \in \Delta(S)$	Stationary distribution of the stochastic dynamic with $m=n$
$P \in \Delta(S)$	Limit distribution of P^{ε} as $\varepsilon \to 0$; used to define long-run equilibrium

We characterize short-run (Nash) equilibria and long-run equilibria of the game. First, we show that there are often two Nash equilibria: one in which all agents cooperate, and another in which all agents defect. Second, we analyze how the relative frequencies of the different punishment strategies $(Q_{IP}, Q_{NE}, Q_{CP}, Q_0)$ affect the likelihood of either equilibrium to emerge and persist in the long run.

3.3. Short-run (Nash) equilibrium

Proposition 1 shows the conditions under which cooperation can be sustained in the short run. Agents do not know the type of agent with whom they are matched. As is standard in the economic literature, we assume that agents have a common prior on the population composition, which corresponds to $(Q_{IP}, Q_{NE}, Q_{CP}, Q_0)$. In the next section 'Long-run equilibrium', we will address the problem of how agents form and update beliefs over time. Exogenous payoff parameters of the game determine the equilibria through their effects on the thresholds θ_C, θ_{NE}, and θ_{CP}.

Proposition 1. (Nash equilibrium)

1. *If $Q_0 > 1 - \frac{n-1}{n}\theta_C$, then in every Nash equilibrium all agents defect.*
2. *If $Q_{IP} \geq \frac{1}{n} + \frac{n-1}{n}\theta_C$, or if $Q_{IP} > \theta_{CP}$ and $Q_{IP} + Q_{CP} \geq \frac{1}{n} + \frac{n-1}{n}\theta_C$, then in every Nash equilibrium all agents cooperate.*
3. *If $Q_0 < 1 - \frac{1}{n} - \frac{n-1}{n} max\{\theta_C, \theta_{CP}\}$, then there exists a Nash equilibrium in which all agents cooperate, and all agents (except non-punishers) punish defectors.*
4. *If $Q_{IP} < \frac{n-1}{n} min\{\theta_C, \theta_{CP}\}$, then there exists a Nash equilibrium in which all agents defect, and only those punishing independently punish defectors.*

The proof of the proposition is provided at the end of this section. The proposition states that, first, if there are many agents who do not punish (Q_0 exceeds a critical threshold), then all agents defect in equilibrium. This result is intuitive: if the likelihood of being punished after defection is sufficiently low, then agents will defect to maximize their expected material payoffs. Second, if there is a high enough level of independent punishment such that cooperation is the payoff maximizing choice for all individuals ($Q_{IP} \geq \frac{1}{n} + \frac{n-1}{n}\theta_C$), then all agents cooperate in equilibrium. Similarly, if there are enough independent punishers such that their behavior triggers punishment by conformist punishers ($Q_{IP} > \theta_{CP}$), and their joint number is high enough to make cooperation the payoff maximizing choice for all agents ($Q_{IP} + Q_{CP} \geq \frac{1}{n} + \frac{n-1}{n}\theta_C$), then all agents cooperate in equilibrium as well.

Third, if there are not sufficiently many non-punishers ($Q_0 < 1 - \frac{1}{n} - \frac{n-1}{n} max\{\theta_C, \theta_{CP}\}$), then there exists a Nash equilibrium in which all agents cooperate and all agents (apart from non-punishers) punish defectors. Fourth, if there are not sufficient independent punishers to make cooperation the payoff maximizing choice (either by themselves, or in unison with conformist punishers), then there exists a Nash equilibrium in which all agents defect.

Together, the third and the fourth statement imply that when both independent punishers and non-punishers occur at relatively low frequencies, cooperation and defection can both emerge as Nash equilibria. The intuition for this result is that, when independent punishers are insufficient to enforce cooperation, but there are potentially enough of the (two types of) conditional punishers to do so, then the fact that agents this latter group conditions their behavior either on the existing levels of cooperation or punishment can make both the defection equilibrium and the cooperation equilibrium possible.

In the following section, we focus on situations where there are multiple equilibria, and examine how (conditional and unconditional) punishment strategies affect which of these equilibria is selected in the long run. In the simulations reported in the main text, we examine the short-run dynamics of our model.

3.4. Long-run equilibrium

In this section we examine the long-run effects of conditional and unconditional punishment strategies on cooperation. We aim to delineate the conditions under which conditional punishment strategies (norm enforcement and conformist punishment) will, in the long run, cause the population to be in or around the cooperation equilibrium for most of the time. Our analysis builds on Kandori et al., (1993) and Young (1993, 2001).

We consider discrete time periods: $t = 0,1,2, \dots, T$. In each period, agents are randomly matched and interact in the two-stage game described in Section 3.1 above. An agent's punishment strategy and the population composition $(Q_{IP}, Q_{NE}, Q_{CP}, Q_0)$ are fixed over time, but agents may update their cooperation and punishment decisions as their beliefs b_c and b_p change. In each period agents react to their beliefs 'myopically' to maximize their expected payoffs in that period.

To be more precise, each period involves two subsequent classes of events:

I. *Updating beliefs.* In each period $t \geq 1$, each agent updates their beliefs with probability *u*, with 0<*u*<1. Belief updating works as follows. The agent randomly samples *m* agents from the population, with $0 < m \leq n$. She counts how many agents in the sample cooperated and would punish according to their strategies in the previous period, and divide the counts by *m*. The results become their beliefs b_c and b_p in the current period.
II. *Responding myopically to beliefs.* An agent cooperates in a period if and only if they have belief $b_p \geq \theta_C$. Punishment decisions are determined according to the agent's type (as specified in Section 3.2 above).

With a high probability, an agent's decisions are implemented according to the rules stated above. With small probability $\varepsilon \geq 0$, however, an agent makes a mistake ("tremble"). A mistake implies that the agent randomly selects a cooperative action or a punishment action. We assume that mistakes are independent across periods, agents, and across cooperation and punishment decisions. Following Kandori et al., (1993) and Young, (1993, 2001), we refer to the dynamic with $\varepsilon > 0$ as *the stochastic dynamic*, and the dynamic with $\varepsilon = 0$ as *the best-response dynamic*.

We first analyze the stochastic dynamic in the case of $m = n$, $T \to \infty$, and $\varepsilon \to 0$ using analytical methods. As previous studies of the same class of stochastic dynamics show (Kandori et al., 1993; Young, 1993, 2001), whether $m < n$ or $m = n$ does not affect the stationary distributions of the dynamics. We also conduct simulations to explore the cases of small sample size m, finite T, and non-negligible ε (see main text and Figures A7-A10).

Our analytical results aim to characterize the set of long-run equilibria. These are the equilibria that have a positive frequency in the stationary distribution of the stochastic dynamic when the probability of mistakes is vanishingly small. A long run equilibrium is formally defined as follows. Let s be a population state specifying the cooperation decision and punishment decision of each agent in the population. Let S denote the set of all population states. Let $P^{\varepsilon} \in \Delta(S)$ denote the stationary distribution of the stochastic dynamic under $\varepsilon > 0$ and $m = n$. The stochastic dynamic is an irreducible Markov chain on the finite state space S. Hence P^{ε} exists and is unique for each ε. We obtain P^{ε} by taking $T \to \infty$. Let $P \equiv \lim_{e \to 0} P^{\varepsilon}$ denote the limit distribution as ε approaches zero. A state s is a *long-run equilibrium* if $P(s) > 0$ (Kandori et al., 1993; Young, 1993, 2001). If a state is a unique long-run equilibrium for sufficiently large n, then it is a *generically unique long-run equilibrium*.

For the sake of exposition and analytical tractability, we restrict our attention to the parameter ranges specified by the assumptions below.

Assumptions.

1) $Q_{IP} < min\{\theta_C, \theta_{CP}\}$, $\theta_{NE} < 1$, and $Q_0 < 1 - max\{\theta_C, \theta_{CP}\}$;
2) *either* (*i*) $Q_{NE} \geq |\theta_C - \theta_{CP}|$ and $Q_{CP} \geq |\theta_C - \theta_{CP}|$, *or* (*ii*) $Q_{NE} \leq |\theta_C - \theta_{CP}|$ and $Q_{CP} \leq |\theta_C - \theta_{CP}|$;
3) *either* (*i*) $\theta_{CP} \leq Q_{IP} + Q_{NE}$ and $\theta_{CP} \leq Q_{IP} + Q_{CP}$, *or* (*ii*) $\theta_{CP} \geq Q_{IP} + Q_{NE}$ and $\theta_{CP} \geq Q_{IP} + Q_{CP}$.

Assumption (1) restricts our attention to cases where the following Nash equilibria both exist (see Proposition 1): *the defection equilibrium* in which all agents defects and only independent punishers punish defectors; and *the cooperation equilibrium* in which all agents cooperate and all agents (except non-punishers) punish defectors. The remaining two assumptions greatly reduce the number of cases we need to consider, but still allow us to obtain the key intuitions from the model. Specifically, Assumption (2) holds that the proportions of conditional punishers (Q_{NE} and Q_{CP}) are both either high or low. Assumption (3) holds that the value of θ_{CP} is either high or low, compared to the number of punishers.

Now we can state the proposition about the long-run equilibrium of the stochastic dynamic. It shows how norm enforcement and conformist punishment interact with independent punishment to affect cooperation in the long run.

Proposition 2. (Long-run equilibrium) *Suppose $m = n$ and Assumption 1 hold. Let $\underline{\theta} \equiv min\{\frac{1}{2}(\theta_C+\theta_{CP}), 2\theta_{NE}+\theta_{CP} - 1\}$ and $\bar{\theta} \equiv max\{\frac{1}{2}(\theta_C+\theta_{CP}), 2\theta_{NE}+\theta_{CP} - 1\}$.*

1. *If $Q_{IP} + Q_{NE} > \bar{\theta}$ and $Q_{IP} + Q_{CP} > \bar{\theta}$, then the cooperation equilibrium is the generically unique long-run equilibrium.*
2. *If $Q_{IP} + Q_{NE} < \underline{\theta}$ and $Q_{IP} + Q_{CP} < \underline{\theta}$, then the defection equilibrium is the generically unique long-run equilibrium.*

The parameters $\underline{\theta}$ and $\bar{\theta}$ are two critical thresholds derived from θ_C, θ_{CP} and θ_{NE}. Fig. A11 illustrates the proposition: together with the independent punishment, conditional punishment ($Q_{IP} + Q_{NE} > \bar{\theta}$ and $Q_{IP} + Q_{CP} > \bar{\theta}$) can support cooperation as the generically unique long-run equilibrium. If the frequencies of independent punishment and conditional punishment are both low ($Q_{IP} + Q_{NE} < \underline{\theta}$ and $Q_{IP} + Q_{CP} < \underline{\theta}$), then the cooperation equilibrium cannot be sustained in the long run.

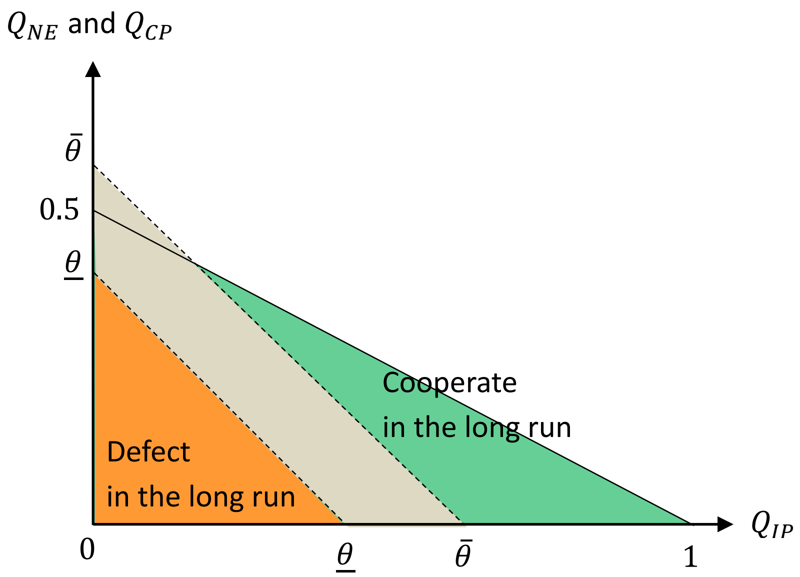


Fig. A11. Illustration of Proposition 2. When $Q_{IP} + Q_{NE} > \bar{\theta}$ and $Q_{IP} + Q_{CP} > \bar{\theta}$, we expect to see the cooperation equilibrium in the long run. When $Q_{IP} + Q_{NE} < \underline{\theta}$ and $Q_{IP} + Q_{CP} < \underline{\theta}$, we expect to see the defection equilibrium in the long run. When $\theta_C = \theta_{NE} = \theta_{CP} = 0.5$, we have $\underline{\theta} = \bar{\theta} = 0.5$.

The proof of Proposition 2 is provided in Section 3.5.2 below. Here we sketch the intuition. When the probability of mistake ε is small, the dynamic stays mostly in the absorbing states (i.e., the strategy profiles that the dynamic will not escape from without mistakes when they are reached). Occasionally, the dynamic jumps from one absorbing state to another state when a certain sequence of mistakes occur. As ε approaches zero, leaving an absorbing state is extremely difficult because even a single mistake would be a rare event. Hence, the long-run equilibrium must be one of the absorbing states. To identify the long-run equilibrium, we count the minimum numbers of mistakes required to transit from one absorbing state to another. The long-run equilibrium, roughly speaking, is the absorbing state that is most difficult to leave and that is relatively easy to transit to from other absorbing states (as measured by the minimum numbers of mistakes required for the transitions).

Our model has two absorbing states: the cooperation equilibrium and the defection equilibrium. To determine which one is the long-run equilibrium, we examine the transition paths that exist between them and identify the transition path which requires the smallest number of mistakes. Fig. A12 shows the relevant transition paths from the defection equilibrium to the cooperation equilibrium. For a transition from the defection equilibrium to the cooperation equilibrium to occur, we require either enough agents to punish defectors by *mistakes* (Paths E1 and E2; 'E' for *E*mergence of cooperation) or enough agents to cooperate by *mistakes* (Path E3). These mistakes then trigger a chain of reactions leading to the cooperation equilibrium. How many mistakes are enough depends on the thresholds θ_C (for agents to cooperate), θ_{CP} (for conformist punishers to punish), θ_{NE} (for norm enforcers to punish) and the numbers of punishers Q_{IP}, Q_{CP} and Q_{NE}.

The minimum number of mistakes required to transit from the defection equilibrium to the cooperation equilibrium is compared with that of transitions in the opposite direction. Fig. A13 shows transition paths from the cooperation equilibrium to the defection equilibrium. Starting the cooperation equilibrium, we require either enough agents to stop punishing defectors by mistakes (Paths B1 and B2; 'B' for *B*reakdowns of cooperation) or enough agents to defect by mistakes (Path B3). The minimum number of mistakes required to lead to the defection equilibrium depends on the thresholds $1-\theta_C$ (for agents to defect), $1-\theta_{CP}$ (for conformist punishers to stop punishing), $1-\theta_{NE}$ (for norm enforcers to stop punishing) and the number of non-punishers Q_0.

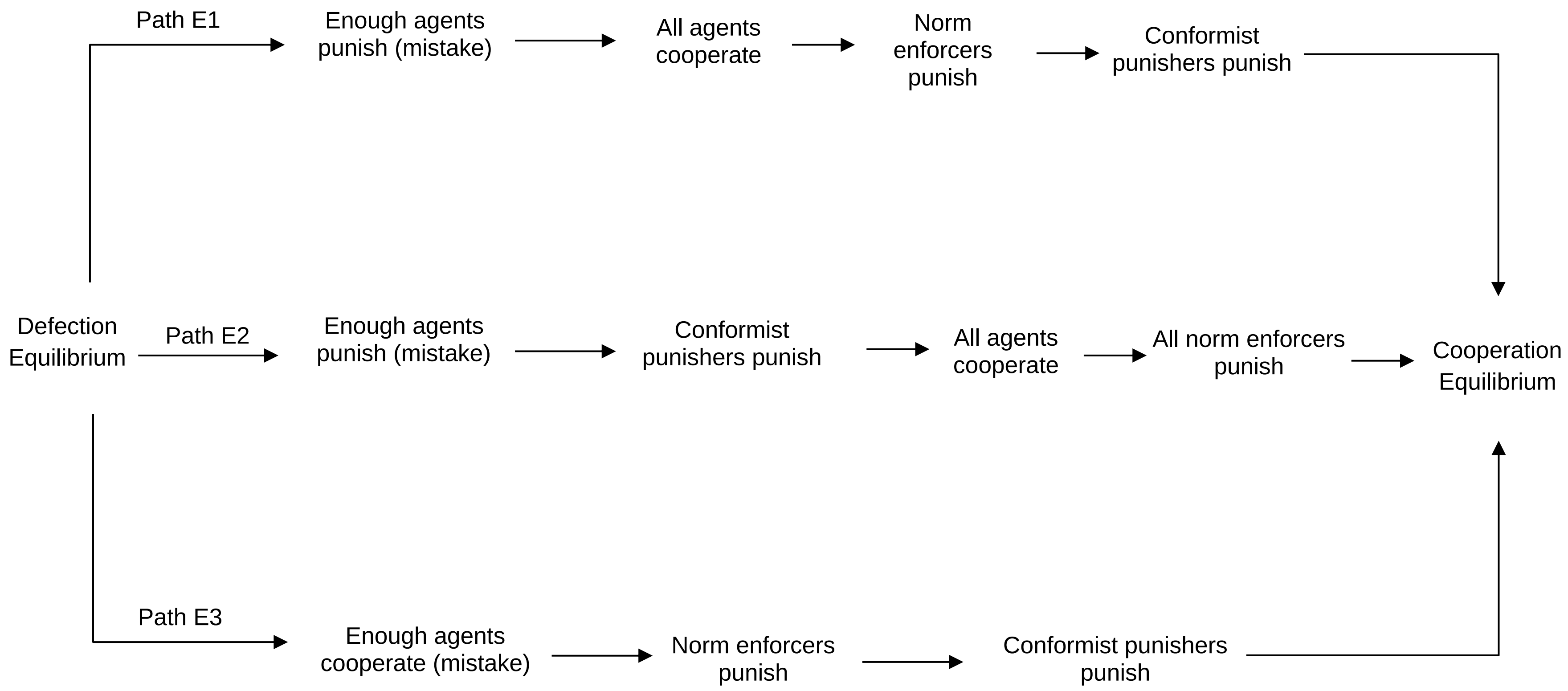


Fig A12. Transition paths from defection equilibrium to cooperation equilibrium

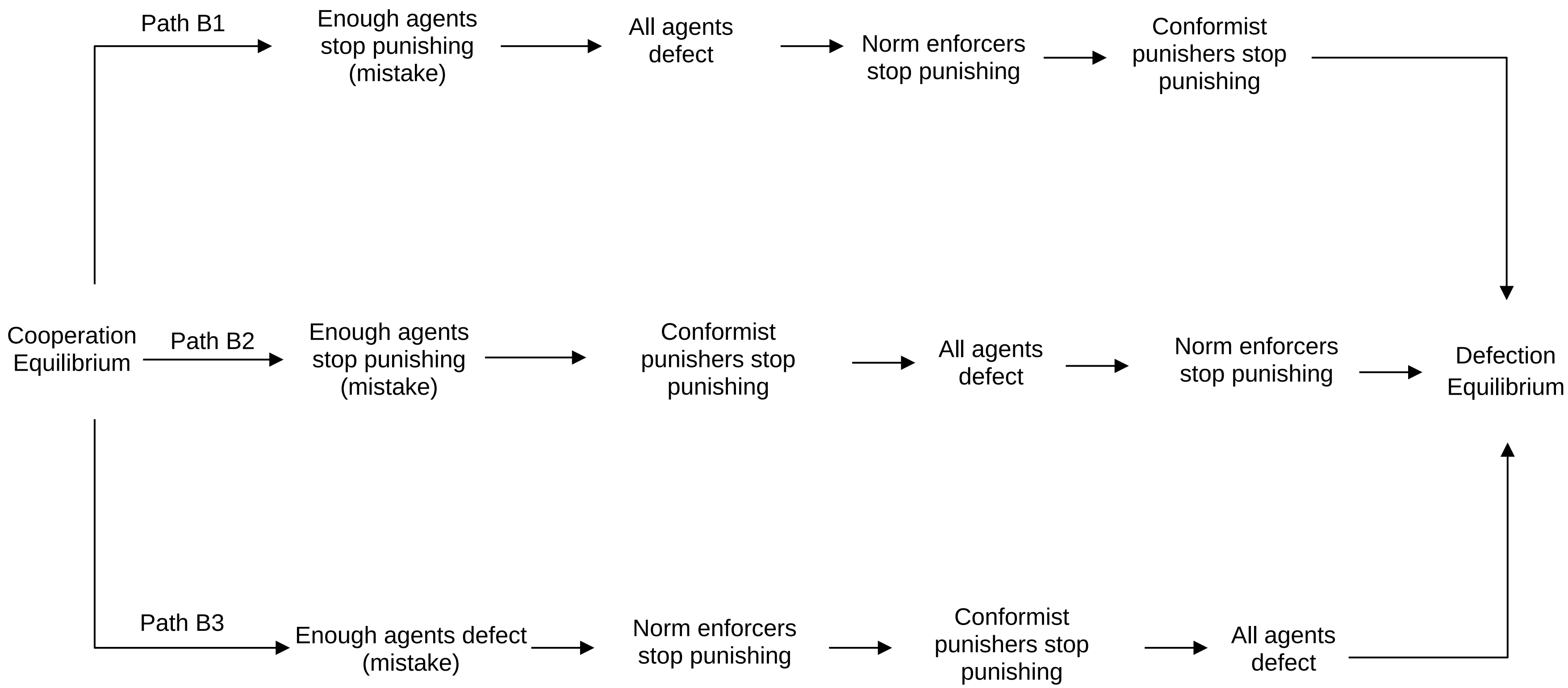


Fig A13. Transition paths from cooperation equilibrium to defection equilibrium

Note that Proposition 2 is silent about the case of $\underline{\theta} < Q_{IP} + Q_{NE} < \bar{\theta}$ or $\underline{\theta} < Q_{IP} + Q_{CP} < \bar{\theta}$. Proposition 3 below provides a precise cut-off condition for the long-run equilibrium in the special case of $\theta_C = \theta_{NE} = \theta_{CP} = \frac{1}{2}$, which is also the set of parameters we use in our simulations presented in the main text. The assumption that all thresholds are equal to a half is somewhat arbitrary. As stated in Section 2 of this Supplement, the threshold θ_C is determined by exogenous payoff parameters. Hence, setting it equal to a half comes down to considering a subset of the potential payoff space. For θ_{NE} and θ_{CP}, however, a threshold of a half makes intuitive sense. For norm enforcement, it is in line with the idea that people will judge the more common behavior as the more moral one, and act to enforce it (Lindström et al., 2018). For conformist punishment, it states that these agents follow the behavior of the majority. Furthermore, our focus here is not on the comparative statics with respect to these thresholds, but rather on how the population composition (with respect to punishment strategies) affects cooperation dynamics. In this regard, the proposition below is illuminating.

Proposition 3. *Suppose $m = n$ and $\theta_C = \theta_{NE} = \theta_{CP} = \frac{1}{2}$. Then*

1. *If $Q_{IP} + \frac{1}{2}(Q_{NE} + Q_{CP}) > \frac{1}{2}$, then the cooperation equilibrium is the generically unique long-run equilibrium;*
2. *If $Q_{IP} + \frac{1}{2}(Q_{NE} + Q_{CP}) < \frac{1}{2}$, then the defection equilibrium is the generically unique long-run equilibrium.*

The proof for Proposition 3 is provided at the end of this section. When $\theta_C = \theta_{NE} = \theta_{CP} = \frac{1}{2}$, we have $\underline{\theta} = \bar{\theta} = \frac{1}{2}$ in the conditions specified in Proposition 2. Proposition 3 further reveals that: if and only if independent punishers and the average number of norm enforcers and conformist punishers together add up to over a half of the population, then the cooperation equilibrium will be the only state that occurs with positive probability $P(s) > 0$ in the long run. That is, the average frequency of norm enforcement and conformist punishment is important to support cooperation in the long run; it is as important as the role played by independent punishment.

Remarks. Economists have used myopic best-response stochastic dynamics to study bargaining norms (Young, 1998), customs in economic contracts (Young & Burke, 2001), evolution of altruism (Eshel et al., 1998), the selection of coordination actions in social networks (Goyal & Vega-Redondo, 2005; Jackson & Watts, 2002), diffusion of innovations (Young, 2009, 2011), and the evolution of cooperation strategies in repeated games (Young & Foster, 1991). In particular, studies (Kandori et al., 1993; Young, 1993, 2001) show that many details of these dynamics do not affect their stationary distributions when

$\varepsilon \to 0$. In particular, the stationary distribution is not affected by the value of the updating probability u as long as $0 < u < 1$, or the sample size m as long as m does not become too small to affect the tipping thresholds, or the probability distribution used to pick actions when making mistakes.

Assuming $u < 1$ means that it will not occur that all agents update simultaneously in a period. If $u = 1$ and ε is small, then besides the cooperation equilibrium and the defection equilibrium, the population can also be trapped in a loop of jumping back and forth between two states: in one, all agents defect and all punish defectors except for the non-punishers; in the other, all agents cooperate but no one would punish defectors except for the independent punishers. We exclude this possibility to focus on the transitions between the cooperation equilibrium and the defection equilibrium characterized by Proposition 1.

3.5. Proofs

3.5.1. Proof of Proposition 1

(1) By contradiction: Suppose there is a (Nash) equilibrium in which some agent cooperates. Then for this agent, $b_p \geq \theta_C$ where b_p is the proportion of those who punish among the other agents. Note $b_p \leq \frac{(1-Q_0)n}{n-1}$, where $(1-Q_0)n$ is an upper bound on the number of agents who punish, and $n-1$ is the number of all other agents. Given $Q_0 > 1 - \frac{n-1}{n}\theta_C$, however, we have $\frac{(1-Q_0)n}{n-1} < \theta_C$, contradicting with $b_p \geq \theta_C$.

(2) We show the contrapositive: Suppose there is an equilibrium in which an agent defects. Then for the agent, $b_p < \theta_C$, where b_p is at least $\frac{Q_{IP}n-1}{n-1}$. Hence $\frac{Q_{IP}n-1}{n-1} < \theta_C$, implying $Q_{IP} < \frac{1}{n} + \frac{n-1}{n}\theta_C$.

Next, suppose $Q_{IP} > \theta_{CP}$. Then both independent punishers and conformist punishers punish. Hence, for each agent, the proportion of those who punish among the others is at least $\frac{(Q_{IP}+Q_{CP})n-1}{n-1}$. By $Q_{IP} + Q_{CP} \geq \frac{1}{n} + \frac{n-1}{n}\theta_C$, we have $\frac{(Q_{IP}+Q_{CP})n-1}{n-1} \geq \theta_C$. Thus, every agent cooperates.

(3) Consider the strategy profile such that all agents cooperate, and all agents (except non-punishers) punish defectors. To show that this is an equilibrium, we check the best-response of each agent. First, consider each agent's cooperation decision. The specified condition implies $b_p \geq \frac{(Q_{IP}+Q_{NE}+Q_{CP})n-1}{n-1} \geq \theta_C$ (*a*). Hence it is each agent's best response to cooperate. Second, given that everyone cooperates, it is each norm enforcer's best response to punish any defector. And

by definition, each independent punisher also punishes. Third, it is each conformist punisher's best response to punish if $\frac{(Q_{IP}+Q_{NE}+Q_{CP})n-1}{n-1} \geq \theta_{CP}$ (*b*). The condition $Q_0 < 1 - \frac{1}{n} - \frac{n-1}{n} max\{\theta_C, \theta_{CP}\}$ implies both (*a*) and (*b*). This establishes the statement.

Consider the strategy profile such that all agents defect, and no agent (except independent punishers) punishes defectors. We prove the statement by checking each agent's best responses. First, $\theta_{IP} < \frac{n-1}{n}\theta_C$ implies $\frac{\theta_{IP}n}{n-1} < \theta_C$. Hence it is each agent's best response to defect. Second, that all agents defect implies that it is each norm enforcer's best response to not punish defectors. Third, $\theta_{IP} < \frac{n-1}{n}\theta_{CP}$ implies $\frac{\theta_{IP}n}{n-1} < \theta_{CP}$. Hence it is each conformist punisher's best response to not punish. This completes the proof of Proposition 1.

3.5.2. Proof of Proposition 2

Preliminaries. First, we introduce necessary terminology for our proof (see, e.g., (Young, 2001) for a more extensive discussion). An *absorbing set (of the best-response dynamic)* is a subset of states $X \subset S$ such that (*i*) if the best-response dynamic starts from a state in X then it stays within X with probability 1, and (*ii*) for any $s, s' \in X$, there is a positive probability of transiting from s to s' within a finite number of periods. If an absorbing set contains only one state, then we call the state an *absorbing state*. A *transition path from s to* s' is a finite sequence of states, $s_1, s_2, \dots, s_K \in S$, with $s_1 = s, s_K = s'$, and $s_k \neq s_{k+1}$ for each $1 \leq k < K$. The *cost* of a transition path, denoted by $cost(s_1, s_2, \dots, s_K)$, is the number of mistakes (choices that are not best responses) that occur along the path. For any real number x, $\lceil x \rceil$ is the lowest integer equal to or greater than x, and we let $x^+ \equiv max\{0, x\}$.

We use stochastic trees to represent minimum transition costs between absorbing sets. A *stochastic tree* is a directed tree with each absorbing set as a vertex. The directed edges in a stochastic tree represent transitions among absorbing sets. Each edge is weighted by the minimum number of mistakes required to transit from one absorbing set to another. An absorbing set is said to be at the root of a stochastic tree if there is no edge (with positive weight) leading from it to other absorbing sets in the tree. The cost of a stochastic tree is the sum of the weights of all its edges. Our proof applies the following theorem:

Young's Theorem (Young, 1993).

1. *A state is a long-run equilibrium only if it is contained in an absorbing set.*

2. *If an absorbing state is at the root of the stochastic tree that strictly minimizes the cost among all stochastic trees, then the state is the unique long-run equilibrium.*

The best-response dynamic in our model has only two absorbing sets: one consisting of defection equilibrium, and the other consisting of the cooperation equilibrium. With abuse of notation, we denote them by D and C, respectively. By Young's theorem, D and C are the only candidates for a long-run equilibrium.

We can construct two stochastic trees: $D \to C$ (a directed line with D and C as its two vertices connected by a unique edge leading from D to C) and $C \to D$. Let $M_{C \to D}$ denote the minimum number of mistakes required to transit from C to D. More precisely, $M_{C \to D}$ is the minimum value of $cost(s_1, s_2, \dots, s_K)$ among the set of all paths $s_1, s_2, \dots, s_K$ with $s_1 = C$ and $s_K = D$. Likewise, $M_{C \to D}$ is the minimum value of $cost(s_1, s_2, \dots, s_K)$ among the set of all paths $s_1, s_2, \dots, s_K$ with $s_1 = D$ and $s_K = C$. By Young's theorem, it suffices to compare $M_{C \to D}$ with $M_{C \to D}$ to determine the long-run equilibrium.

Transition paths. Now we examine transition paths with minimum costs between D and C. Three paths are relevant to determine the minimum cost of transitions from D to C (Fig. A12):

- Path E1: Starting from D at time $t = 0$, if $\theta_C - Q_{IP} < Q_0$, then let $\lceil(\theta_C - Q_{IP})n\rceil$ non-punishers punish defectors *by mistake* at $t = 1$. If $\theta_C - Q_{IP} \geq Q_0$, then let all non-punishers and $\lceil(\theta_C - Q_{IP} - Q_0)n\rceil$ conformist punishers punish by *mistake* at $t = 1$. At $t = 2$, let all agents update their cooperation decision. Then they all cooperate (for brevity, if we do not explicitly mention that agents update their cooperation or punishment decision, then the agents are not selected to update from the last period and do not make any additional mistakes). At $t = 3$, let all norm enforcement agents update their punishment decision: they will all punish. It follows that following $t = 3$, all agents cooperate, and $\lceil(\theta_C + Q_{NE})n\rceil$ agents punish defectors. At $t = 4$, let $\Delta_1^E \equiv \lceil(\theta_{CP} - \theta_C - Q_{NE})^+ n\rceil$ agents who do not punish at $t = 3$ start to punish by *mistake*. Finally, at $t = 5$, let all conformist punishers update their punishment decision. Then by requiring all agents to update both cooperation and punishment decisions at $t = 6$, we reach C. Collecting the mistakes, we obtain the cost of path E1: $cost(E1) = \lceil(\theta_C - Q_{IP})n\rceil + \Delta_1^E$, where $\Delta_1^E = 0$ if $min\{Q_{NE}, Q_{CP}\} \geq |\theta_C - \theta_{CP}|$.
- Path E2: Starting from D at $t = 0$, if $\theta_{CP} - Q_{IP} < Q_0 + Q_{NE}$, let $\lceil(\theta_{CP} - Q_{IP})n\rceil$ non-punishers or norm enforcers punish by *mistake* at $t = 1$. If $\theta_{CP} - Q_{IP} \geq Q_0 + Q_{NE}$, then let all non-punishers, all norm enforcers, and $\lceil(\theta_{CP} - Q_{NE} - Q_0)n\rceil$ conformist punishers punish by *mistake* at $t = 1$, resulting in $\lceil\theta_{CP} n\rceil$ agents punishing. At $t = 2$, let all conformist punishers update their

punishment decision. Then all conformist punishers punish. At $t = 3$, let $\Delta_2^E \equiv \lceil(\theta_C - \theta_{CP} - Q_{CP})^+ n\rceil$ agents who do not punish at $t = 2$ start to punish by *mistake*. At $t = 4$, let all agents update their cooperation decision. Then all agents cooperate. At $t = 5$, let all norm enforcers update their punishment decision and start to punish. By requiring all agents to update both cooperation and punishment decisions at $t = 6$, we reach C. The cost of path E2 is $cost(E2) = \lceil(\theta_{CP} - Q_{IP})n\rceil + \Delta_2^E$, with $\Delta_1^E = 0$ if $min\{Q_{NE}, Q_{CP}\} \geq |\theta_C - \theta_{CP}|$.

- Path E3: Starting from D at $t = 0$, let $\lceil\theta_{NE} n\rceil$ agents cooperate by *mistakes* at $t = 1$. At $t = 2$, let all norm enforcers update their punishment decision. Then all norm enforcers will punish defectors. At $t = 3$, let $\Delta_3^E \equiv \lceil(\theta_{CP} - Q_{IP} - Q_{NE})^+ n\rceil$ agents who did not punish in the last period start to punish by *mistakes*. At $t = 4$, let all conformist punishers update their punishment decision. Then all conformist punishers will punish. At $t = 5$, let all agents update their cooperation decision. Then by requiring all agents to update both cooperation and punishment decisions in $t = 6$, we reach C. The cost of this is path is $cost(E3) = \lceil\theta_{NE} n\rceil + \Delta_3^E$, where $\Delta_3^E = 0$ if $Q_{IP} + Q_{NE} \geq \theta_{CP}$.

Correspondingly, the following three paths are relevant to compute the minimum cost of transiting from C to D (Fig. A13):

- Path B1: Starting from C at $t = 0$, if $\lceil(1 - \theta_C - Q_0)n\rceil \leq Q_{IP}$, let $\lceil(1 - \theta_C - Q_0)n\rceil$ independent punishers stop punishing by *mistake* at $t = 1$. If $\lceil(1 - \theta_C - Q_0)n\rceil > Q_{IP}$, let all independent punishers and $\lceil(1 - \theta_C - Q_0 - Q_{IP})n\rceil$ conformist punishers stop punishing by *mistakes* at $t = 1$. At $t = 2$, let all agents update their cooperation decision. Then all agents now defect. At $t = 3$, let all norm enforcement punishers update punishment decisions and stop punishing. At $t = 4$, let all $\Delta_1^B \equiv \lceil(\theta_C - \theta_{CP} - Q_{NE})^+ n\rceil$ agents who punish in $t = 3$ stop punishing by *mistake*. At $t = 5$, let all conformist punishers update punishment decisions and stop punishing. Then by requiring all agents to update cooperation decisions as well as punishment decisions at $t = 6$, we reach D. The cost of this path is $cost(B1) = \lceil(1 - \theta_C - Q_0)n\rceil + \Delta_1^B$, where $\Delta_1^B = 0$ when $min\{Q_{NE}, Q_{CP}\} \geq |\theta_C - \theta_{CP}|$.

- Path B2: Starting from C at $t = 0$, let $\lceil(1 - \theta_{CP} - Q_0)n\rceil$ agents who punish at $t = 0$ stop punishing by *mistake* at $t = 1$. At $t = 2$, let all conformist punishers update punishment decision. They all stop punishing. At $t = 3$, let $\Delta_2^B \equiv \lceil(\theta_{CP} - \theta_C - Q_{CP})^+ n\rceil$ agents who punish at $t = 2$ stop punishing by *mistake*. At $t = 4$, let all agents update cooperation decisions and start to defect. At $t = 5$, let all norm enforcers update punishment decisions and stop punishing. Then we reach D by requiring all agents update both decisions at $t = 6$. The cost of this path is $cost(B2) = \lceil(1 - \theta_{CP} - Q_0)n\rceil + \varDelta_2^B$, where $\varDelta_2^B = 0$ when $min\{Q_{NE}, Q_{CP}\} \geq |\theta_C - \theta_{CP}|$.

- Path B3: Starting from C at $t = 0$, let$\lceil(1 - \theta_{NE})n\rceil$ agents defect by *mistake* at $t = 1$. At$t = 2$, let all norm enforcers update punishment decisions and stop punishing. At $t = 3$, let $\varDelta_3^B \equiv \lceil(1 - \theta_{CP} - Q_0 - Q_{NE})^+ n\rceil$ agents who punish at $t = 2$ stop punishing by *mistake*. Note $\varDelta_3^B$ can also expressed by $\varDelta_3^B = \lceil(Q_{IP} + Q_{CP} - \theta_{CP})^+ n\rceil$. By requiring all agents to update both decisions at $t = 3$, we reach D. The cost of this path is $cost(B3) = \lceil(1 - \theta_{NE})n\rceil + \varDelta_3^B$, where $\varDelta_3^E = 0$ if $Q_0 + Q_{NE} \geq 1 - \theta_{CP}$.

Simplifying observations. We need to determine the path with minimum cost among the six paths above. This requires solving a set of linear inequalities. Two observations simplify our calculations. First, since we are only concerned with generically unique long-run equilibria, it is both sufficient and necessary for the minimum cost path to have *strictly* lower cost than all transition paths of the opposite direction for *infinitely many* n. A sufficient and necessary condition for this is that there is a finite n under which all relevant inequalities for pairwise cost comparisons hold strictly. This condition is equivalent to having all relevant inequalities holding strictly when we ignore all "$\lceil . \rceil$" brackets, i.e., by ignoring the "least integer greater than" operator. To see the equivalence, first, suppose $\lceil xn\rceil < \lceil yn\rceil$ for some positive n. Then obviously $xn < yn$. Conversely, suppose $xn < yn$ for some positive n. Then $x < y$, and there is some large enough integer r such that $(y - x)r > 1$, implying $xr + 1 < yr$. Thus $\lceil xr\rceil < \lceil yr\rceil$, and $\lceil xn\rceil < \lceil yn\rceil$ for all $n > r$.

Second, after removing all "$\lceil . \rceil$" brackets, the costs of the paths listed above are all multiplications of n. Taking the two observations together, it suffices to consider their *relative costs* $\widehat{cost}(.) \equiv cost(.)/n$ and ignore all "$\lceil . \rceil$" operators. Henceforth we will focus on $\widehat{cost}(.)$ and remove all "$\lceil . \rceil$" operators. The six transition paths and their relative costs $\widehat{cost}$ are summarized in Table A5 below.

Table A5. Transition paths and relative costs.

	Relative costs $\widehat{cost}$
Transition paths from D to C	
E1	$\theta_C - Q_{IP} + \varDelta_1^E$, where $\varDelta_1^E = (\theta_{CP} - \theta_C - Q_{NE})^+$
E2	$\theta_{CP} - Q_{IP} + \varDelta_2^E$, where $\varDelta_2^E = (\theta_C - \theta_{CP} - Q_{CP})^+$
E3	$\theta_{NE} + \varDelta_3^E$, where $\varDelta_3^E = (\theta_{CP} - Q_{IP} - Q_{NE})^+$
Transition paths from C to D	
B1	$1 - \theta_C - Q_0 + \varDelta_1^B$, where $\varDelta_1^B = (\theta_C - \theta_{CP} - Q_{NE})^+$

B2	$1-\theta_{CP}-Q_0+\Delta_2^B$, where $\Delta_2^B=(\theta_{CP}-\theta_C-Q_{CP})^+$
B3	$1-\theta_{NE}+\Delta_3^B$, where $\Delta_3^B=(Q_{IP}+Q_{CP}-\theta_{CP})^+$

Final steps. Three final steps pin down the minimum cost path:

Step 1: If $Q_{IP}+min\{Q_{NE},Q_{CP}\}>2\theta_{NE}+\theta_{CP}-1$ then $\widehat{cost}(E3)<\widehat{cost}(B3)$; if $Q_{IP}+max\{Q_{NE},Q_{CP}\}<2\theta_{NE}+\theta_{CP}-1$ then $\widehat{cost}(E3)>\widehat{cost}(B3)$.

First, suppose $\theta_{CP}\le Q_{IP}+min\{Q_{NE},Q_{CP}\}$. Then $\Delta_3^E=0$, and $\Delta_3^B\ge 0$. Thus, $\widehat{cost}(E3)=\theta_{NE}$, and $\widehat{cost}(B3)=1-\theta_{NE}+Q_{IP}+Q_{CP}-\theta_{CP}$. It follows that

$$\widehat{cost}(E3)\le\widehat{cost}(B3)\Leftrightarrow Q_{IP}+Q_{CP}\ge 2\theta_{NE}+\theta_{CP}-1.$$

Second, suppose $\theta_{CP}\ge Q_{IP}+max\{Q_{NE},Q_{CP}\}$. Then $\Delta_3^E\ge 0$ and $\Delta_3^B=0$. Thus, $\widehat{cost}(E3)=\theta_{NE}+\theta_{CP}-Q_{IP}-Q_{NE}$, and $\widehat{cost}(B3)=1-\theta_{NE}$.Then

$$\widehat{cost}(E3)\le\widehat{cost}(B3)\Leftrightarrow Q_{IP}+Q_{NE}\ge 2\theta_{NE}+\theta_{CP}-1.$$

Taking together, we have the claimed properties.

In the remaining two steps, we write $r_E\equiv min\{\widehat{cost}(E1),\widehat{cost}(E2)\}$ and $r_B\equiv min\{\widehat{cost}(B1),\widehat{cost}(B2)\}$.

Step 2: In the case of $min\{Q_{NE},Q_{CP}\}\ge|\theta_C-\theta_{CP}|$, we have $r_E\le r_B$ if and only if $2Q_{IP}+Q_{NE}+Q_{CP}\ge\theta_C+\theta_{CP}$.

In this case, $\Delta_1^E=\Delta_2^E=\Delta_1^B=\Delta_2^B=0$. Hence, $r_E=min\{\theta_C,\theta_{CP}\}-Q_{IP}$, and $r_B=1-max\{\theta_C,\theta_{CP}\}-Q_0$. It follows that $r_E\le r_B\Leftrightarrow 2Q_{IP}+Q_{NE}+Q_{CP}\ge\theta_C+\theta_{CP}$.

Step 3: In the case of $max\{\theta_C,\theta_{CP}\}\le|\theta_C-\theta_{CP}|$, if $Q_{IP}+min\{Q_{NE},Q_{CP}\}>\frac{1}{2}(\theta_C+\theta_{CP})$, then $r_E<r_B$; if $Q_{IP}+max\{Q_{NE},Q_{CP}\}<\frac{1}{2}(\theta_C+\theta_{CP})$, then $r_E>r_B$.

In this case, first, suppose $\theta_{CP}\ge\theta_C$. Then $\Delta_1^E\ge 0$, $\Delta_2^E=0$, $\Delta_1^B=0$, and $\Delta_2^B\ge 0$. Thus $r_E=\theta_{CP}-Q_{IP}-Q_{NE}$ and $r_B=1-\theta_C-Q_0-Q_{CP}$. Hence,

(8) $$r_E\le r_B\ \Leftrightarrow\ 2Q_{IP}+2Q_{NE}\ge\theta_C+\theta_{CP}.$$

Second, suppose $\theta_{CP}<\theta_C$. Then $\Delta_1^E=0$, $\Delta_2^E\ge 0$, $\Delta_1^B\ge 0$, and $\Delta_2^B=0$. Hence $r_E=\theta_C-Q_{IP}-Q_{CP}$ and $r_B=1-\eta_2-Q_0-Q_{NE}$. Therefore,

(9) $$r_E \le r_B \iff 2Q_{IP} + 2Q_{CP} \ge \theta_C + \theta_{CP}.$$

Collecting (8) and (9), we establish the claim.

To complete the proof, take together Steps 1 to 3. Then we know that if $Q_{IP} + min\{Q_{NE}, Q_{CP}\} > 2\theta_{NE} + \theta_{CP} - 1$ and $Q_{IP} + min\{Q_{NE}, Q_{CP}\} > \frac{1}{2}(\theta_C + \theta_{CP})$, then $\widehat{cost}(E3) < \widehat{cost}(B3)$ and $r_E < r_B$, so that by Young's theorem, C is the generically unique long-run equilibrium. Conversely, if $Q_{IP} + max\{Q_{NE}, Q_{CP}\} > 2\theta_{NE} + \theta_{CP} - 1$ and $Q_{IP} + max\{Q_{NE}, Q_{CP}\} < \frac{1}{2}(\theta_C + \theta_{CP})$, then $\widehat{cost}(B3) < \widehat{cost}(E3)$ and $r_B < r_E$, so that D is the generically unique long-run equilibrium.

This completes the proof of Proposition 2.

3.5.3. Proof of Proposition 3

From Table A5, we know the relative costs of transitions between C and D. Let $\theta_C = \theta_{NE} = \theta_{CP} = 0.5$. Then $\Delta_1^E = \Delta_2^E = \Delta_1^B = \Delta_2^B = 0$, and

$$r_E \equiv min\{\widehat{cost}(E1), \widehat{cost}(E2)\} = \frac{1}{2} - Q_{IP}$$

$$r_B \equiv min\{\widehat{cost}(B1), \widehat{cost}(B2)\} = \frac{1}{2} - Q_0.$$

Therefore, $r_E \le r_B$ if and only if $Q_{IP} \ge Q_0$, which is equivalent to $2Q_{IP} + Q_{NE} + Q_{CP} \ge 1$.

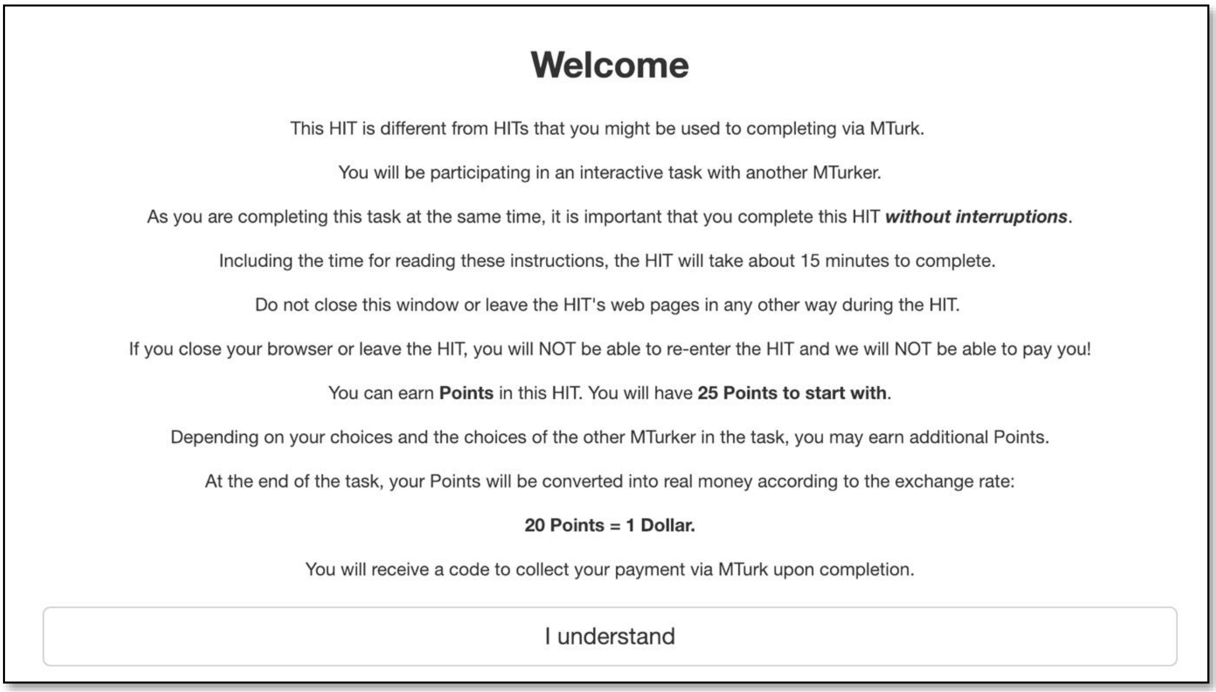
Observe that $r_E, r_B \le \frac{1}{2}$, but $\widehat{cost}(E3), \widehat{cost}(B3) \ge \frac{1}{2}$. Hence, $E3$ and $B3$ are not the paths with strictly minimum costs. Therefore, by Young's theorem, $Q_{IP} > Q_0$, implying $r_E < r_B$, is both necessary and sufficient for C to be the generically unique long-run equilibrium. And $Q_{IP} < Q_0$, implying $r_E > r_B$, is necessary and sufficient for D to be the generically unique long-run equilibrium. This completes the proof of Proposition 3.

4. Experimental Materials

Below we show on-screen instructions as displayed to participants. We start with the CC experiment in which participants could condition punishment of their interaction partner on descriptive norms of cooperation. Then we show the CP experiment, in which participants could condition punishment of their interaction partner on descriptive norms of punishment.

Where necessary we add additional notes and explanations in grey boxes like this one.

4.1. Instructions for the conditional cooperation (CC) experiment



Welcome

This HIT is different from HITs that you might be used to completing via MTurk.

You will be participating in an interactive task with another MTurker.

As you are completing this task at the same time, it is important that you complete this HIT ***without interruptions***.

Including the time for reading these instructions, the HIT will take about 15 minutes to complete.

Do not close this window or leave the HIT's web pages in any other way during the HIT.

If you close your browser or leave the HIT, you will NOT be able to re-enter the HIT and we will NOT be able to pay you!

You can earn **Points** in this HIT. You will have **25 Points to start with**.

Depending on your choices and the choices of the other MTurker in the task, you may earn additional Points.

At the end of the task, your Points will be converted into real money according to the exchange rate:

20 Points = 1 Dollar.

You will receive a code to collect your payment via MTurk upon completion.

I understand

Instructions

You and **another MTurker** will form a pair and participate in this task at the same time.

We will refer to the MTurker in your pair as **your partner**.

You and your partner have received these same instructions. You two will play a game.

The game has two stages: **Stage 1** and **Stage 2**.

Continue

Stage 1

(Page 1/3)

In Stage 1, you and your partner at the same time choose between ***Keep*** and ***Share***.

There are **four possible outcomes**:

If you choose:	and your partner chooses:	then you get:	and your partner gets:
Keep	***Keep***	16	16
Share	***Share***	18	18
Keep	***Share***	25	9
Share	***Keep***	9	25

After you and your partner have each made your choice, you will proceed to Stage 2.
Your earnings from Stage 1 will carry over to Stage 2.

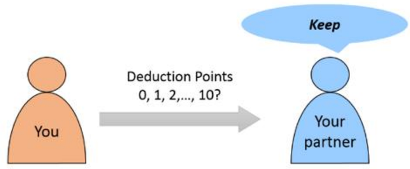
Continue

Stage 2

(Page 2/3)

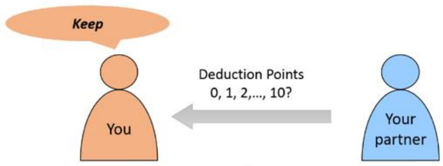
Once you and your partner have each submitted your choice for Stage 1, you proceed to Stage 2.

If your partner chose ***Keep*** in Stage 1, you can choose to assign **up to 10 Deduction Points** to your partner.



For each Deduction Point you assign, **1 Point** will be deducted from **your earnings**, and **3 Points** will be deducted from **your partner's earnings**.

Similarly, if you chose ***Keep*** in Stage 1, your partner can choose to assign up to 10 Deduction Points to you.
For each Deduction Point your partner chooses to assign, 1 Point will be deducted from your partner's earnings and 3 Points will be deducted from your earnings.

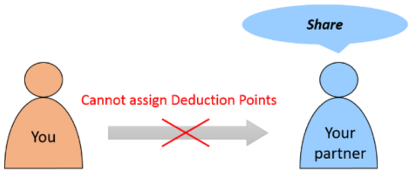


Previous | Continue

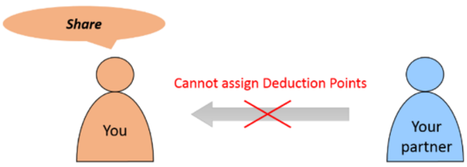
Stage 2

(Page 3/3)

If your partner chose ***Share*** in Stage 1, you **cannot** assign any Deduction Points to your partner.



Similarly, if you chose ***Share*** in Stage 1, your partner **cannot** assign Deduction Points to you.



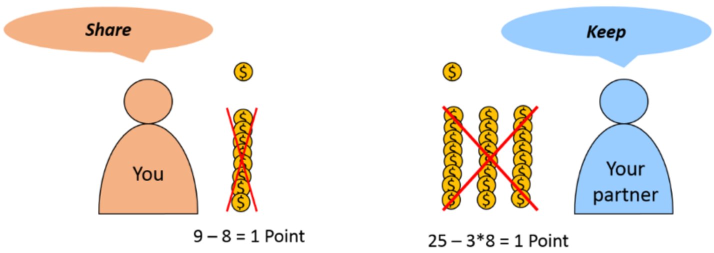
Previous | Continue

Here is an example:

Suppose that, in Stage 1, you choose ***Share*** and your partner chooses ***Keep***.
Then at the end of Stage 1, **you** will have **9 Points** and **your partner** will have **25 Points**.

In Stage 2, suppose that you assign **8 Deduction Points** to your partner. Then at the end of the game:

Your earnings will be **1 Point** (9 points from Stage 1 *minus* 8 points deducted from Stage 2).
Your partner's earnings will be **1 Point** (25 points from Stage 1 *minus* 3 x 8 points deducted from Stage 2).



Previous | Continue

Note that the control questions on the next screen included numerical examples of punishment that do not equalize the payoffs of both players (see below). As a result, participants had seen multiple different examples of possible outcomes of the two-stage game before making their decisions.

Summary

In Stage 1, you and your partner choose between ***Share*** and ***Keep***.

If you choose:	and your partner chooses:	then you get:	and your partner gets:
Keep	***Keep***	16	16
Share	***Share***	18	18
Keep	***Share***	25	9
Share	***Keep***	9	25

If your partner chooses ***Keep***, you can assign up to 10 Deduction Points to your partner in Stage 2. For each Deduction Point you assign, 1 Point will be deducted from your earnings, and 3 Points will be deducted from your partner's earnings.

If your partner chooses ***Share*** in Stage 1, you CANNOT assign Deduction Points to your partner in Stage 2.

Your partner faces the same choices as you do.

Questions to check whether you understand the task

1. Suppose that in Stage 1, you choose *Share* and your partner chooses *Keep*.

(a) How many Points would you earn from Stage 1?

(b) How many Points would your partner earn from Stage 1?

2. Subsequently in Stage 2, you assign 5 Deduction Points to your partner.

(a) How many Points would you have at the end of the game?

(b) How many Points would your partner have at the end of the game?

Previous | Continue

Stage 1

You answered the questions correctly.

Now click the button below to start Stage 1.

Continue

Your choice for Stage 1

Remember:

If you choose:	and your partner chooses:	then you get:	and your partner gets:
Keep	***Keep***	16	16
Share	***Share***	18	18
Keep	***Share***	25	9
Share	***Keep***	9	25

Now please make your choice:

Share

Keep

After you enter your decision, press "Submit" to proceed to Stage 2.
Once you press "Submit", you **cannot** go back and change your choice for Stage 1.

Submit

Stage 2

(Page 1/3)

Your choice for Stage 1 has been recorded. Stage 2 begins now.

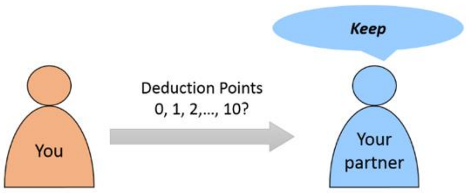
Before we inform you about your partner's choice, imagine that

your partner has chosen *Keep*.

Hence:

You can assign up to 10 Deduction Points to your partner.
For each Deduction Point you assign, 1 Point will be deducted from your earnings, and 3 Points will be deducted from your partner's earnings.

If your partner indeed chose ***Keep***, the Deduction points you assign will be implemented.
If your partner in fact chose ***Share***, the Deduction points you assign will NOT be implemented.



Continue

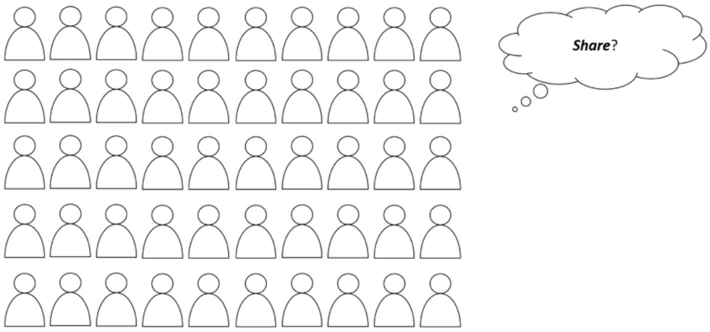
Stage 2

(Page 2/3)

Over 200 MTurkers from the USA have participated in this study before.
They were also paired and made choices between ***Share*** and ***Keep***.

Before you make your decision, we will **randomly select fifty MTurkers (NOT your partner)** from the ones who participated in this study before.

We will tell you how many of the randomly selected **previous MTurkers** (NOT your partner) chose ***Share***.



Previous | Continue

Stage 2

(Page 3/3)

We distinguish between **11 possible situations** regarding how many of the selected **previous MTurkers** (NOT your partner) chose ***Share***. The eleven possible situations are:

Less than 5 percent of them chose ***Share***;
Between 5 and 15 percent of them chose ***Share***;
Between 15 and 25 percent of them chose ***Share***;
Between 25 and 35 percent of them chose ***Share***;
Between 35 and 45 percent of them chose ***Share***;
Between 45 and 55 percent of them chose ***Share***;
Between 55 and 65 percent of them chose ***Share***;
Between 65 and 75 percent of them chose ***Share***;
Between 75 and 85 percent of them chose ***Share***;
Between 85 and 95 percent of them chose ***Share***;
More than 95 percent of them chose ***Share***.

Before we inform you about how many of the randomly selected previous MTurkers (NOT your partner) chose ***Share***, we ask you to consider EACH of the situations above.
You need to indicate, for **EACH** of the situations above, how many Deduction Points you assign to your partner if your partner chose ***Keep***.

After you submit your decisions for all eleven situations, we will let you know which situation actually occurred.
The **Deduction points** you assign in the **actual situation** will be used to calculate the earnings of you and your partner.
Since you do not know yet which situation is the actual one when you make your decisions, this means that you need to consider each of the situations above seriously.

Note:

Your partner is NOT one of these previous MTurkers.
You CANNOT assign Deduction Points to any of these previous MTurkers, and the previous MTurkers CANNOT assign Deduction Points to you or your partner.
The Deduction Points you assign only affect the earnings of you and your partner.

Previous | Next

Questions to check whether you understand the task

1. Is it possible for you to assign Deduction Points to the randomly selected MTurkers (NOT your partner)?

Yes, this is possible.

No, this is NOT possible.

2. Is it possible for the randomly selected MTurkers to assign Deduction Points to your partner?

Yes, this is possible.

No, this is NOT possible.

3. Suppose that, in Stage 1, your partner chose *Keep*. Among the fifty randomly selected MTurkers (NOT your partner), between 75 and 85 percent of them chose *Share*. In this situation, you assign 3 Deduction Points to your partner if your partner chose *Keep*.

(a) How many Points will be deducted from your earnings?

(b) How many Points will be deducted from your partner's earnings?

(c) How many Points will be deducted from the previous MTurkers' earnings because of the Deduction Points you assign?

Previous

Continue

Stage 2

You answered the questions correctly.

Now click the button below to start Stage 2.

Continue

Your choice for Stage 2

(Page 1/2)

Now consider each of the eleven situations and make your decisions.
"**The others**" in the descriptions below are referred to **the fifty randomly selected MTurkers** (NOT your partner).

Less than 5 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 5 and 15 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 15 and 25 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 25 and 35 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 35 and 45 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Next page

Your choice for Stage 2

(Page 2/2)

Now consider each of the eleven situations and make your decisions.
"**The others**" in the descriptions are referred to **the fifty randomly selected MTurkers** (NOT your partner).

Between 45 and 55 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 55 and 65 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 65 and 75 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 75 and 85 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Between 85 and 95 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

More than 95 percent of the others (NOT your partner) chose ***Share***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

To go back and change your decisions for previous situations, press "Previous page" below.
To submit your final decision, press "Submit".

Previous page | Submit

Questionnaire

You have now completed the decision making part of this HIT. Please fill out this brief questionnaire.

Once your partner is ready, we will calculate you and your partner's earnings and display them on your screen.

You will then receive a code to collect your earnings on MTurk.

1. What is your age?

2. What is your gender?

Male

Female

3. Could you briefly describe your reasoning for choosing to either *Share* or *Keep*?

remaining characters 200

4. Could you briefly describe the reasoning you used when allocating Deduction Points? In particular, why did or didn't you make your choices dependent on how many previous MTurkers chose *Share*?

remaining characters 400

Submit

4.2. Stage 2 instructions for the conditional punishment (CP) experiment

Stage 1 of the game was exactly the same as in the CC experiment. We omit it to avoid duplicating information.

Stage 2

(Page 1/3)

Your choice for Stage 1 has been recorded. Stage 2 begins now.

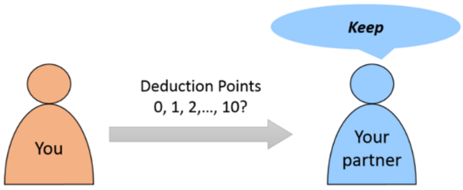
Before we inform you about your partner's choice, imagine that

your partner has chosen *Keep*.

Hence:

You can assign up to 10 Deduction Points to your partner.
For each Deduction Point you assign, 1 Point will be deducted from your earnings, and 3 Points will be deducted from your partner's earnings.

If your partner indeed chose ***Keep***, the Deduction points you assign will be implemented.
If your partner in fact chose ***Share***, the Deduction points you assign will NOT be implemented.



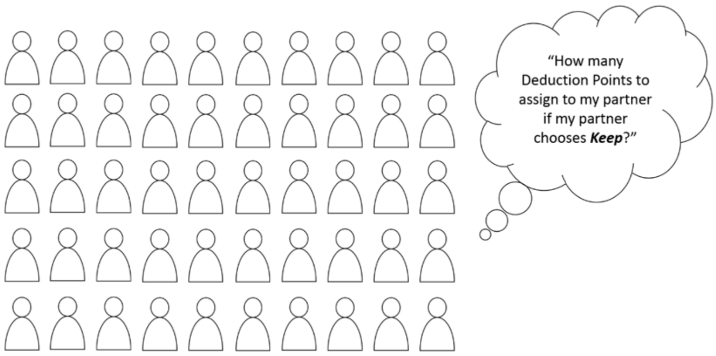
Continue

Stage 2

(Page 2/3)

Over 200 MTurkers from the USA have participated in this study before.
They were also paired and could also assign 0,1,2,…, or 10 Deduction Points to their partners when their partners chose ***Keep***.

Before you make your decision, we will **randomly select fifty MTurkers (NOT your partner)** from the ones who participated in this study before.
We will tell you the **average Deduction Points** that the randomly selected **previous MTurkers** (NOT your partner) assigned to their partners when their partners chose ***Keep***.



Previous | Continue

Stage 2

(Page 3/3)

Since each player can assign up to 10 Deduction Points, there are **eleven possible situations** regarding the **average Deduction Points** that the selected **previous MTurkers** (NOT your partner) assigned to their partners.
Rounding to the nearest integer, the eleven possible situations are:

On average, they assigned 0 Deduction Points;
On average, they assigned 1 Deduction Point;
On average, they assigned 2 Deduction Points;
On average, they assigned 3 Deduction Points;
On average, they assigned 4 Deduction Points;
On average, they assigned 5 Deduction Points;
On average, they assigned 6 Deduction Points;
On average, they assigned 7 Deduction Points;
On average, they assigned 8 Deduction Points;
On average, they assigned 9 Deduction Points;
On average, they assigned 10 Deduction Points.

Before we inform you about the average Deduction Points that the randomly selected MTurkers actually assigned, we ask you to consider EACH of the situations above.
You need to indicate, for **EACH** of the situations above, how many Deduction Points you assign to your partner if your partner chose ***Keep***.

After you submit your decisions for all eleven situations, we will let you know which situation actually occurred.
The **Deduction points** you assign in the **actual situation** will be used to calculate the earnings of you and your partner.
Since you do not know yet which situation is the actual one when you make your decisions, this means that you need to consider each of the situations above seriously.

Note:

Your partner is NOT one of these previous MTurkers.
You CANNOT assign Deduction Points to any of these previous MTurkers, and the previous MTurkers CANNOT assign Deduction Points to you or your partner.
The Deduction Points you assign only affect the earnings of you and your partner.

Previous | Next

Questions to check whether you understand the task

1. Is it possible for you to assign Deduction Points to the randomly selected MTurkers (NOT your partner)?

Yes, this is possible.

No, this is NOT possible.

2. Is it possible for the randomly selected MTurkers to assign Deduction Points to your partner?

No, this is NOT possible.

Yes, this is possible.

3. Suppose that, in Stage 1, your partner chose *Keep*. The fifty randomly selected MTurkers (NOT your partner) on average assigned 5 Deduction Points to their partners when their partners chose *Keep*. In this situation, you assign 3 Deduction Points to your partner when your partner chooses *Keep*.

(a) How many Points will be deducted from your earnings?

(b) How many Points will be deducted from your partner's earnings?

(c) How many Points will be deducted from the previous MTurkers' earnings because of the Deduction Points you assign?

Previous | Continue

Stage 2

You answered the questions correctly.

Now click the button below to start Stage 2.

Continue

Your choice for Stage 2

(Page 1/2)

Now consider each of the eleven situations and make your decisions.
"**The others**" in the descriptions are referred to **the fifty randomly selected MTurkers** (NOT your partner).

On average, **the others assigned 0 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 1 Deduction Point** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 2 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 3 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 4 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

Next page

Your choice for Stage 2

(Page 2/2)

Now consider each of the eleven situations and make your decisions.
"**The others**" in the descriptions are referred to **the fifty randomly selected MTurkers** (NOT your partner).

On average, **the others assigned 5 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 6 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 7 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 8 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 9 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

On average, **the others assigned 10 Deduction Points** when their partners chose ***Keep***. Enter the Deduction Points you assign to your partner below if your partner chose ***Keep***.

To go back and change your decisions for previous situations, press "Previous page" below.
To submit your final decisions, press "Submit".

Previous page | Submit

5. Supplementary References

Eshel, I., Samuelson, L., & Shaked, A. (1998). Altruists, egoists, and hooligans in a local interaction model. In *American Economic Review* (Vol. 88, Issue 1, pp. 157–179). JSTOR. https://doi.org/10.2307/116823

Goyal, S., & Vega-Redondo, F. (2005). Network formation and social coordination. *Games and Economic Behavior*, *50*(4), 178–207. https://doi.org/10.1016/j.geb.2004.01.005

Jackson, M. O., & Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, *106*(2), 265–295. https://doi.org/10.1006/JETH.2001.2903

Kandori, M., Mailath, G. J., & Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, *61*(1), 29–56. https://doi.org/10.2307/2951777

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a “common is moral” heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, *147*(2), 228–242.

Young, H. P. (1993). The evolution of conventions. In *Econometrica* (Vol. 61, Issue 1, pp. 57–84). JSTOR. https://doi.org/10.2307/2951778

Young, H. P. (1998). Conventional contracts. *Review of Economic Studies*, *65*(4), 773–792. https://doi.org/10.1111/1467-937X.00068

Young, H. P. (2001). *Individual strategy and social structure: An evolutionary theory of institutions*. Princeton University Press.

Young, H. P. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. In *American Economic Review* (Vol. 99, Issue 5, pp. 1899–1924). American Economic Association. https://doi.org/10.1257/aer.99.5.1899

Young, H. P. (2011). The dynamics of social innovation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Supplement 4), 21285–21291. https://doi.org/10.1073/pnas.1100973108

Young, H. P., & Burke, M. A. (2001). Competition and custom in economic contracts: A case study of Illinois agriculture. *American Economic Review*, *91*(3), 559–573. https://doi.org/10.1257/aer.91.3.559

Young, H. P., & Foster, D. (1991). Cooperation in the long-run. In *Games and Economic Behavior* (Vol. 3, Issue 1, pp. 145–156). https://doi.org/10.1016/0899-8256(91)90010-C