OPEN

Social cycling and conditional responses in the Rock-Paper-Scissors game

Zhijian Wang[1]*, Bin Xu[2,1]* & Hai-Jun Zhou[3]

[1]Experimental Social Science Laboratory, Zhejiang University, Hangzhou 310058, China, [2]Public Administration College, Zhejiang Gongshang University, Hangzhou 310018, China, [3]State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China.

How humans make decisions in non-cooperative strategic interactions is a big question. For the fundamental Rock-Paper-Scissors (RPS) model game system, classic Nash equilibrium (NE) theory predicts that players randomize completely their action choices to avoid being exploited, while evolutionary game theory of bounded rationality in general predicts persistent cyclic motions, especially in finite populations. However as empirical studies have been relatively sparse, it is still a controversial issue as to which theoretical framework is more appropriate to describe decision-making of human subjects. Here we observe population-level persistent cyclic motions in a laboratory experiment of the discrete-time iterated RPS game under the traditional random pairwise-matching protocol. This collective behavior contradicts with the NE theory but is quantitatively explained, without any adjustable parameter, by a microscopic model of win-lose-tie conditional response. Theoretical calculations suggest that if all players adopt the same optimized conditional response strategy, their accumulated payoff will be much higher than the reference value of the NE mixed strategy. Our work demonstrates the feasibility of understanding human competition behaviors from the angle of non-equilibrium statistical physics.

The Rock-Paper-Scissors (RPS) game is a fundamental non-cooperative game. It has been widely used to study competition phenomena in society and biology, such as species diversity of ecosystems[1–6] and price dispersion of markets[7,8]. This game has three candidate actions R (rock), P (paper) and S (scissors). In the simplest settings the payoff matrix is characterized by a single parameter, the payoff a of the winning action ($a > 1$, see Fig. 1A)[9]. There are the following non-transitive dominance relations among the actions: R wins over S, P wins over R, yet S wins over P (Fig. 1B), so no action is absolutely better than the others.

The RPS game is also a basic model system for studying decision-making of human subjects in competitive environments and the associated social dynamics and non-equilibrium physics. Assuming ideal rationality for players who repeatedly playing the RPS game within a population, classical game theory predicts that individual players will completely randomize their action choices so that their behaviors will be unpredictable and not be exploited by the other players[10,11]. This is referred to as the mixed-strategy Nash equilibrium (NE), in which every player chooses the three actions with equal probability 1/3 at each game round (see Supplementary Notes online). When the payoff parameter $a < 2$ this NE is evolutionarily unstable with respect to small perturbations but it becomes evolutionarily stable at $a > 2$ (see Supplementary Notes online)[12]. On the other hand, evolutionary game theory drops the infinite rationality assumption and looks at the RPS game from the angle of evolution and adaption[13–18]. Evolutionary models based on various microscopic learning rules (such as the replicator dynamics[12,19–21], the best response dynamics[22,23] and the logit dynamics[24,25]) generally predict cyclic evolution patterns for the action marginal distribution (mixed strategy) of each player, especially in finite populations.

Empirical verification of non-equilibrial persistent cycling in the human-subject RPS game (and other non-cooperative games) has been rather nontrivial, as the recorded evolutionary trajectories are usually highly stochastic and not long enough to draw convincing conclusions. Two of the present authors partially overcame these difficulties by using social state velocity vectors[26] and forward and backward transition vectors[27] to visualize violation of detailed balance in game evolution trajectories, but a simple way of quantitatively measuring persistent cyclic behaviors in a highly stochastic trajectory was still lacking. The cycling frequency of directional flows in the neutral RPS game ($a = 2$) was later quantitatively measured in[28] using a coarse-grained counting technique. Cason and co-workers[29] using another cycle rotation index as the order parameter also obtained evidence of persistent cycling in some evolutionarily stable RPS-like games, if players were allowed to update actions asynchronously in continuous time and were informed about the social states of the whole population by some sophisticated 'heat maps'.
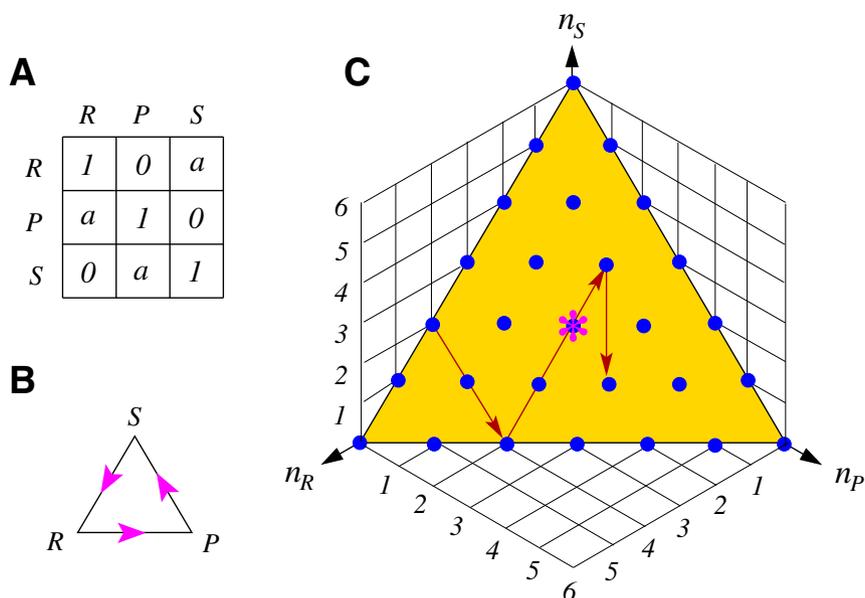
Figure 1 | The Rock-Paper-Scissors game. (A) Each matrix entry specifies the row action's payoff. (B) Non-transitive dominance relations (R beats S, P beats R, S beats P) among the three actions. (C) The social state plane for a population of size $N = 6$. Each filled circle denotes a social state (n_R, n_P, n_S); the star marks the centroid $\mathbf{c}_0$; the arrows indicate three social state transitions at game rounds $t = 1, 2, 3$.

In this work we investigate whether cycling is a general aspect even for the simplest RPS game. We adopt an improved cycle counting method on the basis of our earlier experiences[28] and study directional flows in evolutionarily stable ($a > 2$) and unstable ($a < 2$) discrete-time RPS games. We show strong evidence that the RPS game is an intrinsic non-equilibrium system, which cannot be fully described by the NE concept even in the evolutionarily stable region but rather exhibits persistent population-level cyclic motions. We then bridge the collective cycling behavior and the highly stochastic decision-making of individuals through a simple conditional response (CR) mechanism. Our empirical data confirm the plausibility of this microscopic model of bounded rationality. Our theoretical calculations also demonstrate that, if all the players adopt the same CR strategy and if the transition parameters of this strategy are chosen in an optimized way, this CR strategy will outperform the NE mixed strategy in terms of the accumulated payoffs of individual players, yet the action marginal distribution of individual players is indistinguishable from that of the NE mixed strategy. Our work as a successful attempt of understanding competition dynamics from the perspective of non-equilibrium statistical physics may stimulate future more refined experimental and theoretical studies on the microscopic mechanisms of decision-making and learning in basic game systems[19,30–34].

Results

Experimental system. We recruited a total number of 360 students from different disciplines of Zhejiang University to form 60 disjoint populations of size $N = 6$. Each population then carries out one experimental session by playing the RPS game 300 rounds (taking 90–150 minutes) with a fixed value of a. In real-world situations individuals often have to make decisions based only on partial input information. We mimic such situations by adopting the traditional random pairwise-matching experimental protocol[11]: At each game round (time) t the players are randomly paired within the population and compete with their pair opponent once; after that each player gets feedback information about her own payoff as well as her and her opponent's action. As the experimental session finishes, the players are paid in real cash proportional to their accumulated payoffs (see Methods). Our experimental setting differs from those of two other recent experiments, in which every player competes

against the whole population[9,29] and may change actions in continuous time[29]. We set $a = 1.1, 2, 4, 9$ and 100, respectively, in one-fifth of the populations so as to compare the dynamical behaviors in the evolutionarily unstable, neutral, stable and deeply stable regions.

Action marginal distribution of individual players. We observe that the individual players shift their actions frequently in all the populations except one with $a = 1.1$ (this exceptional population is discarded from further analysis, see Supplementary Notes online). Averaged among the 354 players of these 59 populations, the probabilities that a player adopts action R, P, S at one game round are, respectively, 0.36 ± 0.08, 0.33 ± 0.07 and 0.32 ± 0.06 (mean ± s.d.). We obtain very similar results for each set of populations of the same a value (see Supplementary Table S1 online). These results are consistent with NE and suggest the NE mixed strategy is a good description of a player's marginal distribution of actions. However, a player's actions at two consecutive times are not independent but correlated. As demonstrated in Fig. 2A–2E, at each time the players are more likely to repeat their last action than to shift action either counter-clockwise (i.e., $R \rightarrow P, P \rightarrow S, S \rightarrow R$, see Fig. 1B) or clockwise ($R \rightarrow S, S \rightarrow P, P \rightarrow R$). This inertial effect is especially strong at $a = 1.1$ and it diminishes as a increases.

We notice that at $a \geq 2$, an individual player's probability of making a clockwise action shift is equal to or just slightly different from that of making a counter-clockwise action shift (Fig. 2A–2E). There is no or only very weak cycling behavior at the level of individual players in the evolutionarily neutral ($a = 2$) and stable ($a > 2$) RPS games, in accordance with the NE theory. As shown in Fig. 2F–2J, the action shift statistics of individual players can be well explained by the later introduced conditional response model.

Collective behaviors of the whole population. The social state of the population at any time t is denoted as $\mathbf{s}(t) \equiv (n_R(t), n_P(t), n_S(t))$ with n_q being the number of players adopting action $q \in \{R, P, S\}$. Since $n_R + n_P + n_S \equiv N$ there are $(N + 1)(N + 2)/2$ such social states, all lying on a three-dimensional plane bounded by an equilateral triangle (Fig. 1C). Each population leaves a trajectory on this plane as the RPS game proceeds. To detect rotational flows, we assign for every social state transition $\mathbf{s}(t) \rightarrow \mathbf{s}(t + 1)$ a rotation angle $\theta(t)$, which measures the angle this transition rotates with respect to the centroid
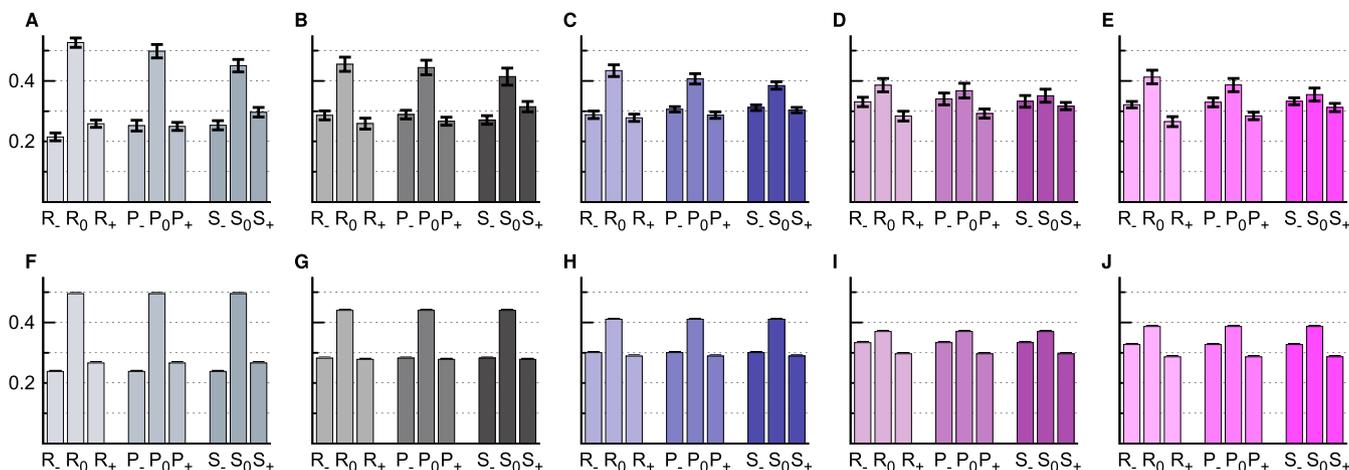
Figure 2 | Action shift probability conditional on a player's current action. If a player adopts action R at one game round, this player's probability of repeating the same action at the next game round is denoted as R_0, while the probability of performing a counter-clockwise or clockwise action shift is denoted, respectively, as R_+ and R_-. The conditional probabilities P_0, P_+, P_- and S_0, S_+, S_- are defined similarly. (A–E) The mean (vertical bin) and the SEM (standard error of the mean, error bar) of each conditional probability obtained by averaging over the different populations of the same payoff parameter $a = 1.1, 2, 4, 9$, and 100 (from left to right). (F–J) The corresponding action shift probability values predicted by the conditional response model using the parameters of Fig. 3F–3J as inputs.

$\mathbf{c}_0 \equiv (N/3, N/3, N/3)$ of the social state plane (see Methods)[28]. Positive and negative θ values signify counter-clockwise and clockwise rotations, respectively, while $\theta = 0$ means the transition is not a rotation around $\mathbf{c}_0$. For example, we have $\theta(1) = \pi/3$, $\theta(2) = 0$, and $\theta(3) = -2\pi/3$ for the three transitions shown in Fig. 1C.

The net number of cycles around $\mathbf{c}_0$ during the time interval $[t_0, t_1]$ is computed by

$$C_{t_0,t_1} \equiv \sum_{t=t_0}^{t_1-1} \frac{\theta(t)}{2\pi}. \qquad (1)$$

As shown in Fig. 3A–3E, $C_{1,t}$ has an increasing trend in most of the 59 populations, indicating persistent counter-clockwise cycling. The cycling frequency of each trajectory in $[t_0, t_1]$ is evaluated by

$$f_{t_0,t_1} \equiv \frac{C_{t_0,t_1}}{t_1 - t_0}. \qquad (2)$$

The values of $f_{1,300}$ for all the 59 populations are listed in Table 1, from which we obtain the mean frequency to be 0.031 ± 0.006 ($a = 1.1$, mean $\pm$ SEM), 0.027 ± 0.008 ($a = 2$), 0.031 ± 0.008 ($a = 4$), 0.022 ± 0.008 ($a = 9$) and 0.018 ± 0.007 ($a = 100$). These mean frequencies are all positive irrespective to the particular value of a, indicating that behind the seemingly highly irregular social state evolution process, there is a deterministic pattern of social state cycling from slightly rich in action R, to slightly rich in P, then to slightly rich in S, and then back to slightly rich in R again. Statistical analysis confirms that $f_{1,300} > 0$ is significant for all the five sets of populations (Wilcoxon signed-rank test, $p < 0.05$). The correlation between the mean cycling frequency $f_{1,300}$ and the payoff parameter a is not statistically significant (Spearman's correlation test: $r = -0.82$, $p = 0.19$, for $n = 5$ mean frequencies; and $r = -0.16$, $p = 0.24$, for $n = 59$ frequencies). We also notice that the mean cycling frequency in the second half of the game ($f_{151,300}$) is slightly higher than that in the first half ($f_{1,150}$) for all the five sets of populations (Supplementary Table S2 online), suggesting that cycling does not die out with time.

A recent experimental work[35] also observed cycling behaviors in a RPS-like game with more than three actions. Evidences of persistent cycling in some complete-information and continuous-time RPS-like games were reported in another experimental study[29]. However, no (or only very weak) evidence of population-level cycling was detected in[29] if action updating was performed in discrete time.

Here and in Ref. 28 we find that even discrete-time updating of actions will lead to collective cyclic motions in the RPS game, and such a population-level behavior is not affected by the particular value of a.

Empirical conditional response patterns. Under the assumption of mixed-strategy NE (i.e., each player chooses the three actions with equal probability at every game round, independent of each other and of the payoffs of previous plays), the social state transitions should obey the detailed balance condition. Therefore the observed persistent cycling behavior cannot be understood within the NE framework. Persistent cycling can also not be explained by the independent decision model which assumes the action choice of a player at one time is influenced only by her action at the previous time (see Supplementary Notes online). Using the empirically determined action shift probabilities of Fig. 2A–2E as inputs, we find that this independent decision model predicts the cycling frequency to be 0.0050 (for $a = 1.1$), -0.0005 ($a = 2$), -0.0024 ($a = 4$), -0.0075 ($a = 9$) and -0.0081 ($a = 100$), which are all very close to zero and significantly different from the empirical values.

The action choices of different players must be mutually influenced. Our empirical data shown in Fig. 3F–3J confirm the existence of such mutual influences. Let us denote by O the performance (output) of a player at a given game round, with $O \in \{W \text{ (win)}, T \text{ (tie)}, L \text{ (lose)}\}$. Conditional on the output O, the probability that this player will decide to shift action clockwise or counter-clockwise or keep the same action in the next play is denoted as O_-, O_+ and O_0 ($\equiv 1 - O_- - O_+$), respectively. Most interestingly, we see from Fig. 3F–3J that if a player wins over her opponent in one play, her probability (W_0) of repeating the same action in the next play is considerably higher than her probabilities (W_- and W_+) of shifting actions. Furthermore, for payoff parameter $a \geq 2$, if a player loses to her opponent in one play, she is more likely to shift action clockwise (probability L_-) than either to keep the old action (L_0) or to shift action counter-clockwise (L_+).

The conditional response model. Inspired by these empirical observations, we develop a simplest nontrivial model by assuming the following conditional response strategy: at each game round, every player review her previous performance $O \in \{W, T, L\}$ and makes an action choice according to the corresponding three

3
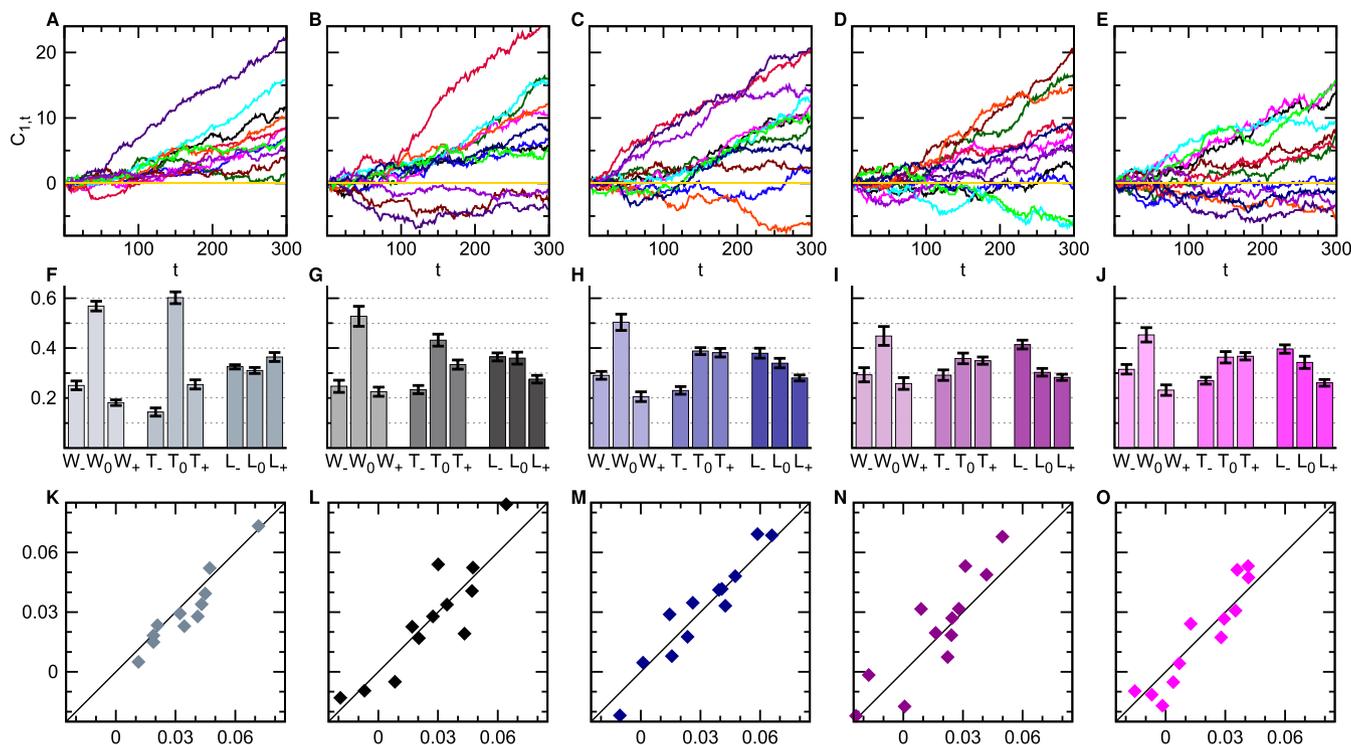
Figure 3 | Social cycling explained by conditional response. The payoff parameter is $a = 1.1, 2, 4, 9$ and 100 from left-most column to right-most column. (A–E) Accumulated cycle numbers $C_{1,t}$ of 59 populations. (F–J) Empirically determined CR parameters, with the mean (vertical bin) and the SEM (error bar) of each CR parameter obtained by considering all the populations of the same a value. (K–O) Comparison between the empirical cycling frequency (vertical axis) of each population and the theoretical frequency (horizontal axis) obtained by using the empirical CR parameters of this population as inputs.

conditional probabilities (O_-, O_0, O_+). This model is characterized by a set $\Gamma \equiv \{W_-, W_+; T_-, T_+; L_-, L_+\}$ of six CR parameters. Notice this CR model differs qualitatively from the discrete-time logit dynamics model[24,25] used in Ref. 28, which assumes each player has global information about the population's social state.

We can solve this win-lose-tie CR model analytically and numerically (see Supplementary Notes online). Let us denote by n_{rr}, n_{pp}, n_{ss}, n_{rp}, n_{ps} and n_{sr}, respectively, as the number of pairs in which the competition being R–R, P–P, S–S, R–P, P–S, and S–R, in one game

round t. Given the social state $\mathbf{s} = (n_R, n_P, n_S)$ at time t, the conditional joint probability distribution of these six integers is expressed as

$$P_{\mathbf{s}}\left[n_{rr}, n_{pp}, n_{ss}, n_{rp}, n_{ps}, n_{sr}\right] =$$

$$\frac{n_R! \, n_P! \, n_S! \, \delta^{n_R}_{2n_{rr}+n_{sr}+n_{rp}} \, \delta^{n_P}_{2n_{pp}+n_{rp}+n_{ps}} \, \delta^{n_S}_{2n_{ss}+n_{ps}+n_{sr}}}{(N-1)!! \, 2^{n_{rr}} n_{rr}! \, 2^{n_{pp}} n_{pp}! \, 2^{n_{ss}} n_{ss}! \, n_{rp}! \, n_{ps}! \, n_{sr}!}, \quad (3)$$

where $(N-1)!! \equiv 1 \times 3 \times \dots \times (N-3) \times (N-1)$ and δ^n_m is the Kronecker symbol ($\delta^n_m = 1$ if $m = n$ and $= 0$ if otherwise). With the help of this expression, we can then obtain an explicit formula for the social state transition probability $M_{cr}[\mathbf{s}'|\mathbf{s}]$ from $\mathbf{s}$ to any another social state $\mathbf{s}'$ (see Methods). We then compute numerically the steady-state social state distribution $P^*_{cr}(\mathbf{s})$ of this Markov matrix[36] and other average quantities of interest. For example, the mean steady-state cycling frequency f_{cr} of this model is computed by

$$f_{cr} = \sum_{\mathbf{s}} P^*_{cr}(\mathbf{s}) \sum_{\mathbf{s}'} M_{cr}[\mathbf{s}'|\mathbf{s}] \theta_{\mathbf{s} \to \mathbf{s}'}, \quad (4)$$

where $\theta_{\mathbf{s} \to \mathbf{s}'}$ is the rotation angle associated with the social state transition $\mathbf{s} \to \mathbf{s}'$, see Eq. (7).

Using the empirically determined response parameters as inputs, the CR model predicts the mean cycling frequencies for the five sets of populations to be $f_{cr} = 0.035$ ($a = 1.1$), 0.026 ($a = 2$), 0.030 ($a = 4$), 0.018 ($a = 9$) and 0.017 ($a = 100$), agreeing well with the empirical measurements. Such good agreements between model and experiment are achieved also for the 59 individual populations (Fig. 3K–3O). In addition, we find the empirically observed inertial effect of Fig. 2A–2E is quantitatively reproduced by the CR model without any fitting parameter (see Fig. 2F–2J).

Because of the rotational symmetry of the conditional response parameters, the CR model predicts that each player's action marginal

Table 1 | Empirical cycling frequencies $f_{1,300}$ of 59 populations

$a = 1.1$	2	4	9	100
$f_{1,300}$	$f_{1,300}$	$f_{1,300}$	$f_{1,300}$	$f_{1,300}$
0.039	0.019	0.033	0.007	0.047
0.023	0.023	0.005	−0.002	0.004
0.005	0.054	0.029	0.053	0.024
0.029	0.034	0.041	0.027	0.051
0.015	−0.010	0.008	0.068	0.027
0.052	0.052	0.042	−0.017	0.031
0.028	0.084	0.069	0.032	0.017
0.034	0.041	−0.022	0.049	−0.017
0.073	−0.013	0.069	0.020	−0.012
0.023	0.017	0.035	−0.022	0.053
0.018	−0.005	0.048	0.018	−0.010
	0.028	0.018	0.032	−0.005
μ 0.031	0.027	0.031	0.022	0.018
σ 0.019	0.029	0.026	0.027	0.025
δ 0.006	0.008	0.008	0.008	0.007

μ: the mean cycling frequency, σ: the standard deviation (s.d.) of cycling frequencies, δ: the standard error (SEM) of the mean cycling frequency ($\delta = \sigma/\sqrt{n_s}$, with sample number $n_s = 11$ for $a = 1.1$ and $n_s = 12$ for $a = 2, 4, 9$ and 100).

distribution is uniform, identical to the NE mixed strategy (Supplementary Notes online). On the other hand, according to this model, the expected payoff g_{cr} per game round of each player is

$$g_{cr} = g_0 + (a - 2) \times (1/6 - \tau_{cr}/2), \quad (5)$$

where $g_0 \equiv (1 + a)/3$ is the expected payoff of the NE mixed strategy, and τ_{cr} is the average fraction of ties among the $N/2$ pairs at each game round, with the expression

$$\tau_{cr} = \sum_{\mathbf{s}} P_{cr}^*(\mathbf{s}) \frac{\sum_{n_{rr},\ldots,n_{rs}} (n_{rr} + n_{pp} + n_{ss}) P_{\mathbf{s}}(n_{rr},\ldots,n_{sr})}{N/2}. \quad (6)$$

The value of g_{cr} depends on the CR parameters. By uniformly sampling 2.4×10^9 instances of Γ from the three-dimensional probability simplex, we find that for $a > 2$, g_{cr} has high chance of being lower than g_0 (Fig. 4), with the mean value of $(g_{cr} - g_0)$ being $-0.0085(a - 2)$. (Qualitatively the same conclusion is obtained for larger N values, e.g., see Supplementary Fig. S1 online for $N = 12$.) This is consistent with the mixed-strategy NE being evolutionarily stable[12]. On the other hand, the four g_{cr} values (for the four cases of $a \neq 2$) determined by the empirical CR parameters and the corresponding four mean payoffs of the empirical data sets all weakly exceed g_0, indicating that individual players are adjusting their responses to achieve higher accumulated payoffs (Supplementary Notes online). The positive gap between g_{cr} and g_0 may further enlarge if the individual players were given more learning time to optimize their response parameters (e.g., through increasing the repeats of the game).

As shown in Fig. 4 and Supplementary Fig. S1 online, the CR parameters have to be highly optimized to achieve a large value of g_{cr}. For population size $N = 6$ we give three examples of the sampled best CR strategies for $a > 2$: $\Gamma_1 = \{0.002, 0.000; 0.067, 0.110; 0.003, 0.003\}$, with cycling frequency $f_{cr} = 0.003$ and $g_{cr} = g_0 + 0.035(a - 2)$; $\Gamma_2 = \{0.995, 0.001; 0.800, 0.058; 0.988, 0.012\}$, with $f_{cr} = -0.190$ and $g_{cr} = g_0 + 0.034(a - 2)$; $\Gamma_3 = \{0.001, 0.004; 0.063, 0.791; 0.989, 0.001\}$, with $f_{cr} = 0.189$ and $g_{cr} = g_0 + 0.033(a - 2)$. For large a these CR strategies outperform the NE mixed strategy in payoff by about 10%. Set Γ_1 indicates that population-level cycling is not a necessary condition for achieving high payoff values. On the other hand, set Γ_3 implies $W_0 \approx 1, L_0 \approx 0$, therefore this CR strategy can be regarded as an extension of the win-stay lose-shift (also called Pavlov) strategy, which has been shown by computer simulations to facilitate coop-
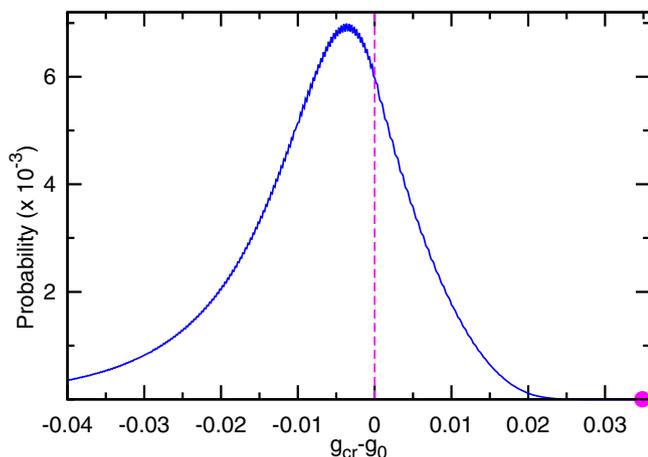
Figure 4 | Probability distribution of payoff difference $g_{cr} - g_0$ at population size $N = 6$. We assume $a > 2$ and set the unit of the horizontal axis to be $(a - 2)$. The solid line is obtained by sampling 2.4×10^9 CR strategies uniformly at random; the filled circle denotes the maximal value of g_{cr} among these samples.

eration in the prisoner's dilemma game[37–40]. We should also emphasize that the empirically observed CR transition parameters (Fig. 3F–3J) still differ considerably from those of the win-stay lose-shift strategy Γ_3.

Discussion

In game-theory literature it is common to equate individual players' action marginal distributions with their actual strategies[11,18]. In reality, however, decision-making and learning are very complicated neural processes[41–45]. The action marginal distributions are only a consequence of such complex dynamical processes, their coarse-grained nature makes them unsuitable to describe dynamical properties[17]. Our work on the finite-population RPS game clearly demonstrates this point. This game exhibits persistent cyclic motions at the population level (but not at the individual player level) which cannot be understood by the NE concept but are successfully explained by the empirical data-inspired CR mechanism. As far as the action marginal distributions of individual players are concerned, the CR strategy is indistinguishable from the NE mixed strategy, yet it is capable of bringing higher payoffs to the players if its parameters are optimized and all players adopt the same CR strategy. This simple conditional response strategy, with the win-stay lose-shift strategy being a special case, appears to be psychologically plausible for human subjects with bounded rationality[46,47]. For more complicated game payoff matrices, we can generalize the conditional response model accordingly by introducing a larger set of CR parameters (see Supplementary Notes online). It should be very interesting to re-analyze many existing laboratory experimental data[9,29,35,48–51] using this extended model. Figure 3 also reveals that the empirical CR parameters and the social-state cycling frequency change with the payoff parameter a. In a following paper we will study the effect of the payoff parameter a to the individual- and population-level behaviors in more detail[52].

The CR model as a simple model of decision-making under uncertainty deserves to be fully explored. For example, different players may have different CR transition parameters and these transition parameters may change with time constantly as a result of learning. We find the cycling frequency is not sensitive to population size N at given CR parameters (see Supplementary Fig. S2 online); and the cycling frequency is nonzero even for symmetric CR parameters (i.e., $W_+/W_- = T_+/T_- = L_+/L_- = 1$), as long as $W_0 \neq L_0$ (see Supplementary Fig. S3 online). The optimization issue of CR parameters is left out in this work. We will investigate whether an optimal CR strategy is achievable through simple stochastic learning rules[42,43,45]. The effects of memory length[53] and population size to the optimal CR strategies also need to be thoroughly studied. On the more biological side, whether conditional response is a basic decision-making mechanism of the human brain or just a consequence of more fundamental neural mechanisms is a challenging question for future studies.

Methods

Experiment. The experiment was approved by the Experimental Social Science Laboratory of Zhejiang University and performed at Zhejiang University in the period of December 2010 to March 2014. The corresponding author confirms that this experiment was performed in accordance with the approved social experiments guidelines and regulations. A total number of 360 undergraduate and graduate students of Zhejiang University volunteered to serve as the human subjects of this experiment. These students were openly recruited through a web registration system. Female students were slightly more enthusiastic than male students in registering as candidate human subjects of our experiment. Since we sampled students uniformly at random from the candidate list, more female students were recruited than male students (among the 360 students, the female versus male ratio is 217 : 143). Informed consent was obtained from all the participating human subjects.

The 360 human subjects (referred to as players in this work) were distributed into 60 populations of equal size $N = 6$. The six players of each population carried one experimental session by playing the RPS game for 300 rounds with fixed payoff parameter a, whose value is chosen from {1.1, 2, 4, 9, 100}. During the game process the players sited separately in a classroom, each of which facing a computer screen.

They were not allowed to communicate with each other during the whole experimental session. Written instructions were handed out to each player and the rules of the experiment were also orally explained by an experimental instructor. The rules of the experimental session are as follows:

(i) Each player plays the RPS game repeatedly with the same other five players.
(ii) Each player earns virtual points during the experimental session according to the payoff matrix shown in the written instruction. These virtual points are then exchanged into RMB as a reward to the player, plus an additional 5 RMB as show-up fee.
(iii) In each game round, the six players of each group are randomly matched by a computer program to form three pairs, and each player competes only with the pair opponent.
(iv) Each player has at most 40 seconds in one game round to make a choice among the three candidate actions "Rock", "Paper" and "Scissors". If this time runs out, the player has to make a choice immediately (the experimental instructor will loudly urge these players to do so). After a choice has been made it can not be changed.

Before the start of the actual experimental session, the player were asked to answer four questions to ensure that they understand completely the rules of the experimental session. These four questions are: (1) *If you choose "Rock" and your opponent chooses "Scissors", how many virtual points will you earn?* (2) *If you choose "Rock" and your opponent chooses also "Rock", how many virtual points will you earn?* (3) *If you choose "Scissors" and your opponent chooses "Rock", how many virtual points will you earn?* (4) *Do you know that at each game round you will play with a randomly chosen opponent from your group (yes/no)?*

During the experimental session, the computer screen of each player will show an information window and a decision window. The window on the left of the computer screen is the information window. The upper panel of this information window shows the current game round, the time limit (40 seconds) of making a choice, and the time left to make a choice. The color of this upper panel turns to green at the start of each game round. The color will change to yellow if the player does not make a choice within 20 seconds. The color will change to red if the decision time runs out (and then the experimental instructor will loudly urge the players to make a choice immediately). The color will change to blue if a choice has been made by the player. After all the players of the group have made their decisions, the lower panel of the information window will show the player's own choice, the opponent's choice, and the player's own payoff in this game round. The player's own accumulated payoff is also shown. The players are asked to record their choices of each round on the record sheet (Rock as R, Paper as P, and Scissors as S).

The window on the right of the computer screen is the decision window. It is activated only after all the players of the group have made their choices. The upper panel of this decision window lists the current game round, while the lower panel lists the three candidate actions "Rock", "Scissors", "Paper" horizontally from left to right. The player can make a choice by clicking on the corresponding action names. After a choice has been made by the player, the decision window becomes inactive until the next game round starts.

The reward in RMB for each player is determined by the following formula. Suppose a player i earns x_i virtual points in the whole experimental session, the total reward y_i in RMB for this player is then given by

$$y_i = x_i \times r + 5,$$

where r is the exchange rate between virtual point and RMB. According to the mixed-strategy Nash equilibrium, the expected payoff of each player in one game round is $(1 + a)/3$. Therefore we set the exchange rate to be $r = 0.45/(1 + a)$ to ensure that, under the mixed-strategy NE assumption, the expected total earning in RMB for a player will be 50 RMB irrespective of the particular experimental session. The value of the payoff parameter a, the numerical value of r, and the above-mentioned reward formula were listed in the written instruction and also orally mentioned by the experimental instructor at the instruction phase of the experiment.

Rotation angle computation. Consider a transition from one social state $\mathbf{s} = (n_R, n_P, n_S)$ at game round t to another social state $\tilde{\mathbf{s}} = (\tilde{n}_R, \tilde{n}_P, \tilde{n}_S)$ at game round $(t + 1)$, if at least one of the two social states coincides with the centroid $\mathbf{c}_0$ of the social state plane, or the three points $\mathbf{s}$, $\tilde{\mathbf{s}}$ and $\mathbf{c}_0$ lie on a straight line, then the transition $\mathbf{s} \rightarrow \tilde{\mathbf{s}}$ is not regarded as a rotation around $\mathbf{c}_0$, and the rotation angle $\theta = 0$. In all the other cases, the transition $\mathbf{s} \rightarrow \tilde{\mathbf{s}}$ is regarded as a rotation around $\mathbf{c}_0$, and the rotation angle is computed through

$$\theta = \text{sgn}_{\mathbf{s} \rightarrow \tilde{\mathbf{s}}} \times \text{acos}\left(\frac{3(n_R \tilde{n}_R + n_P \tilde{n}_P + n_S \tilde{n}_S) - N^2}{\sqrt{\left[3\left(n_R^2 + n_P^2 + n_S^2\right) - N^2\right]\left[3\left(\tilde{n}_R^2 + \tilde{n}_P^2 + \tilde{n}_S^2\right) - N^2\right]}} \right), \quad (7)$$

where $\text{acos}(x) \in [0, \pi]$ is the inverse cosine function, and $\text{sgn}_{\mathbf{s} \rightarrow \tilde{\mathbf{s}}} = 1$ if $[3(n_R \tilde{n}_P - n_P \tilde{n}_R) + N(n_P - n_R + \tilde{n}_R - \tilde{n}_P)] > 0$ (counter-clockwise rotation around $\mathbf{c}_0$) and $\text{sgn}_{\mathbf{s} \rightarrow \tilde{\mathbf{s}}} = -1$ if otherwise (clockwise rotation around $\mathbf{c}_0$).

Statistical Analysis. Statistical analyses, including Wilcoxon signed-rank test and Spearman's rank correlation test, were performed by using stata 12.0 (Stata, College Station, TX).

Transition matrix of the conditional response model. For the conditional response model, the transition probability $M_{cr}[\mathbf{s}'|\mathbf{s}]$ from the social state $\mathbf{s} \equiv (n_R, n_P, n_S)$ at time t to the social state $\mathbf{s}' \equiv (n_R', n_P', n_S')$ at time $(t + 1)$ is expressed as:

$$
M_{cr}[\mathbf{s}'|\mathbf{s}] = \sum_{n_{rr}, n_{pp}, \ldots, n_{sr}} \frac{n_R!\, n_P!\, n_S!\; \delta_{2n_{rr}+n_{sr}+n_{rp}}^{n_R}\, \delta_{2n_{pp}+n_{rp}+n_{ps}}^{n_P}\, \delta_{2n_{ss}+n_{ps}+n_{sr}}^{n_S}}{(N-1)!!\; 2^{n_{rr}} n_{rr}!\; 2^{n_{pp}} n_{pp}!\; 2^{n_{ss}} n_{ss}!\; n_{rp}!\, n_{ps}!\, n_{sr}!}
$$
$$
\times \sum_{n_{rr}^{rr}, \ldots, n_{rr}^{sr}} \frac{n_{rr}!\, T_0^{2n_{rr}^{rr}}\, T_+^{2n_{rr}^{pp}}\, T_-^{2n_{rr}^{ss}}\, (2T_+T_0)^{n_{rr}^{rp}}\, (2T_+T_-)^{n_{rr}^{ps}}\, (2T_0T_-)^{n_{rr}^{sr}}}{n_{rr}^{rr}!\, n_{rr}^{pp}!\, n_{rr}^{ss}!\, n_{rr}^{rp}!\, n_{rr}^{ps}!\, n_{rr}^{sr}!}\, \delta_{n_{rr}^{rr}+\ldots+n_{rr}^{sr}}^{n_{rr}}
$$
$$
\times \sum_{n_{pp}^{rr}, \ldots, n_{pp}^{sr}} \frac{n_{pp}!\, T_-^{2n_{pp}^{rr}}\, T_0^{2n_{pp}^{pp}}\, T_+^{2n_{pp}^{ss}}\, (2T_0T_-)^{n_{pp}^{rp}}\, (2T_+T_0)^{n_{pp}^{ps}}\, (2T_+T_-)^{n_{pp}^{sr}}}{n_{pp}^{rr}!\, n_{pp}^{pp}!\, n_{pp}^{ss}!\, n_{pp}^{rp}!\, n_{pp}^{ps}!\, n_{pp}^{sr}!}\, \delta_{n_{pp}^{rr}+\ldots+n_{pp}^{sr}}^{n_{pp}}
$$
$$
\times \sum_{n_{ss}^{rr}, \ldots, n_{ss}^{sr}} \frac{n_{ss}!\, T_+^{2n_{ss}^{rr}}\, T_-^{2n_{ss}^{pp}}\, T_0^{2n_{ss}^{ss}}\, (2T_+T_-)^{n_{ss}^{rp}}\, (2T_0T_-)^{n_{ss}^{ps}}\, (2T_+T_0)^{n_{ss}^{sr}}}{n_{ss}^{rr}!\, n_{ss}^{pp}!\, n_{ss}^{ss}!\, n_{ss}^{rp}!\, n_{ss}^{ps}!\, n_{ss}^{sr}!}\, \delta_{n_{ss}^{rr}+\ldots+n_{ss}^{sr}}^{n_{ss}}
$$
$$
\times \sum_{n_{rp}^{rr}, \ldots, n_{rp}^{sr}} \frac{n_{rp}!\; \delta_{n_{rp}^{rr}+\ldots+n_{rp}^{sr}}^{n_{rp}}}{n_{rp}^{rr}!\, n_{rp}^{pp}!\, n_{rp}^{ss}!\, n_{rp}^{rp}!\, n_{rp}^{ps}!\, n_{rp}^{sr}!}(W_-L_0)^{n_{rp}^{rr}}(W_0L_+)^{n_{rp}^{pp}}(W_+L_-)^{n_{rp}^{ss}}
$$
$$
\times (W_0L_0 + W_-L_+)^{n_{rp}^{rp}}(W_+L_- + W_0L_-)^{n_{rp}^{ps}}(W_+L_0 + W_-L_-)^{n_{rp}^{sr}} \quad (8)
$$
$$
\times \sum_{n_{ps}^{rr}, \ldots, n_{ps}^{sr}} \frac{n_{ps}!\; \delta_{n_{ps}^{rr}+\ldots+n_{ps}^{sr}}^{n_{ps}}}{n_{ps}^{rr}!\, n_{ps}^{pp}!\, n_{ps}^{ss}!\, n_{ps}^{rp}!\, n_{ps}^{ps}!\, n_{ps}^{sr}!}(W_+L_-)^{n_{ps}^{rr}}(W_-L_0)^{n_{ps}^{pp}}(W_0L_+)^{n_{ps}^{ss}}
$$
$$
\times (W_+L_0 + W_-L_-)^{n_{ps}^{rp}}(W_0L_0 + W_-L_+)^{n_{ps}^{ps}}(W_+L_+ + W_0L_-)^{n_{ps}^{sr}}
$$
$$
\times \sum_{n_{sr}^{rr}, \ldots, n_{sr}^{sr}} \frac{n_{sr}!\; \delta_{n_{sr}^{rr}+\ldots+n_{sr}^{sr}}^{n_{ps}}}{n_{sr}^{rr}!\, n_{sr}^{pp}!\, n_{sr}^{ss}!\, n_{sr}^{rp}!\, n_{sr}^{ps}!\, n_{sr}^{sr}!}(W_0L_+)^{n_{sr}^{rr}}(W_+L_-)^{n_{sr}^{pp}}(W_-L_0)^{n_{sr}^{ss}}
$$
$$
\times (W_+L_+ + W_0L_-)^{n_{sr}^{rp}}(W_+L_- + W_0L_-)^{n_{sr}^{ps}}(W_0L_0 + W_-L_+)^{n_{sr}^{sr}}
$$
$$
\times \delta_{2\left(n_{rr}^{rr}+n_{pp}^{rr}+n_{ss}^{rr}+n_{rp}^{rr}+n_{ps}^{rr}+n_{sr}^{rr}\right)+\left(n_{rr}^{rp}+n_{pp}^{rp}+n_{ss}^{rp}+n_{rp}^{rp}+n_{ps}^{rp}+n_{sr}^{rp}\right)+\left(n_{rr}^{sr}+n_{pp}^{sr}+n_{ss}^{sr}+n_{rp}^{sr}+n_{ps}^{sr}+n_{sr}^{sr}\right)}^{n_R'}
$$
$$
\times \delta_{2\left(n_{rr}^{pp}+n_{pp}^{pp}+n_{ss}^{pp}+n_{rp}^{pp}+n_{ps}^{pp}+n_{sr}^{pp}\right)+\left(n_{rr}^{rp}+n_{pp}^{rp}+n_{ss}^{rp}+n_{rp}^{rp}+n_{ps}^{rp}+n_{sr}^{rp}\right)+\left(n_{rr}^{ps}+n_{pp}^{ps}+n_{ss}^{ps}+n_{rp}^{ps}+n_{ps}^{ps}+n_{sr}^{ps}\right)}^{n_P'}
$$
$$
\times \delta_{2\left(n_{rr}^{ss}+n_{pp}^{ss}+n_{ss}^{ss}+n_{rp}^{ss}+n_{ps}^{ss}+n_{sr}^{ss}\right)+\left(n_{rr}^{ps}+n_{pp}^{ps}+n_{ss}^{ps}+n_{rp}^{ps}+n_{ps}^{ps}+n_{sr}^{ps}\right)+\left(n_{rr}^{sr}+n_{pp}^{sr}+n_{ss}^{sr}+n_{rp}^{sr}+n_{ps}^{sr}+n_{sr}^{sr}\right)}^{n_S'}.
$$

1. Sinervo, B. & Lively, C. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* **380**, 240–243 (1996).
2. Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**, 171–174 (2002).
3. Semmann, D., Krambeck, H.-J. & Milinski, M. Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature* **425**, 390–393 (2003).
4. Lee, D., McGreevy, B. P. & Barraclough, D. J. Learning and decision making in monkeys during a rock-paper-scissors game. *Cogn. Brain Res.* **25**, 416–430 (2005).
5. Reichenbach, T., Mobilia, M. & Frey, E. Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. *Nature* **448**, 1046–1049 (2007).
6. Allesina, S. & Levine, J. M. A competitive network theory of species diversity. *Proc. Natl. Acad. Sci. USA* **108**, 5638–5642 (2011).
7. Maskin, E. & Tirole, J. A theory of dynamic oligopoly, ii: Price competition, kinked demand curves, and edgeworth cycles. *Econometr.* **56**, 571–599 (1988).
8. Cason, T. N. & Friedman, D. Buyer search and price dispersion: a laboratory study. *J. Econ. Theory* **112**, 232–260 (2003).
9. Hoffman, M., Suetens, S., Nowak, M. A. & Gneezy, U. An experimental test of nash equilibrium versus evolutionary stability. In *Proc. 4th World Cong. Game Theory Soc. (GAMES 2012)*, session 145, paper 1 (Istanbul, Turkey, 2012) (Data of access: 05/02/2014).
10. Nash, J. F. Equilibrium points in *n*-person games. *Proc. Natl. Acad. Sci. USA* **36**, 48–49 (1950).
11. Osborne, M. J. & Rubinstein, A. *A Course in Game Theory* (MIT Press, New York, 1994).
12. Taylor, P. D. & Jonker, L. B. Evolutionarily stable strategies and game dynamics. *Math. Biosci.* **40**, 145–156 (1978).
13. Maynard Smith, J. & Price, G. R. The logic of animal conflict. *Nature* **246**, 15–18 (1973).
14. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, 1982).
15. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
16. Nowak, M. A. & Sigmund, K. Evolutionary dynamics of biological games. *Science* **303**, 793–799 (2004).
17. Szabó, G. & Fáth, G. Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).
18. Sandholm, W. M. *Population Games and Evolutionary Dynamics* (MIT Press, New York, 2010).
19. Claussen, J. C. & Traulsen, A. Cyclic dominance and biodiversity in well-mixed populations. *Phys. Rev. Lett.* **100**, 058104 (2008).

20. Roca, C. P., Cuesta, J. A. & Sánchez, A. Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics. *Phys. Life Rev.* **6**, 208–249 (2009).

21. Andrae, B., Cremer, J., Reichenbach, T. & Frey, E. Entropy production of cyclic population dynamics. *Phys. Rev. Lett.* **104**, 218102 (2010).

22. Matsui, A. Best response dynamics and socially stable strategies. *J. Econ. Theory* **57**, 343–362 (1992).

23. Hopkins, E. A note on best response dynamics. *Gam. Econ. Behav.* **29**, 138–150 (1999).

24. Blume, L. E. The statistical mechanics of strategic interation. *Gam. Econ. Behav.* **5**, 387–424 (1993).

25. Hommes, C. H. & Ochea, M. I. Multiple equilibria and limit cycles in evolutionary games with logit dynamics. *Gam. Econ. Behav.* **74**, 434–441 (2012).

26. Xu, B. & Wang, Z. Evolutionary dynamical patterns of 'coyness and philandering': Evidence from experimental economics. In *Proc. 8th Int. Conf. Compl. Sys.* 1313–1326 (Boston, MA, USA, 2011) (Data of access: 15/03/2014).

27. Xu, B. & Wang, Z. Test maxent in social strategy transitions with experimental two-person constant sum 2 × 2 games. *Resul. Phys.* **2**, 127–134 (2012).

28. Xu, B., Zhou, H.-J. & Wang, Z. Cycle frequency in standard rock-paper-scissors games: Evidence from experimental economics. *Physica A* **392**, 4997–5005 (2013).

29. Cason, T. N., Friedman, D. & Hopkins, E. Cycles and instability in a rock-paper-scissors population game: A continuous time experiment. *Rev. Econ. Stud.* **81**, 112–136 (2014).

30. Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).

31. Huang, J.-P. *Econophysics* (Higher Education Press, Beijing, 2013).

32. Ao, P. Emerging of stochastic dynamical equalities and steady state thermodynamics from darwinian dynamics. *Commun. Theor. Phys.* **49**, 1073–1090 (2008).

33. Ao, P., Chen, T.-Q. & Shi, J.-H. Dynamical decomposition of markov processes without detailed balance. *Chin. Phys. Lett.* **30**, 070201 (2013).

34. Zhou, T., Han, X.-P. & Wang, B.-H. Towards the understanding of human dynamics. In Burguete, M. & Lam, L. (eds.) *Science Matters: Humanities as Complex Systems* 207–233 (World Scientific, Singapore, 2008).

35. Frey, S. & Goldstone, R. L. Cyclic game dynamics driven by iterated reasoning. *PLoS ONE* **8**, e56416 (2013).

36. Kemeny, J. G. & Snell, J. L. *Finite Markov Chains; with a New Appendix "Generalization of a Fundamental Matrix"* (Springer-Verlag, New York, 1983).

37. Kraines, D. & Kraines, V. Learning to cooperate with pavlov: an adaptive strategy for the iterated prisoner's dilemma with noise. *Theory Decis.* **35**, 107–150 (1993).

38. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364**, 56–58 (1993).

39. Wedekind, C. & Milinski, M. Human cooperation in the simultaneous and the alternating prisoner's dilemma: Pavlov versus generous tit-for-tat. *Proc. Natl. Acad. Sci. USA* **93**, 2686–2689 (1996).

40. Posch, M. Win-stay, lose-shift strategies for repeated games–memory length, aspiration levels and noise. *J. Theor. Biol.* **198**, 183–195 (1999).

41. Glimcher, P. W., Camerer, C. F., Fehr, E. & Poldrack, R. A. (eds.). *Neuroeconomics: Decision Making and the Brain* (Academic Press, London, 2009).

42. Börgers, T. & Sarin, R. Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **77**, 1–14 (1997).

43. Posch, M. Cycling in a stochastic learning algorithm for normal form games. *J. Evol. Econ.* **7**, 193–207 (1997).

44. Galla, T. Intrinsic noise in game dynamical learning. *Phys. Rev. Lett.* **103**, 198702 (2009).

45. Janacsek, K. & Nemeth, D. Predicting the future: From implicit learning to consolidation. *Int. J. Psychophysiol.* **83**, 213–221 (2012).

46. Camerer, C. Behavioral economics: Reunifying psychology and economics. *Proc. Natl. Acad. Sci. USA* **96**, 10575–10577 (1999).

47. Camerer, C. *Behavioral game theory: Experiments in strategic interaction* (Princeton University Press, Princeton, NJ, 2003).

48. Berninghaus, S. K., Ehrhart, K.-M. & Keser, C. Continuous-time strategy selection in linear population games. *Exper. Econ.* **2**, 41–57 (1999).

49. Traulsen, A., Semmann, D., Sommerfeld, R. D., Krambeck, H.-J. & Milinski, M. Human strategy updating in evolutionary games. *Proc. Natl. Acad. Sci. USA* **107**, 2962–2966 (2010).

50. Gracia-Lázaro, C. *et al.* Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma. *Proc. Natl. Acad. Sci. USA* **109**, 12922–12926 (2012).

51. Chmura, T., Goerg, S. J. & Selten, R. Generalized impulse balance: An experimental test for a class of 3 × 3 games. *Rev. Behav. Econ.* **1**, 27–53 (2014).

52. Wang, Z. & Xu, B. Incentive and stability in the Rock-Paper-Scissors game: An experimental investigation. *e-print arXiv*/1407.1170 (2014).

53. Press, W. H. & Dyson, F. J. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci. USA* **109**, 10409–10413 (2012).

Acknowledgments

Author contributions

Z.W. and B.X. contributed equally to this work. Z.W., B.X. designed and performed experiment; B.X. measured conditional response transition probabilities; H.J.Z. developed analytical and numerical methods; B.X., Z.W., H.J.Z. analyzed and interpreted data; H.J.Z., B.X. wrote the paper.

Additional information