Optimal Cooperation-Trap Strategies for the Iterated Rock-Paper-Scissors Game

Zedong Bi, Hai-Jun Zhou*
State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing, China

Abstract
In an iterated non-cooperative game, if all the players act to maximize their individual accumulated payoff, the system as a whole usually converges to a Nash equilibrium that poorly benefits any player. Here we show that such an undesirable destiny is avoidable in an iterated Rock-Paper-Scissors (RPS) game involving two rational players, X and Y. Player X has the option of proactively adopting a cooperation-trap strategy, which enforces complete cooperation from the rational player Y and leads to a highly beneficial and maximally fair situation to both players. That maximal degree of cooperation is achievable in such a competitive system with cyclic dominance of actions may stimulate further theoretical and empirical studies on how to resolve conflicts and enhance cooperation in human societies.

Introduction
The solution concept of Nash equilibrium (NE) plays a fundamental role both in classic game theory and in evolutionary game theory [1–4]. This concept is developed under the assumption that the players of a game system are sufficiently rational, so that they are able to learn accurately the strategies of the competing players and to optimize their own strategy accordingly. A Nash equilibrium is then a point in the strategy space of the game system such that any single player is unable to achieve better performance by changing her/his own strategy in any arbitrary way.

Many non-cooperative games have only a unique NE. When such a game is played by highly rational players who act to maximize their individual accumulated payoff, it is unavoidable that the system will sooner or later converge to this unique equilibrium situation. Unfortunately, however, it is usually the case that the NE of a non-cooperative game is an unfavorable or even miserable destiny for all the players. Let’s consider the two-player Prisoner’s Dilemma (PD) game as a simple example. The cooperative situation of both players choosing not to confess is much better than the defection situation of both players choosing to confess, but the latter is the unique NE of this game while the former is not [5]. The Nash equilibrium theory therefore predicts that cooperation is unlikely to sustain when rational players face the conflict between self-interest and group benefit. Yet cooperation is actually a ubiquitous phenomenon of human society at all levels, and it is also widely observed in various biological systems. Researchers have been puzzled by these facts very much for many years, and they have proposed a long list of microscopic mechanisms trying to explain the promotion and maintenance of cooperation [6–8].

In this paper we study the issue of cooperation in the iterated two-player Rock-Paper-Scissors (RPS) game, which is a fundamental non-cooperative game with cyclic dominance among its action choices (namely Rock beats Scissors, Scissors beats Paper, and Paper in turn beats Rock), see Fig. 1. This game has been widely used to study competition phenomena in society and biology, especially species diversity and pattern formation (see, for example, references [9–12]). While the NE theoretical framework assumes that the rational players of such a game behave passively in the sense that they try to maximize individual gains by making best responses to the inferred/experienced strategy of the opponent, we assume that one of the players might act more proactively. An intelligent and rational player may ask the following question: how should I design my own strategy so that my rational opponent(s), in best response to me, for sure will adopt certain strategy that is most beneficial to me? In later discussions we refer to such a strategy as a cooperation-trap (CT) strategy, as it has the effect of trapping an opponent in a cooperation state. When optimized, such a CT strategy offers high and maximally fair accumulated payoffs to both players.

In a literature search for related studies, we found that an early paper of Grofman and Pool [13] investigated cooperation in the PD game from the same angle of intelligent design. In this pioneer but largely forgotten paper, the authors proved that a partial Tit-for-Tat strategy [14] has the potential of enforcing cooperation in the two-player iterated PD game. The Win-stay, Lose-shift strategy [15,16] can also be analyzed in a similar way.

The present effort can be regarded as an extension of the Grofman-Pool theory to the iterated RPS game, which has the
additional difficulty of having more than two action choices that are related by a rotation symmetry [see Fig. 1B]. This same theoretical framework may also be applicable to many other two-player iterated non-cooperative games, and it may serve as a guiding principle of designing fair strategies for the purpose of resolving conflicts and enhancing cooperation in human societies.

Results

The Rock-Paper-Scissors game

Consider two players X and Y playing the RPS game for an indefinite number of rounds. At every game round each player can choose one action among three candidate actions: Rock (R) beats Scissors (S), S beats Paper (P), and P in turn beats R.

![Figure 1: The Rock-Paper-Scissors game. (A) The payoff matrix. Each matrix element is the payoff of the row player X's action in competition with the column player Y's action. (B) The cyclic (non-transitive) dominance relationship among the three candidate actions: Rock (R) beats Scissors (S), S beats Paper (P), and P in turn beats R.](image)

The associated optimal EPR of player X is

\[ g_x = p_r p_s r + p_s p_p r + p_r p_p r + \alpha(p_r p_s r + p_s p_p r + p_r p_p r) \]

(1)

If the strategy of player X have the following property that \( p_r > p_s > p_p \), then because action P is strictly the least favored choice of player X, then player Y realizes that it is of his best interest to choose action R in every game round of \( q_r = q_s = 0 \) if \( p_r - p_p > 0 \) but to choose action P in every game round of \( q_r = q_s = 0 \) if \( p_r - p_p < 0 \). In other words, player X traps player Y to stay in a pure strategy which has maximal degree of predictability. Player X of course should choose the strategy parameters \( p_r \) and \( p_p \) to maximize her EPR \( g_x \) under the constraint of not destroying the nice trapping effect of her strategy. It is not difficult to verify the following conclusions:

(2). If the payoff parameter \( a \epsilon [1, (1 + \sqrt{3})/2] \), the optimal CT strategy is

\[ p_r = \frac{a}{2a - 1} - \epsilon, \quad p_p = 0, \quad p_s = \frac{a - 1}{2a - 1} + \epsilon \]  

(3)

(Here and in latter discussions, \( \epsilon \rightarrow 0^+ \) is an arbitrarily small positive value.) The associated maximal EPR for player X is \( g_x^* = a/(2a - 1) \), while player Y is very satisfied with sticking to action R and getting a larger EPR of \( g_y^* = a^2/(2a - 1) \). To give a concrete example, at \( a = 1.1 \) we have \( g_y^* \approx 0.917 \) and \( g_y^* \approx 0.008 \), which are considerably larger than the NE payoff \( g_y = 0.7 \).

Memoryless cooperation-trap strategies

We now develop CT strategies for player X, and begin with the simplest case of memoryless strategies, namely at every game round player X does not consider her and her opponent’s prior actions nor the outcomes of prior plays but chooses actions R, P and S according to the corresponding probabilities \( p_r, p_p, \) and \( p_s \) \((= 1 - p_r - p_p)\), which are fixed by player X at the beginning of the whole game. Without loss of generality we assume that \( p_r \geq p_p \) and \( p_r \geq p_s \), i.e., action R is a favorable choice of X.

As player Y is sufficiently intelligent, he will figure out the strategy of X after a small number of game repeats. Alternatively, with the aim of promoting cooperation from player Y, player X may also explicitly inform Y about her strategy parameters, which are \( p_r \) and \( p_p \), in the present case. And since Y is sufficiently rational, he then for sure will adopt the optimized probabilities \( q_r^*, q_p^*, \) and \( q_s^* \), of choosing the three actions R, P and S.

The EPR \( g_x \) of player X and the optimized EPR \( g_y^* \) of player Y are

\[ g_x = p_r q_r^* + p_s q_s^* + p_p q_p^* + \alpha(p_r q_s^* + p_s q_p^* + p_r q_p^*) \]

(1)

\[ g_y^* = q_r^* p_r + q_p^* p_p + q_s^* p_s + \alpha(q_r^* p_s + q_s^* p_p + q_r^* p_p) \]

(2)

If the strategy of player X have the following property that \( p_r > p_s > p_p \), then because action P is strictly the least favored choice of player X, then player Y realizes that it is of his best interest to choose action R in every game round of \( q_r = q_s = 0 \) if \( p_r - p_p > 0 \) but to choose action P in every game round of \( q_r = q_s = 0 \) if \( p_r - p_p < 0 \). In other words, player X traps player Y to stay in a pure strategy which has maximal degree of predictability. Player X of course should choose the strategy parameters \( p_r \) and \( p_p \) to maximize her EPR \( g_x \) under the constraint of not destroying the nice trapping effect of her strategy. It is not difficult to verify the following conclusions:

(2). If the payoff parameter \( a \in [2, \sqrt{3} + \infty) \), the optimal CT strategy is

\[ p_r = \frac{a}{2a - 1} + \epsilon, \quad p_p = 0, \quad p_s = \frac{a - 1}{2a - 1} - \epsilon \]  

(4)

The associated optimal EPR of player X is \( g_x^* = a(a - 1)/(2a - 1) \), while player Y receives a larger EPR value of \( g_y^* = a^2/(2a - 1) \) by sticking to action P. Notice that when \( a \) is sufficiently large, \( g_y^* \approx a^2/2 - 1/4 \) and \( g_y^* \approx a^2/2 + 1/4 \), which are almost 1.5 times that of the NE payoff \( g_y = 0.7 \).

Figure 2 gives a direct view about how the optimal EPRs of both players and the optimal CT strategy of player X change with \( a \). This optimal memoryless CT strategy indeed offers both players higher accumulated payoffs than the NE mixed strategy does. However, the passive player Y benefits more than the proactive player X. It is then natural for player X to feel that she has sacrificed too much for enforcing cooperation and to declare that such a CT strategy, although better than the NE mixed strategy, is unfair as her opponent earns more by free riding. Furthermore, this CT strategy is worse than the NE mixed strategy in the parameter range of \( a \in [1, 0.663, 7.32] \).

These shortcomings of the memoryless CT strategy can be eliminated by increasing the memory length of the CT strategy.
Cooperation-trap strategies with finite memory length

Recent laboratory experiments carried at Zhejiang University [19] revealed that decision-making of human subjects has strong memory effect, namely the payoffs of the previous game rounds influence considerably a player’s action choices in the following game rounds. For the RPS game, the implications of such conditional response strategies have not yet been fully explored. Here we suggest that the proactive player X can adopt an optimized version of such a strategy to enforce fair cooperation.

When the payoff parameter \( a > 2 \), a play output of win-lose brings payoff \( a \) to the group, while a tie output only brings lower payoff 2. Therefore it is desirable for player X to discourage the occurrence of tie output. For the simplest case of unit memory length, the CT strategy then goes as follows: If player X wins over or loses to player Y in the previous game round, then in the next round she chooses action \( R \) with probability \( p_s \geq 1/2 \) and action \( S \) with probability \( p_t = 1 - p_s \) (she avoid choosing action \( P \), i.e., \( p_P = 0 \)); but if X ties with Y in the previous game round, then in the next round she chooses the three candidate actions with equal probability 1/3. This strategy has only a single parameter \( p_s \). The motivation for player X to adopt the NE mixed strategy after experiencing a tie output is to discourage player Y from choosing action \( R \); although Y might get a higher expected payoff in one game round by choosing action \( R \) rather than action \( P \), the former choice has a high probability of leading to a tie, which will then reduce player Y’s expected payoff to \( g_y \) in the following one or even more game rounds.

On the other hand, when \( 1 < a < 2 \), a play output of tie is better off to the group than a win-lose output. Then player X has the option of implementing a CT strategy to discourage player Y from either winning over or losing to her. Again for the simplest case of unit memory length, the recipe of the CT strategy is: If player X ties with player Y in the previous game round, then in the next round she chooses action \( R \) with probability \( p_s \geq 1/2 \) and action \( S \) with probability \( p_t = 1 - p_s \); but if X either wins over or loses to Y in the previous round, then in the next round she chooses the three candidate actions with equal probability 1/3.

It turns out that the optimal CT strategy of unit memory length has the following quantitative properties:

1. If \( a \in [1.3/2] \), then the optimal value \( p_s^* \) for the choice probability \( p_s \) is \( p_s^* = 1 - \epsilon \), and the optimal EPRs for player X and player Y are equal, \( g_x^* = g_y^* = 1 \).
2. If \( a \in (3/2, 1.649] \), then \( p_s^* = 8a/\left[9a - 3 + (9a - 3)^2 - 48a^2\right] - \epsilon \), and the optimal EPRs for X and Y are, respectively, \( g_x^* = \frac{1 + a(1 - p_s^*)}{4 - 3p_s^*} \) and \( g_y^* = \frac{1}{4 - 3p_s^*} \).
3. If \( a \in (2.905, 6) \), then \( p_s^* = 4a/[3(a - 1) + (9(a - 1)^2 + 24a^2) + \epsilon] \), and the optimal EPRs for X and Y are, respectively, \( g_x^*(1 - p_s^*)a \) and \( g_y^* = p_s^*a \).
4. If \( a \in [6, +\infty) \), then \( p_s^* = 1/2 + \epsilon \), and the optimal EPRs for X and Y are equal, \( g_x^* = g_y^* = a/2 \).

Figure 3 gives a direct view of these properties. Compared with the optimal memoryless CT strategy of Fig. 2, we notice a major qualitative improvement is that this new optimal CT strategy achieves fair outcomes to player X and player Y when \( 1 \leq a \leq 3 \) or \( a \geq 6 \). However, this optimal CT strategy of unit memory length is still not perfect, as it is not applicable for \( a \in [1.649, 2.905] \), and it is not completely fair to the proactive player X for \( a \in (3/2, 1.649) \cap (2.905, 6) \).

To completely eliminate these undesirable features, player X can increase the memory length of her CT strategy and therefore be more non-tolerant to defection. There are many ways of implementing such an idea. When \( a > 2 \), arguably the simplest CT strategy of memory length \( m \) goes as follows: By default player X adopts the mixed strategy \( (p_P, 0.1 - p_P) \) in every game round, namely she chooses action \( R \) with probability \( p_s \geq 1/2 \) and action \( S \) with the remaining probability \( p_t = 1 - p_s \); however if a tie...
occurs in one game round, then player X shifts to the NE mixed strategy \((1/3,1/3,1/3)\) in the next \(m\) game rounds and then shifts back to the default strategy \((p_r,0,1-p_r)\) in the \((m+1)\)-th game round. It is a simple exercise to check that, if

\[
p_r \geq \max \left( \frac{1}{2}, \frac{2a}{\sqrt{4am^2 + A^2 - |A|}} \right), \tag{5}
\]

where \(A \equiv 1 + (1+a)m/3 - 2a\), then player Y will be satisfied with sticking to action \(P\) in every game round. If player X sets the memory length to the smallest positive integer \(m^*\) which reduces Eq. (5) to the trivial requirement of \(p_r \geq 1/2\), then it is optimal for player X to set \(p_r\) to the value \(p_r^* = 1/2\), and the optimal EPRs for player X and player Y are equal, \(G_r = G_p = a/2\). Notice that for \(a\) approaches 2 from above with \(a > 2 + \epsilon\), the required minimal memory length diverges as \(m^* \approx 6/\epsilon\). In other words, it is most difficult to enforce fair cooperation when \(a \approx 2\), see Fig. 4.

If the payoff parameter \(a < 2\), an optimal CT strategy with memory length \(m\) can be constructed following the same line of reasoning as above, namely that player X adopts \(R\) at every game round, but if she loses to player Y in one game round, then she shifts to the NE mixed strategy in the next \(m\) game rounds and then shifts back to the default strategy \((1,0,0)\) in the \((m+1)\)-th game round. We can easily verify that if player X sets the memory length to be \(m > (a-1)/(2-a)\), then it is optimal for player Y to stick to action \(R\) in every game round, and the optimal EPRs for both players are equal, \(G_r = G_p = 1\).

As clearly demonstrated in Fig. 4, for each payoff parameter \(a \neq 2\), an optimal CT strategy with a finite memory length \(m^*\) can be implemented to achieve maximal and fair accumulated payoff for both players. At \(a = 2\), there is no need to adopt a CT strategy, as the NE mixed strategy is itself optimal.

**Discussion**

We have demonstrated in this paper that fair cooperation can be achieved in the two-player iterated RPS game. Such a highly cooperative state brings maximal accumulated payoff to the group, and it is not enforced by external authorities but by the proactive decision of one player to adopt an optimal cooperation-trap strategy. The basic designing principle of such optimal CT strategies should be generally applicable to other two-player iterated non-cooperation games.

For the optimal CT strategies to work, the passive player Y is assumed to be considerably rational so that he adopts a best response strategy to that of his opponent X to maximize his accumulated payoff, while the proactive player X is assumed in addition to be wise enough so that she does not exploit the cooperation state of her opponent too much but is satisfied with a fair share of the total accumulated group payoff. This latter assumption might be a little bit too strong, but maybe it is not strictly necessary as player Y will punish X for defection behaviors.

For the iterated RPS game, it appears to be impossible for the proactive player X to design a CT strategy which brings higher expected payoff per game round to herself than to her opponent. However, this is not a general conclusion. For some other game systems, notably the iterated PD game [13,20], the proactive player X has the option of optimizing her CT strategy to extort her opponent Y. We do not recommend the adoption of such greedy strategies, as the opponent player Y will very likely be frustrated by the defection behaviors of player X and he may then choose not to cooperate even such a choice hurts also himself [21].

When strategic interactions occur in biological systems [9,10], the involved individual animals, insects, bacteria, cells, ... are of course far from being rational or sufficiently intelligent. However the collective decision-making of such agents at the population level, aided by the evolutionary mechanism of mutation and selection, may appear to be very rational. By trial and error, such
systems may develop certain CT-like strategies even without the need of intelligent designing. It would be very interesting to investigate empirically whether CT strategies are actually implemented in some biological systems, such as the formation of symbiosis relationship between two species.

Cooperation in a finite-population RPS game system with more than two players may be much more difficult to achieve than the case of two players. A recent theoretical investigation by one of the present authors [19] suggested that optimized conditional response strategies might offer higher accumulated payoffs to individual players than the NE mixed strategy does. But it is still an open question as to whether high degree of cooperation can also be enforced in a multiple-player iterated RPS game by a number of proactive players. The case of multiple players interacting through a ring topology might serve as the simplest model system to study. We leave such a challenging issue to future investigations.

The iterated two-player RPS game might also serve as a simple system to quantitatively measure the degree of rationality of single human subjects. For example, an experiment can be arranged as follows. A human subject Y plays repeatedly with a fixed opponent X which is actually a computer implementing an optimal CT strategy. But Y does not know that he is playing with a computer and assumes he is playing with another human subject. By analyzing the evolution trajectory of player Y’s action choices, we may quantitative measure the learning behavior of player Y and his tendency of making rational decisions. We are discussing with colleagues about the possibility of carrying out such an experimental study.

Acknowledgments

HJZ thanks Zhijian Wang and Bin Xu for a recent fruitful collaboration on the finite-population Rock-Paper-Scissors game, which inspired the present work greatly, and thanks Jiping Huang for comments on the manuscript.

Author Contributions

Conceived and designed the experiments: HJZ. Performed the experiments: ZB HJZ. Analyzed the data: ZB HJZ. Contributed reagents/materials/analysis tools: HJZ. Wrote the paper: HJZ.

References

1. Nash JF (1950) Equilibrium points in n-person games. Proc Natl Acad Sci USA 36: 48–49.
2. Osborne MJ, Rubinstein A (1994) A Course in Game Theory. New York: MIT Press.
3. Maynard Smith J, Price GR (1973) The logic of animal conflict. Nature 246: 15–18.
4. Weibull JM (1995) Evolutionary Game Theory. Cambridge, MA: MIT Press.
5. Fisher L (2000) Rock, Paper, Scissors: Game Theory in Everyday Life. New York: Basic Books.
6. Axelrod R (1984) The Evolution of Cooperation. New York: Basic Books.
7. Kollock P (1998) Social dilemmas: The anatomy of cooperation. Annu Rev Sociol 24: 183–214.
8. Nowak NA, Highfield R (2011) SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed. New York: Free Press.
9. Sinervo B, Lively C (1996) The rock-paper-scissors game and the evolution of alternative male strategies. Nature 380: 240–243.
10. Kerr B, Riley MA, Feldman MW, Bohannan BJM (2002) Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature 414: 171–174.
11. Reichenbach T, Mobilia M, Frye E (2007) Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. Nature 448: 1046–1049.
12. Jiang LL, Zhou T, Perc M, Wang BH (2011) Effects of competition on pattern formation in the rock-paper-scissors game. Phys Rev E 84: 021912.
13. Grofman B, Pool J (1977) How to make cooperation the optimizing strategy in a two-person game. J Math Sociol 5: 173–186.
14. Rapoport A, Chammah AM (1965) Prisoner’s Dilemma: A Study in Conflict and Cooperation. Ann Arbor, Michigan: University of Michigan Press.
15. Kraines D, Kraines V (1993) Learning to cooperate with pavlov: an adaptive strategy for the iterated prisoner’s dilemma with noise. Theory and Decision 35: 107–150.
16. Nowak M, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tie-for-tat in the prisoner’s dilemma game. Nature 364: 56–58.
17. Taylor PD, Jonker LB (1978) Evolutionarily stable strategies and game dynamics. Mathematical Biosciences 40: 145–156.
18. Sandholm WM (2010) Population Games and Evolutionary Dynamics. New York: MIT Press.
19. Wang Z, Xu B, Zhou HJ (2014) Social cycling and conditional responses in the rock-paper-scissors game. Sci Rep 4: 5830.
20. Press WH, Dyson FJ (2012) Iterated prisoner’s dilemma contains strategies that dominate any evolutionary opponent. Proc Natl Acad Sci USA 109: 10409–10413.
21. Hilbe C, Röhl T, Milinski M (2014) Extortion subdues human players but is finally punished in the prisoner’s dilemma. Nature Commun 5: 3976.