Multi-AI competing and winning against humans in iterated Rock-Paper-Scissors game

Lei Wang1, Wenbing Huang1,2, Yuanpeng Li1,2, Julian Evans1, Sailing He*1,2

1. National Engineering Research Center for Optical Instruments, Centre for Optical and Electromagnetic Research, Zhejiang University, Hangzhou 310058, China
2. Ningbo Research Institute, Zhejiang University, Ningbo 315100, China

Abstract: Predicting and modeling human behavior and finding trends within human decision-making process is a major social sciences’ problem. Rock Paper Scissors (RPS) is the fundamental strategic question in many game theory problems and real-world competitions. Finding the right approach to beat a particular human opponent is challenging. Here we use Markov Chains with set chain lengths as the single AIs (artificial intelligences) to compete against humans in iterated RPS game. This is the first time that an AI algorithm is applied in RPS human competition behavior studies. We developed an architecture of multi-AI with changeable parameters to adapt to different competition strategies. We introduce a parameter “focus length” (an integer of e.g. 5 or 10) to control the speed and sensitivity for our multi-AI to adapt to the opponent’s strategy change. We experimented with 52 different people, each playing 300 rounds continuously against one specific multi-AI model, and demonstrated that our strategy could win over more than 95% of human opponents.
Introduction

The Rock-Paper-Scissors (RPS) game has been widely used to study competitive phenomena in society and biology, such as ecological interactions, the maintenance of biodiversity in ecological systems and price dispersion of markets. There are two general approaches for RPS play, Nash equilibrium randomness and evolving circular exploitation. As a typical example, the payoff $a$, defined as the money incentive of the winning action divided by the money incentive of drawing action, is set to 2 to form a neutral RPS game.

Previous research has found that there is a social circle in human competitive strategy when playing iterated RPS games. In this article we proposed a multi-AI algorithm that can exploit human strategy and win over human in the same iterated RPS games and we conduct experiments with human players to confirm the results.

Our work may stimulate future more refined experimental and theoretical studies on the microscopic mechanisms of decision-making and learning in basic game systems.

Markov chain models are the single models which our multi-AI is composed of. The assumption of a Markov chain is that the current state only depends on a finite number of previous states. Here the iterated RPS game is considered as a Markov process and Markov chains are built throughout the process of 300-rounds of competition. The models are built to exploit attempted circular exploitation patterns.

The simplest discrete-time Markov chain is the first-order Markov chain, where the probability of moving to the next state depends only on the present state and not on the previous states:
\( \Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n), \)

where \( X_1, X_2, X_3 \) are a sequence of random variables, Rock Paper Scissors here.

What you will play in the next round only depends on what you played this round, like a short memory pattern sequence.

Markov chains can be generalized to cases of short-term dependency, by taking into account recent past states in the chain. The m-th order Markov chain\(^{20} \) considers the current state to depend on m previous states, where m is finite, and is a process satisfying

\[
\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_1 = x_1) = \Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \ldots, X_{n-m} = x_{n-m}) \quad \text{for} \ n > m
\]

Here the m-th order Markov chain is like a model with memory length m, which ‘remembers’ the previous m states.

For \( n < m \), all m-th order single Markov chain models will select Rock, Paper, Scissors randomly, with 1/3 probability each.

We conducted experiments on the combination of several selected single models according to their last 5 or 10 rounds best score, we calculate all the selected single model results and select the one with the highest score of last 5 or 10 rounds as the dominant AI output for the next round.

Here for simplicity we use AI-m to denote our single Markov chain model of order m.

Through the experiments we found that different models work best against different human opponents’ competition strategies and the prediction results vary greatly so we build the 1\(^{\text{st}}\) - 5\(^{\text{th}}\) order Markov chain models (i.e. AI-1 to AI-5) with different memory lengths for exploiting different human competition strategies. To make a multi-AI model that can differentiate and adapt to different human opponents, we combine the 1\(^{\text{st}}\)-5\(^{\text{th}}\) order single Markov models and introduce a “focus length” parameter to control
the adaptation speed and sensitivity to form a multi-AI that can adapt to different human strategies and win over most of its opponents. Table 1 illustrates how this multi-AI model competes against a specific player as an example with focus length F=5.

Table 1. An example of our multi-AI algorithm competes against a player when F = 5.

| Player played: | R | S | P | S | S | P | R | R | P | S | S |
|---------------|---|---|---|---|---|---|---|---|---|---|---|
| AI 1 played:  | P | P | S | P | S | S | R | R | R | S | R |
| AI 1 score    | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | 0 | 0 |
| AI 1 Last 5 Rounds Total Score: | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| AI 2 played:  | P | P | S | S | S | R | R | R | R | R | R |
| AI 2 score    | -1 | -1 | 1 | -1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| AI 2 Last 5 Rounds Total Score: | 0 | 1 | 0 | 1 | 1 | 0 | 1 | -1 | 1 | -1 | 1 |
| AI 3 played:  | P | P | P | P | S | P | P | P | R | R | R |
| AI 3 score    | -1 | -1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| AI 3 Last 5 Rounds Total Score: | 0 | -1 | -2 | -2 | -3 | -2 | 1 | 2 | 2 | 1 | 1 |
| AI 4 played:  | R | S | S | P | R | R | S | R | S | R | P |
| AI 4 score    | 0 | 0 | 1 | -1 | -1 | 1 | -1 | 1 | -1 | 1 | 0 |
| AI 4 Last 5 Rounds Total Score: | 0 | 0 | 0 | 1 | -1 | -1 | -4 | -5 | -4 | -5 | -3 |
| AI 5 played:  | P | P | S | S | S | S | R | R | R | R | R |
| AI 5 score    | 1 | -1 | 1 | -1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| AI 5 Last 5 Rounds Total Score: | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| Multi AI played: | P | P | S | P | S | S | R | R | R | R | R |

For the first round, the multi-AI will use the result from AI-1 and it rolls Rock, Paper or Scissors randomly with 1/3 probability each.

Focus length parameter F is set to control the speed and sensitivity for our multi-AI model to adapt to the opponent’s strategy change. Our multi-AI model will look at the recent F rounds of history to decide which single model is currently performing the best and should produce the next output. For the first 4 rounds when our competition data is less than focus length F, which is set to 5 as before, multi-AI will simply consider all rounds to determine its next round’s dominant single AI model. In the specific case of Table 1, the dominant AI for all of the first 4 rounds is AI-1. In round 5, AI-2 has the best cumulative score and thus is the dominant AI.
Table 1 shows how our multi-AI algorithm competes against a specific player when F=5 as an example. The transition matrix for the last AI-2 seeing the player played “PS” in the past 2 rounds is:

|     | R  | P  | S  |
|-----|----|----|----|
| PS  | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |

Thus for the next round, AI-2 has $\frac{1}{3}$ probability to roll Paper, Scissors and Rock.

Table 2 shows an example for the selection between AI-1 and AI-4 when focus length F=5. Although globally AI-1 has a higher score in total (from the first round AI-1 has a total score of 2, but its recent 5 rounds score is 0), AI-4 has a higher score (which is 3 in this case) locally during the recent 5 rounds. Thus the next round multi-AI will pick AI 4’s result as our multi-AI output.

**Table 2. Selection between AI-1 and AI-4 when focus length F = 5.**

|     | W | W | D | Next Round AI |
|-----|---|---|---|----------------|
| AI 1 |   |   |   | AI 4           |
| AI 4 | L | L | L |                 |

The transition matrix for AI-2 after 20 rounds of competition is in Table 3:

**Table 3: The transition matrix for AI-2 after 20 rounds of competition.**

|     | RR | RS | RP | PR | PP | PS | SR | SP | SS |
|-----|----|----|----|----|----|----|----|----|----|
| R   | 0  | 1/18 | 1/18 | 0  | 0  | 1/18 | 1/18 | 3/18 | 1/18 |
| P   | 0  | 0  | 0  | 1/18 | 0  | 2/18 | 0  | 1/18 | 0  |
| S   | 1/18 | 1/18 | 0  | 0  | 0  | 1/18 | 1/18 | 0  | 2/18 |
Results

All experiments are conducted with money incentive. We did the experiment with 52 human subjects recruited at Zhejiang University and used a multi-AI model which has 5 or 10 single length Markov chain models. Figure 1 shows 4 typical results of our multi-AI strategy with a combination of the 1\textsuperscript{st} to 5\textsuperscript{th} order Markov chain models (here focus length F is also set to 5, but it can be any other integer) competition against 4 typical players for 300 rounds.

The total score for the multi-AI against 52 people are shown in Figure 2. Against one player, multi-AI with a combination of 5 models had a score of 151, which is 198 wins,
55 draws and only 47 losses, more than 4 times of human total wins. Fig. 1.d is an example of a player who beat the AI by a margin of 4, and this is not a statistically meaningful margin as there are several lead changes.

![Figure 2: Total scores for multi-AI competing against different players in 300 rounds game.](image)

42 players played against the multi-AI with focus length F=5 and AI-1 to AI-5 (see blue bars in Fig. 2) and 11 players played against the multi-AI with focus length F=10 and AI-1 to AI-10 (see orange bars in Fig. 2). From the overall results in Table 5, we see that our multi-AI algorithm with F=10 give similar scores, but has lower standard deviation than that with F=5.

For simplicity, we let multi-5AI denote our multi-AI model with a combination of the 1st to 5th order Markov chain models (here focus length F is also set to 5, but it can be any other integer), and multi-10AI denote our multi-AI model with a combination of the 1st to 10th order Markov chain models (focus length F is also set to 10).

52 people’s preferences in choosing Rock Paper Scissors in their 300 rounds competition are shown in Table 4. There is a slight preference in choosing Rock, which
is consistent with some previous human vs human studies.

Table 4: 52 people’s preferences in choosing Rock Paper Scissors in 300 rounds competitions.

|       | R        | P        | S        |
|-------|----------|----------|----------|
| MEAN (times) | 106.6098 | 96.29268 | 97.09756 |
| STDEV.S   | 13.08095 | 13.12239 | 12.11851 |

We looked at the performance of individual models within a multi-AI 300 game set. The AI which performs the best against a particular individual varies greatly, but overall AI 2-6 perform better than AI 1 or higher-order Markov chains of longer memory length.

This general trend is consistent with human short memory holding around 7 items\(^\text{21}\).

Table 5: Game results (total scores) of our multi-10AI competing with human in 300 rounds.

|       | A11 | A12 | A13 | A14 | A15 | A16 | A17 | A18 | A19 | A20 | 10 single models average score | multi-10AI score | multi-5AI score |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------------------------------|----------------|----------------|
| Player1 | -13 | 37  | 24  | 20  | 22  | 14  | -6  | -2  | -10 | 3   | 8.9                          | 10             |                |
| Player2 | 2   | 50  | 48  | 42  | 42  | 50  | 61  | 7   | 12  | 5   | 31.9                         | 41             |                |
| Player3 | 46  | 58  | 52  | 53  | 24  | 29  | 11  | -15 | -14 | 12  | 25.6                         | 8              |                |
| Player4 | 26  | 41  | 26  | 55  | 46  | 28  | 20  | 13  | 10  | 36  | 30.1                         | 34             |                |
| Player5 | 35  | 39  | 92  | 107 | 85  | 67  | 88  | 61  | 78  | 70  | 72.2                         | 73             |                |
| Player6 | 28  | 60  | 52  | 55  | 49  | 36  | 14  | 1   | -18 | 12  | 28.9                         | 43             |                |
| Player7 | 56  | 63  | 68  | 65  | 55  | 46  | 54  | -1  | -14 | 17  | 40.9                         | 48             |                |
| Player8 | 28  | 29  | 47  | 29  | 12  | 5   | 21  | 5   | -10 | -5  | 16.1                         | 43             |                |
| Player9 | 32  | 63  | 58  | 72  | 62  | 72  | 22  | 26  | 20  | 15  | 45.2                         | 59             |                |
| Player10| -5  | -34 | -8  | 11  | 16  | -7  | 23  | 19  | 20  | 25  | 2.9                          | 24             |                |
| Player11| 40  | 15  | 33  | 16  | 9   | 24  | 25  | -9  | 6   | 37  | 19.6                         | 9              |                |
| MEAN   | 25.00| 40.09| 44.73| 47.73| 38.36| 33.09| 26.73| 9.55| 7.27| 20.45| 29.30| 35.64| 37.39 |
| STDEVA | 21.61| 23.57| 26.02| 28.36| 21.93| 24.52| 31.69| 20.76| 27.38| 20.86| 19.05| 21.21| 31.12 |

Fig. 3 and Fig. 4 show AI with different memory lengths’ performance against specific human players.
Figure 3: Game results (total scores) of Multi-10AI competing with human in 300 rounds.

Figure 4: Game results (total scores) of multi-AI with Markov chain lengths 1-5 competing with 8 typical players in 300 rounds.
It is hard to build one single model that can exploit every different human’s behavior, and thus we decided to combine the single models to make it able to differentiate and adapt to more different human competition strategies and win over most of its opponents.

Discussion and conclusions

In this paper, we have introduced a multi-AI model that wins over human in iterated RPS games and experimentally confirmed our results. We found that using a single-length Markov model could beat most human players, but not all human players. In single model experiments, we found the model with the best performance varies greatly for different people, which indicates that different people have different patterns. Although different humans have different patterns, and in total the patterns may be very hard to observe and exploit. Human competition behavior indeed has patterns and the patterns are exploitable by using proper simple models (single models that successfully predict this human’s behavior). We have obtained and exploited different human behaviors by building and combining single Markov chain models of different memory lengths and during the competition process it learns and switch to the best prediction model according to its focus length. We have introduced one possible architecture for human AI RPS games competition, and this model could be further improved by e.g., optimizing the voting weights of single Markov chain models, using the first part of the competition data to pre-train multi-AI model and switch to only two or three dominant single models after the pre-training process. Focus length is a hyper parameter and can
be tested by more human experiments for further optimization. After rearranging single models and adjusting “focus length” our model can potentially be improved further. The competition behavior patterns and their successful exploitation may lead our future work to better modeling, predicting and adapting to different specific human’s competition behavior patterns.

**Methods**

**Experiment**

Our methods for experiment mostly follows the RPS social experiments conducted in the period of December 2010 to March 2014 by Zhijian Wang et al.\textsuperscript{13}. The experiment was approved by the Experimental Social Science Laboratory of Zhejiang University and performed at Zhejiang University in the period of July 2019 to September 2019. The first author confirms that this experiment was performed in accordance with the approved social experiments guidelines and regulations. A total number of 52 undergraduate and graduate students of Zhejiang University volunteered to serve as the human subjects of this experiment. These students were openly recruited at the university library through onsite volunteer recruitment. More female students were registered than male students due to the humanity subjects of our experiment (more female students in the related humanity subjects in our university). Since we sampled students uniformly at random from the candidate list, more female students were recruited than male students. Informed consent was obtained from all the participating human subjects.

The 360 human subjects (referred to as players in this work) carried one experimental session by playing the RPS game for 300 rounds with fixed payoff parameter $a = 2$. During the game process the players were seated separately in a classroom, each 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 our computer program.

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. After a choice has been made it cannot be changed.

Before the start of the actual experimental session, the player were asked 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 also chooses “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 an AI opponent?

During the experimental session, the computer screen in front 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.

Data Availability

All 52 players experiment data with all process data of Markov chains we used are provided in the supplementary information.
References

1. O'Dwyer JP. Contests between species aid biodiversity. *Nature* **548**, 166-167 (2017).

2. Grilli J, Barabás G, Michalska-Smith MJ, Allesina S. Higher-order interactions stabilize dynamics in competitive network models. *Nature* **548**, 210-213 (2017).

3. Bergstrom CT, Kerr B. Taking the bad with the good. *Nature* **521**, 431-432 (2015).

4. Allesina S, Levine JM. A competitive network theory of species diversity. *Proceedings of the National Academy of Sciences* **108**, 5638 (2011).

5. Reichenbach T, Mobilia M, Frey E. Mobility promotes and jeopardizes biodiversity in rock-paper-scissors games. *Nature* **448**, 1046-1049 (2007).

6. Lee D, McGreevy BP, Barraclough DJ. Learning and decision making in monkeys during a rock-paper-scissors game. *Cognitive Brain Research* **25**, 416-430 (2005).

7. Semmann D, Krambeck H-J, Milinski M. Volunteering leads to rock-paper-scissors dynamics in a public goods game. *Nature* **425**, 390-393 (2003).

8. Kerr B, Riley MA, Feldman MW, Bohannan BJM. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**, 171-174 (2002).

9. Sinervo B, Lively CM. The rock-paper-scissors game and the
evolution of alternative male strategies. *Nature* **380**, 240-243 (1996).

10. Cason TN, Friedman D. Buyer search and price dispersion: a laboratory study. *Journal of Economic Theory* **112**, 232-260 (2003).

11. Maskin E, Tirole J. A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles. *Econometrica* **56**, 571-599 (1988).

12. Hoffman M, Suetens S, Nowak MA, Gneezy U. An experimental test of Nash equilibrium versus evolutionary stability. In: *Proc. Fourth World Congress of the Game Theory Society (Istanbul, Turkey), session* (2012).

13. Wang Z, Xu B, Zhou H-J. Social cycling and conditional responses in the Rock-Paper-Scissors game. *Scientific Reports* **4**, 5830 (2014).

14. Claussen JC, Traulsen A. Cyclic Dominance and Biodiversity in Well-Mixed Populations. *Physical Review Letters* **100**, 058104 (2008).

15. Castellano C, Fortunato S, Loreto V. Statistical physics of social dynamics. *Reviews of Modern Physics* **81**, 591-646 (2009).

16. Ao P, Chen T-Q, Shi J-H. Dynamical Decomposition of Markov Processes without Detailed Balance. *Chinese Physics Letters* **30**, 070201 (2013).

17. Ao P. Emerging of Stochastic Dynamical Equalities and Steady State Thermodynamics from Darwinian Dynamics. *Communications in Theoretical Physics* **49**, 1073-1090 (2008).

18. Zhou T, Han X-P, Wang B-H. Towards the understanding of
human dynamics. In: Science matters: humanities as complex systems). World Scientific (2008).

19. Grewal JK, Krzywinski M, Altman N. Markov models—Markov chains. Nature Methods 16, 663-664 (2019).

20. Skewes AD, Welch RD. A Markovian analysis of bacterial genome sequence constraints. PeerJ 1, e127 (2013).

21. Lisman JE, Idiart MA. Storage of 7 +/- 2 short-term memories in oscillatory subcycles. Science 267, 1512 (1995).

Acknowledgements

This work was partially supported by the National Key Research and Development Program of China (No. 2018YFC1407500), the National Natural Science Foundation of China (No. 11621101), and the Fundamental Research Funds for the Central Universities. We thank Shuo Li for excellent assistance in arranging the experiments.

Author contributions

S. H. conceived and supervised the work. L. W. developed the algorithm and built the multi-AI. L. W., Y. L. and W. H. performed the experiments and collected data. L.W. processed and analyzed the data. The manuscript was discussed and written by L. W., J. E. and S. H. with comments and input from all authors.

Additional Information

Supplementary Information is available for this paper.
Competing interests

The authors declare no competing interests.