Evolution with Reinforcement Learning in Negotiation

Yi Zou*, Wenjie Zhan, Yuan Shao

School of Management, Huazhong University of Science and Technology, Wuhan, P.R. China

Abstract

Adaptive behavior depends less on the details of the negotiation process and makes more robust predictions in the long term as compared to in the short term. However, the extant literature on population dynamics for behavior adjustment has only examined the current situation. To offset this limitation, we propose a synergy of evolutionary algorithm and reinforcement learning to investigate long-term collective performance and strategy evolution. The model adopts reinforcement learning with a tradeoff between historical and current information to make decisions when the strategies of agents evolve through repeated interactions. The results demonstrate that the strategies in populations converge to stable states, and the agents gradually form steady negotiation habits. Agents that adopt reinforcement learning perform better in payoff, fairness, and stableness than their counterparts using classic evolutionary algorithm.

Citation: Zou Y, Zhan W, Shao Y (2014) Evolution with Reinforcement Learning in Negotiation. PLoS ONE 9(7): e102840. doi:10.1371/journal.pone.0102840

Editor: Matjaz Perc, University of Maribor, Slovenia

Received April 4, 2014; **Accepted** June 23, 2014; **Published** July 21, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The research was supported in part by the National Natural Science Foundation of China (Grant No. 70871045) and the Key Program of National Natural Science Foundation of China (Grant No. 71232001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: mail_zou@126.com

Introduction

Uncertainty in negotiation can be caused by a number of factors, such as fuzzy opponent type, unknown strategies, and deadlines [1]. Hence, one of the central topics in negotiation is how to design agents with higher adaptability for changing circumstances [2,3]. In the environment of incomplete information, the primary concern is how to acquire more information and use it appropriately to reach consensus through concession in negotiation [4–6]. Through feedback, the agents learn how to make future decisions. Because the learning effect in the short term depends strongly on the details of the negotiation process and the characteristics of the opponent, it is difficult to evaluate a learning model through sparse interactions. In addition, because heterogeneous agents behave differently, it is reasonable to assess the performance from a macroscopic view [7,8].

To address the uncertainty, learning in the long term emphasizes strategy selection and adjustment through repeated negotiations using qualitative or quantitative methods. For instance, Eduard Gim'enez-Funes et al. have applied a qualitative approach, case-based reasoning, to find the appropriate strategy by comparing the similarity of current situations to history [9]. Matos et al. have applied a widely used quantitative model, an evolutionary algorithm, to investigate long-term behavior [10]. This approach is derived from biology, with simple rules to evaluate payoff [11]. A number of researchers have investigated strategy evolution in the long term in recent years [12–17]. In the genetic algorithm, agents with higher fitness are passively selected and put into the mating pool to replicate the next generation, without considering the learning behavior of the agents. This approach myopically assesses the performance of agents with payoff only in the current period, while humans learn by weighing both the historical information and the current performance. If the agent represents a person in reality, it is reasonable to incorporate individual learning because humans adjust strategy through experience with initiative.

Fudenberg and Levine have investigated long-term strategy dynamics, including replicator dynamics and reinforcement learning [18]. Because reinforcement learning is a type of individual learning while the evolutionary approach concerns population dynamics, prior literature has generally examined them separately. However, Börgers, T. and R. Sarin [19] find that a type of continuous time reinforcement learning can converge to an equilibrium of replicator dynamics, which indicates some interaction between population dynamics and individual reinforcement learning. Reinforcement learning, with its basis in psychology [20,21], evaluates the reward by weighting historical and current payoffs and has been applied to human strategy adjustment as well as to artificial intelligence [22–24]. Recently, researchers have begun to integrate different learning approaches to determine an agent's optimal strategy in the case of incomplete information. Reinforcement learning is a good fit when information on the opponent and environment is limited. To this end, agents in our model adopt reinforcement learning to calculate the reward of each strategy and then use replicator dynamics to adjust the probability of strategies. We integrate replicator dynamics and reinforcement learning to explore the efficiency, fairness, and strategy convergence in negotiation. In addition to the efficiency and strategy evolution, fairness has also been a concern of many researchers [25,26]. The simulation results indicate that our approach achieves higher reward, shorter negotiation time, and a lower degree of greediness of strategies than the classic evolution model. It is also shown that the weight tradeoff between current and historical experience impacts the negotiation performance and learning effect to a large extent.

Methods

This model is based on the alternate offering mechanism of Rubinstein, in which agents adopt the time-dependent concession function [10]. The agents use reinforcement learning to accumulate experience and update the probability of each strategy using replicator dynamics. The negotiation process is illustrated in Figure 1.

Negotiation rules

The reservation prices R_s and R_b of the sellers and buyers are uniformly distributed between $P_{\min}$ and $P_{\max}$. In each period, seller-buyer pairs are randomly selected and begin to negotiate if $R_s \leq R_b$. The offering intervals for a buyer and a seller are $(P_{\min}, R_b)$ and $(R_s, P_{\max})$, respectively. They alternate in making offers, and the discount rates of sellers and buyers are C_s and C_b, respectively (the first offer is proposed by the buyer by default). If an agent rejects the offer of the opponent, he then proposes his own offer (P_s for the seller and P_b for the buyer), and the accepted price is P. The payoff of the seller is $U_s = (P - R_s) * C_s^t$, and the payoff of the buyer is $U_b = (R_b - P) * C_b^t$, where t is the negotiation time.

Concession functions

The time-dependent strategies for buyers and sellers are $P_b(t)$ and $P_s(t)$, respectively, and the concession rate increases with time. $P_b(t)$ and $P_s(t)$are represented in equation (1) as follows:

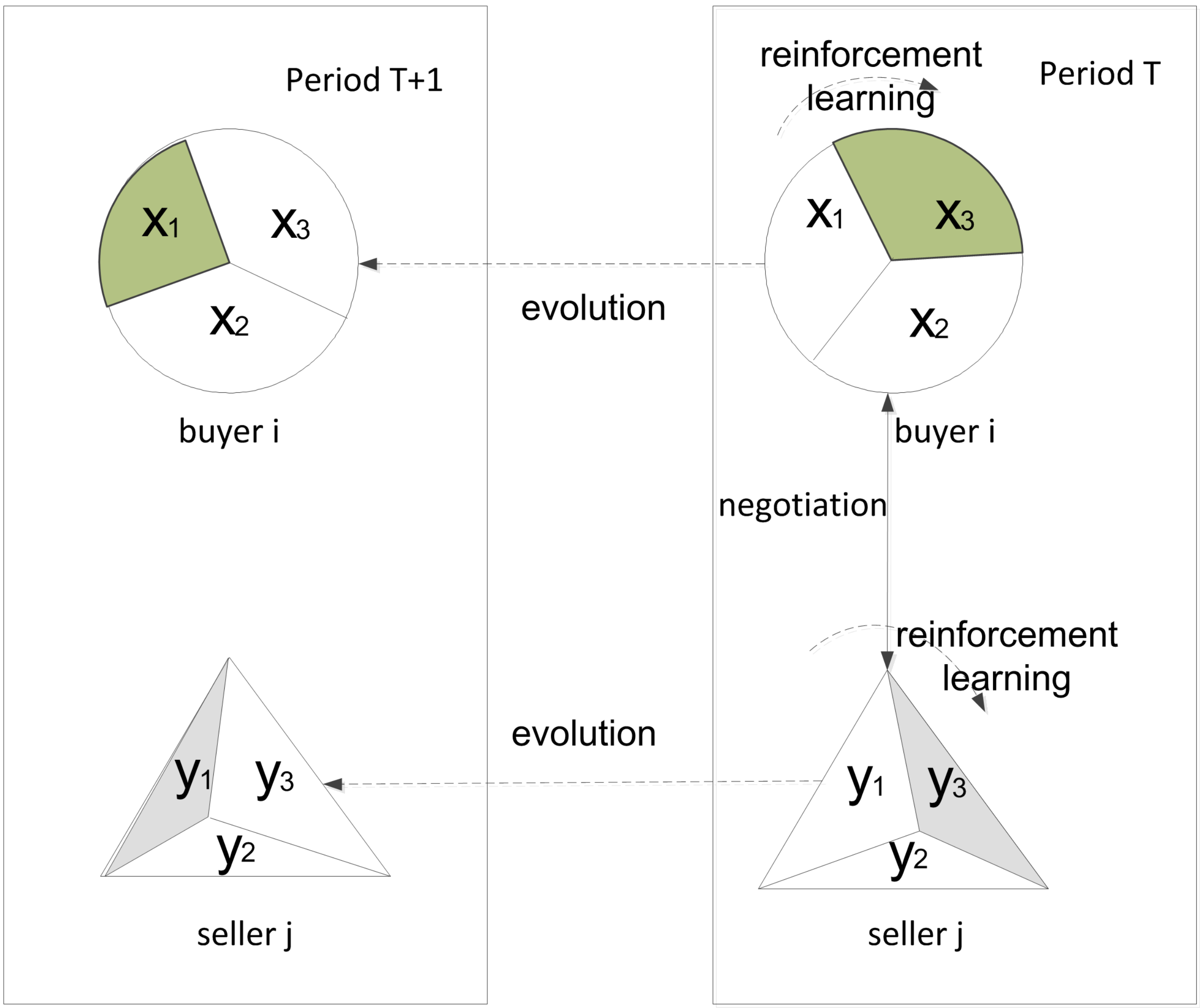

Figure 1. The negotiation model. The circles denote buyers; triangles denote sellers, and ellipses denote environmental factors. The values x^1,x^2,x^3 are probabilities of the frugal, cool-headed and anxious strategies of buyers, respectively, and y^1,y^2,$y^3$are the corresponding strategies for the sellers. The shaded area represents the selected strategy in the current period. At the end of each negotiation period, the buyers (sellers) calculate the reward through reinforcement learning and update x^1,x^2,x^3(y^1,y^2,y^3) accordingly. The process continues until the strategies of all agents in the market converge to a stable state.
doi:10.1371/journal.pone.0102840.g001

$$\begin{cases} P_b(t) = P_{\min} + \dfrac{g * k_b^i * (R_b - P_{\min}) * (t+\Delta)^{\lambda}}{P_{\max} - P_{\min}}, \\ \qquad t \le \left(\dfrac{p_{\max} - p_{\min}}{g * k_b^i}\right)^{\frac{1}{\lambda}} - \Delta \\ P_s(t) = P_{\max} - \dfrac{g * k_s^j * (P_{\max} - R_s) * (t+\Delta)^{\lambda}}{P_{\max} - P_{\min}}, \\ \qquad t \le \left(\dfrac{p_{\max} - p_{\min}}{g * k_s^j}\right)^{\frac{1}{\lambda}} - \Delta, \end{cases} \quad (1)$$

where g denotes the time pressure of negotiators, and $k_b^i, k_s^j (0<k_b^i, k_s^j<1)$ denote the greediness of buyers and sellers, with smaller k_b^i and k_s^j values indicating higher levels of greediness. Three strategies with different degrees of greediness exist for each population: $0<k_b^1<k_b^2<k_b^3<1$, $0<k_s^1<k_s^2<k_s^3<1$. Here, k_b^1 represents the frugal strategy; k_b^2 represents the cool-headed strategy; and k_b^3 represents the anxious strategy for the buyers as well as for the sellers. Δ refers to the initial offer parameter, which creates an offer closer to the reservation price, with a larger Δ; λ refers to the concession type, which is defined as a convex function for buyers and a concave function for sellers when $\lambda > 1$. The buyer and the seller propose offers alternately until an offer is accepted, which occurs when the agent receives a higher payoff than refusing it by proposing his own offer, which is expected to be accepted by the opponent in the next round. The buyer accepts an offer P_t^s when $R^b - P_t^s > C_b(R^b - P_{t+1}^b)$, and a seller accepts P_t^b when $P_t^b - R^s > C_s(P_{t+1}^s - R^s)$.

Learning rules

The agents update their rewards according to feedback based on historical information and eventually develop a negotiation habit [27]. Due to the constantly changing environment, the reference value of information decreases with time, and thus, the agents assign different weights to historical and current payoffs. The reward function is defined as follows:

$$u_t^i = w * u_{t-1}^i + (1-w) * v_t^i, \quad (2)$$

where u_t^i is the average reward of strategyk^iin period t, w is the weight of the historical payoff, and v_t^i is the average current payoff strategy k^i. At first, each agent has a subjective probability for every strategy and thereby chooses one strategy to negotiate. In period t, the agent chooses strategy k^1 with probability x_t^1, k^2 withx_t^2, k^3 with x_t^3, and $x_t^1 + x_t^2 + x_t^3 = 1$. The agent updates the probabilities according to the rewards of each strategy, where a higher reward leads to an increased probability in the next period and vice versa. Until the total negotiation frequency reaches N, the agents adjust their strategy, which is defined as a learning period. The adjustment refers to replicator dynamics, but it is slightly modified to fit the reality:

$$\begin{cases} f_{t+1}^i = \dfrac{u_t^i}{\overline{u_t}} * x_t, \text{ if } |\, x(t+1) - x(t)| > s \\ f_{t+1}^i = x_t + s, \text{ if } \ 0 < x(t+1) - x(t) < s \\ f_{t+1}^i = x_t - s, \text{ if } \ \text{-}s < x(t+1) - x(t) < 0, \end{cases} \quad (3)$$

$$x_{t+1}^i = \frac{f_{t+1}^i}{\sum_{k=1}^{3} f_{t+1}^k}, \quad (4)$$

where $\overline{u_t} = \sum_{i=1}^{3} x_t^i * u_t^i$ is the average payoff, and u_t^i is the payoff of strategy k^i in period t. In Eq.(3), f_{t+1}^i is the temporary probability of strategy i, and $\sum_{i=1}^{3} f_{t+1}^i$does not necessarily equal 1. In Eq.(4), x_{t+1}^i is the normalized result, which ensures $\sum_{i=1}^{3} x_{t+1}^i = 1$. In period t+1, the probability of strategy k^i is proportional to $\frac{u_t^i}{\overline{u_t}}$, which means that the probability increases in the next period when u_t^i is more than $\overline{u_t}$ and decreases in the opposite case. We define the adjustment precision as s with the default value of 0.01. When the change in probability is smaller than s ($|x(t+1) - x(t)| < s$), the adjustment size is s. When $x_t^i < 0.01$, we set $x_t^i = 0$, which means that this strategy has vanished. When $x_t^i > 0.99$, we set $x_t^i = 1$, which means that the agent has converged to this strategy. In addition, US denotes the payoff of the seller without learning, and US^* denotes the payoff with learning. Similarly, UB and UB^* denote the corresponding meanings for the buyer. Fairness without learning is evaluated by $\frac{UB}{US}$, and fairness with learning is evaluated by $\frac{UB^*}{US^*}$.

Experiments

The first experiment investigates the general performance of the agents when $g = 6$, $k^1 = 0.3, k^2 = 0.5, k^3 = 0.7$, and $w = 0.9$. The buyer proposes the initial offer, and the seller continues to negotiate, with the discount rate C_s and C_b changing within the range of 0.5–1 and a minimum adjustment size of 0.025. The parameters in the control group without learning are the same as in the above-mentioned group except that agents in the control group do not adjust the probabilities of strategies.

The second experiment explores the impact of the weights of historical information on the negotiation result and strategy convergence. The adjustment range of w is 0.2–0.9, and the size is 0.1. Other parameters are the same as in the first experiment. We observe the negotiation result and calculate the related variables using the formulas in Table S1.

Results

Performance of the negotiation agents

The positive growth rate of the payoff in most cases means that the learning approach is beneficial to both buyers and sellers. As the discount rate decreases (see Figure 2A), the growth rate gradually increases. The reason may be that a lower discount rate encourages the intention of accepting an offer. The negotiation time is stable and only varies slightly with the change in discount rate, so it is not illustrated in Figure 2. Figure 2B illustrates the fairness with and without learning, where the noticeable difference between the two indicates that learning has changed the division of payoff between buyers and sellers. The distinction of the growth rate of payoffs in Figure 2A between the two populations is also the major reason for the variation of fairness with learning. Initialized with random subjective strategy probabilities, most agents finally

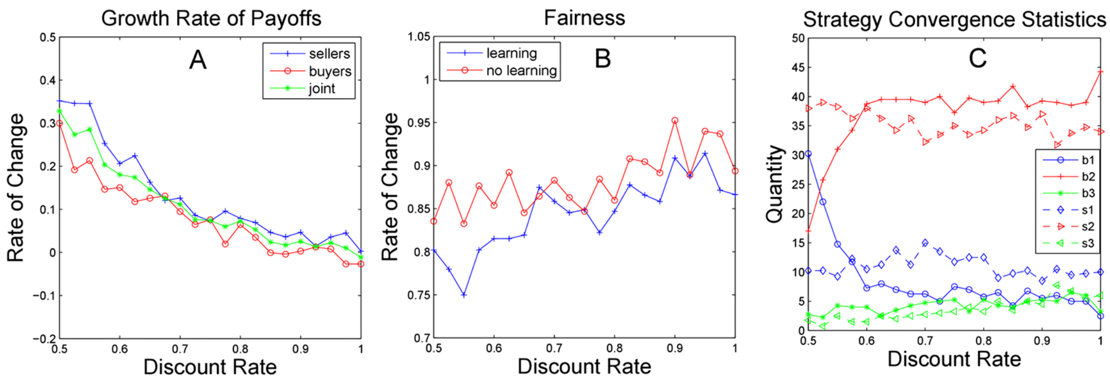

Figure 2. Negotiation performance of the agents. Figure 2A illustrates the growth rate of the average payoff with learning for buyers, sellers, and joint payoff compared to the growth rate without learning. Figure 2B illustrates the fairness comparison between the group with learning and the group without learning. Figure 2C shows the number of agents of each strategy in the convergence results, as every agent converges ultimately to a pure strategy.
doi:10.1371/journal.pone.0102840.g002

converge to a cool-headed strategy. Only a small portion of agents converge to an anxious strategy or frugal strategy after periods of evolution. While the convergence results of the two populations are somewhat different, the patterns of strategy distribution are similar. The result shows that learning has changed the habit of strategy adoption, and most agents become sensible by using a cool-headed strategy despite lacking emotion.

Impact of weight of historical information

In Figure 3A, the joint payoff of buyers and sellers changes only slightly when the weight increases from 0.5 to 0.9. As the weight drops below 0.5, the payoff declines sharply with the decreasing discount rate and reaches a minimum at approximately 0.8. It is obvious that the agents who emphasize historical experience perform better than the agents who ignore it. In Figure 3B, the negotiation time remains stable when the weight is above 0.5 and rises obviously until the weight drops below 0.5. The discount rate in our model indicates that more rounds of negotiation lead to lower payoff and inefficiency of the market. In Figure 3C, there is only a minor fluctuation of fairness when the weight rises above 0.5, and fluctuation becomes noticeable when the weight drops below 0.5. It can be concluded that the profit division of the market becomes unfair if the agents rarely consider prior information.

To summarize, agents who value long-term experience achieve more stable and efficient performance such as joint payoff, negotiation time, and fairness.

Figures 4A, 4B and 4C illustrate the impacts of the weight of historical information on buyers. The figures demonstrate that the convergence strategies remain stable when the weight is above 0.4 but vary significantly when the weight drops below 0.4. As the weight is high, the majority of the market adopts the cool-headed strategy, with the anxious and the frugal strategies as minorities. When the weight decreases, the advantage of the cool-headed strategy declines, while the frugal strategy increases and the anxious one remains stable. The results only change slightly with decreasing discount rate when the weight is high. In contrast, the results fluctuate substantially as the weight reaches a low level.

Figures 4D, 4E and 4F illustrate the impacts of the weight of historical information on the convergence of the frugal, cool-headed, and anxious strategies of the sellers, respectively. The general trend shows that the greediness of the agents rises notably when the weight decreases. Specifically, we find that the frugal strategy increases, while the cool-headed strategy simultaneously decreases. The frugal strategy is the minority when the weight is high and rises gradually when the weight drops, and it proves to be dominant at a high discount rate when the weight drops to approximately 0.2(see Figure 4D). The cool-headed strategy shows

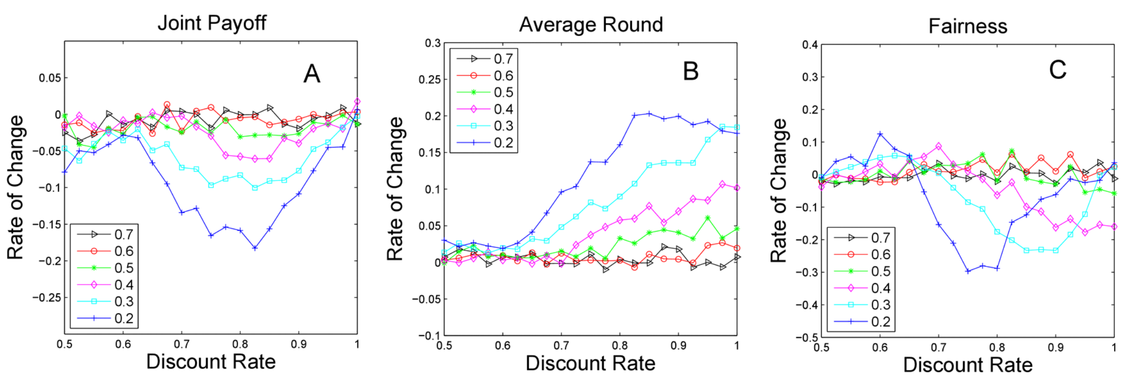

Figure 3. Impact of historical weight on payoff, time and fairness. Figures 3A, 3B and 3C illustrate the change in joint payoff, negotiation time, and fairness with different weights, compared with a weight of 0.8 as the baseline. Because the performance of agents with weights of 0.8 and 0.9 are very similar, the results for a weight of 0.9 are not displayed in Figure 3. To concisely demonstrate the result, we select the representative data with weights between 0.2 to 0.7 based on the benchmark of 0.8.
doi:10.1371/journal.pone.0102840.g003

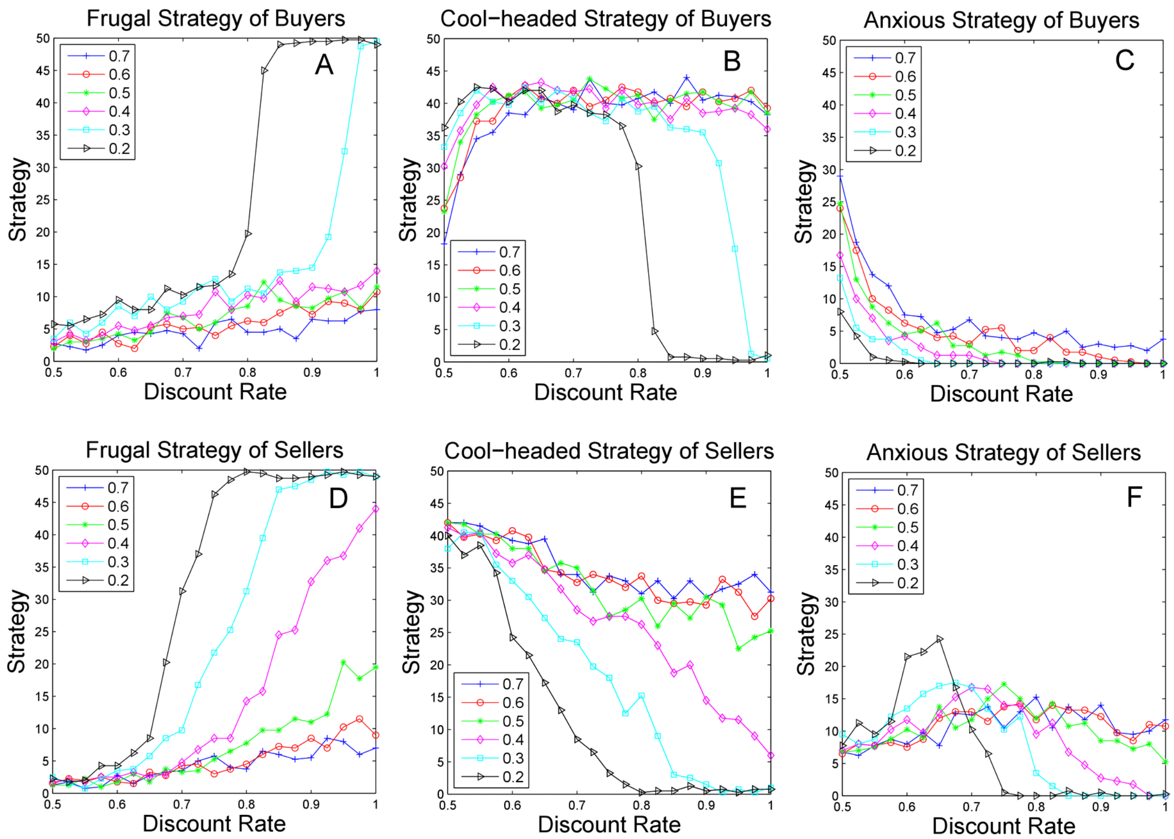

Figure 4. Impact of historical weight on the strategy convergence. Figures 4A, 4B and 4C illustrate the evolution results of the frugal, cool-headed, and anxious strategies of buyers using different weights. The Y axis represents the number of agents using each strategy in the population. Similarly, Figures 4D, 4E and 4F represent the corresponding results for sellers.
doi:10.1371/journal.pone.0102840.g004

the opposite trend to the frugal one and declines as the weight decreases (see Figure 4E). As the discount rate decreases, the frugal strategy decreases, and the cool-headed strategy increases. The cool-headed strategy becomes dominant when the discount rate is very low, while the frugal strategy is a tiny minority. This result indicates that the low discount rate's leading to less greediness in agents may be caused by the time pressure to accept offers more actively. The anxious strategy remains at a low level when the weight is high and fluctuates substantially when the weight is below 0.5 (see Figure 4F). In conclusion, the strategy convergence remains stable and the cool-headed strategy has a noticeable advantage when the weight of historical information is high. The greediness of the whole market rises as the weight drops.

To summarize, the impacts of weight on convergence strategies are different between buyers and sellers. The strategies of sellers vary gradually with the weight, while the strategies of the buyers remain relatively stable when the weight is above 0.5 and change rapidly when the weight falls below 0.5. Because the only distinction between the two populations is the order of proposing the first offer, the asymmetry may arise from this factor. Therefore, the offering order not only affects the division of the payoff between buyers and sellers but also gives rise to the difference in the strategy distribution between them. The agents update their strategy probability after a period of negotiation during which the costs of agents remain stable, and therefore the frugal strategy achieves a higher payoff. The agents use a more greedy strategy when the weight is lower. Our results show the following: (1) myopic adjustment leads to a more frugal strategy with less concession, and (2) the algorithm with reinforcement learning results in a more cool-headed strategy with more efficient performance in the overall population, such as less negotiation time and less fluctuation in fairness.

Discussion

The evolutionary approach is effective in investigating the collective behavior of the population in the long term. However, human learning involves more initiative than biological evolution, and therefore, we have integrated reinforcement learning with replicator dynamics to investigate negotiation behavior. Negotiation strategies are generally complex, and a new strategy type is created by design instead of mutation. In the genetic algorithm, the population of each generation is created by passive selection and reproduction [28–30], but agents representing humans usually will not depart the market even if they suffer from occasional loss in negotiation. Rather, they make decisions using initiative and accumulate experience through multiple periods of negotiation. To this end, this model evaluates the rewards of strategies by assigning weights to historical payoffs as well as to current ones. As a result, our learning pattern incorporating replicator dynamics differs from classic reinforcement learning, which determines the probability of strategies in proportion to the rewards [31]. Reinforcement learning has many models that differ from each other in details such as the probability determination rules. Rajiv Sarin and Farshid Vahid design a simple reinforcement learning model without probability, in which the agents choose the strategy with the highest reward instead of through subjective probability [32]. Borgers believes that individual learning is a process of idea evolution as well as habit formation and proves that a continuous-time reinforcement learning of an individual converges to equilibrium of replicator dynamics [19]. We adopt replicator dynamics as the strategy adjustment rule and incorporate reinforcement learning into the payoff evaluation. The simulation results suggest that agents using this new learning model achieve higher payoff, shorter negotiation time, and more stable fairness than agents using the classic evolutionary approach.

This paper presents two experiments designed to study the general performance of agents and the impacts of the weight of historical information on the negotiation result and convergence strategy. The results indicate that in most cases, learning increases the payoff of both buyers and sellers. Thus, the learning pattern is beneficial to both sides, and the growth rate rises with decreasing discount rate. Ravindra Krovi et al. [33] compare the payoff and fairness when one or two variables are controlled. However, they examine the evolution of offer instead of strategy, which is different from this paper. We evaluate the learning effect from the perspective of the population instead of the individual agent.

In addition to the market efficiency and fairness, we also consider long-term strategy evolution. In our model, all the agents converge to pure strategy and form stable habits. Although heterogeneous agents have different reservation values and initial states regarding strategies, the strategy distribution of the whole market is relatively stable, and the convergent results vary slightly with the initial settings. Noyda Matos et al. [10] allow the agents to use mixed strategies, but the proportion of each strategy is similar, and there is no dominant strategy. In our model, the majority is the cool-headed strategy, which means that the agents become rational by learning in the long run, although we do not consider psychological factors.

The classic evolutionary dynamics without considering historical information, including replicator dynamics and best response dynamics, are myopic adjustment dynamics [34–36]. The strategy adjustment in our model is not myopic because we have incorporated historical information into the evaluation. The simulation results indicate that the learning pattern focusing more heavily on historical information achieves better performance than its myopic counterparts. If agents ignore previous experience, their payoff declines, and simultaneously, negotiation time increases and fairness fluctuates substantially. The weight of prior experience also affects the strategy equilibrium, which includes mostly the cool-headed strategy and minor others.

This model still has a few limitations, such as the simple time-dependent concession function. We will consider more complex strategies such as behavior-dependent and resource-dependent concession functions in the future.

Supporting Information

Table S1 Variables and formulas. Average payoff of buyer, average payoff of seller, joint payoff, average round and fairness are calculated with the corresponding formulas.
(DOCX)

Acknowledgments

We thank Wen Ye for contributing insightful discussion.

Author Contributions

Conceived and designed the experiments: YZ. Performed the experiments: YZ. Analyzed the data: YZ WZ. Contributed reagents/materials/analysis tools: YZ YS WZ. Contributed to the writing of the manuscript: YZ YS.

References

1. Li C, Giampapa J, Sycara K (2006) Bilateral negotiation decisions with uncertain dynamic outside options. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 36: 31–44.
2. Ren F, Zhang M, Sim KM (2009) Adaptive conceding strategies for automated trading agents in dynamic, open markets. Decision Support Systems 46: 704–716.
3. Sim KM, Wang SY (2004) Flexible negotiation agent with relaxed decision rules. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34: 1602–1608.
4. Sim KM, Guo Y, Shi B (2009) BLGAN: Bayesian learning and genetic algorithm for supporting negotiation with incomplete information. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39: 198–211.
5. Williams CR, Robu V, Gerding EH, Jennings NR (2011) Using gaussian processes to optimise concession in complex negotiations against unknown opponents. Proc. of the 21st International Joint Conference on Artificial Intelligence. Barcelona, Spain, 22(1): 432–438.
6. Sycara K, Zeng D (1997) Benefits of learning in negotiation. Proceedings of the AAAI National Conference on Artificial Intelligence. Menlo Park, California, pp. 36–41
7. Couzin ID, Krause J, Franks NR, Levin SA (2005) Effective leadership and decision-making in animal groups on the move. Nature 433: 513–516.
8. Han J, Wang L (2013) Nondestructive intervention to multi-agent systems through an intelligent agent. PloS One 8: e61542.
9. Gimenez-Funes E, Godo L, Rodriquez-Aguilar JA, Garcia-Calves P (1998) Designing bidding strategies for trading agents in electronic auctions. Proceedings of the 3rd International Conference on Multi Agent System. Paris, pp. 136–143.
10. Matos N, Sierra C, Jennings NR (1998) Determining successful negotiation strategies: An evolutionary approach. Proceedings of the 3rd International Conference on Multi Agent System. Paris, pp. 182–189.
11. Smith JM (1982) Evolution and the Theory of Games. Cambridge, UK: Cambridge University Press.
12. Wang Z, Szolnoki A, Perc M (2014) Self-organization towards optimally interdependent networks by means of coevolution. New Journal of Physics 16: 033041.
13. Perc M, Wang Z (2010) Heterogeneous aspirations promote cooperation in the prisoner's dilemma game. PLoS One 5: e15117.
14. Wang Z, Wang L, Yin Z, Xia C (2012) Inferring reputation promotes the evolution of cooperation in spatial social dilemma games. PLoS One 7: e40218.
15. Xia C, Wang Z, Sanz J, Meloni S, Moreno Y (2013) Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks. Physica A: Statistical Mechanics and its Applications 392: 1577–1585.
16. Wang Z, Perc M (2010) Aspiring to the fittest and promotion of cooperation in the prisoner's dilemma game. Physical Review E 82: 021115.
17. Wang Z, Szolnoki A, Perc M (2012) If players are sparse social dilemmas are too: Importance of percolation for evolution of cooperation. Scientific Reports 2:369.
18. Fudenberg D (1998) The theory of learning in games. Cambridge, MA: MIT Press.
19. Börgers T, Sarin R (1997) Learning through reinforcement and replicator dynamics. Journal of Economic Theory 77: 1–14.
20. Sutton RS, Barto AG (1998) Reinforcement learning: An introduction. Cambridge, UK: Cambridge University Press.
21. Holroyd CB, Coles MG (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychological Review 109(4): 679
22. Erev I, Roth AE (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review 88: 848–881.
23. Mahadevan S, Connell J (1992) Automatic programming of behavior-based robots using reinforcement learning. Artificial Intelligence 55: 311–365.
24. Brede M (2013) Costly advertising and the evolution of cooperation. PloS One 8: e67056.
25. Miyaji K, Wang Z, Tanimoto J, Hagishima A, Kokubo S (2013) The evolution of fairness in the coevolutionary ultimatum games. Chaos, Solitons & Fractals 56: 13–18.
26. Szolnoki A, Perc M, Szabó G (2012) Defense mechanisms of empathetic players in the spatial ultimatum game. Physical Review Letters 109: 078701.
27. Young HP (1993) An evolutionary model of bargaining. Journal of Economic Theory 59: 145–168.
28. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Boston, MA: Addison-wesley Publishing Co., Inc.

29. liver JR (1997) A machine-learning approach to automated negotiation and prospects for electronic commerce. Journal of Management Information Systems 13: 83–112.
30. Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. Machine Learning 3: 95–99.
31. Camerer C, Hua Ho T (1999) Experience-weighted Attraction Learning in Normal Form Games. Econometrica 67: 827–874.
32. Sarin R, Vahid F (1999) Payoff assessments without probabilities: A simple dynamic model of choice. Games and Economic Behavior 28: 294–309.
33. Krovi R, Graesser AC, Pracht WE (1999) Agent behaviors in virtual negotiation environments. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 29: 15–25.
34. Lau RY(2005) Towards genetically optimised multi-agent multi-issue negotiations. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Piscataway, NJ, pp. 35c.
35. Lau RY, Tang M, Wong O, Milliner SW, Chen YPP (2006) An evolutionary learning approach for adaptive negotiation agents. International Journal of Intelligent Systems 21: 41–72.
36. Weibull JW (1997) Evolutionary game theory. Cambridge, MA: MIT Press.