Robustness Analysis of Negotiation Strategies through Multiagent Learning in Repeated Negotiation Games

Jianye Hao[1], Siqi Chen[2], Weiss Gerhard[2], Ho-fung Leung[3], and Karl Tuyls[4]

[1] Massachusetts Institute of Technology
jianye@mit.edu
[2] Maastricht University
{siqi.chen, gerhard.weiss} @maastrichtuniversity.nl
[3] The Chinese University of Hong Kong
lhf@cuhk.edu.hk
[4] University of Liverpool
k.tuyls@liverpool.ac.uk

Abstract. Automated negotiation techniques play an important role in facilitating human in reaching better negotiation outcomes, and until now lots of research efforts have been devoted to designing effective negotiation strategies. To evaluate the performance of different strategies, one important evaluation criterion is *robustness*, which is to investigate which negotiating strategies the agents are going to adopt finally if they are given the opportunity to repeatedly negotiate and allowed to change their choices. However the current way of evaluating the robustness suffers from several drawbacks. First, it is assumed that all agents can have access to the global payoff information, which may not be available beforehand in practice. Second, it is based on the single-agent best deviation principle, however, in practice, each agent may change their strategies simultaneously and in any possible rational way. To this end, we firstly propose the *repeated negotiation game* learning framework to evaluate the robustness of different negotiation strategies, in which each agent can adopt any rational learning approach to make decisions without knowing the global payoff information beforehand. In this way, we are able to provide more realistic and fine-grained robustness analysis and more insights in terms of the relative robustness of different negotiating strategies can be revealed from our analytical results.

1 Introduction

Automated negotiation techniques can, to a large extent, alleviate the efforts of human, and also facilitate human in reaching better negotiation outcomes in complex negotiations. To this end, until now lots of state-of-the-art negotiation strategies [8, 6, 15, 7, 5, 12] have been proposed to maximize agents' individual benefits from negotiation by exploiting their opponents as much as possible. In recent years, the international competition - *automated negotiating agents competition (ANAC)* [2, 1] held by researchers from automated negotiation area has

emerged accordingly. This competition provides a general negotiation platform and benchmarks, which enables different negotiation strategies to be evaluated within realistic negotiation environments.

In the current setting of ANAC, the performance of different negotiation strategies are evaluated based on the criterion of *efficiency*, i.e., each strategy's average payoff obtained against the rest of participants over different domains. Efficiency is indeed an important evaluation criterion to consider. However it only reflects the static aspect of negotiation, which assumes that each participant's strategy is fixed beforehand. In real life, it is common to encounter repeated negotiations between multiple parties in many scenarios such as e-commerce negotiation between different sellers and buyers [10]. This thus gives the agents (or people) the opportunity to choose different negotiation strategies against different opponents at different negotiation stages based on the past negotiation performance. Therefore in practice the current efficiency criterion may not be quite useful, since the most efficient strategy in one static negotiation setting may become the most inefficient one in another setting. The efficiency of a negotiation strategy makes sense only when the current negotiation setting is stable, however, it is unclear which strategy will be eventually adopted by each agent and which strategy profile will be the stable one eventually. To this end, an alternative evaluation criterion, *robustness*, was firstly proposed by Baarslag et al. [1] to evaluate the performance of different negotiation strategies from a new perspective based on empirical game-theoretic analysis. In general, the robustness analysis focuses on investigating whether an agent would have the incentive to switch to other negotiation strategies, and which strategy (combination of strategies) would be finally adopted by agents if strategy switching is allowed.

The current way [1] of analyzing the robustness of negotiation strategies suffers from several drawbacks. First, the analysis requires the global information (i.e., average payoff between all pairs of negotiating strategies) to be available beforehand. However, from an individual negotiator's perspective, this kind of information is usually not available beforehand until they have actually participated in the negotiations. Second, the robustness analysis is based on the principle of the *single-agent best deviation*, i.e., only one agent is allowed to change its negotiating strategy in each round. However, in practice, it is highly likely that in each round each agent may change its negotiating strategy simultaneously and in any possible rational way which may not necessarily follow the principle of best deviation. We believe that enabling all agents to choose their negotiating strategies simultaneously and autonomously without knowing the global information can provide more realistic analysis and predictions of the dynamic changes of the agents' negotiating strategies and their relative robustness. It is not clear a priori if the agents are able to converge to a stable strategy profile or which strategy will be mostly likely to be adopted by each agent in such a situation. Third, based on the current robustness analysis approach, it reveals nothing about the relative robustness of each negotiating strategy within the best reply cycle if such a cycle exists.

To tackle the above issues, we first introduce the concept of *repeated negotiation game* to model the n-agent repeated negotiation problem (based on the empirical game-theoretic approach) in which each agent is allowed to choose its negotiating strategy independently in each round. To obtain the set of negotiating strategies, we perform comprehensive simulations among all the state-of-the-art strategies entered into the final rounds of ANAC from 2010 to 2012, and choose the top six strategies as our candidate strategies. It is reasonable to assume that any rational agent would employ certain rational learning approach to decide which negotiating strategy to choose through repeated interactions [11]. In this work, we focus on three representative rational learning approaches from multiagent learning literature: Ficitious Play [4], Q-learning[14], and Win or Learn Fast - policy hill climbing (WoLF-PHC) learning [3]. From our analysis, we are able to gain more insights in terms of the relative robustness of different negotiating strategies compared with the previous robustness analysis based on single-agent best deviation [1].

The remainder of the paper is organized as follows. In Section 2, we review some backgrounds of evaluating the robustness of negotiating strategies and the limitation of previous approach. In Section 3, we describe the repeated negotiation game framework we propose to evaluate the robustness of negotiation strategies. In Section 4, we present the robustness evaluation results of different negotiation strategies under our framework and compare with previous approach. An overview of related work in automated negotiation area is given in Section 5. Lastly we conclude and point out some future work in Section 6.

2   Background and Problem Description

Recently some attention has been given to investigate the dynamic aspect of negotiating strategies by evaluating the *robustness* [1] of the current state-of-the-art negotiating strategies, based on the game-theoretic approach. Since there exist an infinite number of possible negotiation strategies that the agents may take, we cannot apply the standard game-theoretic approach to perform such an analysis by explicitly considering all possible strategies. Therefore, the tool of empirical game theoretic (EGT) analysis is adopted to achieve this goal instead, which is originally developed to analyze the Trading Agent Competition. EGT analysis is a game-theoretic analysis approach based on a set of empirical results. It handles the problem of the existence of infinite possible strategies by assuming that each agent only selects its strategy from a fixed set of strategies and the outcomes for each strategy profile can be determined through empirical simulations. This technique has been successfully applied in addressing questions about robustness of different strategies from various domains including continuous double auction [13], trading strategies in previous years' TAC competitions [9] and different negotiation strategies [1, 15, 5].

Given a set of negotiation strategies, different from the setting of ANAC, each agent is free to select any strategy from this set as its negotiation strategy. For each bilateral negotiation, the corresponding payoff received for each

participating agent is determined as its average payoff against its opponent over all domains, which can be obtained through empirical simulations. Based on the bilateral negotiation outcomes, the average payoff of an agent in any given tournament can be determined by averaging its payoff obtained in all bilateral negotiations against all other agents in the tournament. Specifically, for a given tournament involving a set $\mathcal{P}$ of agents, the payoff $U_p(\mathcal{P})$ obtained by agent p can be calculated as follows,

$$U_p(\mathcal{P}) = \frac{\sum_{p' \in \mathcal{P}, p' \neq p} u_p(p, p')}{|\mathcal{P}| - 1} \tag{1}$$

where $u_p(p, p')$ represents the corresponding average payoff of agent p negotiating against another agent p' which is obtained from simulation results. Note that agent p and p' can use either the same or different strategies.

An agent has the incentive to deviate its current strategy to another one if and only if its payoff after deviation can be statistically improved, provided that all the other agents keep their strategies unchanged. There may exist multiple candidate strategies that an agent has the incentive to deviate to, usually we only consider the best deviation available to that agent in terms of maximizing its deviation benefit [1, 15]. Given a strategy profile, if no agent has the incentive to unilaterally deviate from its current strategy, then this strategy profile is called an *empirical pure strategy Nash equilibrium*. In general, a game may have no empirical pure strategy Nash equilibrium. Another useful concept for analyzing the stability of the strategy profiles is *best reply cycle*, which is a subset of strategy profiles in which, for any strategy profile within this subset, there is no single-agent best deviation path leading to any profile outside the cycle. In other words, in a best reply cycle, all single-agent best deviation paths starting from any strategy profile within itself must lead to another strategy profile inside the same cycle.

Both *empirical pure strategy Nash equilibrium* and *best reply cycle* can be considered as two different interpretations of empirical stable sets to evaluate the *stability* of different strategy profiles. Based on these two concepts, the *robustness* of a strategy is evaluated using the concept of *basin of attraction* of a stable set [13]. The *basin of attraction* of a stable set is defined as the percentage of strategy profiles which can lead to this stable set through a series of single-agent best deviations. Accordingly, a negotiation strategy s is considered to be *robust* if it belongs to a stable set with a large *basin of attraction* [13, 1]. However, as we previously mentioned, the current robustness analysis relies on a number of assumptions such that the analysis results may not be able to accurately reflect the relative robustness of different negotiation strategies in practical negotiation scenarios. Moreover, the current robustness analysis cannot provide a more fine-grained analysis in terms of the relative robustness of strategies within the same stable set when the stable set involves multiple negotiation strategies.

3 Robustness Analysis Framework

3.1 Repeated Negotiation Game

We propose analyzing the robustness of negotiation strategies within the framework of repeated negotiation games. We first define the single-shot *negotiation game* as follows. Given a set N of agents and a set S of negotiation strategies, the negotiation problem among n agents can be modeled as a single-shot normal-form game. Formally it can be represented as a tuple $\langle N, (S_i), (U_i) \rangle$ where

- $N = \{a_1, a_2, \ldots, a_n\}$ is the set of agents.
- S_i is the set of negotiating strategies available to agent a_i.
- U_i is the utility function of agent i as defined in Equation 1, and $U_i(\mathcal{P})$ corresponds to the average payoff agent a_i receives under the current negotiation tournament, where $\mathcal{P}$ is the strategy profile in the current round of negotiation.

Similar to the previous robustness analysis [1], in the *negotiation game* definition, we assume that each agent i may only select negotiating strategies from a set S_i of candidate strategies based on empirical game-theoretic analysis. In this way, we are able to handle the problem of the existence of infinite possible negotiating strategies, which would make the analysis infeasible. To select the set of strategies for our analysis, we first collect all the top 8 strategies that enter into the past 3-year ANAC competitions from 2010 to 2012 (24 strategies in total), and evaluate their relative rankings in terms of efficiency over a large number of negotiation domains through extensive simulations. We find that the top six strategies actually correspond to the top 3 strategies from ANAC 2011 and 2012, which indicates that the most recently developed negotiating strategies are more efficient and thus more likely to be adopted in practice. Therefore, we select the top six negotiating strategies as the set S of candidate strategies for our analysis, i.e., $S = \{\mathcal{G}, \mathcal{H}, \mathcal{I}, \mathcal{A}, \mathcal{C}, \mathcal{O}\}$.[5]

Since the negotiation game is defined based on empirical game theory, we can define the concept of *empirical pure strategy Nash equilibrium* in a similar way to the definition of pure strategy (mixed strategy) Nash equilibrium by using the finite strategy set S to replace the original infinite strategy set.

Definition 1. *An* empirical pure strategy Nash equilibrium *for an n-player negotiation game is a strategy profile* $(a_1^*, a_2^*, \ldots, a_n^*)$ *such that* $\forall i \in N$, *we have*

$$U_i(a_i^*, a_{-i}^*) \geq U_i(a_i, a_{-i}^*), \forall a_i \in S_i \qquad (2)$$

where S_i $(\forall i \in N)$ *is the finite set of strategies we choose to represent the original infinite set of strategies.*

---

[5] These bold letters are the abbreviations for the six negotiating strategies as follows: $\mathcal{G}$ – Gahboninho, $\mathcal{H}$ – HardHeaded, $\mathcal{I}$ – IAMhaggler2011, $\mathcal{A}$ – AgentLG, $\mathcal{C}$ –CUHKAgent, $\mathcal{O}$ – OMAC.

If the agents are allowed to use mixed strategy, then we can naturally define the concept of *empirical mixed strategy Nash equilibrium* similarly.

Definition 2. *An* empirical mixed strategy Nash equilibrium *for an n-player normal-form game is a strategy profile* $(\pi_1^*, \pi_2^*, \ldots, \pi_n^*)$ *such that* $\forall i \in N$, *we have*

$$\bar{U}_i(\pi_i^*, \pi_{-i}^*) \geq \bar{U}_i(\pi_i, \pi_{-i}^*), \forall \pi_i \in \Pi(S_i) \tag{3}$$

where $\bar{U}_i(\pi_i^*, \pi_{-i}^*)$ *is player i's expected payoff under the strategy profile* (π_i^*, π_{-i}^*), *and* $\Pi(S_i)$ *is the set of probability distributions over player i's action space* S_i.

An *empirical mixed strategy Nash equilibrium* (π_1^*, π_2^*) is degenerated to an *empirical pure strategy Nash equilibrium* if both π_1^* and π_2^* are pure strategies.

We consider the general setting of the *repeated negotiation game* where each agent is free to choose its negotiating strategy simultaneously based on the feedback from the previous round. In each round, given the negotiation strategy profile of agents, the negotiation tournament starts and each agent i receives its own average payoff $U_i(\mathcal{P})$ from the current round. We assume that initially each agent has equal probability to select each of the negotiation strategy from its strategy space S_i. We evaluate the relative robustness of different negotiation strategies based on the corresponding probability that each strategy profile can be converged to. Given a particular negotiation strategy s, we define its *basin of attraction* as the accumulated frequency of all strategy profiles that the agents can learn to converge to and also involve strategy s. The robustness of a strategy s is then defined based on its basin of attraction.The higher the basin of attraction of a strategy s is, the more robust strategy s is. For example, consider two negotiation strategy s_1 and s_2, and the agents converge to (s_1, s_1) with probability of 0.8, and (s_2, s_2) with probability of 0.2. The basin of attraction of strategy s_1 and s_2 is 0.8 and 0.2 respectively, and thus we can say strategy s_1 is more robust then strategy s_2. The overall robustness analysis framework can be summarized as follows in Algorithm 1.

Algorithm 1 Overall Robustness Analysis Framework

1: Choose a learning strategy for each agent to determine its negotiation strategy each round
2: **repeat**
3:     Each agent choose its current-round negotiation strategy according to its learning strategy
4:     Each agent update its learning strategy based on the current round outcome.
5: **until** The negotiation game ends
6: Calcualte the basin of attraction of each negotiation strategy
7: Determine the relateive robustness of each negotiation strategy

---

3.2   Learning Strategies

A remaining question is how the agents should select their negotiating strategies each round. We assume that the agents are individually rational and thus each

agent is faced with the task of how to make decisions to increase its individual payoff as much as possible through negotiation given the uncertainty of the negotiating strategies chosen by others. We adopt three representative rational learning strategies from multiagent learning literature: fictitious play learning [4], Q-learning [14] and WoLF-PHC learning [3].[6] All the three learning strategies are rational in that they all aim at learning the policy of maximizing their individual payoffs based on the past feedbacks in different ways. Next we will describe how these three strategies can be applied to select negotiating strategies for agents in the context of repeated negotiation games respectively.

Fictitious Play Learning Under fictitious play, an agent maintains the belief that its opponent makes decisions following a fixed mixed strategy and always chooses actions to maximize its average payoff regarding its current belief each round. In the context of n-player repeated negotiation game, the overall learning rule of fictitious play can be described as follows. Similar to the robustness anal-

---

Algorithm 2 Fictitious Play in Repeated Negotiation Games

1: Initialize the agent's belief about its opponents.
2: **repeat**
3:     Choose the negotiating strategy maximizing its average payoff (random choose one in case of a tie) based on its current belief.
4:     Update its belief based on the current round outcome.
5: **until** The negotiation game ends

---

ysis, fictitious play learning implicitly requires that each agent i should know its own utility over all strategy profiles in advance. Besides, it is well-known that a fictitious play learner may never learn its best strategy due to its erroneous belief of other players always playing fixed mixed strategies. To this end, we modify the original fictitious play by allowing each agent to make explorations occasionally, which thus gives the agents the opportunity to identify other possibly better strategies. Specifically, each fictitious play learner chooses action to maximize its expected payoff based on its current belief with probability $1 - \epsilon$, and make random selection with probability ϵ.

Q-Learning Q-learning [14] is one representative reinforcement learning approach and has received much attention in multiagent learning literature. In the context of repeated negotiation game, each Q-learning agent i holds a Q-value $Q_i^t(s)$ for each negotiating strategy $s \in S_i$, and gradually updates its Q-value $Q_i^t(s)$ for each action s based its own payoff in each round. The Q-value update rule for each action s is as follows:

$$Q_i^{t+1}(s) = \begin{cases} Q_i^t(s) + \alpha_i(U_i^t(\mathcal{P}) - Q_i^t(s)) & \text{if } s \text{ is chosen} \\ Q_i^t(s) & \text{otherwise} \end{cases} \quad (4)$$

---
[6] It is worth noting that any other rational learning strategies could be used here.

where $U_i^t(\mathcal{P})$ is the payoff agent i obtains in round t under current outcome $\mathcal{P}$ by taking action s. Besides, α_i is the learning rate of agent i, which determines how much weight we give to the newly acquired payoff $U_i^t(\mathcal{P})$, as opposed to the old Q-value $Q_i^t(s)$.

In each round t, each agent i chooses its action based on the ϵ-greedy exploration mechanism as follows. With probability $1 - \epsilon$, it chooses the negotiating strategies with the highest Q-value from the set S of candidate strategies, and makes random selection from S with probability ϵ. The value of ϵ controls the exploration degree during learning. It initially starts at a high value and decreased gradually to zero as negotiation goes on.

WoLF-PHC Learning WoLF-PHC learning [3] is a rational learning algorithm which incorporates the " win or learn fast" principle into the basic policy hill-climbing (PHC) algorithm [3]. In PHC algorithm, each round each agent maintains a Q-table for each negotiating strategy and updates it in a way similar to that in Q-learning algorithm, but it also explicitly keeps and updates its mixed strategy policy. Each round each agent's mixed strategy policy is updated in the direction of increasing the probability that the action with the highest Q-value with a fixed learning rate. WoLF-PHC learning algorithm extends the basic PHC algorithm by introducing two different learning rates to update its mixed strategy policy. The principle of WoLF update is that it selects the smaller learning rate to update its mixed strategy policy when it wins, and the larger one to update when it losses. A WoLF-PHC agent evaluates whether it wins or not by comparing its expected payoff by playing its current mixed strategy policy with that by playing the average mixed strategy policy over all previous rounds. If its expected payoff by adopting its current mixed strategy is higher, it means it wins, otherwise it losses. The overall description of the WoLF-PHC algorithm in the context of repeated negotiation game is shown in Algorithm 3.

Similar to Q-learning, one advantage of WoLF-PHC learning is that it also only requires the minimum amount of information for each agent, i.e., each agent only needs to know its payoff obtained from each round of negotiation. However, Q-learning only enables the agents to learn pure strategy policies, while WoLF-PHC enables the agents to learn mixed strategy policies.

4   Experimental Evaluation

4.1   Experimental Settings

The payoff matrix for each pair of negotiating strategies in $S \times S$ is obtained based on extensive simulations over all the possible negotiation domains shown in Table 1. For any negotiation game, given a negotiating strategy profile $\mathcal{P}$, the corresponding payoff $U_i(\mathcal{P})$ for each agent i can be easily calculated based on Equation 1 and the payoffs in Table 1.

We start with the simplest setting, bilateral repeated negotiation, in which only two agents repeatedly negotiate with each other. The second negotiation

---

Algorithm 3 WoLF-PHC Learning in Repeated Negotiation Games

1: Initialize $Q^0(s) = 0$, $\pi^0(s) = \frac{1}{|S|}$, $C = 0$, $\forall s \in S$.
2: Initialize two learning rates ρ_s, ρ_l.
3: **repeat**
4:    Choose a negotiating strategy s from S according to the current mixed strategy policy $\pi^t(s)$ with appropriate exploration.
5:    Update its Q-table based on the payoff obtained in the current round following Equation 4.
6:    Update its average mixed strategy policy as follows,
      $C = C + 1$, $\bar{\pi}^t(s) = \bar{\pi}^{t-1}(s) + \frac{\bar{\pi}^{t-1}(s) - \pi^t(s)}{C}$, $\forall s \in S$
7:    Determine the learning rate ρ^t to update its mixed strategy policy $\pi^t(s)$ as follows,

$$\rho^t = \begin{cases} \rho_s & \sum_{s \in S_i} \pi^t(s) Q^t(s) \geq \sum_{s \in S_i} \bar{\pi}^t(s) Q^t(s) \\ \rho_l & \text{otherwise} \end{cases} \tag{5}$$

8:    Update its mixed strategy policy w.r.t. the Q-table.

$$\pi^{t+1}(s) = \begin{cases} \pi^t(s) + \sum_{s' \neq s} \delta(s') & Q(s) \text{ is the highest} \\ \pi^t(s) - \delta(s) & \text{otherwise} \end{cases} \tag{6}$$

      where $\delta(s) = \min(\pi^t(s), \frac{\rho^t}{|S|-1})$
9: **until** The negotiation game ends

---

setting we consider is three-agent repeated negotiation, and the last setting is to consider six-agent repeated negotiation. For all settings, each agent is allowed to choose any negotiating strategy from the strategy set S. Since the negotiation game itself is symmetric, we only need to care about the number of agents choosing each negotiating strategy. Therefore, in the following analysis, we merge those strategy profiles with the same number of agents choosing each negotiation strategy and treat them as the same outcome to make the results clearer. For example, an outcome $(G : 1, A : 1)$ in a bilateral negotiation means one agent chooses strategy G and the other chooses A, and covers both the strategy profiles of (G, A) and (A, G). The parameter settings for each learning appraoch are listed in Table 2.

Table 1: Payoff matrix for the top six negotiation strategies average over all domains (For each strategy profile, only the row player's payoff is given since the game is symmetric.) The letters in bold are the abbreviations for each strategy.

Payoff	G	H	I	A	C	O
G	0.680	0.520	0.812	0.676	0.580	0.555
H	0.662	0.599	0.757	0.569	0.604	0.549
I	0.622	0.564	0.715	0.595	0.470	0.492
A	0.709	0.590	0.787	0.568	0.664	0.561
C	0.740	0.639	0.826	0.552	0.597	0.590
O	0.697	0.628	0.771	0.551	0.605	0.571

Table 2: Parameter Settings for Each Learning Approach

Payoff Matrix	exploration rate ϵ	Learning Rate α	Learning Rate ρ_s	Learning Rate ρ_l
Fictitious-Play	0.05	N/A	N/A	N/A
Q-learning	0.05	0.6	N/A	N/A
WoLF-PHC	0.05	0.6	0.2	0.5

4.2   Bilateral Repeated Negotiations

In this setting, we first give the robustness analysis results based on the previous robustness analysis approach [1]: there only exists a best reply cycle of $(G : 1, A : 1) \rightarrow (G : 1, C : 1) \rightarrow (C : 1, A : 1)$. In other words, for any initial state, the agents would always converge to and stick in this cycle, however, it says nothing about dynamic details within that cycle (e.g., the probability that each strategy profile would be finally adopted by the agents) and the relative robustness of the strategies within that cycle.

Figure 1(a) shows the dynamics of the average frequency the previous three outcomes can be achieved as a function of the number of rounds when both agents negotiates following fictitious play. This can be considered as the dynamic way of understanding how the three outcomes within the best reply cycle evolve from one to another. Whenever the probability of reaching one outcome is decreased, the probability of reaching its neighbor outcome is increased. We can also observe that the basin of attraction of $(G : 1, A : 1)$ and $(G : 1, C : 1)$ (i.e., the frequencies of reaching these two outcomes) is usually larger that that of the outcome $(C : 1, A : 1)$, which thus may indicate that strategy G is more robust than the other two strategies in practice.

Figure 1(b) and Figure 2(a) show the average probabilities that the previous three outcomes can be achieved as a function of rounds when the agents negotiate following Q-learning and WoLF-PHC respectively. Different from fictitious play, it turns out surprisingly that the basin of attraction of outcome (G: 1, A: 1) (i.e., the probability of reaching it) is much larger than the rest of outcomes in the best reply cycle. We hypothesize that it is because in fictitious play the agents are assumed to know their payoff information for each pair of negotiating strategies beforehand, while this is not allowed in both Q-learning and WoLF-PHC. This result may indicate that in practice it is most likely that any rational agent would choose strategy G to participate in a bilateral negotiation while its (rational) opponent chooses strategy A and vice versa. In other words, strategy C is not as robust as the other two strategies G and A even though all of them are within the best reply cycle. For both Q-learning and WoLF-PHC learning, we can easily calcualte the basin of attraction of these three strategies and thus obtain their relative robustness ranking as follows: $G > A > C$.
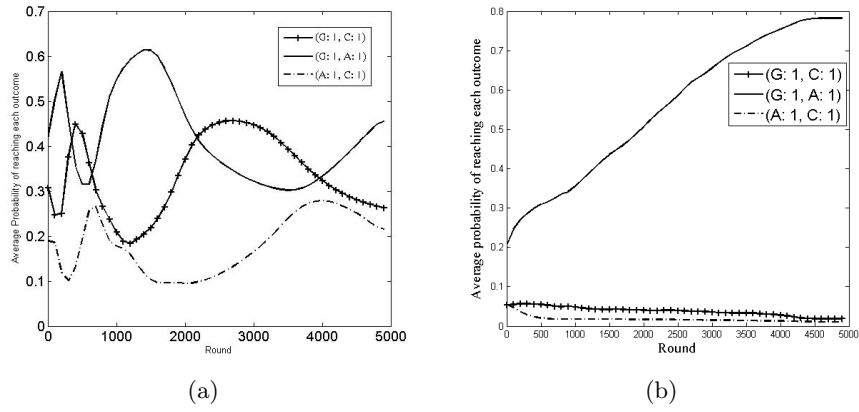
(a)   (b)

Fig. 1: Average probability of reaching each outcome for 2-agent case under a) fictitious play, b) Q-learning

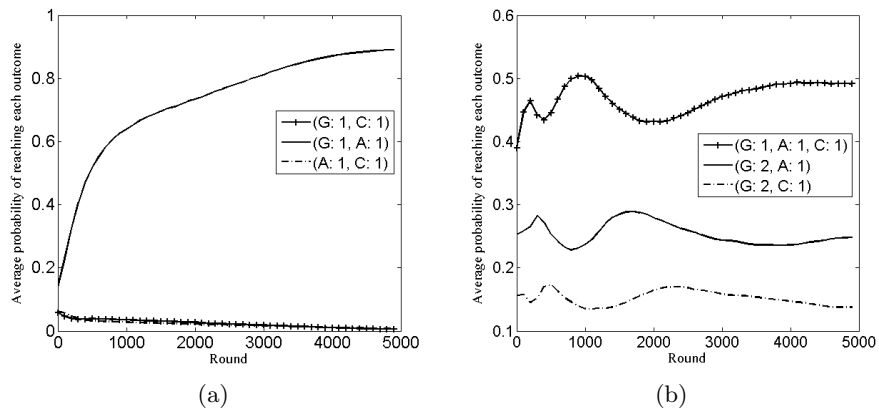(a)   (b)

Fig. 2: (a) Average probability of reaching each outcome for 2-agent case WoLF-PHC learning, (b) Average probability of reaching each outcome for 3-agent case under fictitious play

4.3   Three-agent Repeated Negotiation Tournaments over Six Strategies

Next we increase the number of agents in the negotiation and investigate the case of three-agent repeated negotiations where each agent is allowed to choose any negotiating strategy from S. We first give the analysis results based on the previous robustness analysis approach [1], which indicates that there only exists one best reply cycle: $(G : 2, C : 1) \rightarrow (G : 1, A : 1, C : 1) \rightarrow (G : 2, A : 1)$. This means that all these three strategies are more robust than the rest of strategies.

Figure 2(b) shows the average frequency of reaching the previous three outcomes as a function of rounds when all agents employ fictitious play. We can see that the frequencies of achieving these three outcomes are significantly different $((G : 1, A : 1, C : 1)$ ranks first, $(G : 2, A : 1)$ ranks second and $(G : 2, C : 1)$ ranks last) even though they are all within the best reply cycle. Based on this, we can also calculate the average frequency of each negotiating strategy can be adopted by the agents, and come to the conclusion of their robustness ranking as follows: $G > A > C$.

Figure 3(a) and 3(b) illustrate the average frequency the previous three outcomes can be reached as a function of the number of rounds when the agents employ Q-learning and WoLF-PHC respectively. The results for both cases are similar in that the probability of reaching outcome $(G : 2, A : 1)$ $(> 80\%)$ is much higher than that of reaching the other two outcomes. This indicates that it is most likely that two agents would adopt strategy G while one agent adopt strategy A rather than always cycling around the three outcomes. Besides, this result also further confirms our previous conclusion that strategy C is not as robust as the other two strategies.
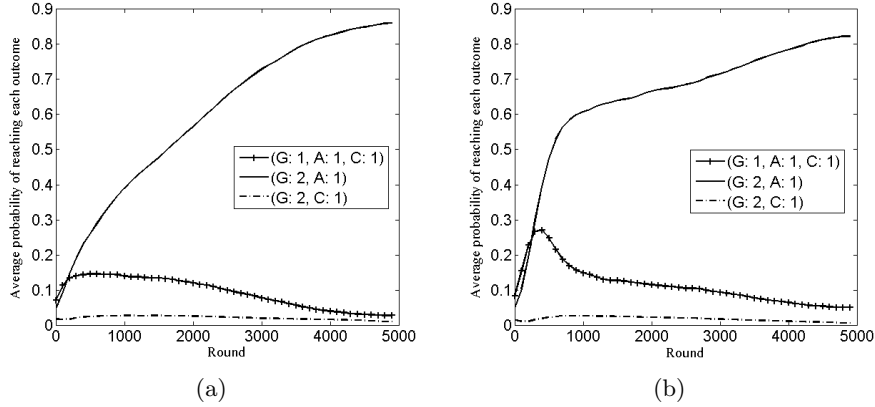

Fig. 3: Averag probability of reaching each outcome for 3-agent case under a) Q-learning, and b) WoLF-PHC learning

4.4 Six-agent Repeated Negotiation Tournaments over Six Strategies

Finally we further increase the number of agents to be equal to the number of negotiating strategies. Based on the previous robustness analysis approach [1], we can know that there only exists one best reply cycle as follows: $(G : 4, A : 1, C : 1) \rightarrow (G : 4, C : 2) \rightarrow (G : 3, A : 1, C : 2) \rightarrow (G : 2, A : 2, C : 2) \rightarrow (G : 3, A : 2, C : 1) \rightarrow (G : 4, A : 2)$, which indicates that the agents will eventually reach this cycle and the three strategies G, A, and C are more robust than the rest of strategies. However, we cannot distinguish the relative robustness among these three negotiation strategies.

We present the dynamics of the frequency of reaching the previous six outcomes when all agents employ fictitious play in Figure 4(a). We can see that the frequencies of reaching each outcome vary dynamically, with $(G : 3, A : 2, C : 1)$ ranks first and $(G : 3, A : 1, C : 2)$ ranks second most of the time (the sum of the probabilities of reaching these two outcomes is about 70%). This means that it is most likely that the agents would frequently change their strategies between A and C during repeated negotiation, which thus indicates that strategy G is more robust than A and C, even though all of them are within the best reply cycle.
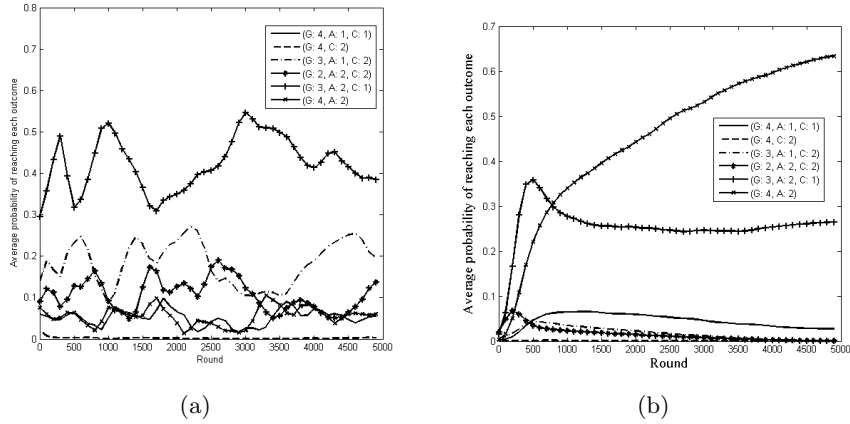

(a)   (b)

Fig. 4: Average probability of reaching each outcome for 6-agent case under a) fictitious play, b) Q-learning

Figure 4(b) and 5 illustrate the dynamics of the average frequency of reaching the previous six outcomes when the agents employ Q-learning and WoLF-PHC respectively. Both figures show similar results that the frequency of reaching outcomes $(G : 4, A : 2)$ and $(G : 3, A : 2, C : 1)$ dominates the rest of outcomes, i.e., the probability of reaching these two outcomes are much higher than the rest

of them. Based on the average probabilities of reaching each outcomes within the best reply cycle, we can also approximately estimate the relative robustness of the three strategies G, A, and C, i.e., $G > A > C$, which is in consistent with the analytical results from previous two cases.
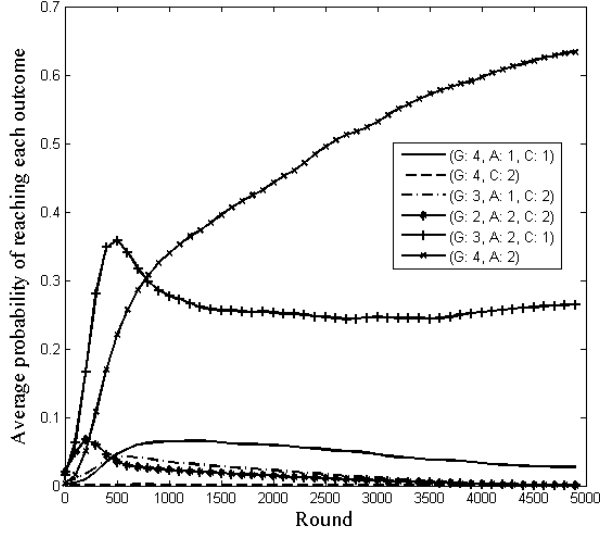

Fig. 5: Average probability of reaching each outcome for 6-agent case under WoLF-PHC learning

4.5   Discussion

From previous analysis results, we can see that our robustness analysis can provide more fine-grained analysis of the relative robustness of different negotiation strategies than the previous approach [1]. If there exists an empirical pure strategy Nash equilibrium, the corresponding interpretation under our analysis framework is that the negotiators learn to converge to that pure strategy equilibrium. If there exists a best reply cycle, for those learning strategies capable of learning a mixed strategy, one reasonable interpretation under our framework could be that the agents are actually learning to converge to the corresponding empirical mixed strategy Nash equilibrium.

From Section 4.2 to 4.4, only the results for one particular learning strategy are shown due to space limitation. However, it is worth noticing that for each tournament setting, the consistent analysis results are obtained for all three learning strategies. Intuitively, our analysis results reflect the relative robustness of different negotiation strategies when the negotiators are rational and have the

freedom to choose their negotiation strategies based on their past experience. We believe that this kind of robustness analysis framework can better reflect the practical multi-agent negotiation scenarios (e.g., e-commerce area) which usually involves repeated negotiations [10], and thus provide more accurate predictions of which negotiation strategies are more likely to be adopted in practice.

5 Related Work

Baarslag et al. [1] firstly propose an alternative evaluation criterion, *robustness*, to evaluate the performance of different negotiation strategies from a different perspective based on empirical game-theoretic analysis. The authors investigate the relative robustness of the top 8 strategies participating in ANAC'11 competition in different tournament settings. Their analysis provides some interesting results. For the bilateral negotiation setting, it is found that the winning strategy in ANAC'11 is not the most robust strategy, and also no pair of agents adopting the same negotiating strategy is stable. For the setting of 8-player tournament with three strategies, it is surprisingly found that the Gahboninho strategy, which is not the winner strategy, seems to be the most robust strategy. For the last setting of 8-players tournament with 8-strategy, it is also found that the non-winner Gahboninho strategy is more robust than the winning strategy HardHeaded. All the previous robustness analysis can provide us with some useful insights about which strategies would be adopted by agents in practice. However, the robustness analysis is limited by the assumption of single-agent best-deviation and the requirement of global payoff information, which may not be realistic in practice. Last, when the agents end up with a best reply cycle, it provides us with little information about the relative robustness of the strategies.

Williams el al. [15] propose a novel concession negotiating strategy, which make concessions based on the observed concession of the opponent and other negotiation constraints under the elapsed real time. They evaluate the efficiency of their strategy against the state-of-the-art strategies under a number of benchmark domains taken from ANAC'10. The simulation results show that their strategy is more efficient in terms of the average payoff obtained over all opponents. The authors also evaluate the robustness of their strategy based on empirical game-theoretic analysis, and focus on the case of 5 players' tournament negotiation with the top 3 strategies. It is found that their strategy is the most robust one among all strategies considered and all agents have the incentive to switch to their strategy eventually. Their robustness analysis follows the work of Baarslag et al. [1], and thus suffers from the same limitations as theirs.

Chen and Weiss [5] propose a novel negotiating strategy, Dragon, which employs sparse pseudo-input Gaussian processes to support more accurate estimations of the opponent's behaviors. In their robustness analysis, they focus on the bilateral negotiation setting among eight strategies, and it is found that the only stable state is when one agent adopts the Dragon strategy while the other agent adopts the TheNegotiator Reloaded strategy from ANAC'12. The result indicates that Dragon strategy is robust compared with others in the bilateral

negotiation setting. However, the robustness analysis also follows the work of Baarslag et al. [1], and thus the same limitations as theirs [1] apply here.

6   Conclusion and Future Work

We introduce the concept of repeated negotiation game and propose employing different rational learning strategies to provide more realistic and fine-grained analysis of the robustness of different negotiating strategies. We make extensive evaluation of the top 6 negotiation strategies participated in the past three-year negotiations under three different negotiation tournament settings. Through our analysis, we show that more insights in terms of the relative robustness of different negotiation strategies can be revealed, which are usually not available from the analysis using the previous approach [1]. As future work, we are going to further generalize the robustness evaluation process in a more formal and systematic way to make the robustness evaluation and analysis of different negotiating strategies more meaningful and efficient.

7   Acknowledgement

References

1. T. Baarslag, K. Fujita, E. H. Gerding, K. Hindriks, T. Ito, N. R. Jennings, C. Jonker, S. Kraus, R. Lin, V. Robu, and C. R. Williams. Evaluating practical negotiating agents: Results and analysis of the 2011 international competition. *Artificial Intelligence Journal*, 198:73–103, 2013.
2. T. Baarslag, K. Hindriks, C. Jonker, S. Kraus, and R. Lin. The first automated negotiating agents competition (anac 2010). *New Trends in Agent-Based Complex Automated Negotiations*, pages 113–135, 2010.
3. M. H. Bowling and M. M. Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, pages 136:215–250, 2003.
4. G. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 1951.
5. S. Q. Chen, H. B. Ammar, K.l Tuyls, and G. Weiss. Optimizing complex automated negotiation using sparse pseudo-input gaussian processes. In *AAMAS'13*, pages 707–714, 2013.
6. R. M. Coehoorn and N. R. Jennings. Learning an opponent's preferences to make effective multi-issue negotiation trade-offs. In *ICEC'04*, pages 59–68, 2004.
7. J.Y Hao and H.F Leung. Abines: An adaptive bilateral negotiating strategy over multiple items. In *WI-IAT'12*, pages 95–102, 2012.
8. K. Hindriks and D. Tykhonov. Opponent modeling in auomated multi-issue negotiation using bayesian learning. In *AAMAS'08*, pages 331–338, 2008.
9. J. Estelle M. P. Wellman, S. Singh, Y. Vorbeychik, and V. Soni. Strategic interactions in a supply chain game. *Computational Intelligence*, 21(1):1–26, 2005.

10. S. Saha, A. Biswas, and S. Sen. Modeling opponent decision in repeated one-shot negotiations. In *AAMAS'05*, pages 397–403, 2005.
11. S. Sen and S. Airiau. Emergence of norms through social learning. In *IJCAI'07*, pages 1507–1512, 2007.
12. J. Y. Hao S. Z. Song, H. F. Leung, and Z. Ming. An efficient and robust negotiating strategy in bilateral negotiations over multiple items. *Engineering Applications of Artificial Intelligence*, 34:45–57, 2014.
13. P. Vytelingum, D. Cliff, and N.R. Jennings. Strategic bidding in continuous double auctions. *Artificial Intelligence*, 172(14):1700–1729, 2008.
14. C. J. C. H. Watkins and P. D. Dayan. Q-learning. *Machine Learning*, pages 3:279–292, 1992.
15. C. R. Williams, V. Robu, E. H. Gerding, and N. R. Jennings. Using gaussian processes to optimise concession in complex negotiations against unknown opponents. In *IJCAI'12*, pages 432–438, 2012.