QESRL: Exploring Selfish Reinforcement Learning for Repeated Quantum Games

Agustin Silva, Omar Gustavo Zabaleta and Constancio Miguel Arizmendi

Instituto de Investigaciones Científicas y Tecnológicas (ICYTE).
Av. Juan B. Justo 4302, Mar del Plata, Argentina.

E-mail: agustinsilva447@gmail.com

Abstract. A novel unified learning algorithm that seamlessly applies to both classical and quantum non-zero sum games is presented. Building upon the exploring selfish reinforcement learning (ESRL) framework previously proposed in the context of classical games, we extend this approach to handle quantum games with imperfect information. A comparison is made between performance and fairness among agents learning using plain QRL vs. QESRL. The latter enables agents to explore and learn periodic policy strategies in quantum games, leveraging the quantization of games to uncover fairer results. By addressing the challenges posed by the expanded strategy space in quantum games, we test the algorithm's scalability by increasing the number of agents. Empirical evidence is provided to showcase its performance and to compare classical and quantum game scenarios. The proposed learning algorithm represents a significant step towards understanding the convergence and optimality of strategies in non-zero sum games across classical and quantum settings, bringing us closer to harnessing the potential of independent reinforcement learning and quantum computing in game theory applications.

1. Introduction

The intersection of game theory, quantum mechanics, and reinforcement learning has sparked a fascinating realm of inquiry, where the dynamics of decision making among rational agents manifest in both classical and quantum settings [1]. In this pursuit, Quantum Exploring Selfish Reinforcement Learning (QESRL), inspired by ESRL [2], stands as a groundbreaking innovation, forging a bridge between the classical and quantum domains. This confluence of disciplines allows us to gain profound insights into strategic interactions, unveiling the intricate fabric that underpins the optimization of agent behaviors in diverse scenarios.

Game theory, a cornerstone of economic and social sciences, provides a framework for analyzing and predicting strategic interactions between agents who each possess their own objectives. A central facet of this theory lies in its depiction of the reward structures that underlie agents' choices, where decisions by one agent impact the gains of others [3]. The ESRL algorithm operates at the nexus of game theory and reinforcement learning, with an aim to optimize strategies that align with each agent's goals while accounting for the complex interplay between individual and collective outcomes.

Quantum mechanics, on the other hand, introduces a realm of uncertainty and entanglement that deviates from the deterministic paradigms of classical mechanics. The advent of quantum game theory has exploited the nuances of quantum states to explore how these phenomena

impact strategic interactions [4]. The foundational concept of entanglement, in which correlated states emerge irrespective of spatial separations, presents a challenge and an opportunity in the realm of quantum game theory [5]. Quantum games have already been used for different applications, from communication networks [6, 7] to economic models [8, 9].

Multi-agent reinforcement learning, a core tenet of artificial intelligence, transcends traditional algorithmic approaches by focusing on agent-environment interactions. By learning from actions and their consequences, agents adapt their strategies over time to maximize cumulative rewards [10]. The ESRL algorithm, a predecessor of QESRL, pioneered this adaptive approach by incorporating exploration, synchronization, and exploitation phases to optimize strategies in classical games.

In the classical realm, there is an extensive bibliography exploring the consequences of agents learning in environments modeled using game theory [11]. On the other hand, there has been little research on the strategies for multi-agent learning in quantum games, most of them coming from the field of quantum evolutionary game theory [12, 13]. The QESRL algorithm provides a framework for delving into quantum environments, allowing agents to navigate the complex landscape of quantum strategy spaces.

This paper embarks on an investigation of the QESRL algorithm's capabilities, analyzing classical and quantum variations of renowned games such as the Battle of the Sexes, the Prisoner's Dilemma, and the Platonia Game. By juxtaposing classical and quantum outcomes, we scrutinize the algorithm's adaptability, convergence dynamics, and the balance between individual and collective rewards. Moreover, we traverse the landscape of fairness and equity by investigating how the QESRL algorithm addresses the challenge of distributing rewards equitably among agents in multi-agent scenarios.

The QESRL algorithm's capacity to navigate the intricate space where strategic interactions meet quantum phenomena holds the promise of reshaping our understanding of cooperative and competitive behaviors. The remainder of this paper is organized as follows. Section 2 briefly describes the classical ESRL algorithm and its implications. Section 3 gives a detailed characterization of the quantum game model and presents the underlying principles of the QESRL algorithm. Section 4 specifies the methodologies used and reports the results of convergence, performance, and fairness of agents using the QESRL algorithm in classical and quantum games. Finally, the work is concluded in Section 5 with a discussion of its consequences.

2. Classical ESRL

ESRL (Exploring Selfish Reinforcement Learning) [2] is an algorithm rooted in learning automata theory, tailored for stochastic, repeated non-zero-sum games. Independent ESRL agents update a probabilistic distribution over actions based on the rewards received. Agents alternate between exploration phases, prioritizing individual optimization, and synchronization phases, coordinating social objectives. Once a fixed period of time has elapsed, where agents have been switching from exploration to synchronization, they start exploiting their preferred strategies learned during the previous phases. In particular, ESRL adapts to different forms of games without a priori knowledge. This innovative approach harmonizes agents sharing very limited information, autonomous learning, and collaborative optimization within a dynamic framework. In the following subsections, we will generally explain the exploration, synchronization and exploitation phases. However, in Section 3.2 we will give a deeper description of them when defining the QESRL algorithm.

2.1. Exploration phase

During the exploration phase of the ESRL algorithm, agents operate autonomously as self-driven learners. Employing learning automata and policy hill-climbing techniques, agents adjust their action probabilities based on the outcomes of their actions and received rewards. This phase

leads to the convergence of the agents towards pure joint Nash equilibria. This self-contained learning process enables agents to adapt to various game scenarios without relying on prior knowledge.

2.2. Synchronization phase
In the synchronization phase, the focus shifts from individual optimization to collaborative decision-making. Agents collectively assess the joint solution attained during the exploration phase, considering its relevance to their individual objectives. This phase introduces limited communication among agents, allowing them to share their performance and contribute to the collective evaluation of the quality of the equilibrium. Through this cooperation, the ESRL algorithm identifies Pareto-optimal Nash equilibria, explores new solution attractors, and combines individual learning experiences to achieve refined and collectively beneficial outcomes.

2.3. Exploitation phase
In this final phase, agents take advantage of the strategies learned in the previous exploration and synchronization phases. After a sufficient amount of time, all agents begin to implement the periodic policy they have learned by coordinating their behavior and alternating between the different joint actions selected by each player. This periodic policy will remain until the end of the episode.

3. Quantum ESRL
The QESRL adapts the ESRL algorithm to be applied in the quantum game model presented in [1].

3.1. Classical and Quantum Games
The goal of game theory is to analyze decision-making systems involving two or more players cooperating or not with each other. An important feature in games is that the reward one player gets depends not only on the action she chooses, but also on other players' actions. It is well known that a game is defined by three elements: players, strategies, and rewards. The first part of this work is based on two player games with two pure strategies each. The rewards are then defined by a 2x2 pay-off matrix. In Table 1a it is possible to observe a general representation of a 2x2 payoff matrix (where [a,c,e,g] and [b,d,f,h] are the rewards of players 0 and 1 respectively). The following tables represent all the games that will be studied in the rest of the article (Battle of the Sexes 1b, Prisoner's Dilemma 1c and Platonia Game 1d), where player 0 can select between row actions and player 1 between column actions and get a reward of $(R_{player0}, R_{player1})$.

\	Player 1		
Player 0	\	A	B
	A	(a ; b)	(c ; d)
	B	(e ; f)	(g ; h)

(a) General matrix representation of a game.

\	Player 1		
Player 0	\	A	B
	A	(10 ; 5)	(2 ; 2)
	B	(0 ; 0)	(5 ; 10)

(b) Battle of the Sexes payoff matrix.

\	Player 1		
Player 0	\	A	B
	A	(6.6 ; 6.6)	(0 ; 10)
	B	(10 ; 0)	(3.3 ; 3.3)

(c) Prisoner's Dilemma payoff matrix.

\	Player 1		
Player 0	\	A	B
	A	(0 ; 0)	(0 ; 10)
	B	(10 ; 0)	(0 ; 0)

(d) Platonia Game.

To study quantum games, we follow the *EWL* [4] protocol for 2 players. The *first* step is to assign a quantum state to each of the possible strategies. In the case of two strategies, for

example, in prisoner's dilemma, *cooperate* $\rightarrow |0\rangle$ and *defect* $\rightarrow |1\rangle$. The *second* step is to create a quantum circuit where each player is assigned a qubit that starts in state $|0\rangle$. The *third* step is to create an entangled state between all players. This is done by applying the entanglement operator $J(\gamma) = cos(\frac{\gamma}{2}) * \mathbb{I}^{\otimes N} + i * sin(\frac{\gamma}{2}) * \sigma_x^{\otimes N}$, as seen in Fig. 1, where $\mathbb{I}$ is the identity matrix, σ_x the Pauli X gate, $N = 2$ represents the number of players, and γ a value determining the amount of entanglement, $\gamma = 0$ being no entanglement at all, and $\gamma = \frac{\pi}{2}$ maximum entanglement.

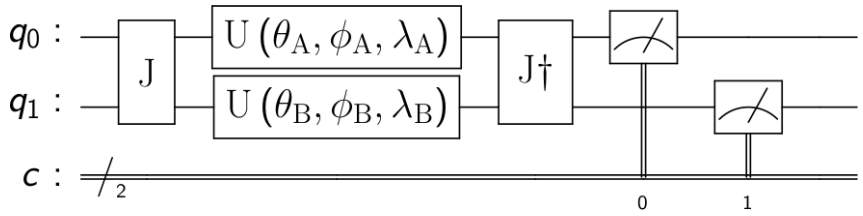


Figure 1: EWL game model for 2 players. Where q_0 and q_1 are the initial quantum states of the players and c is a classical register where the qubits measurements are stored.

In the *fourth* step, every player chooses her most suitable strategy individually and independently. This is done by modifying the state of her own qubit locally. To do this, every player applies one or more one-qubit gates, modifying the state of her qubit. A general one-qubit gate [14] is a unitary matrix that can be represented as:

$$U(\theta, \phi, \lambda) = \begin{pmatrix} \cos(\frac{\theta}{2}) & -e^{i\lambda}\sin(\frac{\theta}{2}) \\ e^{i\phi}\sin(\frac{\theta}{2}) & e^{i(\phi+\lambda)}\cos(\frac{\theta}{2}) \end{pmatrix} \tag{1}$$

We can already highlight the fact that while classic players have only 2 possible pure strategies (e.g. cooperate or defect), quantum players have an infinite number of pure strategies, that is, any combination of real value for the three parameters θ, ϕ and λ. The *fifth* step is to apply the operator $J^\dagger$ (conjugate transpose J) after the strategies of the players. Finally, the *sixth* step consists of measuring the state of the qubits to read the classical output of the circuit and, therefore, the final action of each player. The readouts are used as inputs of the pay-off matrix to determine the players' rewards.

One last thing to add is the fact that we are going to replace the three parameter general one-qubit gate $U(\theta, \phi, \lambda)$ by three one-parameter rotation one-qubit gates $R_X(\varphi_1)R_Y(\varphi_2)R_X(\varphi_3)$, with $R_X(\varphi) = exp(-i\frac{\varphi}{2}X) = \begin{pmatrix} cos(\frac{\varphi}{2}) & -isin(\frac{\varphi}{2}) \\ -isin(\frac{\varphi}{2}) & cos(\frac{\varphi}{2}) \end{pmatrix}$ and $R_Y(\varphi) = exp(-i\frac{\varphi}{2}Y) = \begin{pmatrix} cos(\frac{\varphi}{2}) & -sin(\frac{\varphi}{2}) \\ sin(\frac{\varphi}{2}) & cos(\frac{\varphi}{2}) \end{pmatrix}$. This is possible without losing generality since $U(\theta, \phi, \lambda) = e^{i\alpha}R_{\hat{n}}(\beta)R_{\hat{m}}(\gamma)R_{\hat{n}}(\delta)$ [14]. Having said that, the circuit from Fig. 1 becomes the one from Fig. 2.

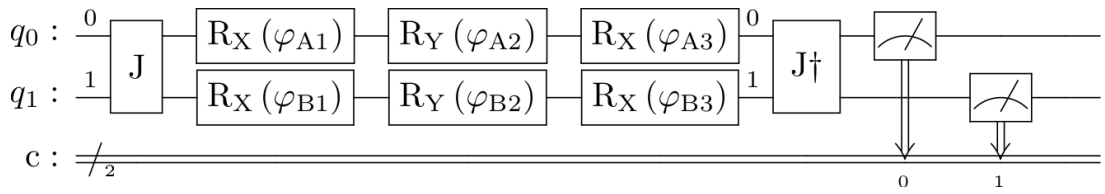


Figure 2: EWL game model for 2 players with rotation gates.

3.2. QESRL descrption

In order to adapt the ESRL to quantum games, we must be able to learn a probability density function over the strategies available to agents in quantum games. Each quantum strategy is

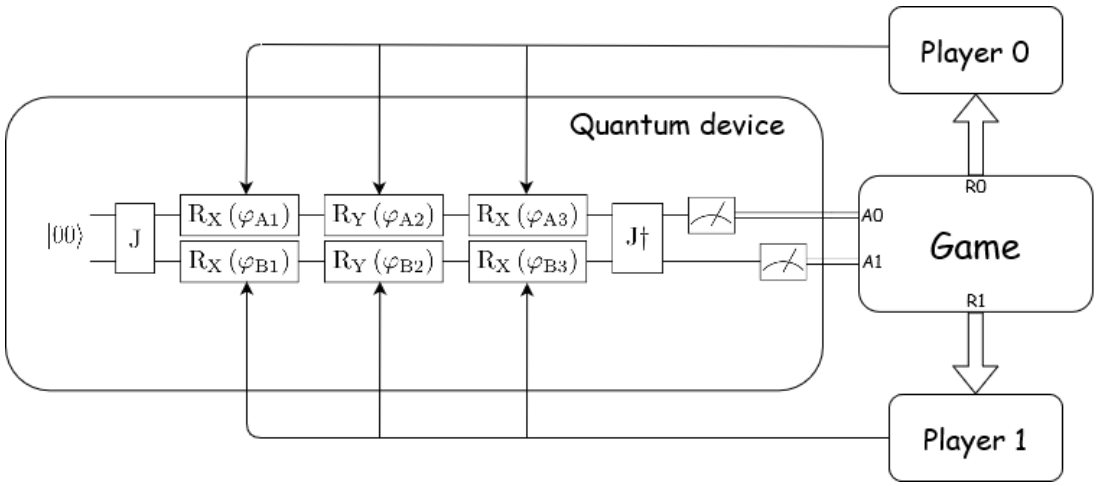


Figure 3: Two agents learning in Quantum Games model.

represented by 3 angles: φ_1, φ_2 and φ_3, which could be in the range of $[0 : 2\pi)$. As previously described in [1], we will discretize the strategy space and create a vector where each element corresponds to a set of angles: $S_i = [\varphi_1; \varphi_2; \varphi_3]$. Therefore, this vector will map the strategy space with the PDF updated in every iteration of the algorithm. In Figure 3, a block diagram of the QESRL algorithm with two learning agents is visualized. The full description of the QESRL exploration and synchronization phases, adapted from the classical ESRL, is described in the rest of this section.

In algorithm 1 it is possible to observe the pseudo-code of the exploration phase of the QESRL, where: α is the learning rate. PDF is a vector representing the probability density function on all the available actions of the player. *Actions* is a vector with all actions currently selected for each player. *Rewards* is a vector with all currently received rewards from each player. *dones* is a vector with a flag indicating if each agent has converged.

In algorithm 2 it is possible to observe the pseudo-code of the synchronization phase of the QESRL, where: *Actions* is a vector with all final actions of each player. *Rewards* is a vector with all final rewards for each player. Finally, $Hist$ is a matrix of N rows and 2 columns, each row representing a different player. Each row has 2 columns, the first representing the preferred joint action in equilibrium and the second its corresponding reward.

4. Results

This section presents two studies performed using the QESRL algorithm. Section 4.1 compares the performance of two agents playing three different games in its classical versus quantum version. Section 4.2 studies how the performance and fairness of agents playing a particular game grow as a function of the number of agents applying plain QRL versus QESRL.

4.1. Classical vs Quantum games

The results of agents utilizing the QESRL algorithm for the Battle of the Sexes, the Prisoner's Dilemma, and the Platonia both in its classical and quantum versions are presented in Figure 4. There are two worth noting considerations before continuing with the results: 1) the vertical black lines in the plots represent the changes in the phases of the algorithm, the last phase always corresponding to the exploitation phase, which will remain indefinitely; 2) the difference between the classical and quantum games is set by modifying the γ parameter of the $J(\gamma)$ operator, $\gamma = 0$ means classical game and $\gamma = \frac{\pi}{2}$ means quantum game. The learning rate was set to $\alpha = 0.001$ for all cases.

Algorithm 1 Exploration phase of QESRL algorithm

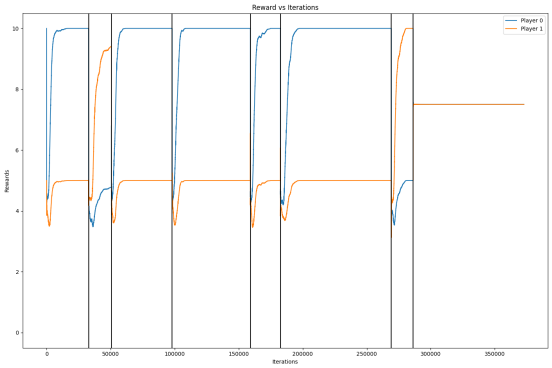
Require: Matrix Game Definition
Require: N ▷ Number of players
Require: A ▷ Number of actions
$\alpha \leftarrow 0.001$ ▷ Learning Rate
$t_{max} \leftarrow 100000$ ▷ Maximum iterations
 # Initialization
for $i = 1$ **to** N **do** ▷ For each agent
Create Agent i ▷ Initialize its PDF uniformly
$Actions[i] \leftarrow sample\_action(PDF)$ ▷ Sample PDF to determine next action
end for
$Rewards \leftarrow Game(Actions)$ ▷ Get all players rewards
 # Main loop
while $(!done\ or\ t_{max})$ **do**
for $i = 1$ **to** N **do** ▷ For each agent
$PDF \leftarrow (1 - \alpha * Rewards[i]) * PDF$ ▷ Update 1 of PDF
$PDF[Actions[i]] \leftarrow PDF[Actions[i]] + \alpha * Rewards[i]$ ▷ Update 2 of PDF
$Actions[i] \leftarrow sample\_action(PDF)$ ▷ Sample PDF to determine next action
$dones[i] \leftarrow done\_checking(PDF)$ ▷ Check agent convergence
end for
$Rewards \leftarrow Game(Actions)$ ▷ Get all players rewards
$done \leftarrow AND(dones)$ ▷ Check if all agents converged
end while

Algorithm 2 Synchronization phase of QESRL algorithm

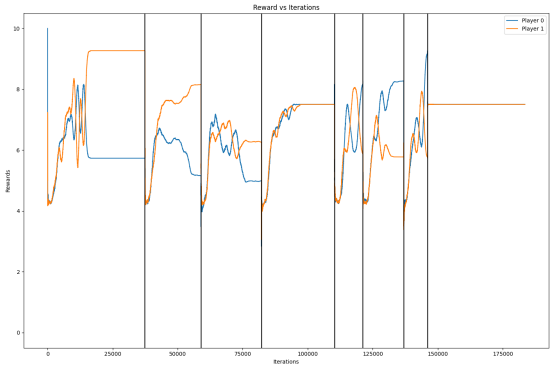
Require: Hist: A list where each agent will store its preferred equilibria and its respected reward.
for $i = 1$ **to** N **do** ▷ For each agent
if $Rewards[i] > Hist[i][1]$ **then**
$Hist[i][0] \leftarrow Actions$
$Hist[i][1] \leftarrow Rewards[i]$
end if
end for

The "Battle of the Sexes" (BoS) depicts coordination hurdles as two players select between actions "A" and "B" to achieve a shared goal, as depicted in table 1b. Differing payoff preferences underscore decision complexities, highlighting challenges in attaining consensus amid distinct priorities. In figure 4a it is possible to observe how two agents playing the classical version of the BoS converge to a reward of $R_{1,2} = 7.5$. In particular, these results confirm that the classical version of QESRL yields the same results as the ESRL algorithm reported in [2]. The BoS game has three Nash equilibriums with rewards: $[R_1, R_2] = [10, 5]$ (pure), $[R_1, R_2] = [5, 10]$ (pure), and $[R_1, R_2] = [3.84, 3.84]$ (mixed). This means that the rewards in the equilibrium achieved by the ESRL (and QESRL) algorithm ($[R_1, R_2] = [7.5, 7.5]$) are higher than the other equilibrium with evenly distributed rewards ($[R_1, R_2] = [3.84, 3.84]$) and have the same total rewards as the other two equilibriums ($[R_1, R_2] = [10, 5]$ and $[R_1, R_2] = [5, 10]$) but now with evenly distributed rewards. The combination of maximum performance and fairness is what makes these algorithms so fascinating.

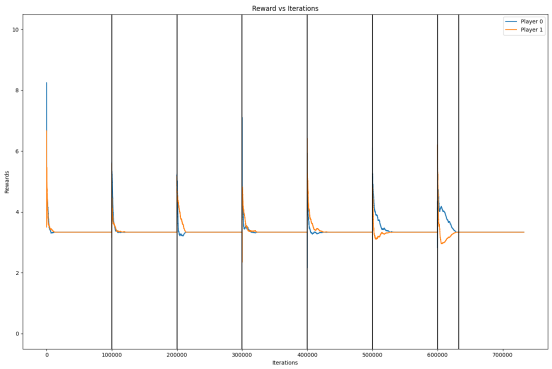
In Figure 4b it can be seen how the rewards of the two agents using QESRL with (quantum)



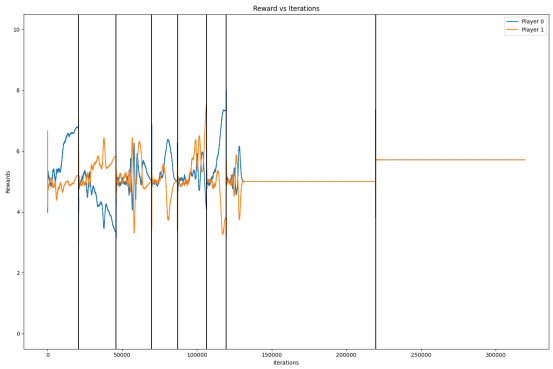
(a) Classical Battle of the Sexes.
QESRL Rewards: Player 0 = 7.5000. Player 1 = 7.5000.



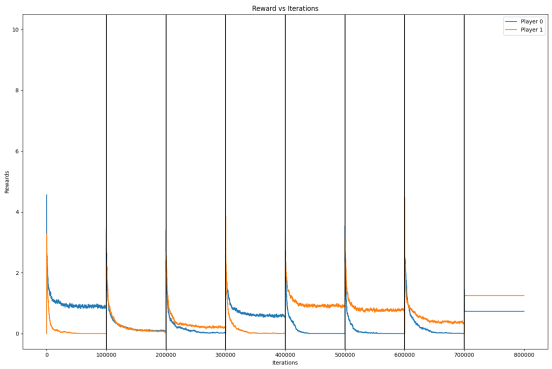
(b) Quantum Battle of the Sexes.
QESRL Rewards: Player 0 = 7.5000. Player 1 = 7.5000.



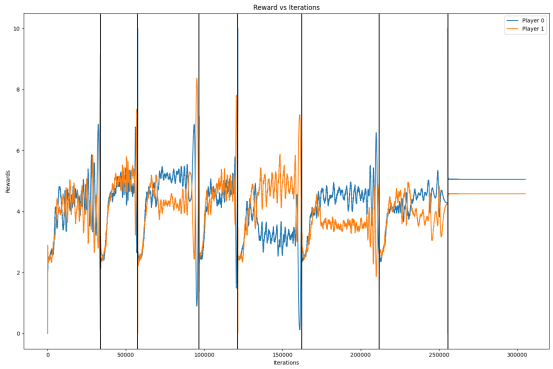
(c) Classical Prisoner's Dilemma.
QESRL Rewards: Player 0 = 3.3333. Player 1 = 3.3333.



(d) Quantum Prisoner's Dilemma.
QESRL Rewards: Player 0 = 5.7113. Player 1 = 5.7113.



(e) Classical Platonia Game.
QESRL Rewards: Player 0 = 0.7322. Player 1 = 1.2500.



(f) Quantum Platonia Game.
QESRL Rewards: Player 0 = 5.0537. Player 1 = 4.5802.

Figure 4: Agents using QESRL algorithm in Classical and Quantum games with 2 players.

entanglement successfully converge the rewards $R_{1,2} = 7.5$. Although it is possible to observe that the dynamics of the rewards of the players before the convergence are more complicated than the classical version (due to their quantumness), both agents achieve a reward identical to

the classical version.

The "Prisoner's Dilemma" (PD) exemplifies cooperation conflicts. Players choose to cooperate or defect, highlighting the tension between self-interest and mutual gain, illustrating the challenges in incentivizing cooperation with conflicting motives as depicted in Table 1c. Figure 4c shows the convergence of the classical QESRL when agents play the Prisoner's Dilemma. It is important to note that the QESRL algorithm converges to the Nash equilibrium, $[R_1, R_2] = [3.33, 3.33]$. Since the classical version of PD only has one Nash equilibrium, it matches with the equilibrium of the QESRL algorithm, again verifying its correct operation.

The rewards of the agents playing the quantum version of the prisoner's dilemma can be observed in Figure 4d. In that graph, it is possible to visualize the proper convergence of the algorithm and to confirm that the rewards of the agents using entanglement $[R_1, R_2] = [5.71, 5.71]$ (quantum game) are higher than those who do not use it $[R_1, R_2] = [3.33, 3.33]$ (classical game).

Finally, The "Platonia Game" (PG) involves N participants pursuing a substantial reward. A single person who claims the prize gets it all, while multiple or none result in no reward for anybody, as shown in table 1d. This scenario unveils the challenges of cooperative decision-making. In Figures 4e and 4f it is possible to observe the rewards of agents playing the same game (PG) using QESRL with or without quantum entanglement. In this particular game, we can observe both the successful convergence of the algorithm and the clear advantage of the quantum setup versus the classical one: $[R_1, R_2] = [5.05, 4.58]$ vs. $[R_1, R_2] = [0.73, 1.25]$, respectively.

4.2. Performance vs Fairness

In this section, we are going to analyze how the QESRL algorithm scales when the number of agents in the game increases. The Platonia Game is very easily extendable to N players by using the following reward table:

Actions		Rest of the N-1 agents	
		Everyone selects 0	At lease one selects 1
Agent i	0	0	0
	1	10	0

Table 1: Rewards of Agent i for the N players Platonia Game.

The games will be played in a quantum setting. The learning rate will be set to $\alpha = 0.001$ and the maximum number of iterations before stopping the simulation when no convergence is achieved will be $t_{max} = 100000$ for all cases. This last specification is very conditional, especially for larger games. As the number of agents increases, the number of iterations required to reach equilibrium also increases. Therefore, it is expected to observe a decay in agent performance as the number of agents grows. However, we will adopt this value because the simulation time increases exponentially as the number of agents grows. This is due to two reasons: 1) more learning agents means updating more probability density functions in each iteration, and 2) the size of the quantum circuit to simulate in each iteration also increases. Otherwise, the simulation would become exponentially more expensive, requiring exponentially more iterations with exponentially longer runtimes for each iteration. In addition, maintaining a fixed number of maximum iterations ensures consistency in the analysis.

Having said that, in Figure 5 (left) it is possible to observe how the performance of plain QRL (which means only the exploration phase of the QESRL algorithm) and QESRL itself evolves as the number of agents increases. Importantly, the performance of the players is comparable

for both algorithms. However, when we look at fairness among the agents' reward distribution, the results are remarkably different.

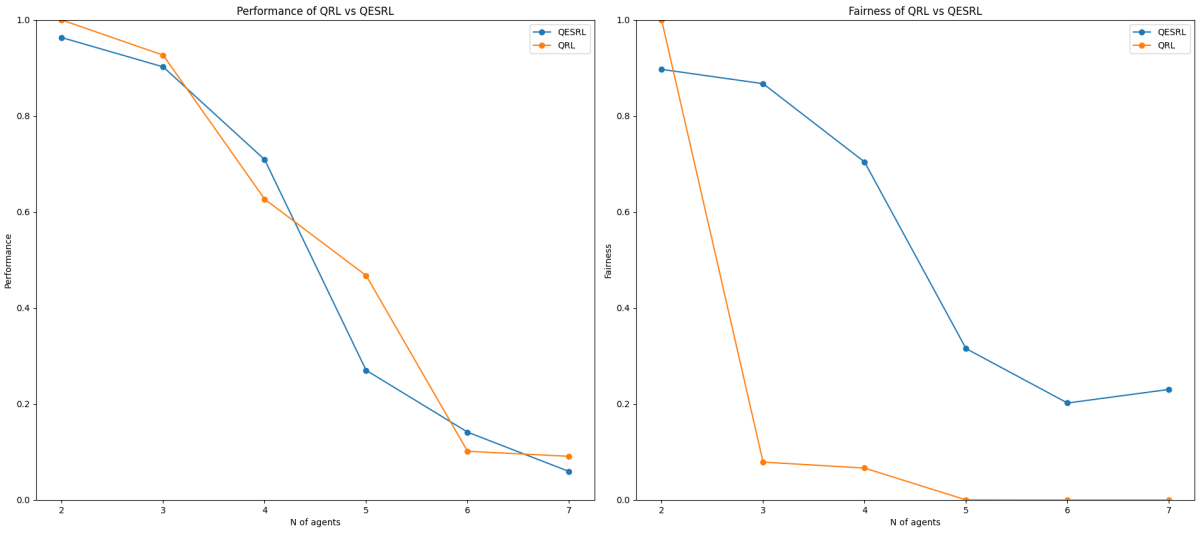


Figure 5: Performance and Fairness of rewards of N agents playing the Platonia Game.

The Gini coefficient, a crucial measure in game theory, evaluates fairness in reward distribution among agents. Ranging from 0 to 1, it signifies equilibrium to imbalance. Computed from the Lorenz curve, it quantifies the disparity between the curve and the ideal fairness line, providing insights into equity within mathematical and physical contexts. The Gini coefficient formula for a set of N rewards is r_i, $G = \frac{2\sum_{i=1}^{N} i r_i}{N\sum_{i=1}^{N} r_i} - \frac{N+1}{N}$. In Figure 5 (right) it is possible to observe the fairness between players $(1-G)$ as a function of the number of agents. In this graph, it is possible to observe that although the performance of the agents is almost the same between plain QRL and QESRL, the fairness of the distribution of rewards among agents is significantly higher for QESRL.

5. Conclusions

This study introduces a groundbreaking approach to address the convergence and optimality of strategies in non-zero-sum games across both classical and quantum settings. The novel QESRL algorithm seamlessly merges the well-established ESRL framework with the complex landscape of quantum game theory, ushering in a new era of exploration and understanding in multi-agent decision-making scenarios.

We investigated the QESRL algorithm in both classical and quantum versions of the Battle of the Sexes, the Prisoner's Dilemma, and the Platonia Dilemma. In the classical Battle of the Sexes, QESRL was consistent with the ESRL outcomes, but when adapted to a quantum setup, it showed more complex convergence dynamics while still achieving comparable rewards. In the classical Prisoner's Dilemma, QESRL mirrored the Nash Equilibrium, but when quantum entanglement is incorporated into the game, QESRL algorithm yielded even higher rewards for players. In the Platonia game, both classical and quantum versions of QESRL converged successfully, but the quantum setup shows a clear advantage, achieving significantly higher rewards. Our evaluation of QESRL across these scenarios and classical-to-quantum transitions has provided foundational insights into its versatility.

In addition, the scalability analysis presented here demonstrates the robustness of the QESRL algorithm to address increasingly complex scenarios with a growing number of agents. On the one hand, plain QRL and QESRL have comparative agent performance in terms of rewards. On the other hand, QESRL has a much higher level of fairness in the reward distribution among agents.

In conclusion, the QESRL algorithm serves as an innovative bridge connecting the classical and quantum decision-making domains. Its robustness, scalability, and capacity to promote fairness within strategic interactions position it as a powerful tool for future research. As the scientific community embarks on the next phase of exploring the intricate interplay between decision theory, quantum mechanics, and reinforcement learning, the QESRL algorithm emerges as a pivotal stepping stone toward unraveling the complexities of strategic interactions across diverse domains.

Acknowledgment
We would like to express our gratitude to the IAEA for providing funding through the SANDWICH TRAINING EDUCATIONAL PROGRAM (STEP) to allow research at the Abdus Salam International Centre for Theoretical Physics.

References

[1] Agustin Silva, Omar Gustavo Zabaleta, and Constancio Miguel Arizmendi. Learning mixed strategies in quantum games with imperfect information. *Quantum Reports*, 4(4):462–475, 2022.
[2] Katja Verbeeck, Ann Nowé, Johan Parent, and Karl Tuyls. Exploring selfish reinforcement learning in repeated games with stochastic rewards. *Autonomous Agents and Multi-Agent Systems*, 14:239–269, 2007.
[3] Tim Roughgarden. Algorithmic game theory. *Communications of the ACM*, 53(7):78–86, 2010.
[4] Jens Eisert, Martin Wilkens, and Maciej Lewenstein. Quantum games and quantum strategies. *Physical Review Letters*, 83(15):3077, 1999.
[5] Faisal Shah Khan, Neal Solmeyer, Radhakrishnan Balu, and Travis S Humble. Quantum games: a review of the history, current state, and interpretation. *Quantum Information Processing*, 17(11):1–42, 2018.
[6] Omar Gustavo Zabaleta, Juan Pablo Barrangú, and Constancio M Arizmendi. Quantum game application to spectrum scarcity problems. *Physica A: Statistical Mechanics and its Applications*, 466:455–461, 2017.
[7] Agustin Silva, Omar G Zabaleta, and Constancio M Arizmendi. Mitigation of routing congestion on data networks: A quantum game theory approach. *Quantum Reports*, 4(2):135–147, 2022.
[8] Edward W Piotrowski and J Sładkowski. Quantum market games. *Physica A: Statistical Mechanics and its Applications*, 312(1-2):208–216, 2002.
[9] Ali Hussein Samadi, Afshin Montakhab, Hussein Marzban, and Sakine Owjimehr. Quantum barro–gordon game in monetary economics. *Physica A: Statistical Mechanics and its Applications*, 489:94–101, 2018.
[10] Yaodong Yang and Jun Wang. An overview of multi-agent reinforcement learning from game theoretical perspective. *arXiv preprint arXiv:2011.00583*, 2020.
[11] Yoav Shoham, Rob Powers, and Trond Grenager. If multi-agent learning is the answer, what is the question? *Artificial intelligence*, 171(7):365–377, 2007.
[12] A Iqbal and AH Toor. Evolutionarily stable strategies in quantum games. *Physics Letters A*, 280(5-6):249–256, 2001.
[13] Ming Lam Leung. Classical vs quantum games: continuous-time evolutionary strategy dynamics. *arXiv preprint arXiv:1104.3953*, 2011.
[14] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.