Deep Reinforcement Learning for Auctions: Evaluating Bidding Strategies Effectiveness and Convergence

Luis Eduardo Craizer[1][a], Edward Hermann[1][b] and Moacyr Alvim Silva[2][c]
[1] *Pontifícia Universidade Católica, 22451-900, Rio de Janeiro, RJ, Brazil*
[2] *Fundação Getulio Vargas, 22250-145, Rio de Janeiro, RJ, Brazil*

Keywords: Auction Theory, Nash Equilibrium, Deep Reinforcement Learning, Multi-Agent Systems.

Abstract: This paper extends our previous work on using deep reinforcement learning, specifically the MADDPG algorithm, to analyze and optimize bidding strategies across different auction scenarios. Our current research aims to empirically verify whether the agents' optimal policies, achieved after model convergence, approach a near-Nash equilibrium in various auction settings. We propose a novel empirical strategy that compares the learned policy of each agent, derived through the deep reinforcement learning algorithm, with an optimal bid strategy obtained via an exhaustive search based on bid points from other participants. This comparative analysis encompasses different auctions, revealing various equilibrium scenarios. Our findings contribute to a deeper understanding of decision-making dynamics in multi-agent environments and provide valuable insights into the robustness of deep reinforcement learning techniques in auction theory.

1 INTRODUCTION

1.1 Problem Statement

Building on our previous work, this study delves deeper into applying deep reinforcement learning (DRL) to improve bidding strategies in various auction formats, specifically focusing on the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm.[1] Our current research seeks to empirically determine whether the optimal policies developed through agent convergence align with a *near-Nash equilibrium* [2] in these auction environments. To this end, we introduce an innovative empirical strategy that compares DRL-derived policies with optimal bidding strategies obtained by exhaustive search for bids from other participants. By examining different auction types, this study aims to evaluate the effectiveness of DRL in achieving equilibrium bidding behaviors and to provide a comprehensive analysis of its adaptability across various auction scenarios. Our findings show that the proposed evaluation method aligns with theoretical expectations of near-Nash equilibrium convergence in several auction settings. However, we also identify instances where agents stabilize without fully converging to optimal strategies, particularly in more complex environments like all-pay auctions. These cases, where agents may tend to bid zero, highlight the importance of metrics that measure deviations from equilibrium. With its challenging equilibrium structure, the all-pay auction format serves as a key motivator for developing this tool to diagnose these deviations and guide future improvements. A significant contribution of this study is the development of a diagnostic tool designed to evaluate agent behavior in auction settings, offering insights into both convergence to equilibrium strategies and deviations from them. By benchmarking agents' strategies against exhaustive search results, the tool provides a practical framework for assessing the robustness of DRL algorithms in multi-agent environments. Beyond its immediate application, this tool holds broader potential by enabling analysis of auction settings where analytical solutions for equilibrium strategies are unknown. Validating its effectiveness in auctions with established theoretical bench-

[a] https://orcid.org/0009-0001-5112-2679

[b] https://orcid.org/0000-0002-4999-7476

[c] https://orcid.org/0000-0001-6519-1264

[1] This section was written with grammatical and lexical revisions made with the help of ChatGPT-3.

[2] "Near-Nash equilibrium" refers to a situation in which the strategies of the players are close to a Nash equilibrium, meaning that while the strategies are not perfectly balanced, they are sufficiently close such that deviations would not significantly improve any player's outcome.

Craizer, L. E., Hermann, E. and Silva, M. A.
Deep Reinforcement Learning for Auctions: Evaluating Bidding Strategies Effectiveness and Convergence.
DOI: 10.5220/0013146400003890
In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 1*, pages 367-376
ISBN: 978-989-758-737-5; ISSN: 2184-433X

marks builds confidence in its applicability to more complex and less-explored scenarios. Ultimately, this approach bridges the gap between theoretical auction models and real-world applications, empowering the study of diverse auction formats in multi-agent learning contexts.

1.2 Related Work

Deep Reinforcement Learning (DRL), which combines deep learning and reinforcement learning, addresses decision-making problems without direct supervision by training agents to maximize cumulative rewards through trial and error in an environment, as described by Sutton and Barto (Sutton, 2018). Significant contributions from OpenAI and DeepMind, including tools like Gymnasium and models like DQN (Mnih et al., 2015), AlphaZero (Schrittwieser et al., 2020), A3C (Mnih, 2016), and PPO (Schulman et al., 2017), have advanced the field considerably. The evolution from single-agent to multi-agent reinforcement learning (MARL) has introduced algorithms like MADDPG and MAPPO, which address non-stationarity and partial observability challenges, showing promise in applications ranging from cooperative multi-robot systems to competitive games. Recent research in auction dynamics has extensively utilized Deep Reinforcement Learning. Studies by Kannan and Luong et al. employ computational agent simulations to explore human decision-making in auctions using DRL algorithms (Kannan et al., 2019) and (Luong et al., 2018). Gemp's work, which simulates all-pay auctions, aligns closely with our research by addressing scenarios where game-theoretic equilibrium analysis is intractable (Gemp et al., 2022). Dütting and Feng contribute to auction theory with neural networks for multi-item auctions, effectively bridging expected and empirical regret gaps (Dütting et al., 2021). Notably, Bichler's NPGA (Neural Pseudo-Gradient Ascent) algorithm estimates equilibrium in symmetric auctions and identifies equilibria in all-pay auctions, focusing on settings without explicit equilibrium functions (Bichler et al., 2021) and (Ewert et al., 2022). Bichler's work is particularly relevant as it tests deviations from neutral to risk equilibrium in human agents, paralleling our study's observations in all-pay auctions and validating DRL's applicability in complex auction environments.

2 BACKGROUND

Auctions, often depicted as glamorous events featuring rare items, actually encompass various formats and purposes.[3] These platforms facilitate the exchange of numerous goods and services, ranging from art to government bonds. Auctions can be classified according to various factors, such as the number of participants, the types of bid, the payment rules, and the nature of the auctioned items. A fundamental distinction is between private value auctions and common value auctions, based on participants' information about the items. In private value auctions, each participant has a personal subjective valuation of the item, influenced by individual preferences or private information. The winner, who submits the highest bid, typically pays an amount that may be less than their valuation, leading to diverse and strategic bidding behaviors. Conversely, in common value auctions, the item's value is consistent across all bidders but not fully known to any participant. The true value depends on external factors that affect all bidders equally, such as the potential for land development in a land auction. Participants must make informed bids based on their assessments and the available information, navigating the uncertainty of the item's true value. Our study focuses particularly on sealed-bid private value auctions, where bids are confidential, and participants aim to maximize their utility by balancing the item's perceived value against the price paid. The specific reward calculations for each auction type will be detailed in the following sections, drawing from fundamental principles outlined in authoritative texts such as (Klemperer, 1999), (Krishna, 2009) and (Menezes and Monteiro, 2008).

2.1 Algorithm Design

This research examines sealed-bid auctions that involve a single item. Here, the auctioneer determines the winning bid from the participating agents N. We conduct n auction rounds to observe the agents' behavior and learning patterns, seeking convergence in their bids for each given value or signal over time. Each player i has a value v_i for the item. In private value auctions, these values differ among participants, while in common value auctions, all values are equal ($v_1 = v_2 = ... = v_N = v$). The profit function for each agent is defined based on their bids: $\pi_i : \mathcal{B} \rightarrow \mathbb{R}$, where $\mathcal{B}$ is the vector space of possible bids $b = (b_1, \ldots, b_N)$ of all agents. For example, in a sealed first-price auction of private values, a (risk-neutral) participant i's profit function is:

$$\pi_i(b_i, b_{\_i}) = \begin{cases} v_i - b_i & \text{if } b_i > \max(b_{\_i}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

[3] This section was written with grammatical and lexical revisions made with the help of ChatGPT-3.

where $b_{\_i}$ represents the bids of other participants, excluding b_i.

2.2 The Rational Bid

Each participant i receives a signal s_i, representing their belief about the item's value. In private value auctions, s_i directly reflects the true value v_i for participant i. Based on this signal, participant i formulates a bid b_i. The expected payoff for participant i is given by:

$$E[u_i|s_i] = \int_B u(\pi(b_i(s_i), y)) f_{b\_i(y|s_i)} dy$$

Here, $f_{b\_i(y|s_i)}$ is the probability density function of the vector y, which contains the bids of other participants given that participant i received signal s_i. If values are independent, the signal does not affect the density function ($f_{b\_i(y|s_i)} = g_{b\_i}(y)$). Participants aim to maximize their expected reward, which requires knowledge of the function $f_{b\_i(y|s_i)}$, dependent on other players' policies.

2.3 Types of Auctions

2.3.1 First Price Auction

In a first-price auction, the participant with the highest bid wins and pays the amount of their bid. The winner's reward is the difference between the item's value and the bid amount, while the other participants receive no reward, as shown below:

$$\Pi_i = \begin{cases} v_i - b_i & \text{if } b_i > \max\limits_{j \neq i}(b_j) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where v_i is player i's valuation, b_i is their bid, and b_j are the bids of other players. We aim to determine the optimal strategy for maximizing expected profit. In a first-price auction with two risk-neutral players with private values independently and identically distributed (i.i.d) in a uniform distribution $[0, 1]$, the bids $\left(\frac{1}{2}v_1, \frac{1}{2}v_2\right)$ form a Nash equilibrium (Shoham and Leyton-Brown, 2008). The optimal bid generally follows the formula, especially for risk-neutral participants, as shown in (Krishna, 2009)

$$b_i^* = \frac{(N-1)v_i}{N}.$$

Interestingly, the optimal strategy in an English auction—a widely used format in real-world settings—is equivalent to that of a first-price auction under certain conditions. In an English auction, participants openly bid in ascending order until only one bidder remains, who then pays the highest bid. This process results in the same equilibrium bidding strategies as the first-price auction when bidders are risk-neutral and possess private values, as mentioned in (Dragoni and Gaspari, 2012). This similarity demonstrates how auction theory provides a unified framework to understand and compare different auction formats commonly used in practice.

2.3.2 Second Price Auction

Also known as a Vickrey auction, named after economist William Vickrey, the second price auction awards the item to the highest bidder, who pays the amount of the second-highest bid. The winner's reward is the difference between their valuation and the second-highest bid, as demonstrated in (Krishna, 2009):

$$\Pi_i = \begin{cases} v_i - b_2 & \text{if } b_i > \max\limits_{j \neq i}(b_j) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where b_2 is the second-highest bid. Regardless of the number of players N in this auction, agents are incentivized to bid their true valuations, reaching a Nash equilibrium where $b_i^* = v_i$ for each player i. Notably, the Dutch auction—a descending-price auction where the auctioneer lowers the price until a participant accepts it—yields the same outcomes as a first-price auction when players are risk-neutral and have private values, as described in (Frahm and Schrader, 1970). While the Dutch auction operates differently from the second-price auction, it shares similar theoretical foundations, resulting in equivalent equilibrium outcomes under certain conditions. This highlights the flexibility of auction theory in comparing various auction formats and understanding their strategic equivalences.

2.3.3 All-Pay Auction

In an all-pay auction, all participants pay their bids regardless of winning, introducing a unique strategic dimension. The highest bidder wins the item, with their reward being the difference between the item's value and bid, while other participants incur the cost of their bids. The payoff function for participant i is:

$$\Pi_i = \begin{cases} v_i - b_i & \text{if } b_i > \max\limits_{j \neq i}(b_j) \\ -b_i & \text{otherwise} \end{cases} \tag{4}$$

The Nash equilibrium strategy for risk-neutral participants in an all-pay auction, considering optimal bid calculation, is:

$$b_i^* = \frac{(N-1)}{N} v_i^N.$$

This formula captures the strategic balance of maximizing expected profit while considering the cost of bids (Riley and Samuelson, 1981).

3 METHODOLOGY

3.1 Training the Agents

Our research investigates the effectiveness of deep reinforcement learning (DRL) algorithms in learning bidding strategies for various auction scenarios.[4] DRL combines reinforcement learning principles with deep learning techniques to enable agents to learn optimal behaviours through interaction with their environment. Agents receive a state representing their current situation, act based on that state, and subsequently change the environment, which provides feedback in the form of rewards, as shown in Figure 1. Using neural networks, DRL algorithms can approximate complex value functions and policy distributions, allowing agents to handle high-dimensional state and action spaces. This flexibility makes DRL particularly suitable for dynamic environments like auctions, where strategies must adapt based on the actions of competing agents. In this study, we specifically employ the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm, a variant of the actor-critic method tailored for continuous action spaces. In MADDPG, each agent has its own actor and critic networks, but the training of the critic networks incorporates the actions and observations of all agents, reflecting the interdependent nature of multi-agent environments, as illustrated in Fig. 2. This setup is well-suited to our auction framework, where each auction round is treated as a single-iteration episode, focusing on developing optimal bidding strategies within that context. We implement a Replay Buffer to ensure stable training, storing the agents' interactions and experiences. This buffer helps mitigate the correlation between consecutive experiences, enhancing the training process. We explore various configurations, including the Combined Experience Replay Buffer, which balances historical and recent experiences, thus adapting to the evolving policies of the agents. Additionally, we introduce dynamic noise into the agents' actions to navigate the exploration-exploitation trade-off, gradually reducing it throughout the training process. This approach facilitates early exploration and later exploitation, fostering adaptability and stability in learning. The training process involves iterative learning, where agents aim to maximize their expected utility across multiple auction instances. At the beginning of each iteration, agents receive a random state, representing their private value, and select actions corresponding to bids. Rewards are assigned based on the auction's payment rules, guiding the agents in refining their policies to optimize expected utility. The actor and critic networks are designed with two layers of 100 neurons each, using sigmoid activation functions. Training parameters include a batch size of 64, an actor learning rate of 0.000025, and a critic learning rate of 0.00025, with a decrease factor of 0.99 to aid learning.

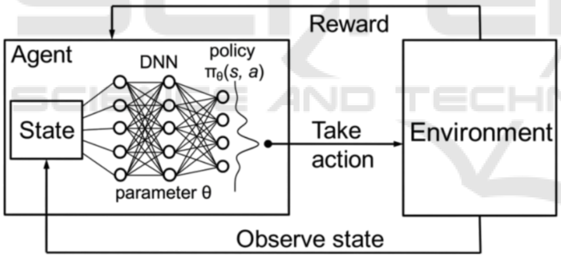


Figure 1: Deep Neural Network architecture in Reinforcement Learning.

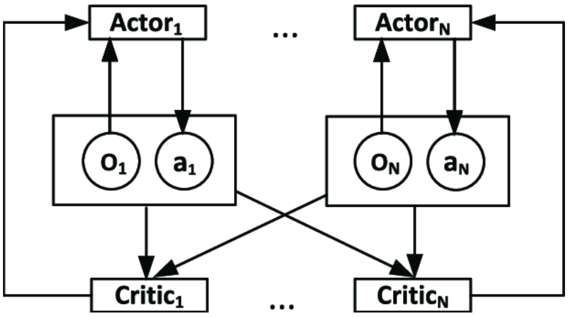


Figure 2: MADDPG Architecture - Figure taken from (Zheng and Liu, 2019).

3.2 Equilibrium Evaluation

To evaluate the effectiveness of the trained agents, we compare the optimal bids generated by the neural network models with those obtained through an exhaustive search strategy. This strategy considers the bid distributions of other participants to determine an agent's optimal bids. Specifically, we calculate the probability of an agent winning the auction by counting the times its bid is higher than the other participants. This measures how frequently the agent's bid would win the auction. Furthermore, we calculate the expected payoff for each private value by summing the expected returns for each agent. We obtain a comprehensive measure of each agent's performance

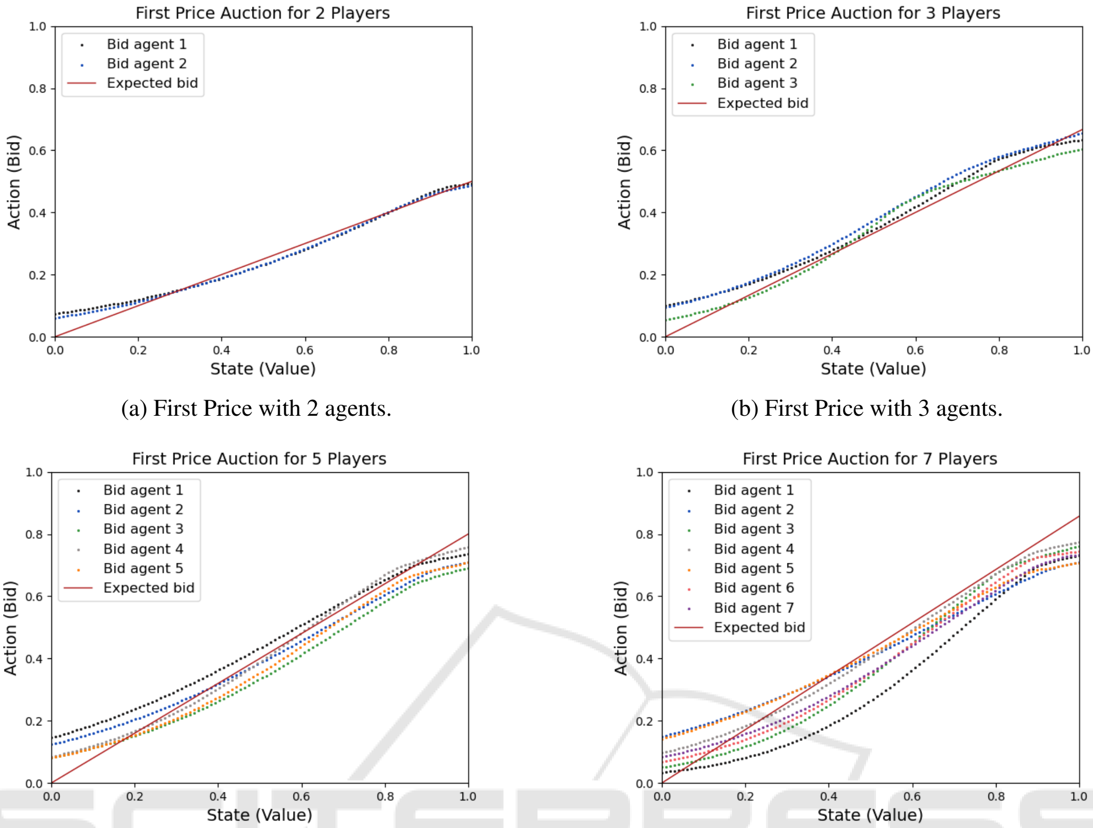
[4] This section was written with grammatical and lexical revisions made with the help of ChatGPT-3.

by integrating these probabilities into the expected return formula. We generate metrics to quantify the differences between the optimal bids produced by the neural network agents and those obtained through exhaustive searches. For precise calculations, we utilize 200 private values for each agent, evenly distributed between 0 and 1, which is adequate to capture the variations in optimal bidding strategies. A minimal difference indicates that the agent's learned policy closely aligns with the optimal bidding strategy. When this difference is sufficiently small, the agents have little incentive to deviate from their current strategy, suggesting that they are approaching a Nash equilibrium. This empirical method is designed to verify equilibrium in auction settings, evaluating the effectiveness of deep reinforcement learning algorithms in achieving equilibrium bidding behaviours. By analyzing these dynamics, we aim to gain valuable insights into the robustness and applicability of the MADDPG algorithm and potentially other deep reinforcement learning algorithms within auction theory and multi-agent environments.

4 RESULTS

In this section, we present and analyze the outcomes of our experiments, demonstrating the efficacy of our proposed method in empirically verifying near-Nash equilibrium convergence across different auction scenarios.[5] To evaluate the results, we measured both the average and maximum deviations between the optimal bids—obtained through exhaustive search—and those achieved by the neural network agents. By analyzing both metrics, we ensure that no agent significantly deviates from the equilibrium, as a high error from even a single agent would suggest an incentive for strategy adjustment. Table 1 provides an overview of the results, capturing key statistics across auction types and different agent counts. It highlights whether the agents converged to their optimal strategies and also includes both the average and maximum error observed for each scenario, giving a comprehensive understanding of the convergence performance. Figures 3, 4, and 5 depict results for first-price, second-price, and all-pay auctions, respectively, showcasing the performance across various agent counts. Each graph displays the expected optimal bid—shown as a red line—and the agents' actual bids, illustrating the extent of convergence in each case. Smaller differences between the actual bids and the red curve represent successful convergence, while larger discrepancies indicate sub-optimal performance. For instance, in the first-price auction with $N = 3$ (Fig. 3b), the optimal bid, derived from the equilibrium formula, is $\frac{2}{3}$ of the private value. The average deviation of 0.046 and maximum error of 0.051 refer to the differences between the bids generated by the neural network agents and those obtained through exhaustive search, as described in Section 3.2. These differences capture how closely the DRL model approximates the optimal strategy. For example, with a private value of 0.6, the optimal bid is 0.4, and a maximum error of 0.051 means that the actual bid would range from 0.349 to 0.451, which closely aligns with the theoretical target. This small deviation suggests that the agent's learned policy closely adheres to the Nash equilibrium, providing little incentive for strategy deviation. In contrast, in scenarios with higher maximum errors (e.g., second-price auctions with $N = 7$, as shown in Fig. 4d), agents exhibit more significant deviations from the Nash equilibrium. Here, a maximum error of 0.206 indicates a wider range of bids, far from the theoretical optimum. For example, a private value of 0.6 could lead to bids varying between 0.394 and 0.806, reflecting a significant departure from the equilibrium strategy. As the number of agents N increases, the complexity of interactions grows exponentially, making it increasingly difficult for the models to converge. This complexity leads to a higher likelihood of divergence from optimal strategies and results in more agents bidding sub-optimally, as observed in cases with larger N. Such challenges are inherent in reinforcement learning models when scaling to higher agent counts, and are typical across many heuristic and optimization methods. Table 1 further illustrates the range of outcomes, with some auction types and agent counts achieving near-perfect convergence, while others show a clear divergence. This comparison helps pinpoint which auction settings and agent configurations tend to converge reliably to equilibrium and which might need further refinement in training or strategy adaptation. Our previous research demonstrated equilibrium in multiple auction scenarios, including first-price and second-price auctions, which are relatively straightforward due to their linear optimal bidding functions. We also observed equilibrium in all-pay auctions; however, for $N > 2$, we occasionally encountered cases where agents converged to a local equilibrium rather than the global Nash equilibrium. For instance, in the case of $N = 3$, we observed both the scenario where all agents played their best responses, reaching the Nash equilibrium, and a distinct local equilibrium. In this local equilibrium, one of the agents consistently bid 0.0 for any private value, effectively opting out of the competition, while

[5] This section was written with grammatical and lexical revisions made with the help of ChatGPT-3.



(a) First Price with 2 agents.

(b) First Price with 3 agents.

(c) First Price with 5 agents.

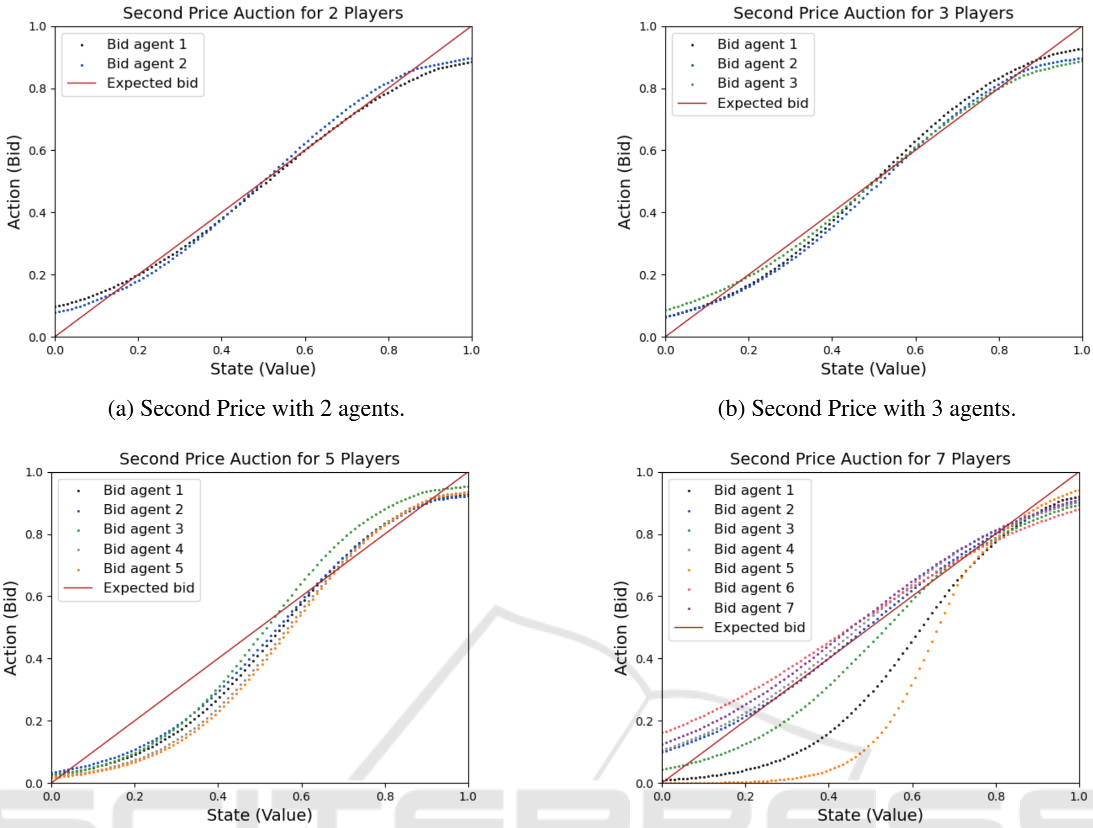
(d) First Price with 7 agents.

Figure 3: First Price Auction.

Table 1: Results.

Auction	N	Avg difference	Max difference
First Price	2	0.034	0.039
First Price	3	0.046	0.051
First Price	5	0.075	0.093
First Price	7	0.087	0.102
Second Price	2	0.024	0.025
Second Price	3	0.031	0.033
Second Price	5	0.054	0.062
Second Price	7	0.111	0.206
All-Pay	2	0.057	0.070
All-Pay	3	0.083	0.107
All-Pay	3	0.122	0.257
All-Pay	4	0.183	0.286
All-Pay	5	0.190	0.286
All-Pay	7	0.230	0.300

the other two agents followed the optimal bidding strategy for $N = 2$, as if the zero-bidding player was absent. This effectively reduced the dynamics of the game to a smaller two-player Nash equilibrium, eliminating the influence of the zero-bidder. This pattern of local equilibrium extended to larger agent counts as well. In these cases, we observed instances where some agents adhered to the Nash equilibrium strategies for smaller games, while others bid 0.0, similar to the $N = 3$ case. For example, in the case of $N = 4$, we often observed two distinct patterns: in one, two agents followed the Nash equilibrium for



(a) Second Price with 2 agents.

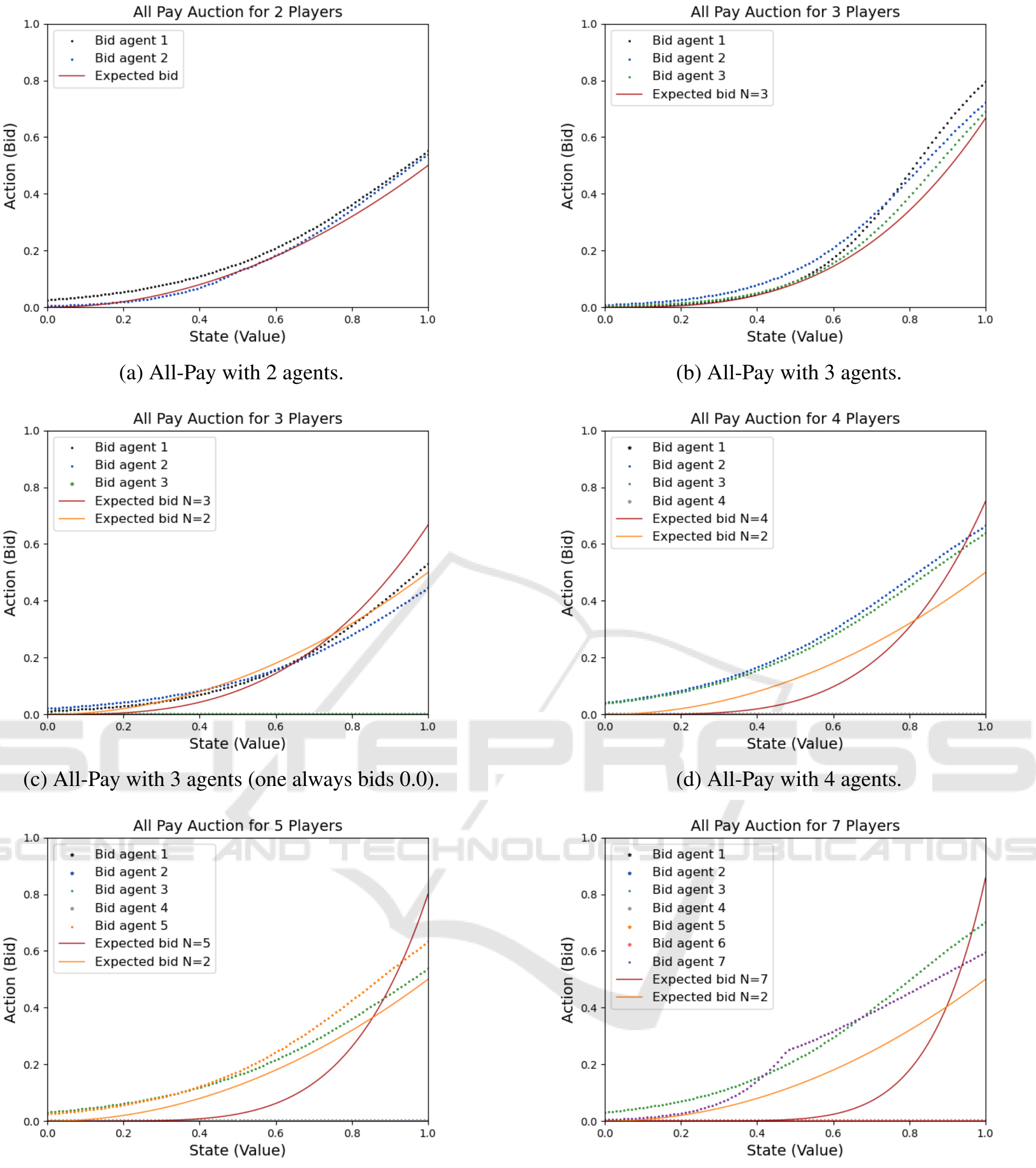
(b) Second Price with 3 agents.

(c) Second Price with 5 agents.

(d) Second Price with 7 agents.

Figure 4: Second Price Auction.

$N = 2$, while the remaining two consistently bid 0.0, effectively splitting the game into two independent two-player auctions; in another, three agents adhered to the Nash equilibrium for $N = 3$, while the last agent bid 0.0, reducing the game to a three-player interaction. Likewise, for $N = 5$, there were instances where three agents followed the Nash equilibrium for $N = 3$, with the remaining two agents bidding 0.0, leading to a division of strategy similar to the smaller cases, and so on. In the all-pay auction with $N = 3$, we obtained two distinct sets of results (Fig. 5b and Fig. 5c). In the first result, where all agents played according to the Nash equilibrium for $N = 3$, we observed an average deviation of 0.083 and a maximum error of 0.107, indicating that the agents closely adhered to the optimal strategy. Conversely, in the second result, where two agents followed the Nash equilibrium for $N = 2$ and the third agent consistently bid 0.0, the average deviation increased to 0.122 and the maximum error to 0.257. This significant deviation from the expected equilibrium for $N = 3$ is consistent with the local equilibrium pattern we discussed earlier, where one agent effectively drops out of the auction, reducing the interaction to a two-player Nash equilibrium for the remaining agents. The same phenomenon is observed in larger auctions, such as those with $N = 5$ and $N = 7$ (Figs. 5e and 5f), where many agents bid 0.0, significantly deviating from the Nash equilibrium. As mentioned earlier, as the number of participants increases, the complexity of the interactions between agents grows, leading to a higher tendency for agents to adopt sub-optimal strategies, such as zero-bid behavior, and causing a breakdown of the equilibrium strategy. Our findings show that the proposed evaluation method not only aligns with theoretical expectations but also provides a clear empirical framework to assess convergence in a variety of auction types. As the number of players increases, the average difference between the neural network-trained strategy and the exhaustive search strategy may smooth out, potentially giving a misleading impression of convergence. While using the maximum error between agents is a good starting point for evaluation, it highlights the need for more robust techniques to ensure accurate assessments in larger and more complex auction scenarios. Looking ahead, we aim to explore alternative approaches to address the challenges of local minimum convergence observed in higher N auctions.



(a) All-Pay with 2 agents.

(b) All-Pay with 3 agents.

(c) All-Pay with 3 agents (one always bids 0.0).

(d) All-Pay with 4 agents.

(e) All-Pay with 5 agents.

(f) All-Pay with 7 agents.

Figure 5: All-Pay Auction results.

One promising strategy we are currently investigating is the transfer learning approach. In this method, we initialize agents with prior knowledge of the optimal bidding strategy derived from smaller player auctions. Although the analytical formula must adapt from N to N, this initialization offers a more advantageous starting point compared to our current random initialization method. By leveraging the learned weights from previous training, we expect this approach to enhance the agents' ability to converge on optimal strategies more effectively in larger auction scenarios.

5 DISCUSSION

The goal of this study is to empirically validate whether agents trained using deep reinforcement learning (DRL) can converge to near-Nash Equilibrium strategies across different auction scenarios.[6] Our method demonstrates strong results for simpler auction types like first-price and second-price auctions, where agent behaviours closely align with theoretical expectations. In these cases, agents effectively learn optimal bidding strategies, evidenced by the small errors between learned policies and the analytical benchmarks. Crucially, the tool we develop excels at detecting errors in complex scenarios, such as all-pay auctions with multiple participants, where deviations from optimal strategies are more common. For example, when agents persist in submitting zero bids for any private value, the tool identifies significantly higher errors, signalling a clear divergence from equilibrium. These results align with human behaviour in similar experimental settings, such as those described in (Dechenaux et al., 2015), where multiple participants bid zero, as illustrated in Figure 6. This raises an important question: are the agents behaving rationally, given that they lack knowledge of the optimal bidding strategies for all-pay auctions? Without clear guidelines on bid optimality, agents appear to adopt a conservative strategy, opting to bid 0.0 to minimize potential losses. This mirrors the behaviour observed in human experiments, where, although some individuals may take higher risks and bid more aggressively, the majority gravitate toward bids close to 0.0. The tool's capacity to detect such behaviours highlights its strength in providing diagnostic insights into learning failures, offering an empirical means to evaluate whether agents adhere to equilibrium strategies. A limitation of our current approach is that it does not address correlated bids between players. This assumption of independence becomes problematic in real-world auctions, where players' bids may depend on one another due to strategic considerations or shared information. Addressing these bid dependencies becomes critical for maintaining computational efficiency as the number of agents increases. However, the main advantage of the tool lies in its ability to validate convergence in auction types with known Nash equilibrium, serving as a robust benchmark. This capability gives us confidence to extend the tool's application to auctions where no analytical solution exists for optimal bidding. By first verifying the tool in settings with established equilibrium strategies, we lay the groundwork for applying it to more complex and less understood auction formats, ultimately broadening its applicability to diverse multi-agent environments.

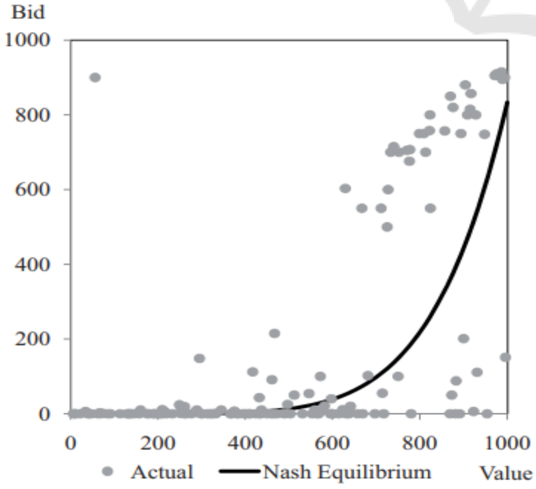


Figure 6: Results from human experiments in All-Pay Auctions - Data extracted from (Noussair and Silver, 2006).

6 CONCLUSIONS

This study presents an empirical method to evaluate whether multi-agent reinforcement learning algorithms can converge to near-Nash equilibrium strategies in auction settings.[7] While other DRL algorithms may also be applied, we chose the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm for this study due to its effectiveness in continuous action spaces and its ability to handle the interdependent nature of multi-agent environments. Our approach is validated in simple auction types, where agents demonstrated effective convergence, and more challenging auctions, like all-pay auctions, where errors were larger, and deviations from optimal strategies were more frequent. A key contribution of this work is the development of a tool that not only identifies when agents align with equilibrium strategies but highlights divergence, offering a clear diagnostic of multi-agent learning outcomes. By comparing agents' strategies against exhaustive search benchmarks, we have provided a practical framework for verifying the robustness of DRL algorithms in auction environments. Future work will extend this method to more complex auction formats, such as multi-unit and multi-stage auctions, where no analytical solution exists for the optimal bid. Additionally, we will investigate the incorporation of risk aversion and strategic dependencies between bids to enhance the model's applicability. To address the challenges of local minimum convergence observed in higher N auctions, one promising strategy we are currently exploring is

[6] This section was written with grammatical and lexical revisions made with the help of ChatGPT-3.

[7] This section was written with grammatical and lexical revisions made with the help of ChatGPT-3.

the transfer learning approach, where agents will be initialized with prior knowledge of optimal bidding strategies derived from smaller player auctions. This initialization will provide a better starting point than random initialization, as the agents will inherit the learned weights from previous training. An important direction for future research is the application of this methodology to scoring auctions, which have significant practical implications, particularly in the Brazilian context. For example, scoring auctions have been used in Brazil to allocate oil exploration rights, as detailed in (Sant'Anna, 2017). In these auctions, bidders submit multidimensional bids, including a monetary bonus and an exploratory program, with a nonlinear scoring rule determining the winner. This format introduces unique challenges and opportunities for modeling and evaluation, as estimating the distribution of primitive variables—such as tract values and exploration commitment costs—enables counterfactual analysis of revenue under alternative bidding schemes. By adapting our tool to this context, we aim to explore its ability to handle the complexities of multidimensional scoring rules and assess its utility in evaluating and optimizing such auction mechanisms. Addressing these complexities will be crucial to advancing our understanding of multi-agent dynamics and improving auction design.

REFERENCES

Bichler, M., Fichtl, M., Heidekrüger, S., Kohring, N., and Sutterer, P. (2021). Learning equilibria in symmetric auction games using artificial neural networks. *Nature machine intelligence*, 3(8):687–695.

Dechenaux, E., Kovenock, D., and Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 18:609–669.

Dragoni, N. and Gaspari, M. (2012). Declarative specification of fault tolerant auction protocols: The english auction case study. *Computational Intelligence*, 28(4):617–641.

Dütting, P., Feng, Z., Narasimhan, H., Parkes, D. C., and Ravindranath, S. S. (2021). Optimal auctions through deep learning. *Communications of the ACM*, 64(8):109–116.

Ewert, M., Heidekrüger, S., and Bichler, M. (2022). Approaching the overbidding puzzle in all-pay auctions: Explaining human behavior through bayesian optimization and equilibrium learning. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 1586–1588.

Frahm, D. G. and Schrader, L. F. (1970). An experimental comparison of pricing in two auction systems. *American Journal of Agricultural Economics*, 52(4):528–534.

Gemp, I., Anthony, T., Kramar, J., Eccles, T., Tacchetti, A., and Bachrach, Y. (2022). Designing all-pay auctions using deep learning and multi-agent simulation. *Scientific Reports*, 12(1):16937.

Kannan, K. N., Pamuru, V., and Rosokha, Y. (2019). Using machine learning for modeling human behavior and analyzing friction in generalized second price auctions. *Available at SSRN 3315772*.

Klemperer, P. (1999). Auction theory: A guide to the literature. *Journal of economic surveys*, 13(3):227–286.

Krishna, V. (2009). *Auction theory*. Academic press.

Luong, N. C., Xiong, Z., Wang, P., and Niyato, D. (2018). Optimal auction for edge computing resource management in mobile blockchain networks: A deep learning approach. In *2018 IEEE international conference on communications (ICC)*, pages 1–6. IEEE.

Menezes, F. and Monteiro, P. (2008). An introduction to auction theory: Oxford university press.

Mnih, V. (2016). Asynchronous methods for deep reinforcement learning. *arXiv preprint arXiv:1602.01783*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533.

Noussair, C. and Silver, J. (2006). Behavior in all-pay auctions with incomplete information. *Games and Economic Behavior*, 55(1):189–206.

Riley, J. G. and Samuelson, W. F. (1981). Optimal auctions. *The American Economic Review*, 71(3):381–392.

Sant'Anna, M. C. B. (2017). Empirical analysis of scoring auctions for oil and gas leases.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., et al. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Sutton, R. S. (2018). Reinforcement learning: An introduction. *A Bradford Book*.

Zheng, S. and Liu, H. (2019). Improved multi-agent deep deterministic policy gradient for path planning-based crowd simulation. *Ieee Access*, 7:147755–147770.