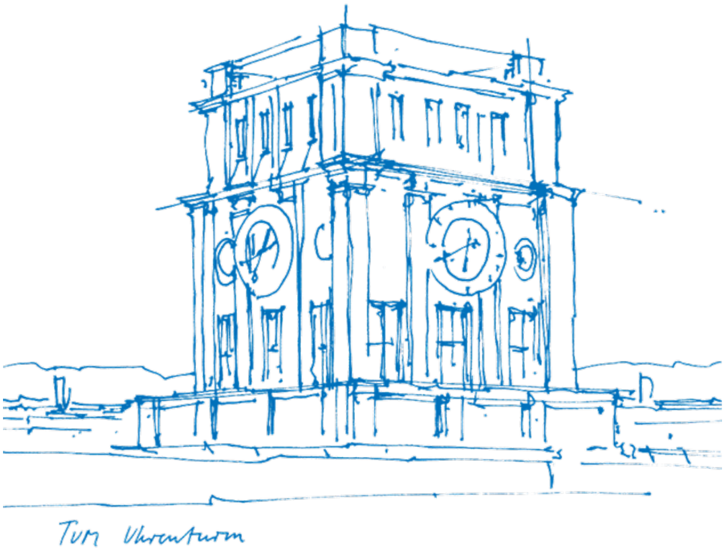
Decision Sciences & Systems
TUM School of Computation, Information and Technology
Technical University of Munich

Learning Continuous and Pure Bayes-Nash Equilibria in Sealed-Bid Auctions

Stefan Heidekrüger



Learning Continuous and Pure Bayes-Nash Equilibria in Sealed-Bid Auctions

Stefan Heidekrüger

TUM School of Computation, Information and Technology
Technische Universität München

Learning Continuous and Pure Bayes-Nash Equilibria in Sealed-Bid Auctions

Stefan Heidekrüger

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Rüdiger Westermann

Prüfer der Dissertation:

1. Prof. Dr. Martin Bichler
2. Prof. Dr. Stephan Günnemann
3. Prof. Kemal Göler, Ph.D.

Die Dissertation wurde am 27.02.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 20.11.2023 angenommen.

To my late grandmother Anna.

Abstract

Economists commonly study markets in their equilibrium state and the properties in equilibrium drive the decisions of market participants and policymakers alike. In markets with asymmetric information, auctions can serve as mechanisms to determine the allocation and prices of traded goods based on the information that bidders choose to share with the auctioneer about their willingness to pay. Auctioneers, like governments allocating wireless spectrum or companies selling advertising inventory, are therefore particularly interested in designing auction mechanisms that achieve certain desiderata, such as allocative efficiency or revenue maximization, assuming bidders act in their own economic interest. However, closed-form characterizations of the equilibrium bidding strategies of such markets are only known for a handful of cases, and finding or even approximating the equilibria numerically is believed to be intractable in the general case.

In this dissertation, we study a learning approach to equilibrium computation in auctions, called *Neural Pseudogradient Ascent (NPGA)* and its applications. Our approach aims to find equilibria in sealed-bid auctions which are modeled as Bayesian games with continuous type and action spaces by following the anticipated gradient dynamics of such games. In NPGA, bidders' strategies are represented as neural networks. This implicitly transforms the market into a finite-dimensional complete-information game. Due to the discrete nature of resource allocation, however, auction mechanisms are inherently non-differentiable, and standard methods, i.e. backpropagation from observed outcomes, cannot be used to train the neural networks. Instead, NPGA relies on the computation of alternative gradient estimates using Evolutionary Strategies.

We theoretically study conditions where such dynamics can be expected to converge to pure-strategy Bayesian Nash equilibria (BNE). Most notably, we show convergence to at least local equilibria under the standard assumption of symmetric auctions. Empirically, we study NPGA in a wide range of auctions, including combinatorial and multiunit auctions. In fact, our study includes all auction settings with previously established analytical closed-form BNE that could be identified in the literature. We find that despite the computational hardness of the equilibrium computation problem, NPGA is able to recover an analytical BNE in all studied settings. This suggests that many common auction formats exhibit additional structure which makes equilibrium computation tractable in practice. NPGA has superior scalability properties to previously described methods and we were also able to compute an approximate BNE in the largest setting where this has been achieved so far, a combinatorial auction with seven bidders and six items.

In applications, we leverage NPGA's novel capabilities as a black-box equilibrium computation tool. For example, we study comparative statics in auction markets where analytical equilibria are yet unknown. Additionally, we develop a novel method for parameter identification in utility functions with an application to the behavioral economics of all-pay auctions.

NPGA leverages massive parallelization via GPU hardware acceleration. As part of this dissertation project, we further developed `bnelearn`, an open-source framework for learning in auctions that contains the largest existing suite of GPU-enabled auction implementations for simulation.

Acknowledgements

There are many people without whom this dissertation project would not have been possible. To them, I would like to express my sincerest gratitude:

My advisor Martin Bichler for his supervision, guidance, approachability, and wisdom, the very close collaboration, and the opportunity to pursue this project.

The members of my dissertation committee for their interest, time and effort.

My coauthors, Paul Sutterer, Nils Kohring, Max Fichtl, and Markus Ewert, for the always constructive and productive collaboration and their ideas and hard work.

All my colleagues at the DSS chair over the last four years, for the great atmosphere, stimulating discussions, fun hours outside of work, and the countless geolocations and bomb defusals. In particular, I would like to thank Stefan Waldherr and my academic mentor Anaëlle Wilczynski for helping me navigate the less academic matters of academia, and Marianne Thanner, Anja Keller and Oliver Jacksch for keeping countless bureaucratic and technical issues off my plate.

My students for sharing their interesting projects with me and teaching me how to be a better mentor. I'd especially like to thank Kevin Falkenstein, Anne Christopher, Iheb Belgacem, Gleb Kilichenko, and Calin Buzetelu, who in their student projects made code contributions to the `bnelearn` software package, or produced valuable insights for our research group.

The Mäandertal and alleinschon crews for their invaluable and loyal friendship, and my bandmates from 203 for helping me clear my mind after long days of research.

My parents Inge and Ralf, for always believing in me and being there for me in all my endeavors. Jochen for all his support and tolerance over the years.

My wife Sharon for being the love of my life, making me laugh every day, making me a better person and for all the sacrifices she has made to support this dissertation project. Thank you!

My daughter Anna for making me feel more love, happiness and pride than I could have ever imagined. For lightening up my long weekends working on this dissertation with your smile, and for sleeping just enough to allow me to finally finish it. I can't wait to continue to see you grow up!

Contents

1 Introduction

Auctions serve an increasingly important function in markets with few participants and asymmetric information. With the ascent of technology-enabled and algorithmically driven marketplaces, auctions are increasingly employed at scale in the real economy. On the one hand, this takes place in rare, but complex and high-stakes auctions, for example in the allocation of wireless spectrum licenses Bichler and Goeree (2017); Milgrom (2021). On the other hand, whole industries rely on millions of individually relatively simple and small-scale auctions every day to match buyers and sellers and determine prices, most notably in computational advertising (Ashlagi et al., 2011). Despite this growing prevalence, the competitive behavior of rational market participants in auctions is not yet fully understood. In a landmark result, the Vickrey-Clarke-Groves (VCG) mechanism (Vickrey, 1961), where winning bidders are charged prices according to the "harm" their presence in the auction causes the other bidders, has been shown to be the unique direct mechanism that is incentive-compatible while also leading to economically efficient outcomes and being individually rational. However, despite this positive result, VCG auctions are inapplicable or intractable in a wide range of settings (Ausubel et al., 2006; Rothkopf, 2007). Consequently, they are not used widely in practice. When forced to give up one of the desiderata above, market designers commonly sacrifice incentive compatibility. As a result, submitting one's true private information may no longer be in the best interest of bidders, who, in turn, are then faced with the problem of choosing a *bidding strategy*.

As economists generally reason about markets in their *equilibrium state*, understanding the equilibria of non-VCG auctions is of utmost importance to the fields of auction theory and market design. Such auctions are commonly modeled as Bayesian games, where equilibria are described by Bayesian Nash equilibria (BNE). While the analytical characterization of such BNE in single-item auctions with *independent private values*—where bidders each independently observe their true valuations of the items for sale—are relatively well understood, few closed-form results exist for markets with either multiple goods, risk-averse bidders, value interdependencies between bidders, or when bidders only have access to partial or noisy information about their own preferences. However, all of these conditions are commonly present in real-world auction markets, and thus understanding strategic behavior in such markets is paramount. Indeed, the 2020 Nobel Memorial Price in Economic Sciences honored Robert B. Wilson and Paul Milgrom for their contributions to the understanding of some of these markets (Nobel Memorial Prize, 2020). Nevertheless, a succinct description of the equilibria of most auction markets remains elusive.

What is more, BNE cannot easily be computed numerically either: The exact computational complexity of computing BNE remains unknown, but it is at least as hard as the computation of Nash equilibria in complete-information games and believed to be intractable in general (see section 2.5).

In the past decade, the "deep learning revolution" has led to remarkable breakthroughs in artificial intelligence: Deep neural networks trained via backpropagation-based stochastic gradient descent methods have shown remarkable performance in a wide range of function approximation tasks in supervised machine learning, unsupervised generative tasks, and reinforcement learning. These advances were especially fueled by data-parallel implementations of the neural network training process that, in turn, were enabled by hardware acceleration on graphics cards (GPU) and later specialized hardware (Krizhevsky et al., 2012). Peculiarly, these results were often achieved in nonconvex settings and despite a theoretical lack of convergence guarantees to global minima and the confirmed general-case hardness of the problem (Blum and Rivest, 1992). This has sparked a renewed interest in the study of learning in multi-agent settings, both in the reinforcement learning and game theory communities.

1.1 Contributions

In this thesis, we investigate a learning approach to numerically compute Bayesian Nash equilibria (BNE) in sealed-bid auctions. Such auctions are commonly modeled as Bayesian games where both the type and action spaces are continuous. As a result, even *pure* strategies are mappings from one continuous space to another, i.e. objects in an infinite-dimensional functional space. Previous numerical approaches have primarily relied on discretization of the game, either in the type space, action space, or both. While such approaches have been shown to be successful in low-dimensional games, their representation size suffers from the curse of dimensionality and they are intractable in auctions with more than a few items unless there are significant symmetries in the game.

Instead, we take an approach inspired by recent advances in reinforcement learning which looks for strategies in the original infinite-dimensional functional space via finite-dimensional function approximation. In particular, we will represent pure strategies by neural networks, although the approach equally applies to other parametric functional forms (such as Gaussian processes or linearizations on finitely many support points). With this approach, we can interpret the problem of finding equilibria in the Bayesian auction game as essentially equivalent to finding Nash equilibria in a continuous, but finite-dimensional, complete-information proxy game over neural network parameters.

Using this representation, we enable the study of the ex-ante *gradient dynamics* in the Bayesian game, which emerge when all agents iteratively make unilateral, infinitesimal strategy improvements based on local feedback in the current behavioral state of the game. Similar dynamics and their convergence behavior have been extensively studied in the literature on finite-dimensional complete information games, as well as in (single-agent) nonlinear optimization and reinforcement learning (see section 2.3). In *differentiable* games, these gradients can be efficiently computed from observed data using established methods like backpropagation. Unfortunately, standard auction formats do not have this property. As a result, nonstandard methods are necessary. A key contribution is the development of an algorithm, called *Neural Pseudogradient Ascent (NPGA)* which approximates ex-ante gradients in Bayesian games from ex-post observational data using an estimator based on *evolutionary strategy* computations, even when the game is ex-post non-differentiable. Publication A (Heidekrüger et al., 2019) discusses the problem of ex-post non-differentiability of auctions, introduces the algorithm and empirically investigates its behavior in finite complete-information games, where, as expected, it behaves similar to analytical gradient ascent, and in single-item first-price auctions, where we observe that it converges to approximate Bayesian Nash equilibria. Publication B (Bichler et al., 2021) contains a more rigorous theoretical analysis of NPGA, and proves (approximate) convergence to *local* BNE in symmetric Bayesian games—a property that is often fulfilled in auctions studied in the economic literature. Such symmetric auctions constitute potential games and can be modeled using a single shared neural network for all players. Empirically, we find convergence to *global* BNE in all studied settings. In Publication C (Bichler et al., 2023a), we additionally investigate the convergence behavior of NPGA in a wide range of *asymmetric* auctions with multiple homogenous or heterogeneous items. We empirically find that the algorithm likewise recovers global equilibria in all studied settings, although this behavior comes without a theoretical guarantee. In original research not included in this dissertation (Heidekrüger et al., 2021c), we also establish a complementary result showing that *monotonicity* of a Bayesian game is a sufficient criterion for convergence of NPGA to the (unique) global BNE, but unfortunately this criterion is hard to verify even in simple auctions.

We argue that the availability of robust numerical equilibrium solvers for continuous Bayesian games will unlock new possibilities for empirical economics research. For example, in Publication B and Publication C, we perform comparative statics in small combinatorial auctions under approximate equilibria computed via NPGA, allowing us to quantitatively investigate the sensitivity of relevant market metrics (like seller revenue or efficiency) to its input parameters (like assumptions on prior distributions, correlation structure, or risk sensitivity of bidders).

Moreover, equilibrium computation techniques may have further applications in behavioral economics: For example, it is a well-known fact, that in real-world settings and lab experiments concerning *all-pay auctions*, human bidders do not follow the analytical equilibria that would result for rational and risk-neutral bidders. Dropping the assumption of risk neutrality, one may hypothesize that players are indeed following an equilibrium but in a market that is dictated by risk-aware utility functions.[1] In Publication D, we present a novel statistical inference method that uses equilibrium computation via NPGA combined with Bayesian Optimization in order to identify parametric utility functions in such a setting.

To facilitate the research in these publications, we have developed a software framework called `bnelearn` for equilibrium learning in auctions (Heidekrüger et al., 2021a). The library is based on `pytorch` (Paszke et al., 2017) and contains a learning algorithm agnostic simulation framework that leverages GPU-hardware acceleration to efficiently compute large numbers of auctions in parallel. To the author's knowledge, `bnelearn` constitutes the largest and fastest repository of sealed-bid auction implementations (including their known equilibria). It may serve as a benchmark suite for future research and algorithm development in equilibrium learning. A non-exhaustive discussion of included auction settings is contained in section 2.5. Additionally, the software may serve as a tool for auction theorists and market design practitioners who may leverage the existing equilibrium learning capabilities to study specific auction markets in detail. `bnelearn` has been released under the GNU General Public License v3.0, and is freely available at `https://github.com/heidekrueger/bnelearn`.

The remainder of this thesis will be structured as follows: In chapter 2, we will introduce the topic, discuss related literature and place our work in its scientific context. We will then present the four original research papers introduced above. Finally, in chapter 3, we will review the most important results from these publications and other original research (Heidekrüger et al., 2021c; Kohring et al.; Heidekrüger et al., 2021b), and discuss remaining open questions and further research opportunities.

[1] Alternatively, it has also often been argued that bidders do, indeed, behave irrationally.

2 Scientific Context

In this chapter, we will introduce foundations from game theory and auction theory to motivate, define and discuss the equilibrium computation problem in auctions and the approach presented in the included publications. This text expects the reader's familiarity with basic concepts from convex optimization and machine learning, such as (stochastic) gradient descent and its convergence characteristics, or simple neural network architectures and training methods like backpropagation. We recommend Boyd and Vandenberghe (2004) and Goodfellow et al. (2016) as references. Section 2.1 introduces necessary game-theoretic concepts before we formally define auctions in section 2.2. We then discuss the concept of learning dynamics in games in section 2.3. The related field of reinforcement learning, with a particular focus on applications to game-theoretic multi-agent settings, is outlined in section 2.4. Finally, in section 2.5, we introduce our main focus in this thesis, the problem of computing Bayesian Nash equilibria in auctions, before ending the chapter with matters relevant to software implementations of such methods in section 2.6.

2.1 Game Theory

In this section, we will briefly introduce the most important game-theoretic concepts that will be relevant to the discussion of learning in sealed-bid auctions. In particular, we will start with the standard notion of complete-information games, before introducing Bayesian games with incomplete information. We will also discuss their respective standard equilibrium concepts.

Definition 2.1.1 (Complete-Information Game)**.** A *Complete Information Game* is a tuple $G = (n, \mathcal{A}, \mathbf{u})$, where

- n is the number of players participating in the game,
- $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ is the set of *action profiles*, with $\mathcal{A}_i$ being the nonempty set of *actions* available to player i, and
- $\mathbf{u} = (u_1, \dots, u_n)$ is the collection of the players' utility functions $u_i : \mathcal{A} \to \mathbb{R}$.

In a game, each player i chooses an action $a_i \in \mathcal{A}_i$. All players choose their action simultaneously.[1] We will call the vector $\mathbf{a} = (a_1, \dots, a_n)$ an *action profile*. Each player's utility function u_i then maps action profiles to the utility, or *outcome*, of the game for player i: $u_i = u_i(\mathbf{a})$. Throughout this thesis, we will assume that these utility functions are von Neumann-Morgenstern (vNM) utility functions (Von Neumann and Morgenstern, 1944), i.e. that players are rational actors and aim to maximize their own utility. Let us now introduce some possible properties of games that will be relevant throughout the discussion.

[1] *Extensive form games*, where players move sequentially, will not be considered in this thesis.

Definition 2.1.2 (Some properties of games)**.** A game $G = (n, \mathcal{A}, \mathbf{u})$ is called a

1. *finite game*, iff $n < \infty$ and $|\mathcal{A}_i| < \infty$ for all $i \in [n]$. Otherwise, G is called *infinite*. Finite complete-information games can be represented in "normal-form" as a multi-matrix of outcomes and are therefore also referred to as *Normal-Form games*.
2. (finite-dimensional and convex) *continuous game*, iff for all players i, the actions sets $\mathcal{A}_i$ are nonempty, compact and convex subsets of $\mathbb{R}^{d_i}$ for some $d_i < \infty$.
3. *differentiable game*, iff it is continuous and additionally, all players' utility functions are continuously differentiable in their own actions, i.e. $\forall i \in [n]$ and $\forall \mathbf{a}_{-i} \in \mathcal{A}_{-i}$, we have
$$u_i(\,\cdot\,, \mathbf{a}_{-i}) \in \mathcal{C}^1(\mathcal{A}_i, \mathbb{R}) \tag{2.1}$$
In differentiable games, we will write $\nabla \mathbf{u}(\mathbf{a}) = (\nabla_{a_i} u_i(\mathbf{a}))_i$ for the concatenation of individual gradient vectors of all players.
4. *zero-sum* game, iff there is a constant $c \in \mathbb{R}$, s.t. for all action profiles $\mathbf{a} \in \mathcal{A}$, we have $\sum_i u_i(\mathbf{a}) = c$.[2]

Strategic Behavior and Nash Equilibria In games, players are primarily faced with the decision of choosing one of their available actions.

Definition 2.1.3 (Strategy)**.** Let $G = (n, \mathcal{A}, \mathbf{u})$ be a complete-information game. A probability distribution $\pi_i \in \Delta\mathcal{A}_i$ over the actions available to player i is called a *(mixed)* strategy of i.

If $\mathbb{P}_{a' \sim \pi_i}(a_i{=}a') = 1$ for some $a_i \in \mathcal{A}_i$, then π_i is called *pure*, and we write $a_i{=}\pi_i$ by abuse of notation.

Due to the fact that players may randomize their choice of actions, game outcomes may be non-deterministic even when players employ fixed strategies. To take this into account, we will often be interested in players' expected utilities rather than the a-posteriori realized utility in hindsight:

Definition 2.1.4 (Expected Utility)**.** Let $G = (n, \mathcal{A}, \mathbf{u})$ be a complete-information game. For a strategy profile $\pi = (\pi_1, \dots, \pi_n)$, we write
$$u_i(\pi) = \mathbb{E}_{\mathbf{a} \sim \pi}[u_i(\mathbf{a})] \tag{2.2}$$
for the expected utility of player i, given that all players j follow their respective strategies π_j.

We will be particularly interested in strategy profiles that lead to an equilibrium state of a game. The classical notion equilibria of such games is due to Nash (1950) and describes a state π^* where no player i can improve her own expected utility by unilaterally deviating from the current strategy profile:

Definition 2.1.5 (Nash Equilibrium)**.** Let $G = (n, \mathcal{A}, \mathbf{u})$ be a complete-information game.

A strategy-profile π^* is called a *Nash equilibrium* of G, iff for all players i, and for all strategies $\pi_i \in \Delta\mathcal{A}_i$, we have
$$u_i(\pi_i, \pi^*_{-i}) \leq u_i(\pi^*_i, \pi^*_{-i}). \tag{2.3}$$

Famously, every (finite or convex continuous) complete-information game admits at least one Nash equilibrium, potentially in mixed strategies.

[2] If $c \neq 0$, then without loss of generality, one may set $u_i^c(\mathbf{a}) = u_i(\mathbf{a}) - c$ to achieve a strategically equivalent zero-sum game with $c = 0$.

Bayesian Game Theory Complete-information games assume that any information available to *any* player, is also available to *all* other players, i.e. everything about the game is common knowledge. This model cannot accurately capture settings of strategic interaction in the presence of partial or asymmetric information. Several extensions to the complete-information setting exist. *Imperfect* information refers to the fact that players may have uncertainty about the state of nature (e.g. due to partial observability or noisy measurements), or (in extensive-form games with sequential moves) about actions that other players have already taken. Under *incomplete* information, on the other hand, players have uncertainty about other players' (and possibly even their own!) utility functions, i.e. their preferences over game outcomes. Auctions are typically modeled as games of *incomplete information* using the formulation of Bayesian games, where players' *types* capture the unique information available to them. It should be noted, however, that (a) incomplete and imperfect information are not mutually exclusive, and (b) formal models incorporating either concept may sometimes be mathematically equivalent, so the distinction may not be clear in practice and is often a matter of interpretation.

Definition 2.1.6 (Bayesian Game, Harsanyi (1968))**.** A *Bayesian game* is given by a tuple $G = (n, \mathcal{V}, \mathcal{A}, F, \mathbf{u})$, where n and $\mathcal{A}$ are defined analogously to complete information games, and

- $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is the set of *type profiles*, with $\mathcal{V}_i$ being the set of possible types for player i.
- F is a *prior probability distribution* over $\mathcal{V}$ that is common knowledge between all players. (We write F_{v_i} for the marginal distribution of the component v_i representing the type of player i.),
- $\mathbf{u} = (u_1, \dots, u_n)$ is the collection of players' individual utility functions $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$.

When all type-spaces $\mathcal{V}_i$ are finite, the game is called *finite-type*, when all $\mathcal{V}_i$ are (convex) subsets of $\mathbb{R}^{d_i}$, the game is called *continuous-type*.

Bayesian games are usually interpreted as follows: In the *ex-ante* phase, players are only aware of the information which is common knowledge, most importantly the prior type distribution F. Then, types $v \sim F$ are drawn, and each player i is informed of her own type v_i only. In this *interim* stage, players need to make a strategic decision about choosing their action $a_i \in \mathcal{A}_i$. The name "Bayesian game" refers to the fact that players may employ Bayesian belief updating about other players' information: Given information about v_i, i's information about other players' types $\mathbf{v}_{-i}$ is now best represented by the posterior $F_{\mathbf{v}_{-i}|v_i}$. Finally, in the *ex-post* phase, when players have submitted their actions $\mathbf{a}$, the outcome of the game and the achieved utilities can be observed. The literature differs on whether players may observe others' actions and utilities ex-post, or only their own individual utility u_i. The latter case will be sufficient in our context. The notions of strategies, expected utility, and Nash equilibria generalize to Bayesian games as follows:

Definition 2.1.7 (Strategies in Bayesian Games)**.** In a Bayesian game, a function $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$, is called a pure strategy of player i. We will denote by $b_i = \beta_i(v_i)$ the action chosen under β_i given type v_i.

Definition 2.1.8 (Expected Utility in Bayesian Games)**.** Let $G = (n, \mathcal{V}, \mathcal{A}, F, \mathbf{u})$ be a Bayesian game, and β_{-i} be the strategy profile of all players but i.

1. Given a realization $v_i \in \mathcal{V}_i$, and an action $a_i \in \mathcal{A}_i$, player i's *interim expected utility* $\overline{u}_i$ when playing action a_i is given by

$$\overline{u}_i(a_i; v_i, \beta_{-i}) = \mathbb{E}_{\mathbf{v}_{-i} \sim F_{\mathbf{v}_{-i}|v_i}}[u(a_i, \beta_{-i}(\mathbf{v}_{-i}))] \tag{2.4}$$

2. Given only a strategy β_i of i, player i's *ex-ante expected utility* $\tilde{u}_i$ is given by

$$\tilde{u}_i(\beta_i;\beta_{-i}) = \mathbb{E}_{v_i\sim F_{v_i}}[\overline{u}_i(\beta_i(v_i);v_i,\beta_{-i})] = \mathbb{E}_{\mathbf{v}\sim F}[u(v_i,\beta_i(v_i),\beta_{-i}(\mathbf{v}_{-i}))] \tag{2.5}$$

Definition 2.1.9 (Bayesian Nash Equilibrium)**.** Let $G = (n, \mathcal{V}, \mathcal{A}, F, \mathbf{u})$ be a Bayesian game. A strategy profile β^* is an *(ex-ante) ε-approximate Bayesian Nash equilibrium (ε-BNE)*, iff for all players i and for all pure strategies $\beta'_i \in \mathcal{A}^{\mathcal{V}}$, we have

$$\tilde{u}_i(\beta'_i,\beta^*_{-i}) \leq \tilde{u}_i(\beta^*_i,\beta^*_{-i}) + \varepsilon. \tag{2.6}$$

A 0-BNE will simply be called *Bayesian Nash Equilibrium*.

It should be noted that in the auction theory literature, Bayesian Nash Equilibria are often defined at the *interim* stage, rather than ex-ante (e.g. Krishna, 2009; Bosshard et al., 2020). The former notion is slightly stronger, as it also demands low exploitability in rare (and even zero-probability) valuation profiles. As such, the ϵ in an *interim* ϵ-BNE can be considered as the exploitability in the *worst-case* v_i, whereas ex-ante, ϵ is an *average-case* exploitability measure. For exact BNE ($\epsilon = 0$), the notions are equivalent for finite-type games, and equivalent almost surely in the general case. In this thesis, we will focus our attention on ex-ante equilibria in pure, continuous strategies. The problem of computing such BNE, with a particular focus on auction games, will be discussed in section 2.5.

2.2 Sealed-Bid Auctions

Auctions are mechanisms for resource allocation most commonly used in sparse markets with asymmetric information, where price discovery is difficult otherwise. Here, we will focus on *sealed-bid* auctions, in which all bidders simultaneously submit a single (possibly vector-valued) *bid*.[3] Additionally, we will limit the description to *one-sided* auctions where a single seller—or an agent acting on the seller's behalf—acts as an auctioneer to sell one or multiple goods to multiple potential buyers who act as *bidders* in the auction. (W.l.o.g., this may include *procurement* auctions, where the roles of buyers and sellers are reversed: The buyer aims to select one or multiple potential suppliers whose bids represent the price at which they are willing to sell the item(s).) In our context, we will assume the auction mechanism to be fixed, and we will only consider strategic interactions among the bidders. Again, it's noteworthy that other formulations exist where the auctioneer herself is considered a player in the game (e.g. aiming to set revenue-maximizing reserve prices). Furthermore, auctions can also serve as two-sided market mechanisms, where both buyers and sellers act as bidders. While we do not explicitly mention such markets below, they can nevertheless be captured mathematically by the definitions we give below, e.g. by allowing negative prices.

We'll give a definition for the general *combinatorial* sealed-bid auction, in which m heterogeneous goods are to be sold to n bidders.

Definition 2.2.1 (Sealed-Bid Combinatorial Auction)**.** A *Sealed-Bid Auction* with n bidders and m heterogeneous goods consists of:

[3] It should be noted that other *sequential* auction mechanisms also exist. They are particularly relevant in the context of combinatorial auctions with a large number of bundles, like spectrum auctions, where XOR bidding and VCG-price computations are intractable. The interested reader is referred to Bichler and Goeree (2017) or Nobel Memorial Prize, 2020 for details.

- The space of *feasible bids* $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$,
- an *allocation rule* $X : \mathcal{A} \longrightarrow \{0,1\}^{n \times m}$, where for a bid profile $\mathbf{b} \in \mathcal{A}$, $x_{ij} = (X(\mathbf{b}))_{ij} = 1$ indicates that good j is allocated to bidder i. We will write $x_i = (X(\mathbf{b}))_i$ for the bundle of goods allocated to player i.[4]
- a *pricing rule* $\mathbf{p} : \mathcal{A} \longrightarrow \mathbb{R}^n$, where $p_i = \mathbf{p}(\mathbf{b})_i$ is the monetary payment bidder i has to pay to the auctioneer.

In the most general case of an XOR bidding language, bidders submit separate nonnegative bids for each possible *bundle* of items $K \in X_i(\mathcal{A}) \subseteq [m]$ that may be allocated to them, resulting in individual bid sets of $\mathcal{A}_i = \mathbb{R}^{2^{[m]}}$ whose dimensionality is exponential in the number of items. As this is intractable for auctions with a large number of goods, alternative bid languages have been proposed that reduce the bidding complexity, e.g. XOS (Nisan, 2000), or FUEL (Bichler et al., 2022). Alternatively, sequential auction formats in which bidders gradually reveal information (Porter et al., 2003) may be used. However, both of these approaches will necessarily lead to mathematically modified games with reduced expressiveness of bidders' action spaces, and we will not consider them here.

When the goods are homogenous rather than heterogeneous (so-called *multi-unit* auctions), or bidders' preferences are inherently restricted to certain bundles (Goeree and Lien, 2016; Ausubel and Baranov, 2019), this structure can be exploited to further reduce the dimensionality (Krishna, 2009, Ch. 13). For a formal description, see Section S.1.3 of Publication B below. From the auctioneer's point of view, it should be noted that determining the outcome $(X, \mathbf{p})$ of an auction instance may be nontrivial and may involve solving a sequence of optimization problems, in particular, computing X is called the *winner determination problem* or *allocation problem*, and computing $\mathbf{p}$ is called the *pricing problem*.

The strategic interaction of bidders in auction markets is commonly modeled as a Bayesian game (Krishna, 2009; Bichler, 2017). Under the standard assumptions of *private-values* with XOR bidding, for which we will formally define the game here, each bidder i is assumed to observe their *private valuation* $v_i(K)$ for each bundle K that they may feasibly win in the auction. A more general treatment beyond private value settings will be presented in Publication B.

Definition 2.2.2 (Bayesian Auction Game)**.** Let $(\mathcal{A}, X, \mathbf{p})$ be an auction mechanism on m items and n bidders, where $\mathcal{A}_i \subseteq \mathbb{R}^{2^{[m]}}$ is the set of possible bids for player i. Further, let $\mathcal{V}_i = \mathcal{A}_i$ be the space of possible *private valuations* for player i, let F be a joint prior distribution over the players' valuations, and let all players have *quasi-linear* utility functions:

$$u_i(v_i; b_i, b_{-i}) = v_i(x_i(\mathbf{b})) - p_i(\mathbf{b}) \tag{2.7}$$

The resulting Bayesian game $G = (n, \mathcal{V}, \mathcal{A}, F, \mathbf{u})$ then describes the strategic interaction of bidders in the auction.

Note that this model assumes the availability of a common prior F which gives bidders some notion about how their strategic opponents' private information—as well as their own—may be distributed. These are strong assumptions that may not always hold in practice. Particularly in markets with few participants and goods that are sold seldomly or only once, forming reasonable prior beliefs about opponents' willingness to pay may be unrealistic. What's more, even observing one's own valuations

[4] Depending on the context, it may be advantageous to interpret x_i either as a one-hot-encoded vector in $\{0,1\}^m$ or as a subset of $[m]$. By slight abuse of notation, we will use the symbol x_i for either of these interchangeably.

may be impossible, noisy, or costly. (As an example, consider the *Mineral Rights* setting, in which all bidders observe a noisy estimate of an item's true *common value* (Krishna, 2009, Example 6.1). Achieving or improving such an estimate may require costly and lengthy business activity, such as surveying an oil field with limited access or information rights, or creating a go-to-market strategy and a business plan based on a possible acquisition of a wireless spectrum license.) An additional assumption made in auction games is that they inherently go beyond mere vNM-utility functions and assume *transferable utility* via the pricing mechanism.

When the action spaces $\mathcal{A}_i$ are identical to the bidders' valuation spaces $\mathcal{V}_i$, as in the definition above, the auction is called a *direct mechanism*. In such games, the bidders' decisions can be interpreted as either truthfully revealing their private information, or strategically misreporting their observations: most commonly this means bid-shading, i.e. underreporting one's true willingness to pay in order to reach a lower price. However, as we will see, *overbidding* may also be rational behavior in some settings. A market designer may have several desiderata about what constitutes a "good" auction rule. The Vickrey-Clarke-Groves mechanism (VCG, Vickrey, 1961) has famously been shown to be the unique auction mechanism that simultaneously fulfills the following three desiderata:

Definition 2.2.3 (VCG Desiderata)**.** Let $(\mathcal{V}, X, \mathbf{p})$ be a direct auction mechanism.

- *Allocative efficiency*: under truthful bidding, the allocation of items should be socially optimal in terms of the total achieved valuation of the winners. Let the social welfare of an allocation K of the m items to the n bidders be given by $w(K, [n]) \equiv \sum_{i\in[n]} v_i(k_i)$, then allocative efficiency holds, iff for all $\mathbf{v} \in \mathcal{V}$, we have

$$X(\mathbf{v}) \in \arg \max_{K\in\{0,1\}^{n\times m}} w(K, [n]). \tag{2.8}$$

- *Strategyproofness*, also referred to as *truthfulness* or *(dominant strategy) incentive compatibility*: Revealing their true observations v should constitute a dominant strategy: For all bidders i, strategies β_i and opponent bids $\mathbf{b}_{-i}$, one desires

$$\overline{u}_i(v_i; v_i, \mathbf{b}_{-i}) \geq \overline{u}_i(v_i; \beta_i(v_i), \mathbf{b}_{-i}). \tag{2.9}$$

  As a direct result, truthful bidding $\beta^*(v) = v$ constitutes a BNE in VCG auctions.

- *(Ex-post) individual rationality*[5]: no potential bidder should be made worse off by participating in the auction and reporting her true valuation, i.e. for all i, v_i and $\mathbf{b}_{-i}$:

$$u_i(v_i; v_i, \mathbf{b}_{-i}) = v_i\left(x_i(v_i, \mathbf{b}_{-i})\right) - p(v_i, \mathbf{b}_{-i}) \geq 0. \tag{2.10}$$

In a VCG-auction, an efficient allocation is chosen and the price the winners pay is equal to the *social cost* of their presence to other bidders, i.e. the amount that other bidders' social welfare decreases compared to an identical auction where bidder i does not participate:

Definition 2.2.4 (VCG prices)**.** In the Vickrey-Clarke-Groves mechanism, prices are given by

$$p_i^{VCG}(\mathbf{b}) = b_i(x_i(\mathbf{b})) - \Big(w\left(X(\mathbf{b}), [n]\right) - \max_K w\left(K, [n] \setminus \{i\}\right)\Big) \tag{2.11}$$

[5] Some authors relax this, instead requiring only rationality in expectation for every interim stage.

Depending on the market designer, other goals may be desirable, for example, revenue-maximization (Myerson, 1981) or core-stability (Ausubel and Baranov, 2019; Day and Milgrom, 2008; Day and Cramton, 2012), but we will not discuss them in detail here. However, one more practically motivated desideratum of auctions will be most relevant to our discussion: *Tractability*. In general combinatorial auctions, solving the Winner Determination Problem (WDP), i.e. computing an efficient allocation of the items is NP-hard, and the VCG algorithm needs to compute multiple rounds of these. This quickly leads to intractability for larger auctions, particularly when there are many items.

It should be noted that intractability is not the only drawback of the VCG auction: It also generally leads to low prices which are often perceived as unfair by losing bidders whose willingness to pay may be higher than the charge incurred by the winners. Moreover, the desiderata above are only fulfilled for bidders with quasi-linear (i.e. risk-neutral) utilities, and its outcomes are not core-stable. For a thorough discussion of VCG's shortcomings, we refer to Rothkopf (2007) and Ausubel et al. (2006).

In any case, VCG auctions are rarely employed in practice, in favor of auction mechanisms that are easier to implement. As such mechanisms are, however, usually no longer strategyproof, truthful bidding no longer constitutes a BNE.

2.3 Learning Dynamics in Games

In his original paper introducing the equilibrium notion, Nash (1950) also proved the existence of at least one NE in every finite or convex-continuous complete-information game. However, this proof relies on the Kakutani fixed point theorem (1941) and is nonconstructive. As a result, the *computation* of such equilibria has been a sought-after question in economics, mathematics, and computer science, and has sparked the interdisciplinary field of algorithmic game theory (Nisan, 2007).

In traditional economics, formal models of competitive settings are usually studied in their *equilibrium state*, and a standard—although, as we will see, generally incorrect—assumption by economists is that markets will attain an equilibrium state eventually because otherwise, some market participants would run arbitrage until this was the case. This motivates a class of potential equilibrium computation methods through *online learning* (Fudenberg and Levine, 1999): The game is played repeatedly, and after each iteration t, all agents update their strategies according to some well-defined learning rule based on the *history* of outcomes that have been observed at times 1 to $t-1$. The emergent behavior of the learning agents over time then describes the *dynamics* of the game.

However, such methods do not converge to Nash equilibria in the general case. Even in simple 2- or 3-player games, they may fail to converge to singular action profiles and instead eventually reach a recurrent distribution (discussed below), exhibit chaotic behavior (Sato et al., 2002), or even reach a non-equilibrium singular point (Fudenberg and Levine, 2009). In fact, it has been shown that there can be *no* learning rule that converges to Nash equilibria in arbitrary games (Benaïm et al., 2012). Beyond learning approaches, computing Nash equilibria is no easier: In fact, the problem of finding NE in complete information games has been shown to be PPAD-complete (Daskalakis et al., 2009), giving little hope of computing Nash equilibria efficiently in general games. As a result, the concept of the Nash equilibrium as the central solution concept of games has itself been criticized as lacking predictive power or even significance due to the inability to tractably compute it. As Kamal Jain eloquently put it: "If your laptop cannot find it, neither can the market."[6] Nevertheless,

[6] The exact origin of the phrase could not be verified, but Jain is thusly quoted by Papadimitriou in Ch. 1 of Nisan (2007).

studying learning dynamics remains an interesting and relevant topic: On the one hand, learning dynamics *do* find Nash equilibria in many relevant cases. Investigating when and whether dynamics converge to Nash is an interesting problem in itself. Furthermore, the study of dynamics itself has produced other solution concepts, such as Correlated Equilibria, which we will discuss below. Recently, Papadimitriou and Piliouras (2019) went one step further and argued that the dynamics themselves should be considered the relevant outcome and the *meaning* of the game, suggesting *Markov-Conley Chains* (MCC) as a novel solution concept. Oversimplified, these can be thought of as the limiting distribution of learning dynamics that allow for some stochasticity or errors. Most recently, a novel evaluation method based on MCCs, called *AlphaRank* (Omidshafiei et al., 2019), has been introduced, that is able to capture relative agent strength in games without relying on equilibria.

We will now discuss some of the most important classes of learning dynamics, in particular Best-Response Dynamics and No-Regret Dynamics with a particular focus on their convergence behavior.

Best-Response Dynamics The oldest and best-known class of dynamics are those where agents update to some notion of *best response* in each iteration.

Definition 2.3.1 (Best Response)**.** A strategy π_i^* is a *best response* (BR) of i to strategy profile π_{-i}, iff

$$\pi_i^* \in \mathrm{BR}(\pi_{-i}) \equiv \arg\max_{\pi_i \in \Delta\mathcal{A}} u_i(\pi_i, \pi_{-i}) \tag{2.12}$$

The first such method, introduced by Cournot (1838), directly lets players choose a *pure* strategy a_i^t as a best response to other players' last-iteration play $\mathbf{a}_{-i}^{t-1}$. *Fictitious Play* (FP, Brown, 1951) instead considers a pure strategy best response to the empirical action distribution of other players as observed in all earlier iterations of play: $a_i^t \in \mathrm{BR}\left((\mathbf{a}_{-i}^{\tau})_{\tau=1..t-1}\right)$. FP generally does not converge in last-iterate actual play or even in its empirical distribution. *Whenever* its empirical distribution converges, however, it forms a mixed Nash equilibrium.(Fudenberg and Levine, 1999, Ch. 2). A further discussion on these properties as well as some extensions of FP that allow mixed strategies are described in Publication A. When FP diverges in distribution, a common pattern is that one observes repeating cycles whose periodicity increases exponentially over time. One approach to alleviate this aims to prevent such cycling by *smoothing* the updates in each iteration, usually by playing a convex combination of the current strategy π^t and the computed best response strategy (Shamma and Arslan, 2005; Heinrich et al., 2015; Bosshard et al., 2017). Another severe drawback of best-response algorithms is that it may be computationally expensive or even intractable to compute best responses in a given iteration (Bosshard et al., 2020; Heinrich et al., 2015; Daskalakis and Syrgkanis, 2016).

No-regret dynamics and Correlated Equilibria While convergence of dynamics to Nash equilibria is generally unobtainable, there exist relaxed equilibrium notions that are efficiently computable in complete-information games. One such notion is the Coarse Correlated Equilibrium, which is closely linked to a class of learning algorithms that are based on regret minimization. We will introduce these concepts below, before discussing the convergence behavior of no-regret dynamics.

Definition 2.3.2 ((Coarse) Correlated Equilibrium (Aumann, 1974))**.** Let $G = (n, \mathcal{A}, \mathbf{u})$ be a complete information game. A distribution over outcomes $\mathcal{D} \in \Delta\mathcal{A}$ is called a

- *Coarse Correlated Equilibrium (CCE)* , iff in expectation over action profiles $\mathbf{a} \sim \mathcal{D}$, no player i has an incentive to deviate from the prescribed action profile $\mathbf{a} = (a_1, \dots, a_i, \dots a_n)$:

$$\forall i \in [n] \quad \forall a'_i \in \mathcal{A}_i \qquad \mathbb{E}_{\mathbf{a}\sim\mathcal{D}}[u_i(a_i, a_{-i})] \geq \mathbb{E}_{\mathbf{a}\sim\mathcal{D}}[u_i(a'_i, a_{-i})] \tag{2.13}$$

- a *Correlated Equilibrium (CE)*, iff this remains true even after learning of their own prescribed action a_i:

$$\forall i \in [n] \quad \forall a_i, a'_i \in \mathcal{A}_i \qquad \mathbb{E}_{\mathbf{a}\sim\mathcal{D}}[u_i(a_i, a_{-i}) \,|\, a_i] \geq \mathbb{E}_{\mathbf{a}\sim\mathcal{D}}[u_i(a'_i, a_{-i}) \,|\, a_i] \tag{2.14}$$

Importantly, such an equilibrium distribution $\mathcal{D}$ may prescribe outcomes that are *correlated* between bidders – the expectations are taken over *outcomes*, not individual strategies. For example, in a game with two players and action sets $\mathcal{A}_1$=$\mathcal{A}_2$=$\{d, w\}$ ("drive" and "wait"), a (coarse) correlated equilibrium $\mathcal{D}$ may be supported on the outcomes (d, w) and (w, d) without ever choosing outcomes (d, d) or (w, w). A common interpretation is that such behavior can be induced by a "correlation device" that *recommends* an action to each player. For example, in the game above, a traffic light at an intersection may perform such a function.

CE and CCE are closely related to the notion of *regret minimization* (Blum and Mansour, 2007) in learning dynamics and (single-agent) online optimization. Assume that agents play the game repeatedly and choose actions $a_i^t \in \mathcal{A}_i$ at times $t = 1, 2, \dots, T$. Then the *(external) regret* of agent i's sequence of actions (and, equivalently, of her learning rule) is given by the amount of utility she "lost" by not having played the best constant action in hindsight:

Definition 2.3.3 (Regret (Hannan, 1958))**.** Let $(\mathbf{a}^t)_t$ be a sequence of action profiles $\mathbf{a}^t \in \mathcal{A}$. Then it's total *external regret* at time T is given by

$$\text{Reg}_i\left(\left(\mathbf{a}^t\right)_{t=1..T}\right) = \max_{a'_i \in \mathcal{A}_i} \sum_{t=1}^{T} u\left(a'_i, \mathbf{a}^t_{-i}\right) - u\left(a^t_i, \mathbf{a}^t_{-i}\right) \tag{2.15}$$

A stronger notion is *internal* or *swap-regret* (Foster and Vohra, 1997), where rather than just considering the best constant action in hindsight, it considers possible deviations of the form $f \colon \mathcal{A}_i \longrightarrow \mathcal{A}_i$ that allow substituting each action a_i with a fixed different action $f(a_i)$ at all times t that a_i has occurred.

Definition 2.3.4 (No-Regret)**.** An action profile sequence $(\mathbf{a}^t)$ is called *regret-free*, *no-regret*, or *Hannan consistent*, iff all players eventually stop accumulating regret, i.e. as $T \longrightarrow \infty$:

$$\text{Reg}_i\left(\left(\mathbf{a}^t\right)_{t=1..T}\right) = o(T) \tag{2.16}$$

No-(external-)regret sequences converge to CCE (Aumann, 1974), no-swap-regret sequences converge to CE (Foster and Vohra, 1997). In fact, there are many efficiently computable learning algorithms that exhibit no regret. Some of these are the *Multiplicative Weights Update (MWU)* algorithm and variants (for a survey, see Arora et al., 2012), (continuous time) replicator dynamics (Smith, 1982; Nisan, 2007, Ch. 8), projected gradient ascent (also referred to as *Generalized Infinitesimal Gradient Ascent (GIGA)* in the context of learning in differential games, Zinkevich (2003)) and a related class of algorithms called *Follow the Regularized Leader* (FTRL, McMahan, 2011). FTRL includes algorithms like online mirror descent (Nemirovski et al., 2009) and *dual averaging* (Nesterov, 2009). Hartline et al. (2015) extend the notion of CCE to Bayesian games, although their analysis is restricted to finite types.

When do no-regret dynamics converge to Nash? An important line of research has focused on investigating under which conditions learning dynamics converge to Nash equilibria. In this paragraph, we will outline some important results with connections to learning in auctions. In the formal notation below, we will focus on gradient dynamics, but the notions apply equally to other no-regret learning dynamics.

One sufficient condition for the convergence of no-regret dynamics to a (necessarily unique) Nash equilibrium is *strict monotonicity* of the game. First introduced by Rosen (1965) under the name *diagonal strict concavity*, a monotonic game can be thought of as the multi-agent learning analog of a strictly (quasi-)concave objective function in single-agent optimization.

Definition 2.3.5 (Monotonicity, (Rosen, 1965))**.** A differentiable game is *strictly monotonic* iff for all pairs of action profiles $\mathbf{a}, \mathbf{b} \in \mathcal{A}$ with $\mathbf{a} \neq \mathbf{b}$, we have

$$\langle \nabla_{\mathbf{a}} \mathbf{u}(\mathbf{a}) - \nabla_{\mathbf{b}} \mathbf{u}(\mathbf{b}), \mathbf{a} - \mathbf{b} \rangle < 0 \tag{2.17}$$

Monotonicity directly implies the concavity of all individual utility functions $u_i(\cdot, \pi_{-i})$ for any opponent strategy profile π_{-i}. Monotonic games admit a unique Nash equilibrium, and it can be shown that no-regret dynamics converge to this NE in finite-dimensional, differentiable complete-information games (Mertikopoulos and Zhou, 2019; Ui, 2008). A similar result holds for *ex-post differentiable* Bayesian games (Ui, 2016). In original research not included in this dissertation (Heidekrüger et al., 2021c), we show that this result generalizes to ex-ante gradient dynamics like NPGA (see section 2.5) in ex-post non-differentiable Bayesian games, under the quite strong assumption of special convex neural network architectures that preserve concavity of the objective. However, whether monotonicity holds in a given auction cannot be determined easily, as it would require verifying eq. (2.17) for all types $\mathbf{v} \in \mathcal{V}$, which is nontrivial and may very likely be intractable. As such, this result can be understood to show *correctness* of NPGA, in the sense that it will find the unique BNE in monotonic games, but it remains uninformative about NPGAs performance in relevant auction games.

A second class of games important in our discussion is that of potential games.

Definition 2.3.6 (Potential Game, (Monderer and Shapley, 1996))**.** A game $G = (n, \mathcal{A}, \mathbf{u})$ is called an (exact) *potential game*, if there exists a function $\phi : \mathcal{A} \longrightarrow \mathbb{R}$, such that for all $\mathbf{a}, \mathbf{a}' \in \mathcal{A}$ and for all $i \in [n]$:

$$u_i(a_i; \mathbf{a}_{-i}) - u_i(a_i'; \mathbf{a}_{-i}) = \phi(a_i; \mathbf{a}_{-i}) - \phi(a_i'; \mathbf{a}_{-i}). \tag{2.18}$$

In the definition above, ϕ is the *potential* or *Lyaponov function* of the game, and computing an equilibrium is equivalent to solving a (single-decision maker) optimization problem with objective function ϕ. Local minima of ϕ become local saddle points in the underlying game, i.e. local Nash equilibria. When ϕ is (strictly) convex, the unique global Nash equilibrium of the game can thus be efficiently computed via standard methods from nonlinear optimization. In fact, gradient dynamics in the game will be compatible with gradient descent on the potential function (Neyman, 1997; Mazumdar et al., 2020) and lead to (at least local) Nash equilibria in potential games. In Publication B, we apply this result to non-differentiable Bayesian games: We show that symmetric auctions can be interpreted as potential games and that our method NPGA converges to approximate Bayesian Nash equilibria in such games when relying on model-sharing between agents to induce symmetric training.

Beyond these properties, Nash-convergence of gradient dynamics, in particular, has been studied extensively by Singh et al. (2000) for 2x2 Normal Form games, and by Letcher et al. (2019) for two-player differentiable games. In these classes of games, the exact convergence characteristics of gradient dynamics are more thoroughly understood.

2.4 Reinforcement Learning

A parallel, related line of research deals with training agents in multi-agent settings via reinforcement learning (RL). We will briefly cover the most important aspects of standard, single-agent reinforcement learning before outlining multi-agent extensions and how they relate to this thesis. In (single-agent) reinforcement learning, an agent faces a sequential decision-making problem that is typically formalized as a *Markov Decision Process (MDP)*: In each state s_t of such a process, an agent must choose an action a_t, which will yield the next step s_{t+1} and (possibly) a reward r_{t+1} according to some (generally unobserved) transition probability distribution $\mathbb{P}(s_{t+1}, r_{t+1} \mid s_t, a_t)$.

The agent is then faced with the task of finding an optimal policy π that maximizes its long-term expected (and possibly discounted) reward.

$$J(\pi) \equiv \mathbb{E}_{(s,a)\sim\pi}[\sum_{t=0}^{\infty} \gamma^t R_t(s_t, a_t)] \tag{2.19}$$

In reinforcement learning, this is done by repeatedly interacting with the environment and observing state transition tuples $(s_t, a_t, r_{t+1}, s_{t+1})$. In practice, one often considers *Partially Observable MDPs (POMDP)*, where the state-observations by the agent may be noisy or incomplete. Analogously to the games in section 2.1, when action spaces are finite, it is common practice in RL to consider policies to be distributions over actions for a given state, and one commonly writes $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$, where $\pi(s, a) = \mathbb{P}(A_t{=}a|S_t{=}s)$. Importantly, this allows for stochasticity in the agent's actions which is usually required to explore previously unseen regions of the state space. When action spaces are continuous, the two prevalent approaches are to either learn *deterministic policies* $\pi : \mathcal{S} \to \mathcal{A}$ where the policy itself usually takes some finite-dimensional parametric form $\pi = \pi_\theta$ (Silver et al., 2014), or to assume a parametric action-distribution and learn its parameters. For example, in a continuous scalar action space, one may restrict oneself to Gaussian policies and sample $a \sim \pi(s) = \mathcal{N}(\mu_\pi(s), \sigma_\pi(s))$, and then learn the mean $\mu_\pi(s)$ and the variance $\sigma^2_\pi(s)$ directly as functions of $s \in \mathcal{S}$. The former approach is closely related to our approach of learning parametric continuous pure strategies in Bayesian games with continuous type and action spaces.

Reinforcement Learning Algorithms For single-agent reinforcement learning in MDPs, there are several well-established learning algorithms that provably (asymptotically) converge to an optimal policy (Sutton and Barto, 2018), although tractability, sample-efficiency, and precision are often a challenge in practice. Such methods roughly split into *value iteration* methods and *policy iteration* methods, although there is some overlap. In value iteration, the agent learns to predict the expected achievable future reward, the *value*, of a state s, or of a state-action pair (s, a). Actions are then chosen in a way to maximize the expected value at time $t + 1$ via dynamical programming methods, most commonly by solving a Bellman equation. In the *policy iteration* regime, the policy representation itself (e.g. probability vectors $(\pi(a|s))_{a\in\mathcal{A}}$ in the finite action case) are updated in each iteration.

An important class of algorithms is based on *policy gradients* and directly aims to estimate $\nabla_\pi J(\pi)$, where the gradient ∇_π is to be understood with respect to the explicit or parametric representation of π. In the finite-action setting (Williams, 1992), an unbiased estimator of this gradient can be efficiently computed from samples of state transitions $(s^t, a^t, r^{t+1}, s^{t+1})$. In the continuous setting with deterministic actions, it is likewise possible under some regularity conditions (Silver et al., 2014). Most relevant to this dissertation, this requires the objective function J to be differentiable in the agent's actions. Given such estimates, one may then perform stochastic gradient ascent to maximize the expected return. Extensions to this regime, like Trust Region Policy Optimization (TRPO, Schulman et al., 2015) or Proximal Policy Optimization (PPO, Schulman et al., 2017), add regularization in order to stabilize the training process by reducing variance.

Recently, a particular focus has been on *deep reinforcement learning*, where the value function, the policy, or both, are represented via neural networks. Such methods have led to several breakthroughs in both discrete and continuous control tasks (Mnih et al., 2015; Wurman et al., 2022).

Multi-Agent Reinforcement Learning *Multi-Agent Reinforcement Learning (MARL)* studies extensions to RL in which multiple agents interact with the environment and all agents' actions influence the transition probabilities to the next state and reward vector. This may give rise to both *cooperative* or *competitive* multi-agent settings: Fully cooperative settings are commonly modeled as *Decentralized Partially Observable Markov Decision Processes (Dec-POMDP)* in which all agents share a common objective function. Applications include the control of Multi-Agent Systems (Shoham, 2009), or cooperative tabletop games like Hanabi (Bard et al., 2020). In such settings, one often takes a *centralized training, decentralized execution* view, which allows access to additional data during the training process which may not be locally available to agents in an online setting. When tasks are not fully-cooperative but instead involve competition, the problem can be formally modeled as a *Markov Game*: In this extension of a POMDP, agents generally each observe their own individual rewards. The system can then be interpreted as a special case of an extensive form game that adheres to the Markov Property.[7]

Competitive MARL has most prominently been applied to zero-sum games, where it has led to several breakthroughs such as superhuman performance in tabletop and video games, such as Go (Silver et al., 2016, 2017; Schrittwieser et al., 2020), StarCraft II (Vinyals et al., 2019).[8] Due to inherent symmetries in most of these applications and the nature of zero-sum games, particularly the applicability of the minimax theorem (v. Neumann, 1928), the methods in these breakthroughs have largely relied on results from single-agent learning theory, and, in the case of team-games, the introduction of explicitly specified or implicitly learned *communication* protocols between allied agents. Nevertheless, there is a growing body of literature that explicitly informs MARL by Game Theory and vice versa. For example, several learning methods explicitly take the effects of interaction with other agents into account (Lowe et al., 2017; Letcher et al., 2019; Foerster et al., 2017), or explicitly state learning equilibria as the *goal* of multi-agent learning (Heinrich et al., 2015; Heinrich and Silver, 2016; Hennes et al., 2020; Letcher et al., 2019). Many of the questions studied in this field have direct analogs in Economics and Auction Theory in particular (Heidekrüger et al., 2021b), such as evaluating

[7] For a concise formal description, the interested reader is referred to Conitzer and Sandholm (2008).

[8] Superhuman performance has recently also been reached in large imperfect information games like Texas-Hold'em Poker (Brown and Sandholm, 2019a,b), but the methods employed here rely on regret-minimization techniques (Compare section 2.3) rather than RL.

agent strength in non-zero-sum environments (Omidshafiei et al., 2019; Lanctot et al., 2017). Zheng et al. (2020) study a setting closely related to market design, where a socially optimal tax policy is learned from the behavior of market participants who are themselves learning agents.

We will not explicitly model the equilibrium computation problem in sealed-bid auctions as a Markov Game – since we do not need to worry about sequential decisions and path dependency. Nevertheless, sealed-bid auctions can be thought of as special cases of Markov Games with episode length 1, and many results from (MA)RL apply to it directly. Beyond theoretical algorithmic results, RL particularly informs the implementation details of our approach.

2.5 Equilibrium Computation in Auctions

In this section, we will look formally at the problem of equilibrium computation in auctions, compare and contrast the problem with those studied in complete information games and the MARL literature, and deduce implications for efficient implementations of learning in auctions. Afterward, we will briefly discuss the computational complexity of the problem. Then, we will discuss analytical equilibrium solutions that have been derived, as well as numerical methods that have been proposed in the literature. Based on these discussions, we will briefly motivate our approach and methodology taken in the original research publications included in this thesis.

Formally, we are seeking the Bayesian Nash equilibria of auction games as defined in section 2.2. Compared to the question of computing Nash equilibria in continuous complete-information games, the problem at hand is further complicated in Bayesian auction games that generally, and typically, have continuous types and actions. As such, even pure strategies β_i constitute infinite-dimensional objects in a functional space: $\beta_i \in \mathcal{A}^{\mathcal{V}}$. Beyond this fact, another complication is given by the fact that the utility functions $\mathbf{u}$ in auction games are typically non-differentiable. This is a consequence of the discrete nature of the allocations X: Assuming quasi-linear utilities $u_i(v_i, b_i, \mathbf{b}_{-i}) = v(x_i(\mathbf{b})) - p_i(\mathbf{b})$, the valuation component will be a step-function in b_i. As a result, standard methods for gradient computation are inapplicable, as discussed in Section 2.4 of Publication A.

Computational Complexity of Bayesian Nash Equilibria In complete-information games, the existence of at least one Nash equilibrium is guaranteed, but computing such a Nash equilibrium has been shown to be PPAD-complete, even in the 2-player case (Daskalakis et al., 2009; Nisan, 2007, Ch. 2). Deciding whether a *second* Nash exists in a given game has been shown to be NP-complete. As such, even in finite complete-information games, finding Nash equilibria is believed to be a hard problem.

The exact complexity of computing BNE in Bayesian games has not yet been conclusively determined, but it must be at least as hard as that of computing NE in complete-information games. So far, any attempts at determining the computational complexity of BNE have focused on the finite-type case only, yet found discouraging results: Deciding whether a *pure strategy* BNE exists in a finite (type-and-action) Bayesian game is known to be NP-complete (Conitzer and Sandholm, 2008), even in symmetric 2-player Bayesian games with uniformly distributed valuations. For a certain type of private-value Simultaneous Second-Price Auction (SiSPA) on multiple heterogeneous items with discrete types, (Cai and Papadimitriou, 2014) have established that finding a BNE is PP-hard. Daskalakis and Syrgkanis (2016) note that in similar SiSPA auctions with XOS-bidding, even no-regret learning and thus the computation of Bayesian CCE is NP-complete for discrete types and discrete (or discretized) action spaces because discrete-action regret-minimization methods suffer from the exponential (or worse) growth of available actions in the number of items. They instead consider *envy-freeness* as an efficiently computable relaxation of regret-freeness that yields a similar price of anarchy. Due to these hardness results, one cannot expect BNE to be computable in the general case, particularly in continuous-type Bayesian games.

Analytical Solutions Inherently, only the Vickrey-Clarke-Groves mechanism is strategyproof for any number of items and an arbitrary number of risk-neutral, rational bidders. For all other mechanisms (or without quasi-linear utilities), however, the analytical derivation of Bayesian Nash equilibria in auctions becomes tedious and nontrivial, even in relatively simple markets.

The traditional approach relies on fully specifying a setting (e.g. the allocation and pricing rule, prior distributions, number of players and items, etc.) and then explicitly describing the equilibrium state, which will result in a system of *variational inequalities (VIs)*, which can be described by partial differential equations. In general, solving or even explicitly *stating* the system of VIs is intractable. Nevertheless, in some settings this approach has led to success and the discovery of one or more BNE in closed-form: For example, in independent private value auctions with a single item, risk-neutral bidders and symmetry assumptions on the priors *and* the equilibrium strategy, the VIs reduce to a single ordinary differential equation (ODE) that can be solved explicitly, yielding the unique[9] symmetric pure-strategy equilibrium for a range of payment functions, including first-price, second-price, and all-pay (Krishna, 2009, Part I). For first-price auctions, this analysis holds even when bidders are risk-averse. When dropping the symmetry assumption on all bidders, results become more sparse. In a first-price auction with $n = 2$ and asymmetric uniform priors, a unique closed-form BNE is known when overbidding is disallowed (Kaplan and Zamir, 2012; Plum, 1992). When allowing overbidding, there are known setting where *multiple* (at least 3) BNE can be derived (Kaplan and Zamir, 2015).

Beyond single-item auctions or independent private values, there are few additional cases where analytical derivations have been successfully applied, but these usually come with stronger restrictions, e.g. on the number of players or simplified valuation or bid spaces through significant a-priori restrictions on bidders' demand sets or homogeneity of items. The included empirical experiments into the convergence of NPGA (Publications A to C) cover all continuous-type-and-action auctions with known equilibria that the author is aware of. Publication A covers simple symmetric single-item auctions. Publication B, including the supplementary material, extends this analysis to markets with multiple items. Publication C additionally studies further settings with (irreducible) asymmetries. In addition to analytically "solved" settings, we also compute approximate equilibria in larger settings without analytical solutions, including the newly proposed LLLLRRG setting (combinatorial with dual-minded bidders, $n = 7$, $m = 6$), the largest setting for which this has been achieved thus far. Publication D includes an application to single-item *all-pay* auctions, where other proposed equilibrium learning methods are known to fail to converge to BNE (see below). Furthermore, efficient parallelized implementations of all studied markets have been incorporated into the `bnelearn` software package (Heidekrüger et al., 2021a).

[9] In some cases, there may actually be a class of outcome-equivalent equilibria: For example, in second-price auctions with asymmetric priors, a "strong" bidder i may bid arbitrarily high when v_i exceeds $\max \operatorname{supp} F_{\mathbf{v}_{-i}}$. We would treat this behavior as equivalent to only bidding up to her own valuation because choosing $b_i > v_i$ will never change the resulting allocation or prices.

Numerical Algorithms We will now introduce numerical methods for equilibrium computation and introduce the high-level approach of our method, *Neural Pseudogradient Ascent (NPGA)*.

Two recent approaches explicitly perform learning dynamics directly in Bayesian auction games, although on discretized or linearized strategies: Bosshard et al. (2020, 2017) discretize the type and action spaces into a finite number of support points and then perform (smoothed) best-response dynamics on pure strategies in the interim state for every possible valuation v_i. The bidding strategy is then assumed to be the linear interpolation of the best responses at the sampled support points. The method successfully recovers the equilibria in the (single-minded) LLG settings discussed above and was the first to compute a close approximation of a BNE in a larger combinatorial first-price auction with 6 bidders and items, referred to as LLLLGG. Another recent approach likewise uses discretization of the type and action spaces, but learns interim-distributional strategies via *Simultaneous Online Dual Averaging* (SODA, Fichtl et al., 2022), i.e. a probability vector $\mathbb{P}(a_i|v_i)$ at every sampled v_i. The resulting distributions place high probability on the interim action corresponding to the analytical pure strategy Nash equilibrium. While both methods perform well in low-dimensional Bayesian games, their discretizations are subject to a combinatorial explosion when increasing the game size, particularly in the number of items, as the memory requirement of the discretization grows double-exponentially in m with an XOR bidding language.

Hartline et al. (2015) consider no-regret learning in Bayesian games with explicitly finite types, formally introduce the notion of Bayesian Coarse Correlated Equilibria, and analyze the price of anarchy, i.e. the difference in social welfare between learned outcomes and the efficient offline solution. Armantier et al. (2008) takes a similar viewpoint to ours and interprets the BNE strategies of a Bayesian game as the NE actions of a (possibly infinite-dimensional) complete-information game. Given a parametrized restricted strategy space Σ^r, they note that one can view a Nash equilibrium β^{*r} in $(n, \Sigma^r, \tilde{u})$ as an approximation to a BNE β^* in $(n, \mathcal{V}, \mathcal{A}, F, u)$ and show that for a sequence of ever-finer grained restrictions $(\Sigma^r)_r$ with $\Sigma^r \to \Sigma$, the equilibrium in the restricted space will converge to a BNE in the original game: $\beta^{*r} \to \beta^*$. In empirical work, they consider piecewise linear bidding functions in two-player single-item first-price auctions and small discriminatory-price multi-unit auctions with $m = n = 4$ and observe quick convergence to the analytical BNE as the number of support points grow.

In contemporaneous but independent work from ours, Li and Wellman (2021) also consider evolutionary strategies and neural networks for learning in auctions, although in a different way from the work presented in this dissertation: Rather than performing agent-based learning dynamics, they note that in symmetric auctions, regret-minimization itself can be interpreted as a two-player zero-sum game of between the agent i who aims to minimize his regret and a fictitious adversary aiming to find a best response β_{-i} to exploit i. Evolutionary Strategies as a black-box optimization method can then be used to implement such a best-response oracle in the inner loop, as well as finding a regret-minimizing pure strategy β_i in an outer loop. The authors also consider an extension based on Empirical Game-Theoretic Analysis (Wellman, 2006; Tuyls et al., 2018) in which *mixed* strategy bidding functions are modeled via mixtures of pure strategies: Pure strategies are successively added to a (normal-form) meta-game, and the mixed Nash equilibrium in this meta-game then serves as the mixed strategy function in the original Bayesian game. Their method recovers analytical BNE in symmetric single-item first- and second-price auctions and achieves low regret in moderately-sized settings with multiple homogenous goods. The authors note that their method fails to recover the analytical BNE in single-item all-pay auctions, and the convergence behavior of the method is not yet fully understood.

Our approach, Neural Pseudogradient Ascent (NPGA), on the other hand, parametrizes bidders' strategies via neural *policy networks* and follows a *deterministic policy gradient* approach, in order to follow the ex-ante (projected) gradient dynamics. Traditional policy gradient approaches based on backpropagation fail in this setting, as explained in detail in Publication A. Instead, NPGA relies on evolutionary strategy gradient computation (Salimans et al., 2017) in order to sample unbiased ex-ante gradient estimates despite the auction's ex-post non-differentiability. An in-depth theoretical discussion of the NPGA algorithm is presented in Publication B. NPGA provably converges to at least local pure-strategy BNE in symmetric auctions, and to the unique global pure-strategy BNE in monotonic auctions. This will be discussed in detail in chapter 3.

In further original research not included in this dissertation (Kohring et al.), we additionally empirically studied the (stochastic) dynamics of agents who adapt their strategies using Particle-Swarm optimization (PSO) methods as an alternative gradient-free black-box optimization method. The method achieves qualitatively similar results to NPGA in the settings presented in Publication B but incurs a significantly larger memory footprint.

2.6 Software Frameworks for Learning in Games

Several open-source frameworks for learning in games and MARL are available. On the one hand, there are several frameworks for game-theoretic analysis of *finite* complete information games. Tools like *Game Theory Explorer* (Savani and von Stengel, 2015) or *Gambit* (McKelvey et al., 2016) allow users to define, visualize and solve finite complete-information games, both in strategic and extensive-form, via both a GUI or a programming interface. On the other hand, multi-agent learning may be implemented using common reinforcement learning libraries. PettingZoo (Terry et al., 2020) is a multi-agent extension to the popular reinforcement-learning framework OpenAI Gym (Brockman et al. (2016), now succeeded by *Gymnasium*, Farama Foundation (2021)), that enables the drop-in study of (single-agent) RL algorithms in Markov Games, as well as specialized MARL algorithms. Ray RLlib (Liang et al., 2017) is a framework for scalable distributed reinforcement learning that supports Markov games natively. However, both of these frameworks focus on single-agent reinforcement learning and are not geared towards primarily game-theoretic use cases.

DeepMind's OpenSpiel (Lanctot et al., 2019) is an ecosystem of algorithms and games that covers a wider area of the learning in games literature beyond MARL. The library includes original and third-party implementations of many complete and imperfect information games. Most games in OpenSpiel leverage performant implementations in C++. However, since the library focuses explicitly on extended-form games with unknown and varying episode lengths, it does not support batch-vectorized game environments which play a crucial role in our work. Furthermore, OpenSpiel does not support continuous state or action spaces.

Bayesian auction games can be modeled as Markov Games but have several concrete properties that warrant the development of a specialized framework. On the one hand, sealed-bid auctions are direct mechanisms. Using RL terminology, the episode length is always 1. As a result, much of the theoretical underpinning of MDPs, Markov games, and learning methods developed for them does not apply to the one-simultaneous-action case, and even the game implementations in the popular frameworks above would contain much superfluous overhead. On the other hand, the fact that the episode length of 1 is (a) fixed, and (b) does not allow any path dependencies within a single episode of the game, makes it possible to efficiently batch-vectorize the game implementation and leverage hardware acceleration to simulate many game instances in parallel. To this end, the `bnelearn` framework (Heidekrüger et al., 2021a) was developed during the course of this dissertation project to facilitate the included original research.[10]

[10] At the time of writing, both Gymnasium and RLlib have recently added support for batch-vectorized environments, but their implementations are primarily geared towards large-scale parallelization across a compute cluster, rather than batching on individual GPUs. Additionally, batch-vectorization in these frameworks is sometimes incompatible with other important features and incurs significant overhead in programming complexity.

Publication A: Computing Approximate Bayes-Nash Equilibria through Neural Self-Play

Peer-Reviewed Conference Paper

Title: Computing Approximate Bayes-Nash Equilibria through Neural Self-Play

Authors: S. Heidekrüger, P. Sutterer, M. Bichler

In: Workshop on Information Technology and Systems (WITS19), Munich, Germany, 2019.

Abstract: Understanding market dynamics means understanding and predicting the behaviour of the market participants. Nash equilibria have proven to be an effective means in this regard. Unfortunately, computing equilibria in a complete information or Bayesian game is computationally hard. We introduce a learning rule based on neural networks that we call Neural Self-Play. This rule is able to compute approximate Nash equilibria for many normal form games as well as for incomplete-information games with continuous type- and action-space, i.e., sealed-bid single-item auctions. Leveraging GPU hardware architecture, which allows for parallelized computation of large matrices, Neural Self-Play finds approximate Bayesian Nash equilibria in first-price sealed bid auctions with 10 players within 10s of minutes.

Contribution of thesis author: development and implementation, empirical analysis (auction settings), writing and revising the manuscript, project management

Reference: Heidekrüger et al. (2019)

Computing Approximate Bayes-Nash Equilibria through Neural Self-Play

Stefan Heidekrüger[*] Paul Sutterer[†] Martin Bichler[‡]

Department of Informatics
Technische Universität München

Abstract

Understanding market dynamics means understanding and predicting the behaviour of the market participants. Nash equilibria have proven to be an effective means in this regard. Unfortunately, computing equilibria in a complete information or Bayesian game is computationally hard. We introduce a learning rule based on neural networks that we call Neural Self-Play. This rule is able to compute approximate Nash equilibria for many normal form games as well as for incomplete-information games with continuous type- and action-space, i.e., sealed bid single-item auctions. Leveraging GPU hardware architecture, which allows for parallelized computation of large matrices, Neural Self-Play finds approximate Bayesian Nash equilibria in first-price sealed bid auctions with 10 players within 10s of minutes.

1 Introduction

Market design has received increasing attention in the information systems literature (Bichler et al. 2010). For market designers, it is important to understand equilibrium behavior of market participants to predict market outcomes and potential strategic problems. While early literature on general equilibrium theory focused on competitive equilibria and assumed players to be non-strategic price takers, auction theory assumes strategic agents and uses the Nash equilibrium concept to study the price formation process (Nash et al. 1950). More precisely, auction theory models auctions as Bayesian games and analyzes the Bayes-Nash equilibria of players.

Unfortunately, for many markets we do not know the Bayes-Nash equilibrium strategy. For example, Bayes-Nash equilibrium strategies for simple combinatorial first-price sealed-bid auction are still unknown, except for restricted environments (Kokott et al. 2019). Different assumptions on the common prior distribution, the risk aversion of the players, or the number

[*] `stefan.heidekrueger@in.tum.de`

[†] `paul.sutterer@tum.de`

[‡] `bichler@in.tum.de`

of players and objects all play a role, and the analytical derivation of equilibrium strategies can be very challenging, often without a closed-form solution if at all possible.

In this paper, we introduce Neural Self-Play (NSP), a method that numerically derives Bayes-Nash equilibria. In experiments, we focus on environments where we know the analytical solution and show that NSP closely approximates the analytical equilibrium strategy. This bears the promise that it can provide such a solution for markets where we cannot derive analytical solutions. While earlier literature either stems from artificial intelligence or game theory, equilibrium computation becomes increasingly important as a tool in market design and other areas of information systems research. This also contributes to the overall theme of the workshop: markets for policy making and sustainability.

1.1 Related Literature

Nash equilibria (NE) are a central solution concept in non-cooperative game-theory. Informally, in a Nash equilibrium no agent has an incentive to deviate, given the current behaviour of all other agents. Therefore, once a NE is found, it is a stable state. However, finding NE is hard. Actually, it is known to be PPAD complete already for 2-player normal-form games (Daskalakis et al. 2009) and it is hard to approximate (Rubinstein 2016).

There exist a number of learning rules which try to find NE, two of the most frequently used are Fictitious Play (FP) (Brown 1951) and Smooth Fictitious Play (SFP), a variant of the first. The idea of FP is an iterative pre-play process in which each player plays a best response to the opponents' expected play, based on past observations. FP applies to games of complete information, such as normal form games, as well as to games of incomplete information. While FP works fine for many games, its direct application fails whenever a game has continuous type- and action-space, as in auctions.

The problem of numerically computing approximate NE in auctions with continuous type- and action-spaces has previously been studied by Bosshard et al. (2017). Bosshard et al.'s algorithm is shown to compute verifyable approximate equilibria in the general setting. They discretize and transform the Bayesian auction game into a normal form game and compute a pointwise best response. Afterwards, they apply a (smoothed) best response update in the original continuous game by interpolating the discrete solution such that it guarantees an upper bound on the utility loss. While their method is shown to converge, the complexity of calculations in the required discretization grows exponentially with the number of players and thus becomes intractable for games with many players or multidimensional type or action spaces.

1.2 Contributions

In this study, we introduce Neural Self-Play (NSP), a learning rule implementing players' strategies as neural networks, and using evolutionary strategies to update the networks parameters. We first test the algorithm's performance on normal form games of complete information with discrete type- and action-spaces. The algorithm performs similar to FP and SFP and is able to find pure Nash equilibria (PNE) as well as mixed Nash equilibria (MNE) in empirical frequencies. While FP and SFP are only able to work with discrete type- and action-spaces, NSP also works with continuous settings. We test the algorithm on games of incomplete information with continuous type- and action-spaces, i.e. on sealed bid single-item auctions. NSP is able to find approximate Bayes-Nash equilibria (BNE) in all performed experiments within minutes. It is able to find BNE for settings with many bidders and scales well even for an increasing number of parameters.

The remainder of this paper is structured as follows: First, in Section 2, we introduce preliminaries as well as NSP and related learning rules. We then present empirical results of applying NSP to normal form and auction games in Section 3 before concluding with a summary of our findings in Section 4.

2 Methodology

In this study, we apply different learning rules for finding Nash equilibria (NE), namely Fictitious Play (FP), Smooth Fictitious Play (SFP), Mixed Fictitious Play (MFP) which are well-studied tabular methods. In addition, we introduce a new algorithm for equilibrium learning based on neural networks that we call Neural Self-Play (NSP). Before we describe these learning rules, let us briefly introduce a few terms.

Games in normal form (complete information, discrete type- and action space) are defined by a tuple: $G = (N, \mathcal{A}, u)$ where $N = \{1, ..., n\}$ describes the set of players; $\mathcal{A} = \mathcal{A}_1 \times ... \times \mathcal{A}_n$ describes the set of action profiles, with $\mathcal{A}_i$ being the set of actions available to player i; and $u = (u_1, ..., u_n)$ is the joint utility function where $u_i : \mathcal{A}_i \to \mathbb{R}$ describes the payoff (utility) function for each player.

Games of incomplete information are described by a quintuple: $G = (N, \mathcal{A}, V, p, u)$. N and $\mathcal{A}$ are as above, with $\mathcal{A}_i$ potentially being continuous sets $\mathcal{A}_i \subset \mathbb{R}$; $V = V_1 \times ... \times V_n$ is the set of type profiles. At the beginning of the game, each player i is informed of her own type $v_i \in V_i$ only (private information). Just as $\mathcal{A}_i$, the V_i are (potentially continuous) subsets of $\mathbb{R}$.[1] $p(v)$ defines a prior probability distribution over type profiles that is assumed to be common knowledge. The payoff (utility) function is now determined by $u_i : \mathcal{A} \times V_i \to \mathbb{R}$, i.e. players' utilities depend on all players' actions but only their own type.

In each game, after receiving the private type information, each player i chooses her strategy according to some (possibly stochastic) strategy $\pi : V_i \to \Delta\mathcal{A}_i$ that maps to a probability distribution over possible actions.[2] All the learning rules described here have in common that the underlying game is played repeatedly—in theory indefinitely—while players observe each other's behavior and adjust their strategies π_i ("learning") in order to ultimately find an equilibrium in the game, i.e. a state where no player can improve their own expected utility by changing their strategy π_i any further. Throughout this paper, we denote by the index $-i$ a profile of types, actions or strategies for all players but player i.

2.1 Fictitious Play

FP was first introduced by (Brown 1951). It can be seen as a process of pre-play by each player to learn more about the game's dynamics. In FP, each player starts with initial beliefs about the other players' strategies and updates these beliefs based on the observations of

[1] Private information may also be multidimensional, but we restrict ourselves to the scalar setting here.

[2] When π_i is known to be deterministic, i.e. return an action a_i with probability 1, we will use the following abuse notation: $\pi_i(v_i) = a_i$.

played actions throughout the process. At each step every player i computes her expected utility u_i for any possible action in $\mathcal{A}_i$, given the current beliefs of opponents play σ_{-i}, and chooses the action $a_i \in \mathcal{A}_i$ that maximizes it, i.e. plays a best response:

$$a_i = \arg\max_{a \in \mathcal{A}_i} \mathbb{E}\left[u_i(a, \sigma_{-i})\right]$$

After each round, players update their beliefs about other players' strategies using Bayesian updating. As the actual play can only converge to pure Nash equilibria (PNE) due to the way actions are chosen, it oscillates in games with only mixed Nash equilibria (MNE). However, the empirical distribution of historical actions may still converge in such games (Fudenberg and Levine 1999, p.42 - 45) and is thus usually considered when speaking about convergence of FP. While FP does not converge in general (Shapley 1964), it has been shown to converge for some general settings such as constant sum games (Robinson 1951) or games that are solvable through iterated elimination of strictly dominated strategies (Nachbar 1990). For details on convergence guarantees of FP, we refer the interested reader to any text book on game theory, e.g. Fudenberg and Levine (1999).

2.2 Smooth Fictitious Play

Smooth Fictitious Play (SFP) is based on FP but differs in that SFP does not deterministically play a best response, but adds randomness to the decision process. In our implementation this is achieved by applying the softmax function to the expected utilities of each action and sampling an action according to the resulting probability distribution. We further apply a temperature parameter τ that controls the level of smoothing, i.e. the degree of indifference between actions. For $\tau \to \infty$, players will be completely indifferent between actions; as $\tau \to 0$, the players probability of playing the best response action approaches 1. Usually, τ is initialized with 1 and decreases with each step. The probability of player i to play an action a, given the beliefs of opponents playing σ_{-i}, is then given by:

$$\mathbf{Pr}\left(a_i \mid \sigma_{-i}\right) = \frac{e^{\frac{u_i(a_i,\sigma_{-i})}{\tau}}}{\sum_{r_i \in A_i} e^{\frac{u_i(r_i,\sigma_{-i})}{\tau}}},$$

where we dropped the expectation around $u_i(\cdot, \sigma_{-i})$ for ease of notation.

SFP can be motivated in multiple ways, among them are: the randomization represents private information about the utility function of a player; and the introduction of randomization allows agents to be less exploitable. In contrast to FP, the actual play in SFP (or the probability for the actions according to players' strategies π) is in principle able to converge to MNE (Fudenberg and Levine 1999, p.131 - 156).

2.3 Mixed Fictitious Play

Mixed Fictitious Play (MFP) is an adjustment of SFP in which players do not sample an action but can "play" mixed strategies that are observed by others. This adjustment makes MFP purely fictitious, i.e. a mind experiment, since players cannot actually play a probability but would have to decide on an action in practice. The advantage is faster convergence due to lack of noise introduced by sampling. This method is thus only suited for finding potential NE of a game but not a method for players to learn to reach the equilibrium strategy through repeated playing of an actual game.

2.4 Neural Self-Play

We propose Neural Self-Play (NSP) as an alternative iterative learning rule that is applicable both to normal form games and continuous-action continuous-type Bayesian games. In NSP, we model players' strategies using neural networks. In each step, players consider their opponents to be stationary in their current strategy (as opposed to updating beliefs over historical play as in FP and variants). The general idea of NSP is that players apply a small update to their neural network parameters θ that will lead to an improvement in utility.

The canonical way of implementing this idea would be applying a gradient ascent algorithm via backpropagation. In fact, this method is called Policy Gradients in (single-agent) reinforcement learning and has been previously studied in multi-agent normal form games where it is called Infinitesimal Gradient Ascent (IGA, Singh et al. 2000; Bowling and Veloso 2002). However, in the following, we demonstrate that this approach fails in the setting of auctions and instead propose to use an alternative training algorithm based on Evolutionary Strategies, before introducing the specific model architectures that we use in this study.

2.4.1 Infinitesimal Gradient Ascent

In Infinitesimal Gradient Ascent (IGA), each player adjusts their own strategy in the direction of the gradient of their utility function when considering opponents fixed at their current strategies[3]:

$$\pi_i^{t+1} := \pi_i^t + \alpha \nabla_{\pi_i} u_i(\pi_i^t, \pi_{-i}^t)$$

In 2x2 normal form games, this simple learning rule has been shown to either converge to a NE or end up in cycling behaviour where each player's average utilities converge to those in a NE (Singh et al. 2000). However, IGA relies on knowledge of the analytical joint gradient dynamics and assumes that the joint utility function is differentiable everywhere. This makes the learning rule unsuitable for continuous-type, continuous-action Bayesian games as these can involve nontrivial discontinuities as we discuss below. To rectify this, our approach differs from IGA mainly in the way gradients are computed, particularly in two aspects:

On the one hand, IGA assumes analytical knowledge of—or an efficient way to compute with arbitrary precision—the global gradient vector field in each step. This assumption, however, becomes impracticable in infinite information-state spaces, as no closed-form description might exist or be known. We thus forgo this assumption and instead rely on stochastic estimation of the gradients: In each iteration, we play a *batch* of games, i.e. draw a batch of valuation profiles from the players' prior distributions and calculate players' current strategy utilities in each of the valuation profiles. Due to the parallel nature of these calculations, we can leverage modern hardware accelerators such as GPUs to perform these batched operations at no additional cost in computation time. We then aim to calculate the gradients for this stochastic joint utility function with respect to each player, which in expectation will approximate the gradient dynamics of the full Bayesian Game.

[3] The sceptical reader might wonder how ∇_{π_i} is defined. π_i could either be a tabular vector of action-probabilities, a parametrized function, etc.; the gradient should be understood with respect to the respective representation determining π_i. We use abuse of notation here to illustrate the concept in a general way.

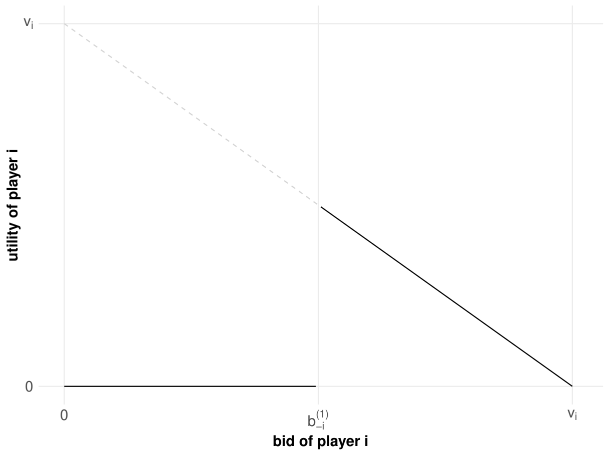


Figure 1: Utility function $u_i(b_i)$ in First Price Sealed Bid Auction for stationary opponent bids b_{-i} with highest opponent bid $b_{-i}^{(1)}$. For a given current bid b_i, the gradient $\nabla u_i(b_i)$ will be zero whenever the player is not winning the item, and negative whenever she is. Thus, when all players update their strategies using gradients, they will eventually all bid zero, as the winner in each round learns to bid less while the losers do not change their strategy.

2.4.2 Evolutionary Strategy Pseudo-gradients

On the other hand, even when available, the exact gradients on this sample may not lead to proper learning, so we rely on pseudo-gradients computed via an Evolutionary Strategy algorithm instead. Exact gradients are problematic because for a fixed valuation and opponent strategy profile (v_i, π_{-i}), player i's utility may be discontinuous in her action. Clearly, such a discontinuity is relevant to playing optimally, but neither the left-sided nor the right-sided derivatives will contain information about its presence, as outlined in Figure 1. The standard method of training neural networks, stochastic gradient ascent (SGA)[4] via backpropagation, calculates exact gradients with respect to the training data[5]; thus, using backpopagation in the multi-agent setting is simply an implementation of IGA on neural network strategies, leading to the problems described above and making it unsuitable in our setting.

Recently, Evolutionary Strategies (ES) have been proposed as an alternative to backpropagation for gradient estimation in neural networks and applied with some success in reinforcement learning (Salimans et al. 2017). In ES, the parameter vector θ of the model is perturbed randomly P times, for example by adding P i.i.d. zero-mean, σ^2-variance Gaussian noise terms ε_p. The resulting P perturbed neural networks are then evaluated with respect to their "fitness" F_p and the model is ultimately updated using a weighted average of the P noise vectors ε_p with more desirable perturbations being weighted higher than less desirable ones: $\theta^{t+1} = \theta^t + \alpha\frac{1}{P}\sum_{m=1}^{P} F_m \varepsilon_p$. While Salimans et al. (2017) mainly motivate this alternative update with the need for large scale parallelization across CPU clusters and computational deficiencies of backpropagation, the method also exhibits an important property that is crucial in our context:

[4] In the Machine Learning and Nonlinear Programming literature the method is commonly known as stochastic gradient descent (SGD). Nevertheless, we will use the maximization formulation here.

[5] Unless impaired by numerical precision.

The ES pseudo-gradient is in expectation identical to the analytical infinitesimal gradient for $\sigma \to 0$; however, in pactice, a small but strictly positive value for σ is used. The resulting *finite* perturbations solve the problem of inconsistent gradient signals at discontinuities of the utility: If an agent is 'barely' losing an auction, a small perturbation resulting in a higher bid will also result in the agent winning the auction, thus providing a positive pseudo-gradient signal. We therefore propose using neural networks trained via ES rather than backpropagation in the multi-agent continuous-action setting whenever the marginal utility functions may not be differentiable or even continuous in action-space.

In our implementation, we extend the basic ES algorithm from Salimans et al. (2017) with two common practices from Reinforcement Learning and Optimization by (a) using the player's utility in the previous iteration as a baseline parameter to reduce variance in the fitness function and (b) replacing the pseudo-gradient update with a momentum update in order to smoothen the learning trajectories. A complete description of Neural Self-Play with Evolutionary Strategy training is given in Algorithm 1.

2.4.3 Representing Strategies: The policy network

In NSP, each agent's strategy is given by a *policy model* that maps her types v_i to (a distribution over) actions and that is represented by a neural network with a parameter vector θ: $\pi_i(\cdot) = \pi_{i,\theta}(\cdot)$.

In the normal form game setting, we implement the policy model as follows: Since we have complete information, there are no information sets and no structure for the neural network to learn. Thus, the policy model for each player consists of a single weight vector representing logits for each possible action, $\theta \in \mathbb{R}^{|\mathcal{A}_i|}$. The logits are then normalized by a softmax function to achieve a vector that can be interpreted as probability distribution: $\mathbf{Pr}\left(a\right) = \frac{e^{\theta_a}}{\sum_{j\in\mathcal{A}_i} e^{\theta_j}}$. This can be interpreted as a no-hidden-layer feed-forward neural network with a constant scalar input of 1 and an output layer with weight vector θ, no bias parameters and a softmax activation function. A schematic of this network is shown in Figure 2a. Actions are then sampled from the resulting distribution.

In the Bayesian, continuous-type-and-action setting, we instead restrict ourselves to deterministic policies: The input to the neural network will be a vector representing the player's private information, the output will be a vector in action space. In the setting of sealed bid single-item auctions, both the input (private valuations v_i) and outputs (bids b_i) happen to be scalars. The deterministic action is then given by $b_i = \pi_{i,\theta}(v_i)$. To map inputs to outputs, we may use an arbitrary neural network architecture; in this study, we restrict ourselves to fully-connected feed-forward networks with two hidden layers, which were sufficient to yield desired results. Advanced network architectures such as recurrent neural networks or attention mechanisms may be required to extend this technique to settings with temporal structure (such as ascending auctions), but we leave this investigation to future work.

For the reported experiments we use *SeLU* activations (Klambauer et al. 2017) in the hidden layers, and a ReLU activation function in the ouput layer. While the ReLU activation in the output layer fulfills a structural role in ensuring non-negative bids, SeLU was chosen in the hidden layers because we found it to be most robust in producing good results. The network architecture used in auction games is illustrated in Figure 2b.

Algorithm 1: Neural Self-Play with Evolutionary Strategy training

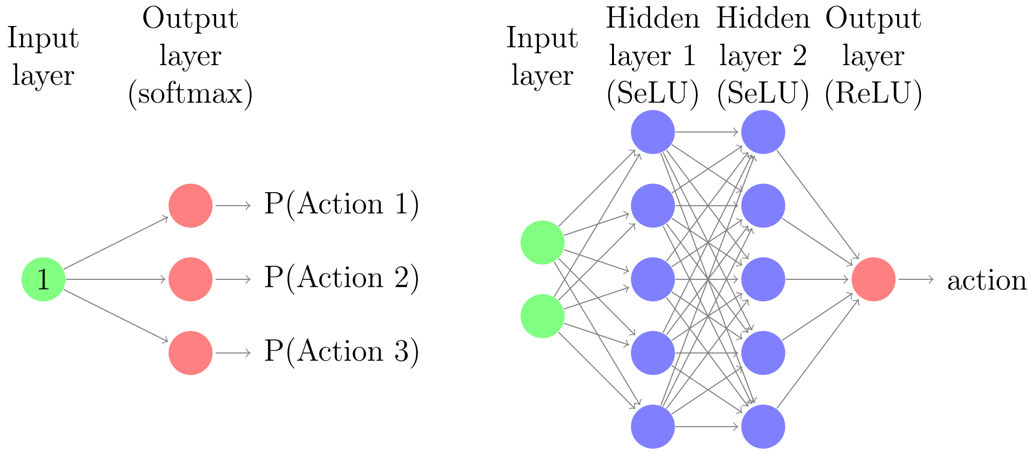
```
Input: players $i \in [N]$ with initial policy $\pi_i^0 := \pi_{i,\theta_i^0}$, defined by a model
       architecture and initial parameter vector $\theta_i^0$;
       batch size $K$; learning rate schedule $(\alpha^t)_{t\geq 1}$; friction parameter $\beta \in [0,1)$;
       ES population size $P$; ES noise stddev $\sigma$
For each player, initialize momentum buffer $m_i^0 = 0$
for $t := 1, 2, \ldots$ do
    For each player $i$, sample a batch of valuations $v_{k,i}$ for $k \in [K]$
    Calculate joint utility in current strategy profile:
        $$u^{t-1} := \frac{1}{K}\sum_k u\left(\pi^{t-1}(v_k)\right)$$
    for each player $i$ do
        Sample $P$ perturbations of player $i$'s current policy model:
            $$\tilde{\pi}_p := \pi_{i,\tilde{\theta}_p}, \qquad \text{with } \tilde{\theta}_p := \theta_i^{t-1} + \varepsilon_p, \quad \varepsilon_p \sim \mathcal{N}(0, \sigma^2 I) \text{ iid. } \forall p \in [P]$$
        Evaluate the fitness of perturbations by playing a batch vs current opponents:
            $$F_p := \frac{1}{K}\sum_k u_i\left(\tilde{\pi}_p(v_{k,i}), \pi_{-i}^{t-1}(v_{k,-i})\right) - \underbrace{u_i^{t-1}}_{\text{baseline}}$$
        Calculate ES pseudo-gradient as fitness-weighted perturbation noise:
            $$\nabla^{ES} := \frac{1}{\sigma^2 P}\sum_p F_p \varepsilon_p$$
        Perform a momentum update on the current policy:
            $$m_i^t := \beta m_i^{t-1} + \nabla^{ES}$$
            $$\theta_i^t := \theta_i^t + \alpha^t m_i^t$$
            $$\pi_i^t := \pi_{i,\theta_i^t}$$
    end
end
```

3 Empirical Results

We study the learning rules above in two settings, namely complete information normal form games and incomplete information single item sealed-bid auctions.

3.1 Normal Form Games

We consider a number of very common normal form games, starting with 2 players and 2 actions, namely Prisoners Dilemma (PD), Battle Of the Sexes (BoS), and Matching Pennies (MP). We also consider a game with 3 players and 2 actions, namely the Jordan-Game (JG) that has been considered a challenge for FP and its variants (Jordan et al. 1993). We run 10 replications of each game with randomly drawn initial beliefs that are identical for each learning rule. In each replication, each learning rule performs 5000 (learning) steps. The



(a) Normal Form Game Policy Network

(b) Auction Game Policy Network

Figure 2: neural network architectures used for normal form games and auctions.

temperature τ (SFP and MFP) is initialized with 1, updated every 10 steps with 0.9 times the previous value and held constant at a minimum of 0.2. NSP is performed with a batch size of $K = 2^{10}$, ES noise parameter $\sigma = 5$, and $P = 10$ ES perturbations per step.

3.1.1 2 Player, 2 Action

The following games are described by $N = \{1, 2\}$ and $\mathcal{A}_1 = \mathcal{A}_2 = \{1, 2\}$. Due to restricted space, we do not present each payoff matrix but only name the Nash equilibria (NE). In PD, the only NE is both players playing action 2 (PNE). In BoS, there are two PNE (both players play 1 or both players play 2) and one MNE, where player 1 has a 60% probability of playing action 1 and a 40% probability of playing action 2, while the probabilities are reversed for player 2. In MP, the unique MNE is that both players have a 50% probability of playing action 1. Figure 3 illustrates the learning process for these three games and the four learning rules. Since there are only two actions and the behaviour of player 2 is very similar to that of player 1, we display only the actual probability of player 1 playing action 1 at any learning step (columns 1-3). The fourth column displays the empirical distribution of historical probabilities of player 1 playing action 1. Let us describe the results now.

FP (row 1) quickly converges to PNE in PD and BoS (column 1-2) while it oscillates between actions in scenarios of only MNE, here MP (column 3), as described in the previous section. However, the empirical distribution (column 4) converges to the MNE. On the other hand, SFP (row 2) has a tendency of playing mixed and therefore takes about 150 steps to finally play PNE in PD and BoS (column 1-2). However, in games of only MNE, SFP converges to the MNE in actual play and not only in the empirical distribution (column 3 and 4). Actually, Fudenberg and Kreps (1993) established global convergence to a Nash distribution in 2×2 games with a unique mixed-strategy equilibrium. MFP (row 3) generally behaves like SFP, however converges much faster and much smoother, especially in MP (column 3 and 4). NSP (row 4) behaves similarly to both FP and SFP. In PD and BoS (column 1 and 2), it is similar to SFP and needs 200 steps in both to converge to the PNE. In MP (column 3), it is similar to FP and actual play cycles between the actions. However,

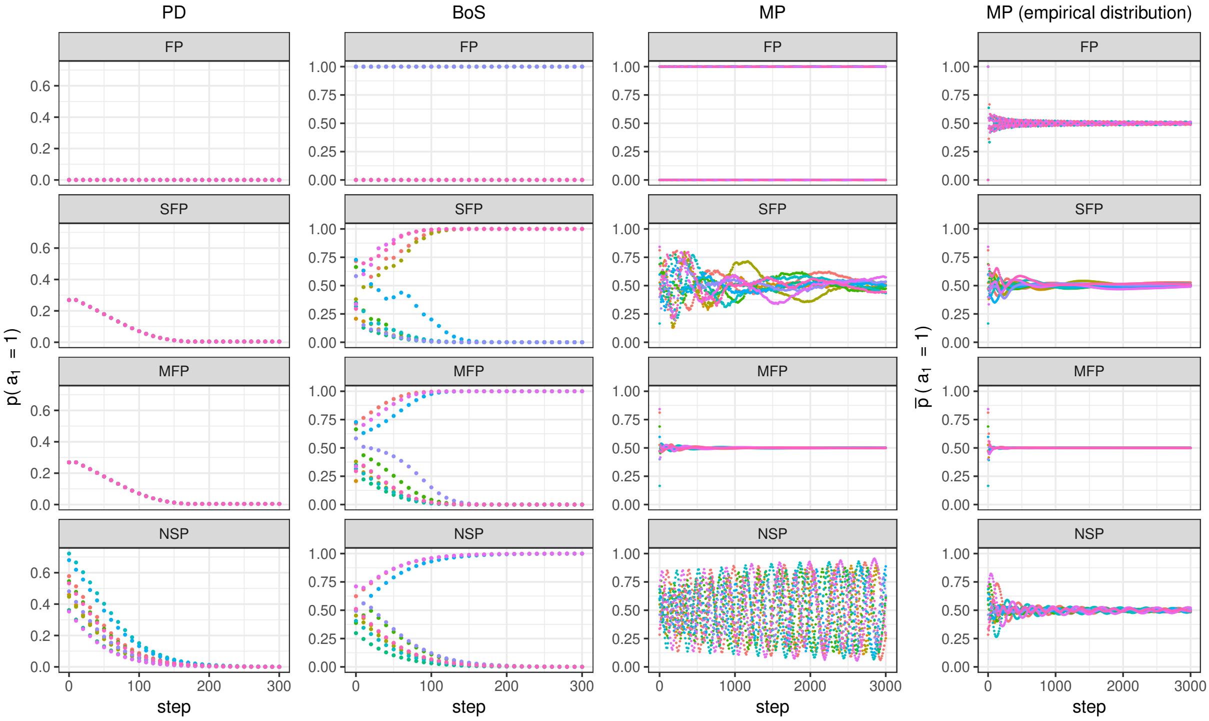


Figure 3: Learning process of player 1 to play action 1 for the four learning rules on three common 2 player 2 actions normal form games. The actual probability to play action 1 at each learning step is shown in column 1-3 and the empirical distribution of historical probabilities in column 4.

this cycling is much smoother than in FP. As in FP, the empirical distribution (column 4) converges to the MNE.

3.1.2 3 Player, 2 Action

While all previous games considered only 2 players, the Jordan-Game (JG) is defined by $N = \{1,2,3\}$ and $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{A}_3 = \{1,2\}$. In this game, player 1 wants to choose an action different to that of player 2 ($u_1 = 1$, else $u_1 = 0$), player 2 wants to choose an action different to that of player 3 ($u_2 = 1$, else $u_2 = 0$), and player 3 one that is different to player 1 ($u_3 = 1$, else $u_3 = 0$). The only NE is for all players to play each action with a probability of 0.5 (MNE). Figure 4 shows the probability of actual play (row 1) and the empirical distribution of historical probabilities (row 2) for each learning rule (columns 1-4) in the JG.

FP (column 1) oscillates in actual play (row 1). However, in contrast to MP, even the empirical distribution (row 2) does not converge but cycles around the equilibrium. Only in one repetition the empirical distribution is perfectly in the MNE. Here, the initial beliefs are such that each player beliefs all other players play action 1, and therefore each player plays action 2. In the next step, each player updates their beliefs and now beliefs all other players play action 2, and therefore each player plays action 1, etc.. Note that while all players play the MNE in the empirical distribution, the actual payoff is 0 at all times. SFP (column 2)

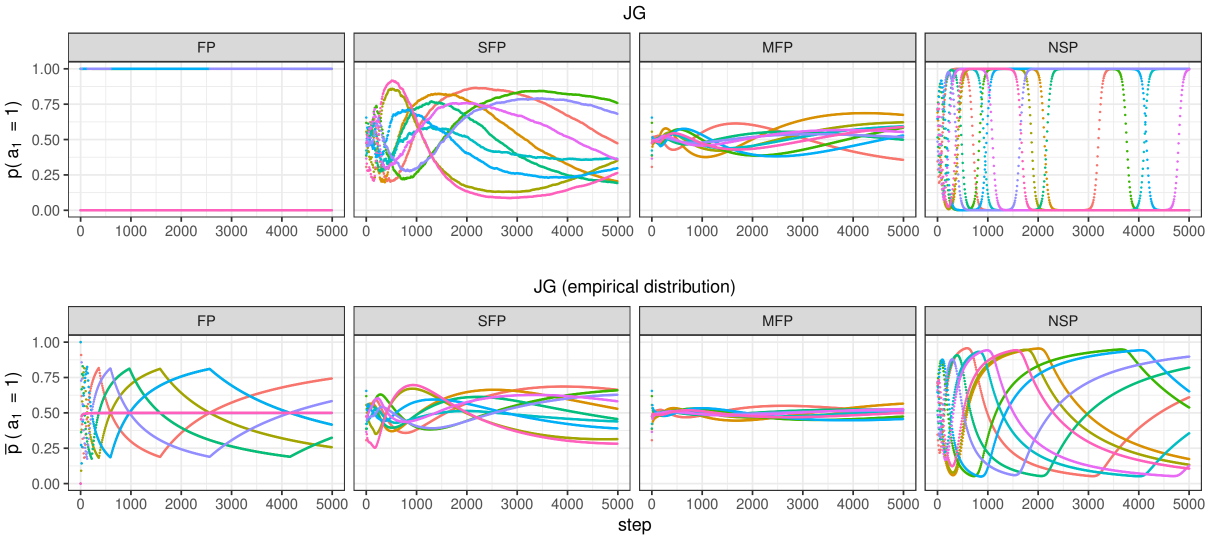


Figure 4: Convergence of actual and historical probability of player 1 to play action 1 for the four learning rules on Jordan Game

does not converge either but both actual play (row 1) and the empirical distribution (row 2) cycle around the equilibrium. This has also been shown by Benaım and Hirsch (1999). The results are the same for MFP (column 3). However, the cycles around the equilibrium are much closer here which is in line with previous results of SFP and MFP. NSP (column 4) performs similar to FP but with even larger cycles of the empirical distribution around the equilibrium. Both would not be suited to find equilibria in this game.

3.2 Single-Item Sealed-Bid Auctions

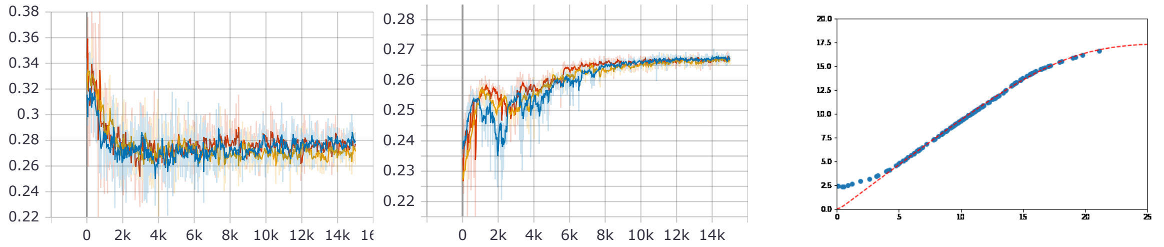
We study NSP behaviour in two types of auctions: First Price Sealed Bid (FPSB) auctions and Second Price Sealed Bid Auctions (also called Vickrey auctions).[6] The latter is well known to be incentive compatible, thus bidding truthfully constitutes a BNE for any combination of valuation distributions. NSP learned a close approxmation to the truthful strategy after just a few 100s of iterations in all Vickrey settings we performed (uniform and normal distributed types with up to 10 players). We thus omit detailed quantitative results for Vickrey auctions for brevity and instead focus on the more challenging case of FPSB auctions.

In FPSBs, analytical Bayes-Nash equilibria (BNE) are known for n players with arbitrary but symmetric prior valuations (Menezes and Monteiro 2005) as well as for 2 players with asymmetric uniform valuation distributions with a stronger and a weaker player (Plum 1992). We ran experiments in the symmetric settings with uniform and normal distributed valuations for 2, 3, 5 and 10 players each. In this setting, we take advantage of the symmetry and implement NSP with model sharing, i.e. symmetric agents share a common parameter vector θ. In this way, the for-loop in line 5 of algorithm 1 will only have to be computed once

[6] We also implemented NSP with policy gradient training via backpropagation and found that, in fact, the problematic behaviour described in Section 2.4.1 always emerges in practice.

Auction	**Valuation Priors**	**n**	**runs**	**iters**	**runtime (mins)**	**Utility in BNE**	**Utility NSP self play**	**Utility NSP vs BNE**	**Relative utility loss vs BNE (%)**
First Price	**Uniform Symmetric**	2	5	2000	7.05 (0.13)	1.667	1.659 (0.007)	1.665 (0.001)	0.083 (0.038)
		3	5	2000	7.83 (0.14)	0.833	0.827 (0.008)	0.832 (0.001)	0.208 (0.100)
		5	5	3000	13.9 (0.10)	0.333	0.335 (0.006)	0.332 (2. e-4)	0.280 (0.079)
		10	3	3000	15.5 (0.15)	0.091	0.100 (0.006)	0.089 (0.001)	2.289 (1.114)
	Uniform Asymmetric	2	6*	5000	18.1 (0.32)	weak: 0.969	0.901 (0.025)	0.958 (0.005)	1.160 (0.518)
						strong: 5.069	5.102 (0.046)	5.033 (0.007)	0.699 (0.149)
	Normal Symmetric	2	5	5000	10.7 (1.13)	2.779	2.639 (0.074)	2.758 (0.013)	0.778 (0.560)
		3	5	5000	24.5 (1.39)	1.401	1.390 (0.057)	1.398 (0.018)	0.876 (1.313)
		5	6*	10000	67.6 (5.89)	0.668	0.676 (0.013)	0.667 (0.001)	0.103 (0.149)
		10	3	15000	56.5 (1.56)	0.269	0.275 (0.013)	0.267 (0.002)	0.861 (0.565)

Table 1: Results of Neural Self-Play in FPSB auctions. For each metric, we report the mean (and standard-deviation) of multiple runs as indicated. See Table **??** for experiment hyperparameters. **Note:* In the uniform asymmetric and normal symmetric 5-player settings, one run each failed to learn (at least one player bidding constant zero). In these cases, reported results are calculated over the remaining 5 runs.



(a) Trajectory of NSP utility in self-play (left) and vs the analytical Bayes-Nash equilibrium (middle) over three runs of 15k iterations each. Opaque lines have been smoothed exponentially, actual values indicated with transparency.

(b) Sampled bids (y-axis) vs valuations (x) of learned strategy (yellow run in (a), 99.2% efficiency). BNE indicated in red.

Figure 5: NSP in 10-player FPSB with symmetric-normal valuation priors

in each time step, giving considerable speedups, especially in settings with many players. In the asymmetric setting, the weak and strong player naturally have distinct models. Each experiment was run on a single Nvidia Geforce RTX 2080Ti GPU with batch sizes chosen as large as possible such that the experiment would fit into GPU-memory.

In all of these settings, we measure players' utilities in self-play as well as when unilaterally playing against the known analytical BNE in each observation. We observe convergence of the players' utilities to those achieved in the BNE for both notions of utility in all considered settings: With rudimentary manual hyperparameter tuning, we achieve more than 97.5% efficiency in all 9 FPSB settings and more than 99% in all but two. Detailed FPSB results are presented in Table 1. Figure 5 shows selected learning behaviour and resulting policy in the settings of 10 player symmetric normal valuations.

In asymmetric settings, an interesting phenomenon can sometimes be observed in early training: Initially, one player i will often randomly play a bid strategy that dominates all other players (i.e. i wins all auctions in the batch) but that is nevertheless below the equilibrium bid level. i will then adjust to bid less globally, while other players increase

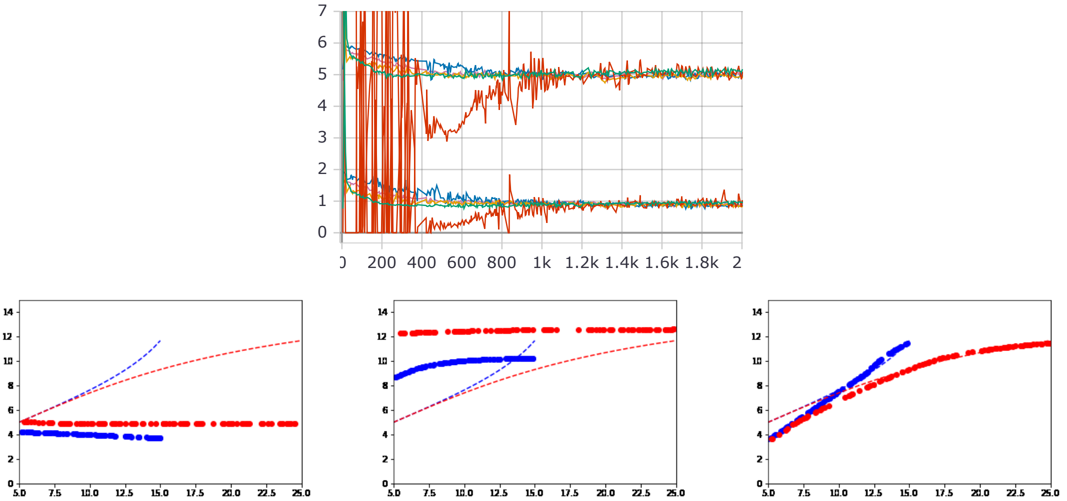


Figure 6: (top) NSP utilities (self-play) in first 2000 iterations of 5 repetitions in asymmetric valuation setting showing both the strong (BNE utility: 5.07) and weak player (0.97). (bottom) NSP strategies corresponding to the red trajectory on (top). Bidding strategies of weak (blue) and strong(red) player as learned after 200, 400 and 2000 iterations (f.l.t.r.).

their bid when they get close to winning. This is precisely the behaviour where in IGA, the ’losing’ agent will fail to adjust their bids upwards, resulting in all players bidding 0 after a while. In NSP with ES, however, we can see that due to the upward correction of the losing bidder, the level where winning and losing players ’flip’ adjusts upward over time until it reaches the equilibrium level. An example of this can be seen in Figure 6: The erratic behaviour in the red trajectories corresponds to this phenomenon and results in oscillations between achieved utilities much higher than in equilibrium (when ’winning’) and 0 (when ’losing’). Ultimately, the level of bids where these flips happen rise to amounts similar to the equilibrium at which point players learn to coordinate, each player wins a fraction of the auctions in each batch.

As expected in deep learning settings, we find that NSP behaviour is sensitive to the choice of hyperparameters in terms of runtime and performance. In our experiments, hyperparameters were chosen and tuned manually and should by no means be considered optimal for their respective settings. In particular, the choice of learning rate α and friction β were found to be of high importance. A too high learning rate (and β) can lead to oscillations around the optimum without convergence. In extreme cases large update steps even lead to a player submitting all-zero bids in one iteration. This results in a behaviour similar to a ’dead ReLU’ in backpropagation, where ES can no longer produce valid pseudo-gradient information and the player will bid constant-zero in all following iterations. On the other hand, small learning rates naturally lead to very slow convergence, especially in the setting with many (5, 10) players. A detailed overview of hyperparameters used in the experiments can be provided upon request.

It should further be noted, that even for small ϵ, an ϵ-BNE might be arbitrarily distant from an exact BNE in type-action space (compare Bosshard et al. 2017). As such, we plotted players’ policy functions over time for inspection. We often see behaviour where agents do not conform to the equilibrium strategy for low valuations, particularly in settings with high number of players (compare Figures 5b and 6 (bottom right)). This results from the fact that in equilibrium, a player with a low random valuation will almost never win an auction, and even if she does, the utility gained will be minuscule. In fact, analysis of learning behaviour shows that agents learn the correct bid-level for their highest valuation levels very quickly, then fine-tune the shape of the policy.

4 Conclusion

Nash equilibria are popular means to predict market participants' behaviour and predict market outcomes. Unfortunately, computing Nash equilibria is extremely difficult, in fact PPAD complete. In this study, we propose a new learning rule based on neural networks that we call Neural Self-Play. First, we show that this learning rule can compete with common learning rules like Fictitious and Smooth Fictitious Play in normal form games. While these common learning rules require discrete type- and action-space, we show that Neural Self-Play is able to find Bayes-Nash equilibria in auction games with continuous type- and action-space, i.e. sealed bid single-item auctions. We leverage the potential of GPUs to parallelize computations and find that Neural Self-Play scales well with an increasing number of parameters, finding approximate Bayes-Nash Equilibria in auction settings with 10 players within 10s of minutes on a single GPU.

After demonstrating the ability of Neural Self-Play to find Bayesian Nash equilibria in sealed bid single-item auctions, we plan to consider more complex auction designs in future research, including combinatorial and sequential auctions. For these complex auctions Neural Self-Play could benefit from more advanced architectures like recurrent neural networks that provide some sort of advanced memory ability. We plan to compare the results to those of Bosshard et al. (2017) who implemented a variant of FP for combinatorial auctions with continuous type- and action-space but whose method is run-time limited for settings with more than a few players or items.

References

Michel Benaım and Morris W Hirsch. 1999. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior* 29, 1-2 (1999), 36–72.

Martin Bichler, Alok Gupta, and Wolfgang Ketter. 2010. Research commentary—designing smart markets. *Information Systems Research* 21, 4 (2010), 688–699.

Vitor Bosshard, Benedikt Bünz, Benjamin Lubin, and Sven Seuken. 2017. Computing bayes-nash equilibria in combinatorial auctions with continuous value and action spaces. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI). Melbourne, Australia.*

Michael Bowling and Manuela Veloso. 2002. Multiagent learning using a variable learning rate. *Artificial Intelligence* 136, 2 (April 2002), 215–250.

George W Brown. 1951. Iterative solution of games by fictitious play. *Activity analysis of production and allocation* 13, 1 (1951), 374–376.

C. Daskalakis, P. Goldberg, and C. Papadimitriou. 2009. The Complexity of Computing a Nash Equilibrium. *SIAM J. Comput.* 39, 1 (Jan. 2009), 195–259.

Drew Fudenberg and David M Kreps. 1993. Learning mixed equilibria. *Games and Economic Behavior* 5, 3 (1993), 320–367.

Drew Fudenberg and David K. Levine. 1999. *The theory of learning in games* (2. ed.). MIT Press series on economic learning and social evolution, Vol. 2. MIT Press, Cambridge.

James S Jordan et al. 1993. Three problems in learning mixed-strategy Nash equilibria. *Games and Economic Behavior* 5, 3 (1993), 368–386.

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. 2017. Self-Normalizing Neural Networks. *arXiv preprint, arXiv:1706.02515* (June 2017).

Gian-Marco Kokott, Martin Bichler, and Per Paulsen. 2019. The beauty of Dutch: Ex-post split-award auctions in procurement markets with diseconomies of scale. *European Journal of Operational Research* 278, 1 (2019), 202–210.

Flavio M Menezes and Paulo Klinger Monteiro. 2005. *An introduction to auction theory.* OUP Oxford.

John H Nachbar. 1990. “Evolutionary” selection dynamics in games: Convergence and limit properties. *International journal of game theory* 19, 1 (1990), 59–89.

John F Nash et al. 1950. Equilibrium points in n-person games. *Proceedings of the national academy of sciences* 36, 1 (1950), 48–49.

Michael Plum. 1992. Characterization and computation of Nash-equilibria for auctions with incomplete information. *International Journal of Game Theory* 20, 4 (1992), 393–418.

Julia Robinson. 1951. An iterative method of solving a game. *Annals of mathematics* (1951), 296–301.

Aviad Rubinstein. 2016. Settling the complexity of computing approximate two-player Nash equilibria. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS).* IEEE, 258–265.

Tim Salimans, Jonathan Ho, Xi Chen, Szymon Sidor, and Ilya Sutskever. 2017. Evolution strategies as a scalable alternative to reinforcement learning. *arXiv preprint arXiv:1703.03864* (2017).

Lloyd Shapley. 1964. Some topics in two-person games. *Advances in game theory* 52 (1964), 1–29.

Satinder Singh, Michael Kearns, and Yishay Mansour. 2000. Nash Convergence of Gradient Dynamics in Iterated General-Sum Games. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)* (June 2000).

Publication B: Learning equilibria in symmetric auction games using artificial neural networks

Peer-Reviewed Journal Paper

Title: Learning equilibria in symmetric auction games using artificial neural networks

Authors: M. Bichler, M. Fichtl, S. Heidekrüger, N. Kohring, P. Sutterer

In: Nature Machine Intelligence **3**, 687–695 (2021).

Abstract: Auction theory is of central importance in the study of markets. Unfortunately, we do not know equilibrium bidding strategies for most auction games. For realistic markets with multiple items and value interdependencies, the Bayes Nash equilibria (BNE) often turn out to be intractable systems of partial differential equations. Previous numerical techniques have relied either on calculating pointwise best responses in strategy space or iteratively solving restricted subgames. We present a learning method that represents strategies as neural networks and applies policy iteration on the basis of gradient dynamics in self-play to provably learn local equilibria. Our empirical results show that these approximated BNE coincide with the global equilibria whenever available. The method follows the simultaneous gradient of the game and uses a smoothing technique to circumvent discontinuities in the ex post utility functions of auction games. Discontinuities arise at the bid value where an infinite small change would make the difference between winning and not winning. Convergence to local BNE can be explained by the fact that bidders in most auction models are symmetric, which leads to potential games for which gradient dynamics converge.

Contribution of thesis author: design, implementation and optimization of the learning algorithm and simulation framework, theoretical analysis (mathematical model, Proof of Proposition 1, stochastic derivations), empirical analysis (combinatorial auction settings), writing and revising the manuscript and Supplementary Information

References: Bichler et al. (2021)

nature machine intelligence

ARTICLES

https://doi.org/10.1038/s42256-021-00365-4

Learning equilibria in symmetric auction games using artificial neural networks

Martin Bichler ✉, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring and Paul Sutterer

Auction theory is of central importance in the study of markets. Unfortunately, we do not know equilibrium bidding strategies for most auction games. For realistic markets with multiple items and value interdependencies, the Bayes Nash equilibria (BNEs) often turn out to be intractable systems of partial differential equations. Previous numerical techniques have relied either on calculating pointwise best responses in strategy space or iteratively solving restricted subgames. We present a learning method that represents strategies as neural networks and applies policy iteration on the basis of gradient dynamics in self-play to provably learn local equilibria. Our empirical results show that these approximated BNEs coincide with the global equilibria whenever available. The method follows the simultaneous gradient of the game and uses a smoothing technique to circumvent discontinuities in the ex post utility functions of auction games. Discontinuities arise at the bid value where an infinite small change would make the difference between winning and not winning. Convergence to local BNEs can be explained by the fact that bidders in most auction models are symmetric, which leads to potential games for which gradient dynamics converge.

The literature on machine learning largely focuses on single-agent learning. Multi-agent learning has become more popular recently due to the advent of generative adversarial networks and applications in complex competitive game playing[1–3]. Although complete-information games have seen some progress, equilibrium learning for incomplete-information (also known as Bayesian) games with continuous action spaces is in its infancy. For complete-information games, the worst-case complexity of finding Nash equilibria is known[4], and a number of learning algorithms have been developed for finding equilibria in specific normal-form games such as zero-sum games[5–7]. Auctions arguably form the best-known and practically most relevant application of Bayesian games, central to modern economic theory[8,9] and with a multitude of applications in the field. The derivation of Bayes Nash equilibrium (BNE) strategies for the first- and second-price sealed-bid auction in the independent private values model led to a comprehensive theoretical framework for the analysis of single-item auctions, a landmark result of economic theory[10,11].

Although single-item auctions in this model are well understood, we only know equilibrium strategies for very few multi-item auction environments. For example, no explicit characterization of BNE strategies is known for first-price sealed-bid auctions of multiple homogeneous goods (multi-unit auctions), nor for first-price sealed-bid combinatorial auctions in which bidders can submit bids on packages of goods[11]. Value interdependencies turn out to be even more challenging[12]. In fact, very little is known about BNE strategies in standard auction formats with multiple objects for sale and value interdependencies. Even for single-object auctions, the specification of equilibria can end up in a system of partial differential equations and no closed-form solution is available[13]; however, such environments are important to understand. In fact, the Nobel Memorial Prize in Economic Sciences that was awarded to Paul Milgrom and Robert B. Wilson in 2020 highlighted their contribution to auctions with interdependent values[14].

Numerical techniques to compute BNEs can be very valuable. Although there has been substantial recent work on imperfect-information finite-dimensional extensive-form games such as Poker or other card games[15–18], relatively few papers focus on continuous-type and -action Bayesian games such as auctions. The few initial attempts make strong restrictions such as finite action spaces, single-object auctions, or independent private values with uniform priors and quasilinear utilities[19–25]. The motivation for such restrictions is the computational hardness of equilibrium computation.

We know of the existence of a mixed Nash equilibrium for finite, complete-information games and that computation is PPAD hard[4]. For Bayesian games with continuous types and actions, we neither know whether (possibly mixed) BNEs exist in the general case nor do we know how hard they are to find if they exist. Cai and Papadimitriou[26] showed that finding a BNE in simultaneous auctions for individual items and bidders with independent private values is already hard for PP, a complexity class above the polynomial hierarchy and close to PSPACE, and we know little about the complexity of finding BNEs in other multi-item auctions. Even approximating equilibria in these auction games is NP hard[26].

The theory of learning in games examines what kind of equilibrium arises as a consequence of a process in which agents are trying to maximize their own payoff by adapting to the actions played by other learning agents[27]. Research on equilibrium learning has largely focused on complete-information normal-form games. So far there is no comprehensive characterization of games that are learnable, but there are some important results. For example, it is well-known that no-regret dynamics converge to a coarse correlated equilibrium in arbitrary finite games[28–31] in their average history of play. Coarse correlated equilibria encompass the set of correlated equilibria. The latter is a non-empty convex polytope that in turn contains the convex hull of the game's Nash equilibria such that we get Nash equilibria ⊂ correlated equilibria ⊂ coarse correlated equilibria. By contrast to correlated equilibria, coarse correlated equilibria may contain strictly dominated (pure) strategy profiles with positive probability. This means that although CCEs are learnable via no-regret algorithms, they are a rather weak solution concept[32]. The question is therefore when learning dynamics converge to a Nash equilibrium. A different relaxation of Nash equilibria is given by local equilibria[33] that only

Department of Computer Science, Technical University of Munich, Garching, Germany. ✉e-mail: bichler@in.tum.de

require stability when allowing agents to make infinitesimal—rather than arbitrary—adjustments to their strategies.

Bayesian auction games have received little attention in equilibrium learning until recently. Given how hard it is to find BNEs even in simple simultaneous single-item auctions in the worst case[26], it is far from obvious that no-regret dynamics can find a BNE in continuous-type and -action Bayesian games. Recent work used deep learning for auction design[34–37] but it did not attempt to find BNEs in auctions. Challenges in computing NEs in general-sum games have also led to alternative solution concepts[38]. Apart from this, artificial intelligence and machine learning are increasingly used to predict strategic behaviour of humans[39] or outcomes of auctions in the field[40], as well as for other problems in automated market design, for example, discovery of socially optimal tax policies[41].

We introduce neural pseudogradient ascent (NPGA) as a method to learn ex ante equilibrium bid functions in symmetric Bayesian auction games with continuous-type and action-spaces. The method is generic in that it allows for different types of value interdependencies and utility functions (for example, accommodating risk aversion). Neural networks are used to represent the bid functions of the players, and the agents learn via self-play. Unfortunately, using neural self-play in this environment is not straightforward: although we assume the expected utility of the players (over the distribution of other players' types) are differentiable in the chosen action, a key challenge is that in auctions, their ex post utilities (which are based on specific realizations of types) have discontinuities. Only the latter, however, can be directly observed in the data generated from self-play. As a result, standard ways of gradient computation (that is, backpropagation from the observed data) fail and would result in constant-zero bids by all bidders. We address this problem by deriving pseudo-gradients via evolutionary strategy optimization rather than exact gradients via standard learning methods.

Given the computational hardness of BNE computations in general Bayesian auction games[26], it is not obvious that gradient-ascent schemes such as ours would converge to BNEs. To prove convergence of NPGA to local equilibria, we leverage the fact that the vast majority of auction games described in the literature assume symmetric bidders and equilibrium bid functions[11]. This leads to a potential game, and gradient dynamics converge to local Nash equilibria in potential games. Although there can also be asymmetric equilibria, such equilibria are often unnatural and the symmetry assumption encompasses a very large set of interesting auction environments. An example of such an asymmetric equilibrium is given in a second-price auction when one player bids the upper bound of the distribution whereas all of the others bid constant zero, independent of their respective private valuations.

In our experiments we illustrate NPGA via a combinatorial auction in the local–local–global (LLG) model[42], which has received considerable attention due to the use of core-selecting combinatorial auctions for spectrum sales worldwide[43]. In the LLG model, core-selecting auctions with risk-neutral bidders are known to be economically inefficient. It is one of the few multi-object auction models in which correlation among bidder valuations has been investigated analytically with quasilinear utility functions, but this is not the case for risk aversion. Yet such multi-object environments with interdependencies and non-quasilinear utility functions have not been explored in the scarce literature on equilibrium computation. Using NPGA, we can show that risk aversion mitigates the inefficiencies that arise in the equilibrium of risk-neutral bidders, while correlation among the bidders' valuations has little impact. This result is of independent interest to policymakers. In the Supplementary Information we discuss further experiments in a number of additional environments to demonstrate the versatility of the method.

To apply NPGA, we neither need to specify the equilibrium as a system of differential equations, nor do we need to derive complex conditional type distributions in settings with interdependencies. As a result, NPGA provides a convenient method to explore symmetric sealed-bid auction models and study the BNEs that arise with different types of interdependencies, distributional assumptions or different levels of risk aversion.

The algorithm

We will now introduce the necessary notation before stating the algorithm and discussing its convergence properties.

Notation. An incomplete-information or Bayesian game is given by a sextuplet $G = (\mathcal{I}, \mathcal{V}, \mathcal{O}, \mathcal{A}, f, \mathbf{u})$. Here $\mathcal{I} = \{1, \ldots, n\}$ denotes the set of agents participating in the game. The joint probability density function $f : \mathcal{V} \times \mathcal{O} \to \mathbb{R}_{\geq 0}$ describes an atomless prior distribution over agents' types, given by tuples (o_i, v_i) of observations and valuations. We make no further restrictions on f, thus allowing for arbitrary correlations; f is assumed to be common knowledge and we will denote its marginals by f_{v_i}, f_{o_i} and so on; its conditionals by $f_{v_i|o_i}$ and so on; and its associated probability measure by F. Agent i's private observation is then given as a realization $o_i \in \mathcal{O}_i$, with $\mathcal{O} = \mathcal{O}_1 \times \cdots \times \mathcal{O}_n$ being the set of possible observation profiles. Similarly, $\mathcal{V}$ denotes the set of true but possibly unobserved valuations. Crucially, we make this distinction to model interdependencies in settings beyond purely private values or purely common values. Based on o_i, the agent chooses an action or bid, $b_i \in \mathcal{A}_i$, and the set of possible action profiles is given by $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$. For each possible action and valuation profile, the vector $\mathbf{u} = (u_1, \ldots, u_n)$ of F-integrable, individual (ex post) utility functions $u_i : \mathcal{A} \times \mathcal{V}_i \to \mathbb{R}$ assigns the game outcome to each player. Ex ante (before the game), agents neither possess observations nor valuations, only knowledge about f. In the interim stage, agents also observe o_i that provide (possibly partial or noisy) information about their own v_i. Full access to the outcomes $\mathbf{u}(\mathbf{v}, \mathbf{b})$ is given only after taking actions (ex post). In our formulation, we do not assume explicit ex post access to any values (for example, $v_i, \mathbf{v}_{-i}, \mathbf{b}_{-i}$) beyond the outcome u itself. An index $-i$ denotes a partial profile of all agents but agent i.

Taking an ex ante view, players are tasked with finding strategies $\beta_i : \mathcal{O}_i \to \mathcal{A}_i$ that map observations to bids. We denote the resulting spaces of individual and joint pure strategies by $\Sigma_i \equiv \mathcal{A}_i^{\mathcal{O}_i}$ and $\Sigma \equiv \prod_i \Sigma_i$, respectively. Note that even for pure strategies, the spaces Σ_i are infinite dimensional unless $\mathcal{O}_i$ are finite (in which case they are finite-dimensional but remain infinite for continuous $\mathcal{A}_i$). We will slightly restrict ourselves to square-integrable strategies and equip Σ_i with the inner product $\langle \cdot, \cdot \rangle_{\Sigma_i} : \Sigma_i \times \Sigma_i \to \mathbb{R}$, $(\alpha, \beta) \mapsto \mathbb{E}_{\mathbf{o}} \sim f_{\mathbf{o}} \left[\alpha(\mathbf{o})^T \beta(\mathbf{o}) \right]$ and the norm $\| \beta \|_{\Sigma_i} \equiv \sqrt{\langle \beta, \beta \rangle_{\Sigma_i}}$ such that they form Hilbert spaces[44].

The primary Bayesian games we will consider are sealed-bid auctions on m indivisible items. In general combinatorial auctions, we thus have a set $\mathcal{K}$ of possible bundles of items and the valuation- and action-spaces are therefore of dimension $|\mathcal{K}| = 2^m$. We always have $o_i = v_i$ in the private values setting, whereas in the common values setting there is some unobserved constant $v_c = v_1 = \cdots = v_n$, where o_i can be considered noisy measurements of v_c. Mixed settings are likewise possible. In any case, based on bid profile $\mathbf{b}$, an auction mechanism will determine two things: (1) an allocation $\mathbf{x} = \mathbf{x}(\mathbf{b}) = (x_1, \ldots x_n)$, which constitutes a partition of m where bidder i is allocated the bundle x_i; and (2) a price vector $\mathbf{p}(\mathbf{b}) \in \mathbb{R}^n$, where the component p_i is the monetary amount bidder i has to pay to receive x_i. Formally, one may consider the individual allocations to be one-hot-encoded vectors $x_i \in \{0, 1\}^{|\mathcal{K}|}$. In the standard risk-neutral model, u_i values are then described by quasilinear (QL) payoff functions $u_i^{\mathrm{QL}}(v_i, \mathbf{b}) = (x_i(\mathbf{b}) \cdot v_i - p_i(\mathbf{b}))$, that is, by how much a player values their allocated bundle minus the price they have to pay. An extension to this basic setting includes risk aversion (RA). Here we model risk-aversion via utilities $u_i{}^{\mathrm{RA}} = \left(u_i{}^{\mathrm{QL}}\right)^\rho$

where $\rho \in (0,1]$ is the risk attitude; $\rho = 1$ describes risk neutrality, where smaller values lead to strictly concave, risk-averse transformations of u_i^{QL}. Risk aversion is an established way to explain why bidders in field studies of single-object first-price sealed-bid (FPSB) auctions bid higher than their risk-neutral counterparts in analytical BNEs[45].

For fixed-strategy profiles $\boldsymbol{\beta} \in \Sigma$, we can extend the notion of utility to the interim and ex ante stages and use this to characterize the Nash equilibria of Bayesian games: although other agents follow $\boldsymbol{\beta}$, we define agent i's interim utility as the expected utility of choosing an action b_i conditioned on o_i:

$$\bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}) = \mathbb{E}_{v_i, \mathbf{o}_{-i}|o_i}\left[u_i(v_i, b_i, \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i}))\right]. \tag{1}$$

We will also introduce the interim utility loss $\bar{\ell}$ that is incurred by not playing a best response b_i':

$$\bar{\ell}_i(o_i; b_i, \boldsymbol{\beta}_{-i}) = \sup_{b_i' \in \mathcal{A}_i} \bar{u}_i(o_i, b_i', \boldsymbol{\beta}_{-i}) - \bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}). \tag{2}$$

Then, an (interim) ϵ-Bayes Nash equilibrium (ϵ-BNE) is a strategy profile $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_n^*) \in \Sigma$ such that no agent can improve their own interim utility by more than $\epsilon \geq 0$ by unilaterally deviating from $\boldsymbol{\beta}^*$; thus, the following holds in an ϵ-BNE:

$$\forall i \in \mathcal{I},\, o_i \in \mathcal{O}_i: \quad \bar{\ell}_i\left(o_i; \beta_i^*(o_i), \boldsymbol{\beta}_{-i}^*\right) \leq \epsilon. \tag{3}$$

For $\epsilon = 0$, we will call the BNE exact, or simply drop the ϵ prefix. We will also need the ex ante utility (defined as $\tilde{u}_i(\beta_i, \boldsymbol{\beta}_{-i}) = \mathbb{E}_{o_i \sim f_{o_i}}[\bar{u}_i(o_i, \beta_i(o_i), \boldsymbol{\beta}_{-i})]$), which can be interpreted as the expected utility over all of f for a particular β_i against fixed opponents $\boldsymbol{\beta}_{-i}$. Similarly, we will define ex ante loss $\tilde{\ell}_i(\beta_i, \boldsymbol{\beta}_{-i})$ and ex ante ϵ-BNEs analogously to equations (2) and (3). Note that now we can interpret the ex ante state of the Bayesian game as a complete-information game $\tilde{G} = (\mathcal{I}, \Sigma, \tilde{\mathbf{u}})$ with an infinite-dimensional action space Σ that is identical to the strategy space of the Bayesian game. Every exact (interim) BNE also clearly constitutes an exact ex ante BNE. The reverse holds almost surely, that is, any ex ante equilibrium fulfills equation (3), except possibly on a set $O \subset \mathcal{O}$ with $F(O) = 0$. To see this, one may consider the equations $0 = \tilde{\ell}_i(\boldsymbol{\beta}^*) = \mathbb{E}_{o_i}\left[\bar{\ell}_i(o_i; \beta_i^*(o_i), \boldsymbol{\beta}_{-i}^*)\right]$ and the fact that $\bar{\ell}_i(o_i, \boldsymbol{\beta}) \geq 0$ by definition. Importantly, this almost sure equivalence of ex ante and (interim) BNEs holds for $\epsilon = 0$ but not for strictly positive ϵ: given an ex ante κ-BNE, equation (3) (with $\epsilon = \kappa > 0$) must only hold in expectation but may be violated with strictly positive probability. To delineate this difference between ex ante and interim approximate equilibria, we will write κ and ϵ to denote their respective approximation bounds.

Due to the known computational hardness of computing NEs and BNEs, one is often interested in relaxations of equilibria that may be easier to find in some circumstances. For example, in local BNEs, the loss requirement is relaxed to only consider best responses from a neighbourhood of the equilibrium strategy profile: we call $\boldsymbol{\beta}^*$ a local ex ante BNE if there exists an open set $\emptyset \neq W_i \subset \Sigma_i$ such that $\beta_i^* \in W_i$ and $\tilde{u}_i(\beta_i^*, \boldsymbol{\beta}_{-i}^*) \geq \tilde{u}_i(\beta_i', \boldsymbol{\beta}_{-i}^*)$ for all agents i and all alternative strategies $\beta_i' \in W_i$. If all utility functions u_i are strictly concave in i's action, the game admits a unique global BNE[46] and no other local BNEs.

Smoothness of the (ex post) utilities is a standard assumption in the analysis of Bayesian games[46], but this is commonly violated in auctions due to the discrete nature of $\mathbf{x}$. Instead let us introduce a weaker notion of smoothness at the interim stage, which lends itself for theoretical analysis while being consistent with auction games.

Definition 1 (interim-smooth Bayesian game). We call a Bayesian game with continuous types $\mathcal{V}_i \times \mathcal{O}_i$ and actions $\mathcal{A}_i \subseteq \mathbb{R}^K$ interim smooth if: (1) the interim utilities $\bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i})$ are continuously differentiable with respect to their second argument for each $i \in \mathcal{I}$ and any $o_i \in \mathcal{O}_i$, $\boldsymbol{\beta}_{-i} \in \Sigma_{-i}$; (2) all partial derivatives are uniformly bounded by a finite constant $Z < \infty$:

$$\forall i, o_i, \boldsymbol{\beta}_{-i}, b_i, k \in [K]: \quad \left\| \frac{\partial \bar{u}_i}{\partial b_{ik}}(o_i, b_i, \boldsymbol{\beta}_{-i}) \right\| \leq Z; \tag{4}$$

and (3) the ex post utilities are F-square-integrable: there exists $S < \infty$, such that for all $i \in \mathcal{I}$, $\boldsymbol{\beta} \in \Sigma$:

$$\mathbb{E}_{v_i, \mathbf{o}}\left[u_i\left(v_i, \beta_i(o_i), \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i})\right)^2\right] \leq S \tag{5}$$

To see why the assumption of interim differentiability is justified, consider that ex post utilities in auctions are generally piecewise smooth. Non-differentiability only occurs at the bid profiles in which the auctioneer is indifferent between multiple possible $\mathbf{x}$. In theory, one could therefore interpret the interim expected utility as a lottery over many smooth ex post utility functions that each describe a particular $\mathbf{x}$. The choice probabilities for these are given by $P(\mathbf{x}|b_i, o_i, \boldsymbol{\beta}_{-i})$, bidder i's Bayesian belief that $\mathbf{x}$ will be chosen if they bid b_i. If $\boldsymbol{\beta}_{-i}$ are continuous and f is atomless, these probabilities—and therefore the interim expected utilities as a whole—are smooth in b_i.

In interim-smooth Bayesian games, we write $\nabla \bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}) \equiv (\partial \bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i})/\partial b_{ik})_k$ and call it the interim payoff gradient. Furthermore, when G is interim-smooth, the ex ante gradients $\nabla_{\beta_i} \tilde{u}_i(\beta_i, \boldsymbol{\beta}_{-i}) \in \Sigma_i$ are also guaranteed to exist and given by the Gateaux derivatives in the Hilbert spaces Σ_i.

Finally, symmetric models are prevalent in auction theory[11]. We will call a Bayesian game symmetric if all players' $i, j \in \mathcal{I}$ marginal prior-type distributions are identical (but not necessarily independent), that is, $f_{v_i, o_i} = f_{v_j, o_j}$, as are their individual utilities (almost surely, up to tiebreaking): $u_i(\beta_i, \boldsymbol{\beta}_{-i}) = u_j(\beta_j; \boldsymbol{\beta}_{-i})$, with probability 1. The literature primarily discusses[11] equilibria that are likewise symmetric, that is, where $\boldsymbol{\beta}^* = (\beta_1^*, \beta_1^*, \ldots \beta_1^*)$. We will refer to auctions that are both symmetric and interim-smooth as symmetric and smooth auction games.

NPGA. Our numerical technique to learn BNEs, NPGA, is based on neural networks and repeated self-play, in which players continually update strategies in response to observed game outcomes, that is, all agents follow the game dynamics. By game dynamics, we mean the vector field of the simultaneous gradients of the ex ante utility functions of all players. The goal will be to find an ex ante BNE $\boldsymbol{\beta}^*$ for a continuum of observations $\mathbf{o}$ that bidders can draw. In other words, we search for a profile of equilibrium bid functions in infinite-dimensional spaces. We will first introduce the procedure in the general case before showing convergence for symmetric and smooth auction games in the 'Convergence' section.

We start by taking the infinite-dimensional, complete-information game interpretation $\tilde{G} = (\mathcal{I}, \Sigma, \tilde{\mathbf{u}})$ mentioned in the previous section. To implement gradient ascent in the Hilbert space Σ, we replace the bid functions by neural networks called policy networks that are parametrized by finite-dimensional parameter vectors $\theta_i \in \Theta_i \subseteq \mathbb{R}^{d_i}$. This lets us define a finite-dimensional approximation of $\tilde{G}$, which we will call the proxy game.

Definition 2 (proxy game). Let $G = (\mathcal{I}, \mathcal{V}, \mathcal{O}, \mathcal{A}, f, \mathbf{u})$ be a Bayesian game with ex ante utilities $\tilde{u}_i$ and let its strategy functions be implemented by neural networks: $\beta_i(o_i) \equiv \pi_i(o_i; \theta_i)$, where θ_i are the networks' parameters chosen from finite-dimensional vector spaces $\Theta_i \subseteq \mathbb{R}^{d_i}$. Set $\boldsymbol{\Theta} \equiv \prod_i \Theta_i$ and (with slight abuse of notation) write $\tilde{u}_i(\theta_i, \boldsymbol{\theta}_{-i}) \equiv \tilde{u}_i\left(\pi_i(\cdot; \theta_i), \boldsymbol{\pi}_{-i}(\cdot; \boldsymbol{\theta}_{-i})\right)$. We then call the resulting finite-dimensional complete-information game on parameters, $\Gamma = (\mathcal{I}, \Theta, \tilde{\mathbf{u}})$, the proxy game of G.

Common neural network architectures have been shown to be able to approximate any sufficiently regular function arbitrarily well[47]; thus, this choice of function approximation enables the learning of a wide variety of bid functions with minimal structural constraints. Neural networks also demonstrably achieve good performance in machine learning settings with very high-dimensional input vectors, as is the case in larger auctions with many items. Using neural networks we therefore effectively reduce the problem from finding an infinite-dimensional vector in Σ to finding finitely many (d_i) weights and biases of the neural networks, and we can now perform gradient ascent in the finite-dimensional parameter spaces.

Each agent aims to maximize the objective function of their network, which is given by $\tilde{u}_i$ and estimated via the empirical sample mean of ex post utilities of a batch of H auctions, where H is a large integer: after playing a batch of games, agents observe their utility, estimate its gradient with respect to θ_i and apply an update to θ_i that is expected to lead to an increase in utility.

Traditionally, gradient estimates in neural networks are computed via backpropagation; however, training neural networks in auction games is challenging as the ex post utility functions of individual auctions are discontinuous, leading to a failure to backpropagate gradients through the empirical objective. We solve this problem by leveraging an evolutionary strategy (ES) optimization technique that effectively smoothes the objective[48,49]. This allows us to derive an adequate estimate of the ex ante payoff gradients even under ex post non-smoothness.

Algorithm 1 (NPGA using ES gradients).

```
Input: agents $i \in \mathcal{I}$ with initial policies $\beta_i^0 := \pi_i(\cdot;\theta_i^0)$ induced
       by initial parameters $\theta_i^0$; ES population size $P$; ES noise
       standard deviation $\sigma$; learning rate $\eta$; batch size $H$
for $t := 1, 2, \ldots$ do
    Sample a batch $(\mathbf{v}_h, \mathbf{o}_h)_{h=1,\ldots,H}$ of valuation and obseravtion
    profiles from the prior $f$
    Calculate joint utility in current strategy profile:
        $\tilde{\mathbf{u}}^{t-1} := \frac{1}{H}\sum_h \tilde{\mathbf{u}}\left(\mathbf{v}_h, \boldsymbol{\beta}^{t-1}(\mathbf{o}_h)\right)$
    for each agent $i \in \mathcal{I}$ do
        Sample $P$ perturbations of agent $i$'s current policy:
            $\pi_{i;p} := \pi_i(\cdot;\theta_p)$
        with $\theta_p := \theta_i^{t-1} + \varepsilon_p$ where $\varepsilon_p \approx \mathcal{N}(0, \sigma^2 I)$ i.i.d. for all
        $p \in \{1, \ldots, P\}$
        For each $p$, evaluate the fitness of $\theta_p$ by playing against current opponents:
            $\varphi_p := \frac{1}{H}\sum_h u_i\left(v_{h,i}, \pi_{i;p}(o_{h,i}), \boldsymbol{\beta}_{-i}^{t-1}(\mathbf{o}_{h,-i})\right) - \underbrace{\tilde{u}_i^{t-1}}_{\text{baseline}}$
        Calculate ES pseudogradient as fitness-weighted perturbation noise:
            $\nabla^{\mathrm{ES}}\tilde{u}_i^{t-1} := \frac{1}{\sigma^2 P}\sum_p \varphi_p \varepsilon_p$
        Perform a gradient update step on the current policy:
            $\Delta\theta_i^t := \eta^t \nabla^{\mathrm{ES}}\tilde{u}_i^{t-1},$
            $\theta_i^t := \theta_i^{t-1} + \Delta\theta_i^t,$
            $\beta_i^t := \pi_i(\cdot;\theta_i^t)$
    end
end
```

We provide the pseudocode of NPGA in *Algorithm 1*. At each time-step t, every agent $i \in \mathcal{I}$ receives a noisy estimate $\hat{\nabla}\tilde{u}_i$ of their individual (ex ante) payoff gradient at the current strategy profile. The noise is an artefact of limited-precision Monte Carlo sampling over $\mathcal{V}$ and $\mathcal{O}$. The agents simultaneously take a step along this gradient estimate to determine the strategies for the next stage and continue playing.

Convergence. In our experimental results below and the Supplementary Information, we find that NPGA always converges very close to the global ϵ-BNE, which was surprising at first given the known results about non-convergence of gradient play to Nash equilibria in general[50], and the locality of gradient-based learning. Non-convergence can be due to conflicting utility functions of players. For example, even in simple two-player zero-sum games with one-dimensional actions, the simultaneous gradient may cycle around the Nash equilibrium[51].

A few observations help explain why NPGA converges to an approximate BNE in a wide range of auction games. First, the vast majority of models studied in the literature are symmetric auction games with symmetric equilibria (see the 'Notation' section). As a result, we no longer need to learn multiple bid functions for each bidder in NPGA, but merely a single symmetric bid function $\beta_1 \in \Sigma_1$ that optimizes the single ex ante utility function $\tilde{u}_1(\beta_1, \ldots, \beta_1)$, which serves as a potential function of the game. Any maximum β_1^* of this potential function directly yields a symmetric pure strategy ex ante BNE $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_1^*)$. For the finite-dimensional proxy game, we can formalize the claim in the following section.

Definition 3 (potential game). A complete-information game $\Gamma = (\mathcal{I}, \Theta, \tilde{\mathbf{u}})$ is an (exact) potential game[52] if there exists a potential function $\phi : \Theta \to \mathbb{R}$, s.t. for all $i \in \mathcal{I}$, $\theta_i, \theta_i' \in \Theta_i$ and $\boldsymbol{\theta}_{-i} \in \Theta_{-i}$, it holds that

$$\tilde{u}_i(\theta_i, \boldsymbol{\theta}_{-i}) - \tilde{u}_i(\theta_i', \boldsymbol{\theta}_{-i}) = \phi(\theta_i, \boldsymbol{\theta}_{-i}) - \phi(\theta_i', \boldsymbol{\theta}_{-i}). \qquad (6)$$

When the auction game is symmetric and we additionally enforce symmetric strategies by sharing a common neural network architecture $\pi(\cdot)$ and common parameter vector $\theta_i \equiv \theta_1$ among all players (symmetric NPGA); it is easy to see that with $\phi \equiv \tilde{u}_1$, the proxy game is an exact potential game. Gradient play provably converges to a pure local Nash equilibria in finite-dimensional, continuous potential games[33]. This leads us to the following proposition.

Proposition 1. In any symmetric and smooth auction game, symmetric NPGA with appropriate gradient update step sizes almost surely converges to a local ex ante κ-BNE.

A formal proof can be found in the Methods.

Empirical evaluation

We illustrate the versatility of NPGA in the context of combinatorial auctions in the well-known LLG environment, which has been an important model for the discussion about spectrum auction formats[43,53]. The NPGA model allows us to analyse how correlation and risk aversion impact the outcome in equilibrium. There are many other interesting environments one can explore. In the Supplementary Information we present further results for single-object auctions with different types of value interdependencies (including common values models), small and larger mult-unit auctions, and a larger combinatorial auction setting with eight items and six bidders. Note that even for a multi-unit auction with three items and bidders, no analytical solutions are known anymore. For single-object, multi-unit and combinatorial auctions with only a few bidders, as reported below, NPGA computes equilibria within hundreds of iterations, each taking a few seconds or less. Larger settings such as multi-unit FPSB auctions with four units and bidders or combinatorial auctions with five items and six bidders reported in the Supplementary Information converged to an approximate BNE with estimated relative utility loss of less than 1% within 15 min; however, the runtime depends on the specific model analysed (for example, the prior distribution, the number of bidders and the auction format).

The LLG model. The LLG model consists of two objects $\{1, 2\}$, two local bidders $i \in \{1, 2\}$ and one global bidder $i = 3$, with each only interested in one specific bundle (of the single object i (locals) or both objects (global)[42]. We will simply denote the valuation of each bidder's single bundle by $v_i \in \mathbb{R}$. We consider a private values (but not independent private values) setting with $o_i = v_i$, which allows for correlation. The situation is akin to spectrum sales in countries with regional spectrum licences such as Australia or Canada, where local telecoms compete against operators who provide their services nationwide, and governments have used core-selecting combinatorial auctions. The core of an auction game describes the set of outcomes such that no coalition of bidders (and possibly the auctioneer) can profitably deviate. Core-selecting auction mechanisms enforce this notion of stability by their choice of prices. Although there are hardly any game-theoretical analyses of combinatorial auctions, this model is simple enough to allow for the derivation of analytical results[54]. It was shown that with independent private values and risk-neutral bidders, core-selecting payment rules lead to considerable inefficiencies in equilibrium[42] in combinatorial auctions. The two local bidders attempt to free ride on each other. If one bidder bids less, the other has to bid more to overbid the global bidder. Due to incomplete information, both local bidders could bid too low in total and fail to outbid the global bidder, even if their combined valuations are higher than those of the global bidders. This results in an inefficient outcome. This fact has been used as an argument against core-selecting combinatorial auctions[43].

It is interesting to understand equilibria with different assumptions. For example, it is reasonable to believe that bidder valuations in spectrum auctions are correlated, because telecoms face the same downstream market. The model was recently analysed with different types of correlation[54]; however, with standard core-selecting payment rules, it turns out that correlation alone cannot mitigate the efficiency and revenue loss encountered with independent private values. Risk aversion has not yet been analysed, although it plays a role in the revenue ranking of single-object auctions. By contrast to single-object auctions, it has been unclear how risk-aversion plays out in equilibrium. If one local bidder knows that the other is risk averse and might thus bid higher, they might bid even lower as a result of this knowledge. The environment is not symmetric as there are two local bidders and a global bidder. However, the global bidder has a simple dominant strategy to bid truthfully under certain core-selecting payment rules. The gradient dynamics of the global player's network will then stably approach this dominant strategy regardless of the local bidders' behaviour, and the two local bidders can indeed be considered symmetric whenever $f_{v_1} = f_{v_2}$ and thus form a 'local potential game'. NPGA can therefore be expected to converge to a BNE despite the environment's asymmetry.

Ausubel and Baranov[54] investigate two models of correlation among local bidders' private values and derive analytical BNEs, which we will use as a baseline in our experiments. Let us define the joint prior f to be the five-dimensional uniform distribution of a latent random variable $\boldsymbol{\omega} \sim \mathcal{U}[0, 1]^5$. Then let $v_3 = 2\omega_3$ be the valuation of the global bidder and

$$v_1(\omega) = w\omega_4 + (1 - w)\omega_1, \quad v_2(\omega) = w\omega_4 + (1 - w)\omega_2 \tag{7}$$

be the valuations of the local bidders where the weight w is a random variable depending on ω_5 only. The valuations of the local bidders can be thought of as a linear combination of an individual component ω_i and a common component ω_4. Now given an exogenous correlation parameter $\gamma \in [0, 1]$, Ausubel and Bananov[54] propose two different ways to choose w such that $\mathrm{corr}(v_1, v_2) = \gamma$: the Bernoulli weights model:

$$w(\omega) = \begin{cases} 1 & \text{if } \omega_5 < \gamma, \\ 0 & \text{else}, \end{cases} \tag{8}$$

and the constant weights model (which does not require w_5):

$$w(\omega) = \begin{cases} \frac{\gamma - \sqrt{\gamma(1-\gamma)}}{2\gamma - 1} & \text{if } \gamma \neq 1/2, \\ 1/2 & \text{else}. \end{cases} \tag{9}$$

They analytically derive the unique symmetric BNE strategies for multiple bidder-optimal core-selecting payment rules including the nearest-zero (NZ), nearest-VCG (NVCG, named after the Vickrey–Clarke–Groves (VCG) payments) and nearest-bid (NB) rule in the Bernoulli weights model. These rules all choose efficient $\mathbf{x}$ (according to the submitted bids), but select different price vectors $\mathbf{p}$ from the set of core-stable outcomes. For example, the NVCG rule picks the point in the core that minimizes the Euclidean distance to the (unique) VCG payments. Similarly, the NZ point takes the origin of the coordinate system as a reference point, whereas the NB rule minimizes the distance to the vector of submitted bids $\mathbf{b}$. Only the NVCG rule has been used in spectrum sales so far. Apart from these core-selecting payment rules, we will also report the results in FPSB auctions, for which no analytical BNEs are known, as these are used in some spectrum sales[43], and in the VCG mechanism, which is not core-stable but always prescribes truthful bidding as a BNE.

Evaluation criteria. Let us discuss how we will evaluate any learned $\boldsymbol{\beta}$ to certify that it indeed constitutes an (approximate) equilibrium. This evaluation is entirely independent of the learning process of NPGA and tries to answer the question of how good a given strategy is. Whenever we encounter a setting where an analytical equilibrium $\boldsymbol{\beta}^*$ is known, we draw on it for direct comparison. In this case, we sample the BNE utility of each player, $\hat{u}_i(\boldsymbol{\beta}^*) \approx \tilde{u}_i(\boldsymbol{\beta}^*)$, as well as the utility β_i played against the BNE, $\hat{u}_i(\beta_i, \boldsymbol{\beta}^*_{-i}) \approx \tilde{u}_i(\beta_i, \boldsymbol{\beta}^*_{-i})$, with a batch size of 2^{22}. We then report the resulting relative utility loss:

$$\mathcal{L}_i(\boldsymbol{\beta}_i) = 1 - \frac{\hat{u}_i(\beta_i, \boldsymbol{\beta}^*_{-i})}{\hat{u}_i(\beta^*_i, \boldsymbol{\beta}^*_{-i})}. \tag{10}$$

We also report the probability-weighted r.m.s.e. of β_i and β^*_i in the action space, which approximates the L_2 distance $\| \beta_i - \beta^*_i \|_{\Sigma_i}$ of these two functions:

$$L_2(\beta_i) = \left(\frac{1}{n_{\text{batch}}} \sum_{o_i} \left(\beta_i(o_i) - \beta^*_i(o_i) \right)^2 \right)^{\frac{1}{2}}. \tag{11}$$

This metric circumvents the drawback of $\mathcal{L}_i$ that even a strategy with a loss very close to zero could be arbitrarily far from the actual BNE in strategy space.

When no analytical BNE is available for certification of the learned bid function, we aim to compute the ex ante utility loss $\tilde{\ell}_i(\beta_i, \boldsymbol{\beta}_{-i}) = \sup_{\beta'_i \in \Sigma_i} \tilde{u}_i(\beta'_i, \boldsymbol{\beta}_{-i}) - \tilde{u}_i(\beta_i, \boldsymbol{\beta}_{-i})$. Evaluating this supremum exactly in function space Σ_i is not tractable and approximations are computationally expensive. Our estimator $\hat{\ell}_i$ of $\tilde{\ell}_i$ relies on finding approximate interim best responses. To do so, we place an equidistant grid indexed with $w = 1, \ldots, n_{\text{grid}}$ over the action space $\mathcal{A}_i$ ranging from zero to the maximum valuation for all dimensions. For o_i and each of the alternative bids b_w, we evaluate the interim utility $\bar{u}_i(o_i, b_w, \boldsymbol{\beta}_{-i})$ against the current opponent strategy profile. This is challenging as it requires access to the distribution of i's true valuation and the opponents' observations, both conditioned on

Table 1 | Convergence results of NPGA in risk-neutral combinatorial LLG auctions with a correlation of $\gamma = 0.5$ among local bidders' valuations. We report mean and s.d. of experiments over ten runs

Auction game	L_2	$\mathcal{L}$	$\hat{\mathcal{L}}$
LLG Bernoulli NZ	0.011 (0.005)	0 (0)	0.007 (0.007)
LLG Bernoulli VCG	0.008 (0.003)	0.001 (0)	0.007 (0.005)
LLG Bernoulli NVCG	0.016 (0.016)	0 (0)	0.008 (0.007)
LLG Bernoulli NB	0.021 (0.021)	0.001 (0)	0.009 (0.008)
LLG Bernoulli FPSB	–	–	0.010 (0.008)
LLG constant NZ	–	–	0.011 (0.010)
LLG constant VCG	–	–	0.008 (0.007)
LLG constant NVCG	–	–	0.011 (0.012)
LLG constant NB	–	–	0.013 (0.015)
LLG constant FPSB	–	–	0.009 (0.006)

o_i (see equation (1)). For n_{batch} samples of o_i and n_{batch} samples of $v_i, \mathbf{o}_{-i}|o_i$ for each o_i, we then have

$$\hat{\ell}_i(\boldsymbol{\beta}) = \frac{1}{n_{\text{batch}}} \sum_{o_i} \max_{w} \lambda_i(o_i, b_w, \boldsymbol{\beta}) \tag{12}$$

with λ_i being the estimated expected utility gain by deviating from playing according to β_i to playing action b':

$$\begin{aligned} \lambda_i(o_i, b', \boldsymbol{\beta}) = \tfrac{1}{n_{\text{batch}}} \sum_{v_i, \mathbf{o}_{-i}|o_i} \Big(& u_i\left(v_i, b', \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i})\right) \\ & - u_i\left(v_i, \beta_i(o_i), \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i})\right)\Big). \end{aligned} \tag{13}$$

For an increasing number of samples and alternative actions, this estimate converges to $\tilde{\ell}_i$. Our estimate for ϵ in an ex ante ϵ-BNE is then $\epsilon \equiv \max_i \hat{\ell}_i$.

The conditional distribution $v_i, \mathbf{o}_{-i}|o_i$ is rarely available upfront. For simple cases one can derive the analytical distributions and draw samples; however, in most programming environments, one is only able to sample from very basic (pseudo)random numbers such as the uniform or normal distribution. For more complicated multivariate conditional distributions, we use the conditional distribution method (for details, see Supplementary Section 3). Based on these estimates, we can compute a relative ex ante utility loss without access to the analytical BNEs:

$$\hat{\mathcal{L}}_i(\boldsymbol{\beta}) = 1 - \frac{\hat{u}_i(\boldsymbol{\beta})}{\hat{u}_i(\boldsymbol{\beta}) + \hat{\ell}_i(\boldsymbol{\beta})}. \tag{14}$$

This metric is the average loss incurred by not playing a best response but instead playing the strategy learned via NPGA. Note that we do not need to make any assumption about the utility function or independence of valuations for this estimator.

Due to the multiple levels of Monte Carlo sampling, the estimator $\hat{\mathcal{L}}_i$ has a higher variance than those that rely on an analytical BNE $\boldsymbol{\beta}^*$, even when the performance of NPGA itself is not affected. Our reported estimates are based on $n_{\text{grid}} = 2^{10}$ possible bids for each sampled interim state using a batch size of $n_{\text{batch}} = 2^{12}$, thus each estimate of $\hat{\mathcal{L}}$ is based on $n_{\text{grid}} \cdot n^2_{\text{batch}} = 2^{34}$ simulated auctions. To sample that many games efficiently, both NPGA and our evaluation procedures leverage parallelization on GPU hardware. Certification of BNEs is a challenge in all computational approaches to equilibrium computation. A thorough discussion for environments with standard quasilinear utility functions and independent private values are provided in ref. [23].

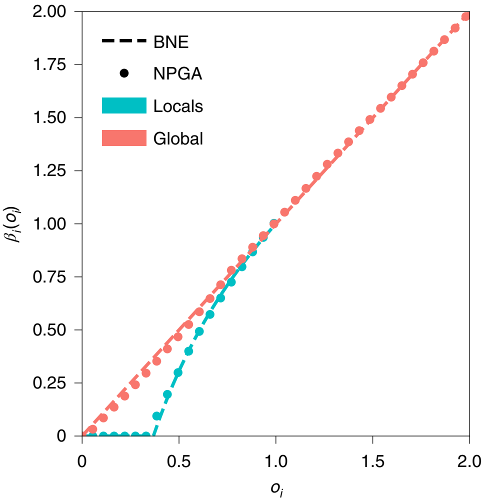


Fig. 1 | Bid functions in the LLG auction with the nearest-zero core payment rule. Bidders are independent and risk neutral. The strategies learned by NPGA (dotted) almost perfectly recover the analytical equilibrium strategies (dashed).

Results. Let us first provide the aggregate convergence results in Table 1, which almost perfectly reproduce the BNE found in ref. [54]. The utility loss is small in all environments and so is the L_2 difference to the analytical BNE wherever it is known. Figure 1 shows the analytical BNE bid function and the NPGA result for a specific setting as an illustrative example. Note that in the FPSB auction, the global bidder does not have a dominant strategy and yet we uncover his equilibrium strategy in spite of the environment being asymmetric.

Next we look at risk aversion. Figure 2 shows that with higher risk aversion, the market efficiency denoted by $\mathcal{E}$ increases for both correlation models in a similar way. Correlation of the local bidders does not influence $\mathcal{E}$ with the widespread VCG nearest payment rule at a precision of ±1% of $\mathcal{E}$. For the highest level of risk aversion of $\rho = 0.1$, $\mathcal{E}$ rose to about 98% from about 84% under risk neutrality; thus, although higher correlation of valuations does not lead to higher $\mathcal{E}$, risk aversion mitigates the efficiency loss, which is important to know for spectrum sales by governments. A similar result has previously been found for an ascending core-selecting auction with a specific tie-breaking rule[55], but the analysis could not yet be extended to the general sealed-bid case.

Similarly, the approximate revenue of the seller can be analysed. In Figure 3 we observe a strong, steady increase of the seller revenue $\mathcal{R}$ with increasing risk aversion and a slight increase with decreasing correlation between the local bidders. Different levels ρ and varying strengths of γ are plotted in the Bernoulli correlation model in the LLG setting with the NVCG payment rule. Results are similar for the constant weights correlation model. Increasing risk aversion has substantial positive impact on revenue, which is important to know for policymakers.

Discussion

Auction theory—and game theory in general—is often very sensitive to model assumptions. Although the results of early studies on auctions in the symmetric independent private values model with quasilinear bidders provided important insights, the assumptions are very restrictive[56]. Value interdependencies and changes in the

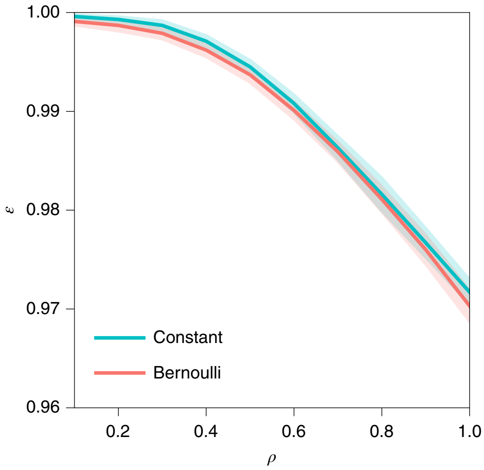


Fig. 2 | The empirical impact of risk-aversion on market efficiency. We depict the market efficiency $\mathcal{E}$ in approximate equilibrium calculated via NPGA for different levels of bidders' risk aversion. The mean (line) and s.d. (shaded bands) of ten runs for each risk-level are depicted.

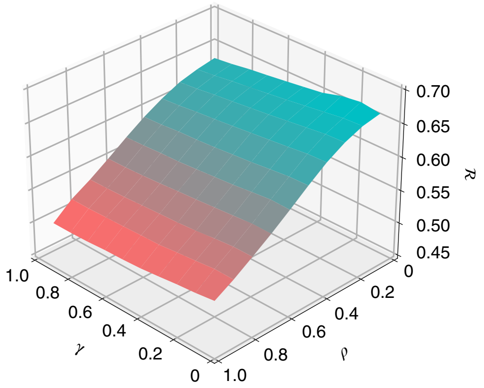


Fig. 3 | The effect of bidder correlation and risk attitudes on seller revenue. The seller revenue $\mathcal{R}$ in approximate equilibrium of LLG auctions with nearest-VCG payments and correlated bidders is shown. The underlying BNEs for each combination of the risk parameter ρ and the correlation strength γ between local bidders have been computed via NPGA.

utility function can have substantial impact on the resulting equilibrium bidding strategies. Although simple single-object auctions in the independent private values model are relatively well understood, we do not know equilibrium bidding strategies for most environments involving multiple objects, interdependencies and different levels of risk aversion to this day.

With NPGA we introduce a numerical technique to compute approximate equilibria in these Bayesian games and show that we converge to a local equilibrium quickly and with high precision. The method can provide a convenient tool for analysts to explore new environments or perform sensitivity analysis with various behavioural assumptions, different priors and value interdependencies. The Supplementary Information provides further experiments to illustrate the versatility of the method.

It is all but clear that gradient dynamics as in NPGA can find global or even local BNEs in auction games. For much simpler min–max games that play an important role in machine learning techniques such as generative adversarial networks, we cannot expect gradient dynamics to find an equilibrium[57]. Convergence of NPGA to approximate local BNEs relies on insights about the symmetry assumptions of bidders in most of the auction models in the literature and their relation to potential games. These assumptions provide the necessary structure for gradient dynamics to converge to local equilibria, and explain our results. Beyond the study of equilibria in games, our techniques can possibly contribute to automated and empirical mechanism design[58,59].

Methods

Proof of Proposition 1. Let G be a symmetric and smooth Bayesian auction game. Per definition, all players in such games have the same marginal type distributions and individual utility functions. Furthermore, assume the auction mechanism to be anonymous: the identity and order of bidders almost surely have no influence on the allocation and payments (tiebreaking on a nullset notwithstanding). Assume that all players play the same strategy β_i. Then, the symmetric ex ante utility function $\tilde{u}_i(\beta_i, \ldots, \beta_i)$ is a potential function and $\tilde{G}$ is a potential game. The same holds for the finite-dimensional proxy game Γ. To use this symmetry, we restrict all players to use the same neural network $\pi(\cdot, \theta)$ with a shared parameter vector $\theta \in \mathbb{R}^d$. Let us first remark that the restriction to symmetric strategies does not alter the gradient vector field in any way, as symmetric strategy profiles also have symmetric gradients.

We draw on a known result that gradient-play with appropriate (summable but not square-summable) step sizes converges almost surely to a local Nash equilibrium in finite-dimensional continuous potential games (see Corollary 4.2 of ref. [33]). It thus remains to be shown that (1) NPGA implements gradient-play in the proxy game Γ and thus finds a local Nash equilibrium θ^* of the proxy game, and (2) that this Nash equilibrium of the proxy game Γ—which restricts the strategy space to neural networks expressible by Θ—is indeed also a BNE of the original unrestricted game G. To show (1) and (2) below, we will rely on some auxiliary lemmata. The proofs of these lemmata are of a technical nature and can be found in Supplementary Section 2. In the following, for a given neural network $\pi(\cdot, \theta)$, we denote its utility and loss in G by $\tilde{u}(\theta)$, $\tilde{\ell}(\theta)$ and in Γ by $\tilde{u}^{\Gamma}(\theta)$, $\tilde{\ell}^{\Gamma}(\theta)$, respectively, where we drop the indices i due to symmetry.

To prove (1), one would need to show that the gradient estimates computed by NPGA have finite variance and at most a small bias with regard to the true gradients of the proxy game Γ. This is not necessarily the case, but let us set $\tilde{u}_i^{\sigma}(\theta_i, \boldsymbol{\theta}_{-i}) \equiv \mathbb{E}_{\varepsilon \sim \mathcal{N}(0,\sigma^2 I)}[\tilde{u}_i(\theta_i + \varepsilon, \boldsymbol{\theta}_{-i})]$ call $\Gamma^{\sigma} = (\mathcal{I}, \Theta, \tilde{\mathbf{u}}^{\sigma})$ the smoothed proxy game and define $\tilde{\ell}^{\sigma}$ analogously. Then Γ^{σ} is likewise a symmetric potential game and we obtain the lemmata described next.

Lemma 1. The gradient estimates ∇^{ES} in NPGA are unbiased and have finite mean squared error with respect to the smoothed utilities $\tilde{u}_i^{\sigma}$ of the game Γ^{σ}.

Lemma 2. For any $\theta \in \Theta$, the loss in Γ is bounded by that in Γ^{σ}:

$$\tilde{\ell}^{\Gamma}(\theta) \ \leq \ \tilde{\ell}^{\sigma}(\theta) + 2\mathrm{ZL}\sqrt{d}\sigma$$

where Z is the partial derivative bound from *Definition 1*, d is the number of parameters in the neural network, σ is the standard deviation of the ES perturbations, and the constant L is a property of the neural network architecture π, describing its regularity. By *Lemma 1*, NPGA implements exact gradient play in Γ^{σ} and thus finds a local Nash equilibrium θ^* of that game via the result in ref. [33]. By *Lemma 2*, any Nash equilibrium of Γ^{σ} is an approximate Nash equilibrium of Γ.

For the latter (2), the universal approximation theorem[47] guarantees that a sufficiently large neural network architecture can approximate every $\beta_i \in \Sigma_i$ with arbitrary precision δ. This yields another error bound:

Lemma 3. Let the neural network π be sufficiently expressive, that is for any $\beta_i \in \Sigma_i$ one can find $\theta \in \Theta$ such that $\| \beta_i - \pi(\cdot, \theta)\|_{\Sigma_i} \leq \delta$. Then the loss of θ in G is bounded by that in Γ: $\tilde{\ell}(\theta) \leq \tilde{\ell}^{\Gamma}(\theta) + Z\delta$.

In summary, NPGA almost surely converges to an (approximate) local Nash equilibrium $\boldsymbol{\theta}^*$ of Γ^{σ}, which, by application of local versions of *Lemma 2* and *Lemma 3*, retains a (local) ex ante loss of at most $\kappa = Z(\delta + 2\mathrm{L}\sqrt{d}\sigma)$, thus constituting a κ-BNE of G. In practice, one may choose the parameters δ (via the neural network architecture and size d) and σ sufficiently small such that the error vanishes.

Neural network architecture and hyperparameters. In our implementation, we use fully connected policy networks with two hidden layers of ten nodes each, using SeLU activation in the hidden layers and a ReLU activation function in the output layer. These simple networks are sufficient for the settings here, but even single-layer nets work with a slight decrease in performance. Instead of standard gradient ascent, we apply the Adam optimization algorithm[60] with standard parameters. In each iteration we generate 64 perturbations of the network π_i for ES gradient estimation, using zero-mean Gaussian noise with a standard deviation of $\sigma = 1/d_i$ (as suggested in ref. [49]). We use batch sizes of 2^{17} chosen such that the largest settings would fit into available GPU memory. In the presence of asymmetries or multiple items, degenerate initializations (for example, when some players never win) can impede convergence. To alleviate this and improve comparability, we force close-to-truthful initializations by pre-training the networks towards the truthful strategy using supervised learning (RMSE-loss, 500 steps of vanilla stochastic gradient descent). We did not perform setting-specific hyperparameter tuning to allow for comparable results. There are possibilities to improve the performance of our results when tuning the hyperparameters for a specific environment.

We implemented the auctions using the PyTorch framework[61] with a focus on computing many auctions in parallel. Unless noted otherwise, all experiments were performed on a single consumer-grade Nvidia GeForce RTX 2080Ti GPU with 1,000 iterations for the single-item auctions and 2,000 iterations for the large setting with correlated values ($n=10$) and the multi-unit auctions, where each experiment was run ten times.

Data availability

All data analyses in this study are based exclusively on data generated by our custom simulation framework (see Code Availability). Raw simulation artefacts (all-iteration logs and trained models) will be made available by the corresponding author on request. Source data are provided with this paper.

Code availability

The source code of our simulation framework[62], including instructions to reproduce all models and datasets referenced in this study, is freely available at https://github.com/heidekrueger/bnelearn, licensed under GNU-GPLv3.

Received: 16 December 2020; Accepted: 7 June 2021;
Published online: 9 August 2021

References

1. Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365**, 885–890 (2019).
2. Daskalakis, C., Ilyas, A., Syrgkanis, V. & Zeng, H. Training gans with optimism. Preprint at https://arxiv.org/abs/1711.00141 (2017).
3. Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362**, 1140–1144 (2018).
4. Daskalakis, C., Goldberg, P. & Papadimitriou, C. The complexity of computing a nash equilibrium. *SIAM J. Comput.* **39**, 195–259 (2009).
5. Brown, G. W in *Activity Analysis of Production and Allocation* (ed. Koopmans, T. C.) 374–376 (Wiley, 1951).
6. Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *Proc. 20th International Conference on Machine Learning* 928–936 (ICML, 2003).
7. Bowling, M. Convergence and no-regret in multiagent learning. In *Advances in Neural Information Processing Systems* 209–216 (NIPS, 2005).
8. Milgrom, P. R. & Weber, R. J. A theory of auctions and competitive bidding. *Econometrica* **50**, 1089–1122 (1982).
9. Klemperer, P. Auction theory: a guide to the literature. *J. Econ. Surveys* **13**, 227–286 (1999).
10. Vickrey, W. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* **16**, 8–37 (1961).
11. Krishna, V. *Auction Theory* (Academic, 2009).
12. Bergemann, D. & Morris, S. Robust implementation in direct mechanisms. *Rev. Econ. Stud.* **76**, 1175–1204 (2009).
13. Campo, S., Perrigne, I. & Vuong, Q. Asymmetry in first-price auctions with affiliated private values. *J. Appl. Econom.* **18**, 179–207 (2003).
14. Janssen, M. C. Reflections on the 2020 Nobel Memorial Prize awarded to Paul Milgrom and Robert Wilson. *Erasmus J. Philos. Econ.* **13**, 177–184 (2020).
15. Heinrich, J. & Silver, D. Deep reinforcement learning from self-play in imperfect-information games. Preprint at https://arxiv.org/abs/1603.01121 (2016).
16. Lanctot, M. et al. A unified game-theoretic approach to multiagent reinforcement learning. In *Proc. 31st International Conference on Neural Information Processing Systems* (NIPS, 2017).
17. Brown, N., Lerer, A., Gross, S. & Sandholm, T. Deep counterfactual regret minimization. In *Proc. 36th International Conference on Machine Learning* 793–802 (PMLR, 2019).
18. Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365**, 885–890 (2019).
19. Reeves, D. M. & Wellman, M. P. Computing equilibrium strategies in infinite games of incomplete information. *Proc. 20th Conference on Uncertainty in Artificial Intelligence* (UAI, 2004).
20. Naroditskiy, V. & Greenwald, A. *Using Iterated Best-response to Find Bayes-Nash Equilibria in Auctions* 1894–1895 (AAAI, 2007).
21. Rabinovich, Z., Naroditskiy, V., Gerding, E. H. & Jennings, N. R. Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types. *Artif. Intell.* **195**, 106–139 (2013).
22. Bosshard, V., Bünz, B., Lubin, B. & Seuken, S. Computing Bayes-Nash equilibria in combinatorial auctions with continuous value and action spaces. In *Proc. 26th International Joint Conference on Artificial Intelligence* 119–127 (IJCAI, 2017).
23. Bosshard, V., Bünz, B., Lubin, B. & Seuken, S. Computing Bayes-Nash equilibria in combinatorial auctions with verification. *J. Artif. Intell. Res.* **69**, 531–570 (2020).
24. Feng, Z., Guruganesh, G., Liaw, C., Mehta, A. & Sethi, A. Convergence analysis of no-regret bidding algorithms in repeated auctions. Preprint at https://arxiv.org/abs/2009.06136 (2020).
25. Li, Z. & Wellman, M. P. Evolution strategies for approximate solution of Bayesian games. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 35, 5531–5540 (AAAI, 2021).
26. Cai, Y. & Papadimitriou, C. Simultaneous Bayesian auctions and computational complexity. In *Proc. 15th ACM Conference on Economics and Computation* 895–910 (ACM, 2014).
27. Fudenberg, D. & Levine, D. K. Learning and equilibrium. *Annu. Rev. Econ.* **1**, 385–420 (2009).
28. Jafari, A., Greenwald, A., Gondek, D. & Ercal, G. On no-regret learning, fictitious play, and Nash equilibrium. *Proc. 18th International Conference on Machine Learning* 226–233 (ICML, 2001).
29. Stoltz, G. & Lugosi, G. Learning correlated equilibria in games with compact sets of strategies. *Games Econ. Behav.* **59**, 187–208 (2007).
30. Hartline, J., Syrgkanis, V. & Tardos, E. No-regret learning in Bayesian games. In *Advances in Neural Information Processing Systems* (eds Cortes, C. et al.) Vol. 28, 3061–3069 (NIPS, 2015); http://papers.nips.cc/paper/6016-no-regret-learning-in-bayesian-games.pdf
31. Foster, D. J., Li, Z., Lykouris, T., Sridharan, K. & Tardos, E. Learning in games: robustness of fast convergence. In *Advances in Neural Information Processing Systems* 4734–4742 (NIPS, 2016).
32. Viossat, Y. & Zapechelnyuk, A. No-regret dynamics and fictitious play. *J. Econ. Theory* **148**, 825–842 (2013).
33. Mazumdar, E., Ratliff, L. J. & Sastry, S. S. On gradient-based learning in continuous games. *SIMODS* **2**, 103–131 (2020).
34. Dütting, P., Feng, Z., Narasimhan, H., Parkes, D. & Ravindranath, S. S. Optimal auctions through deep learning. In *International Conference on Machine Learning* 1706–1715 (PMLR, 2019).
35. Feng, Z., Narasimhan, H. & Parkes, D. C. Deep learning for revenue-optimal auctions with budgets. In *Proc. 17th International Conference on Autonomous Agents and Multiagent Systems* 354–362 (AAMAS, 2018).
36. Tacchetti, A., Strouse, D., Garnelo, M., Graepel, T. & Bachrach, Y. A neural architecture for designing truthful and efficient auctions. Preprint at https://arxiv.org/abs/1907.05181 (2019).
37. Weissteiner, J. & Seuken, S. Deep learning-powered iterative combinatorial auctions. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 34, 2284–2293 (AAAI, 2020).
38. Morrill, D. et al. Hindsight and sequential rationality of correlated play. Preprint at https://arxiv.org/abs/2012.05874 (2020).
39. Hartford, J. S. *Deep Learning for Predicting Human Strategic Behavior*. Ph.D. thesis, Univ. British Columbia (2016).
40. Ghani, R. & Simmons, H. Predicting the end-price of online auctions. In *Proc. International Workshop on Data Mining and Adaptive Modelling Methods for Economics and Management* (CiteSeer, 2004).
41. Zheng, S. et al. The AI economist: improving equality and productivity with AI-driven tax policies. Preprint at https://arxiv.org/abs/2004.13332 (2020).
42. Goeree, J. K. & Lien, Y. On the impossibility of core-selecting auctions. *Theoretical Econ.* **11**, 41–52 (2016).
43. Bichler, M. & Goeree, J. K. *Handbook of Spectrum Auction Design* (Cambridge Univ. Press, 2017).
44. Debnath, L. et al. *Introduction to Hilbert Spaces with Applications* (Academic, 2005).
45. Bichler, M., Guler, K. & Mayer, S. Split-award procurement auctions-can bayesian equilibrium strategies predict human bidding behavior in multi-object auctions? *Prod. Oper. Manag.* **24**, 1012–1027 (2015).
46. Ui, T. Bayesian nash equilibrium and variational inequalities. *J. Math. Econ.* **63**, 139–146 (2016).
47. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4**, 251–257 (1991).
48. Wierstra, D. et al. Natural evolution strategies. *J. Mach. Learn. Res.* **15**, 949–980 (2014).
49. Salimans, T., Ho, J., Chen, X., Sidor, S. & Sutskever, I. Evolution strategies as a scalable alternative to reinforcement learning. Preprint at https://arxiv.org/abs/1703.03864 (2017).
50. Benaím, M., Hofbauer, J. & Sorin, S. Perturbations of set-valued dynamical systems, with applications to game theory. *Dyn. Games Appl.* **2**, 195–205 (2012).
51. Letcher, A. et al. Differentiable game mechanics. *J. Mach. Learn. Res.* **20**, 1–40 (2019).
52. Monderer, D. & Shapley, L. S. Potential games. *Games Econ. Behav.* **14**, 124–143 (1996).
53. Bünz, B., Lubin, B. & Seuken, S. Designing core-selecting payment rules: a computational search approach. In *Proc. 2018 ACM Conference on Economics and Computation* 109 (ACM, 2018).

54. Ausubel, L. M. & Baranov, O. Core-selecting auctions with incomplete information. *Int. J. Game Theory* **49**, 251–273 (2019).
55. Guler, K., Bichler, M. & Petrakis, I. Ascending combinatorial auctions with risk averse bidders. *Group Decis. Negot.* **25**, 609–639 (2016).
56. Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B. & Zame, W. R. The limits of ex post implementation. *Econometrica* **74**, 585–610 (2006).
57. Daskalakis, C., Skoulakis, S. & Zampetakis, M. The complexity of constrained min–max optimization. In *Pro. 53rd Annual ACM SIGACT Symposium on Theory of Computing* 1466–1478 (STOC, 2021).
58. Vorobeychik, Y., Reeves, D. M. & Wellman, M. P. Constrained automated mechanism design for infinite games of incomplete information. In *Proc. 23rd Conference on Uncertainty in Artificial Intelligence* 400–407 (UAI, 2007).
59. Viqueira, E. A., Cousins, C., Mohammad, Y. & Greenwald, A. Empirical mechanism design: designing mechanisms from data. In *Proc. 35th Uncertainty in Artificial Intelligence Conference* 1094–1104 (PMLR, 2020).
60. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2015).
61. Paszke, A. et al. Automatic differentiation in pytorch. In *31st Conference on Neural Information Processing Systems* (NIPS, 2017).
62. Heidekrüger, S., Kohring, N., Sutterer, S. & Bichler, M. *bnelearn: A Framework for Equilibrium Learning in Sealed-Bid Auctions* (Github, 2021); https://github.com/heidekrueger/bnelearn

Acknowledgements

We are grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; grant no. BI 1057/1-9). We thank V. Bosshard, B. Lubin, P. Mertikopoulos, P. Milgrom, S. Seuken, T. Ui, F. Maldonado and participants of the NBER Market Design Workshop 2020 for valuable feedback on earlier versions.

Author contributions

M.B. conceived and supervised the project and contributed to the overall study design, theoretical analysis of NPGA and writing the manuscript. M.F. contributed to the theoretical analysis of NPGA. S.H. contributed to the design, implementation and optimization of the algorithm and simulation framework, the theoretical analysis, and to the writing of the manuscript. N.K. contributed to the optimization of the algorithm, the design, implementation and optimization of the simulation framework, the theoretical and empirical analysis, and the writing of the manuscript. P.S. contributed to design and implementation of the algorithm and simulation framework, and the empirical analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42256-021-00365-4.

Correspondence and requests for materials should be addressed to M.B.

Peer review information *Nature Machine Intellligence* thanks Pierre Baldi, Neil Newman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature portfolio

https://doi.org/10.1038/s42256-021-00365-4

Supplementary information

Learning equilibria in symmetric auction games using artificial neural networks

In the format provided by the
authors and unedited

Supplementary Information: Learning Equilibria in Symmetric Auction Games using Artificial Neural Networks

Martin Bichler*, Max Fichtl, Stefan Heidekrüger, Nils Kohring, Paul Sutterer
Department of Computer Science, Technical University of Munich, 85748 Garching, Germany, bichler@in.tum.de

This supplementary material includes (a) the description and discussion of further experimental results to further substantiate the empirical claims made in the main paper, (b) formal statements and proofs of two auxiliary technical lemmata that are used in the proof of Proposition 1 of the main paper, and (c) mathematical derivations of the conditional distributions $o_j|o_i$ of type signals that are required for evaluation of candidate strategies in the settings where players' types are correlated.

S.1. Additional Experiments

In this section, we illustrate selected auction models to demonstrate the versatility of NPGA and its performance in larger auction models. A comprehensive analysis of the scalability of NPGA is challenging, because the runtime depends very much on the specifics of a model, the prior distribution, the number of bidders, their utility functions, the auction format and whether symmetry is itself learned or enforced a-priori. However, the following results of different auction games provide a better understanding of this question.

The independent private values model is the standard model and has been analyzed extensively in the literature [4]. More challenging environments are auctions with value interdependencies where known BNE strategies are rare. We first discuss standard single-item auction models in the independent private values model, but increase the number of bidders to study runtime and solution quality with different priors. Next, we investigate single-object auctions with value interdependencies, before we discuss multi-unit auctions, and a larger version of the combinatorial LLG model with more items and bidders. The notation follows the main paper. Note that in all our experiments we ended up in the same BNE even if NPGA was run repeatedly with different initialization, which suggests that the equilibria found are global and not local BNE. This is consistent with the well-known observation that in optimization of neural networks one is often able to find global optima even though theoretical guarantees only extend to local optimality.

Table S.1 Results of NPGA learning in single-item first-price auctions with symmetric bidders. We show average and standard deviation over ten runs.

Auction game	Bidders	$\mathcal{L}$	sec/iter
Uniform risk-neutral $\mathcal{U}(0,10)$ $\rho=1$	**2**	0.0001 (0.0009)	0.31
	3	0.0017 (0.0006)	0.40
	5	0.0034 (0.0020)	0.46
	10	0.0084 (0.0110)	0.73
Uniform risk-averse $\mathcal{U}(0,10)$ $\rho=0.5$	**2**	0.0011 (0.0004)	0.46
	3	0.0006 (0.0003)	0.52
	5	0.0012 (0.0011)	0.63
	10	0.0100 (0.0068)	0.92
Gaussian risk-neutral $\mathcal{N}(15,100)$ $\rho=1$	**2**	0.0015 (0.0011)	0.31
	3	0.0037 (0.0043)	0.39
	5	0.0129 (0.0135)	0.44
	10	0.0314 (0.0212)	0.68

S.1.1. Single-Object Auctions with Independent Private Values

We first ran experiments on single-object auctions with analytically known BNE, i. e. with uniform and Gaussian distributed valuations for 2, 3, 5 and 10 bidders each. In the uniform-prior case, we consider risk-neutral $\rho = 1$ and risk-averse $\rho = 0.5$ bidders, for Gaussian priors we only consider risk-neutral bidders. Table S.1 presents the utility loss incurred when playing a learned strategy against the analytical BNE after 20,000 iterations. In order to assess runtime, we report the time per iteration, because it varies depending on the number of bidders and the prior distribution. Although we ran all settings for the same total number of iterations regardless ob difficulty, we observe fast convergence for uniformly distributed valuations within a few hundreds of iterations. For normal distributed valuations, learning is slower, as illustrated in Figure S.1, yet the utility loss is low and stable after 18,000 iterations. It is harder to get high precision in the tails of the value distribution which are rarely sampled.

S.1.2. Single-Item Auctions with Interdependencies

Next, we report the performance of NPGA in single-object auctions with different types of interdependencies. The most well-known examples of interdependencies are the common value model (with conditionally independent observations $o_i|v$) and the affiliated value model for single-item auctions by the 2020 Nobel laureates Robert B. Wilson [8] and Paul Milgrom [5].

The *common value model* is also known as the “mineral rights” model [4, Example 6.1]. We explore the second-price auction in an environment where there is one pure common value v that is the same for all agents. Three bidders $i \in \{1, 2, 3\}$ share a common $\mathcal{U}[0, 1]$-distributed value for the item of interest. Conditioned on this value, the observations o_i are uniformly, and independently, distributed on the interval from zero to two-times the common value. Formally, we can define the joint prior probability density function f as the four-dimensional uniform distribution over $\Omega = [0, 1]^4$. For a draw

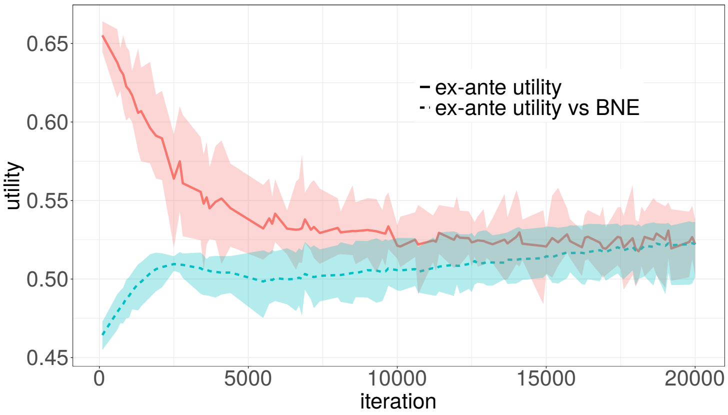


Figure S.1 **Learning curve of NPGA in a 10 player Gaussian first-price auction. Utility of learning opponents against each other (solid red line) and NPGA utility of learning opponents individually evaluated against the analytical BNE strategy (dashed blue line). Line and shaded area indicate mean, minimum, maximum of 10 runs.**

$\omega \sim \mathcal{U}(\Omega)$ we set each player's valuation to $v_i(\omega) = \omega_4$ and each observation to be $o_i(\omega) = 2 \cdot \omega_i \cdot \omega_4$. Notice, all agents have the same value (or type), but they only learn their value if they win the auction. In this model, the symmetric BNE strategy profile can be stated in closed form as

$$\beta_i^*(o_i) = \frac{2o_i}{2+o_i}. \tag{S.1}$$

For this setting, all functions required for the calculation of the utility loss from equation (10) of the main paper can be derived analytically, thus allowing for precise sampling.

In the *affiliated values model* the individual observations are correlated. In the model [4, Example 6.2] with two bidders $i \in \{1,2\}$, we can set $\Omega = [0,1]^3$ and again with $\omega \sim \mathcal{U}(\Omega)$ the observations are given by

$$o_i(\omega) = \omega_i + \omega_3 \tag{S.2}$$

where both bidders have a common value of $v(\omega) = \frac{1}{2}(\omega_1 + \omega_2) + \omega_3$. The symmetric BNE strategy is to bid truthfully under the second-price payment rule, and to follow $\beta_i^*(o_i) = \frac{2}{3}o_i$ for first-price payments.

Table S.2 shows that for single-item auctions with affiliated or common values, NPGA closely approximates the BNE. The true utility loss $\mathcal{L}$ is very low, and so is the L_2 norm of the bid function learned via NPGA compared to the analytical BNE bid function. The more conservative values of $\hat{\mathcal{L}}$ of 11% compared to $\mathcal{L}$ in the common values settings are due to numerical instabilities in the calculated actual utility to estimated BNE utility ratio: The agents have a near-zero utility in these specific games and the values are estimated on a smaller sample size thus having a higher variance.

Table S.2 Single-item auctions. Mean and standard deviation of experiments over ten runs each. Missing entries are due to a lack of an analytical BNE strategy.

Auction game	L_2	$\mathcal{L}$	$\hat{\mathcal{L}}$
Affiliated values	0.018 (0.009)	0.002 (0.001)	0.013 (0.004)
Common value	0.009 (0.002)	0.000 (0.000)	0.025 (0.013)
Common value $n = 10$	–	–	0.068 (0.063)

Note that $\mathcal{L}$ is even negative sometimes, which is an artefact of limited measurement accuracy at the batch size of 2^{22} games played against equilibrium opponents.

The numbers in Table S.2 assume risk neutrality. We do not report further details on different risk attitudes, because they lead to similar level of efficiency and revenue: Efficiency is always close to 100% and revenue is approximately 0.35 and 0.80 in the common value setting and the affiliated values setting, respectively. Overall, NPGA achieved high precision in a large number of single-item auction environments analyzed with different prior distributions, beyond the ones reported here.

S.1.3. Multi-Unit Auctions with and without Interdependencies

Multi-unit auctions in which bidders compete for $m > 1$ homogeneous units are wide-spread in practice. The standard payment rules for selling multiple units include "pay-your-bid" (first-price), Vickrey-pricing, and uniform-pricing (all items are sold at the same price). In each of the auctions, the items are awarded to the bidders corresponding to the p-highest bids. Each bid-component corresponds to the bidders' willingness to pay for one additional unit.

Even for the IPV model, equilibria are only known for small and stylized settings [4]. For example, there is no closed-form solution for the first-price or uniform pricing rule, except for the independent private values model and $n, m \leq 2$. Before we discuss interdependencies, we analyze the standard symmetric multi-unit auctions with independent private valuations and larger number of items and bidders. We will follow the common practice to draw the valuations $v_i \in [0, 1]^m$ for all units uniformly from the unit interval and sort them in decreasing order, to account for marginally decreasing valuations in the number of units. A detailed introduction to these standard multi-unit auctions can be found in [4, Chapter 13].

Table S.3 provides the results for multi-unit auctions with risk-neutral bidders, independent private values, and different auction formats. We provide the results with independent private values as a baseline, before we look at interdependent values. Again, missing entries in the table are due to a lack of an analytical BNE strategy for the respective environments. The estimated relative utility loss $\hat{\mathcal{L}}$, consistently decreases to about 1% within 15 minutes. For the VCG $m = n = 4$ auction we observe a higher L_2 but low $\mathcal{L}$ and $\hat{\mathcal{L}}$ values. This is due to the Monte Carlo estimation of the estimated utility: Agents never win all four items during the learning phase, and therefore do not bid for the last item, even though they should just bid truthful in theory.

Table S.3 **Multi-unit auctions. Mean and standard deviation of experiments over ten runs each.**

Auction game	L_2	$\mathcal{L}$	$\hat{\mathcal{L}}$
FPSB $m = n = 2$	0.077 (0.009)	0.021 (0.005)	0.033 (0.005)
Uniform $m = n = 2$	–	–	0.000 (0.000)
VCG $m = n = 2$	0.029 (0.002)	-0.000 (0.000)	0.006 (0.003)
FPSB $m = n = 4$	–	–	0.072 (0.011)
Uniform $m = n = 4$	–	–	0.000 (0.000)
VCG $m = n = 4$	0.143 (0.037)	0.006 (0.002)	0.015 (0.012)

Interdependencies have received little attention in the literature on multi-unit auctions. Several incentive-compatible mechanisms were proposed for the multi-unit case with interdependencies [2, 6], but BNE strategies for wide-spread first-price auctions are unknown. Here, we report the results of a specific environment where valuations are equal to the observations, but there is correlation among the valuations. The correlation then comes from a shared component that is weighted with a private component in the following way:

$$v_i = o_i = \gamma\omega_{n+1} + (1-\gamma)\omega_i. \tag{S.3}$$

Here, $\gamma \in [0,1]$ is the correlation strength and $\omega_i, \omega_{n+1} \in [0,1]^m$ are the private and a public component, respectively, that are once again uniform random variables as in the IPV model.

Similar to the analysis of combinatorial auctions in the LLG model, it is interesting to look at comparative statics wrt. risk aversion and correlation of bidder valuations. Let us first look at the revenue $\mathcal{R}$ that the seller can expect for different levels of risk aversion of the bidders and payment rules (first-price, VCG, uniform) in multi-unit auctions with independent private values. Risk is modeled by the risk parameter $\rho > 0$, where $\rho = 1$ corresponds to risk neutrality. NPGA can handle risk aversion without modifications to the algorithm, just by changing the utility functions appropriately. Figure S.2 shows the revenue for the common payment rules with different levels of risk aversion. The zero revenue for the uniform pricing scheme is to be expected for risk-neutral bidders, because of tacit collusion and demand reduction in equilibrium [1].

In Figure S.3 we analyze the impact of correlation on the multi-unit FPSB auction with risk-neutral bidders. Bars mark the standard deviation over four runs. An interesting phenomenon occurs at high levels of correlation. For $\gamma < 0.8$ the bidders roughly bid half the value for both units. For γ larger than this threshold, the bidders collude and only bid a small amount on winning one unit and zero on the additional unit. Thus the revenue drops to very small amounts for the seller. This phenomenon can also be seen in slightly different correlation models, even if less pronounced. For example, if we draw two valuations and then use a linear combination depending on the correlation strength, we get a similar result. The extreme case of a perfect correlation gives an intuitive explanation. If two

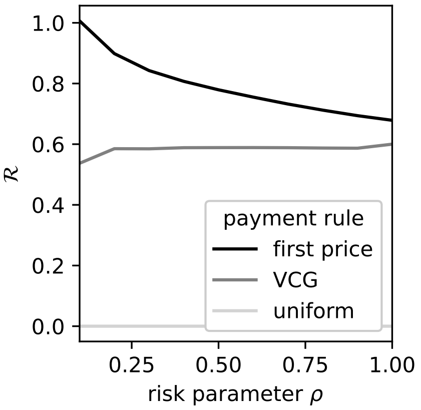


Figure S.2 **Comparison of the revenue $\mathcal{R}$ in a $n = m = 2$ multi-unit auctions for different levels of risk aversion.**

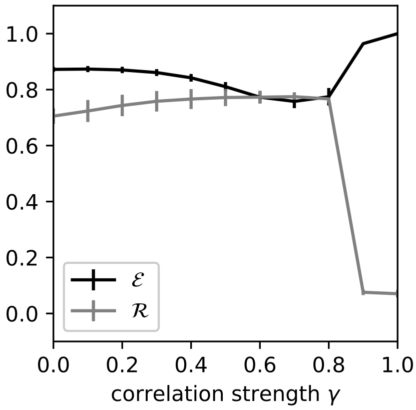


Figure S.3 **Impact of correlation on revenue and efficiency in a $n = m = 2$ correlated multi-unit FPSB auction.**

bidders have exactly the same value they win only half of the time with random tie breaking. In this extreme case, they fare better if they tacitly collude and bid only a low price on one of the items and zero on the other. Both bidders are symmetric and any higher bid price would only reduce their revenue at a 50% chance of winning. The phenomenon illustrates the value of comparative statics in game-theoretical analysis and how NPGA can help analysts study different auction institutions and model assumptions.

S.1.4. Larger Combinatorial Auctions

Finally, we used the well-known LLG model for combinatorial auctions in the main paper, but we also expanded this environment to more items and more local bidders to understand the impact on runtime. Again, for the local bidders to win, the total sum of all their bids must thus exceed the amount of the global bid. For a fair competition in the experiments, we increased the valuation of the global bidder such that in expectation she has the same valuation as the local bidders combined. The bidding strategies are in line with those observed for the LLG model with two items and three bidders only. Figure S.4 depicts that the market efficiency slightly decreases with more local bidders from about 97% in the original LLG setting with two local bidders, to about 95% in the setting with five local bidders and a correlation of $\gamma = 0.5$. When increasing the number of bidders to a total of $n = 6$ (one global and five local bidders) NPGA is still able to learn in these larger markets as fast as the model with only three bidders.

S.2. Auxiliary Lemmata

For our formal analysis in the proofs below, we assume that any neural network architecture is (a) sufficiently regular and (b) achieves universal approximation:

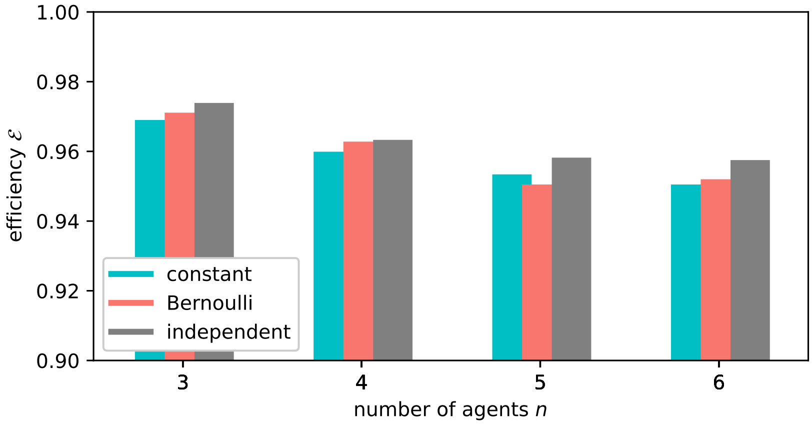


Figure S.4 **Efficiency in combinatorial auctions with increasing numbers of items and bidders. For the two correlated models, i. e. the constant weights model and the Bernoulli weights model, a correlation strength of $\gamma = 0.5$ is assumed and the selected pricing rule for all settings is nearest-VCG.**

Definition 1 (NPGA policy network). An NPGA policy network $\pi_i : \mathcal{O}_i \times \Theta_i \to \mathcal{A}_i$ is a neural network, with the following properties:

1. *Lipschitz-continuous dependence of the network on its parameters:* The network π_i depends Lipschitz-continuously on the parameters θ_i in the following sense: There exists some $L > 0$, such that for all $i \in \mathcal{I}$ and θ_i, θ_i' we have

$$\mathbb{E}_{o_i}\left[\|\pi_i(o_i,\theta_i) - \pi_i(o_i,\theta_i')\|\right] \leq L\|\theta_i - \theta_i'\|. \tag{S.4}$$

2. *Approximability of Σ_i by Θ_i:* There exists some $\delta > 0$, such that for all $i \in \mathcal{I}$ and $\beta_i \in \Sigma_i$ there exist parameters $\theta_i \in \Theta_i$, such that

$$\mathbb{E}_{o_i}\left[\|\beta_i(o_i) - \pi_i(o_i,\theta_i)\|\right] \leq \delta. \tag{S.5}$$

Let us now prove the three auxiliary lemmata from the main text.

Lemma 1. *The gradient estimates ∇^{ES} in NPGA are unbiased and have finite mean squared error with respect to the smoothed utilities $\tilde{u}_i^\sigma$ of the game Γ^σ.*

Proof: We consider the smoothed ex-ante utility $\tilde{u}_i^\sigma$. For fixed $\sigma > 0$, we have

$$\tilde{u}_i^\sigma(\theta_i,\theta_{-i}) := \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\sigma^2 I)}[\tilde{u}_i(\theta_i+\varepsilon,\theta_{-i})].$$

This is equal to the convolution of $\tilde{u}_i$ with a Gaussian kernel in the i-th coordinate. As was noted by [7], its (exact) gradient with respect to θ_i is thus given by

$$\nabla_{\theta_i}\tilde{u}_i^\sigma(\theta_i,\theta_{-i}) = \frac{1}{\sigma}\,\mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}[\varepsilon(\tilde{u}_i(\theta_i+\sigma\varepsilon,\theta_{-i}) - \tilde{u}_i(\theta_i,\theta_{-i}))].$$

By the substitution $\varepsilon' = \sigma\varepsilon$, we see by the transformation formula that

$$\nabla_{\theta_i} \tilde{u}_i^\sigma(\theta_i, \theta_{-i}) = \frac{1}{\sigma^2}\, \mathbb{E}_{\varepsilon \sim \mathcal{N}(0,\sigma^2 I)}[\varepsilon(\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i}))].$$

If we approximate this term by taking P independent samples $\varepsilon_p \sim \mathcal{N}(0, \sigma^2 I)$, we get

$$\nabla_{\theta_i} \tilde{u}_i^\sigma(\theta_i, \theta_{-i}) \approx \frac{1}{P\sigma^2} \sum_p \varepsilon_p(\tilde{u}_i(\theta_i + \sigma\varepsilon_p, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i})).$$

In the same way, we can approximate $\tilde{u}_i$ by sampling H observation and valuation profiles v_h with respect to the distribution the valuations are drawn from:

$$\tilde{u}_i(\theta_i + \sigma\varepsilon_p, \theta_{-i}) \approx \frac{1}{H} \sum_h u_i(v_{h,i}, \pi_i(o_{h,i}, \theta_i + \sigma\varepsilon_p), \pi_{-i}(o_{h,-i}, \theta_{-i})).$$

The combination of these approximations is exactly how ∇^{ES} is computed in Algorithm 1:

$$\begin{aligned}\nabla^{ES} \tilde{u}_i(\theta_i, \theta_{-i}) = \frac{1}{PH\sigma^2} \sum_p \varepsilon_p \sum_h \big(& u_i\left(v_{h,i}, \pi_i(o_{h,i}, \theta_i + \sigma\varepsilon_p), \pi_{-i}(o_{h,-i}, \theta_{-i})\right) \\ &- u_i\left(v_{h,i}, \pi_i(o_{h,i}, \theta_i), \pi_{-i}(o_{h,-i}, \theta_{-i})\right)\big).\end{aligned}$$

Since we sample independently and with respect to the original distributions, the approximation is in expectation equal to the true gradient. Thus, the approximation is unbiased with respect to the smoothed utilities $\tilde{u}_i^\sigma$. ∇^{ES} also has finite mean squared error: Define

$$X_{p,h} = \varepsilon_p \left(u_i(v_{h,i}, \pi_i(o_{h,i}, \theta_i + \varepsilon_p), \pi_{-i}(o_{h,-i}, \theta_{-i})) - u_i(v_{h,i}, \pi_i(o_{h,i}, \theta_i), \pi_{-i}(o_{h,-i}, \theta_{-i}))\right).$$

Because of equation (5) in Definition 1 in the paper (smooth Bayesian game), we have

$$\mathbb{E}_v[u_i(v_{h,i}, \pi_i(o_{h,i}, \theta_i + \varepsilon_p), \pi_{-i}(o_{h,-i}, \theta_{-i}))^2] \leq S$$

and

$$\mathbb{E}_v[u_i(v_{h,i}, \pi_i(o_{h,i}, \theta_i), \pi_{-i}(o_{h,-i}, \theta_{-i}))^2] \leq S.$$

This implies $\mathbb{E}[X_{p,h}^2] \leq 4S\,\mathbb{E}[\|\varepsilon\|^2] = 4Sd_i\sigma^2$, where we used the inequality $(a-b)^2 \leq 2a^2 + 2b^2$. Since $\nabla^{ES}\tilde{u}_i(\theta_i, \theta_{-i}) = \frac{1}{PH\sigma^2}\sum_{p,h} X_{p,h}$, we have that

$$\begin{aligned}\mathbb{E}\left[\nabla^{ES}\tilde{u}_i(\theta_i, \theta_{-i})^2\right] &= \frac{1}{P^2H^2\sigma^2}\,\mathbb{E}\left[\left(\sum_{p,h} X_{p,h}\right)^2\right] = \frac{1}{\sigma^2}\,\mathbb{E}\left[\left(\sum_{p,h} \frac{X_{p,h}}{PH}\right)^2\right] \\ &\leq \frac{1}{PH\sigma^2}\,\mathbb{E}\left[\sum_{p,h} X_{p,h}^2\right] \leq \frac{1}{PH\sigma^2} 4PHd_i\sigma^2 S = 4Sd_i < \infty.\end{aligned}$$

Consequently, our gradient estimate has finite mean squared error.

Lemma 2. *Consider the utility loss $\tilde{\ell}_i^{\Gamma}(\theta_i, \theta_{-i})$ of agent i with respect to the utility function $\tilde{u}_i$ in the finite-dimensional game Γ, and the utility loss $\tilde{\ell}_i^{\sigma}(\theta_i, \theta_{-i})$ with respect to the smoothed utility $\tilde{u}_i^{\sigma}$ in the game Γ^{σ}. Then*

$$\tilde{\ell}_i^{\Gamma}(\theta_i, \theta_{-i}) \leq \tilde{\ell}_i^{\sigma}(\theta_i, \theta_{-i}) + 2ZL\sqrt{d_i}\sigma.$$

Proof: We start by bounding the difference between the utilities of the game Γ and the game Γ^{σ}. To be precise, we prove the following bound for arbitrary strategies θ:

$$|\tilde{u}_i(\theta_i, \theta_{-i}) - \tilde{u}_i^{\sigma}(\theta_i, \theta_{-i})| \leq ZL\sqrt{d_i}\sigma \tag{S.6}$$

By definition, $\tilde{u}_i^{\sigma}(\theta_i, \theta_{-i}) = \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, \sigma^2 I)}[\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i})]$. Since $\tilde{u}_i(\theta_i, \theta_{-i}) = \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, \sigma^2 I)}[\tilde{u}_i(\theta_i, \theta_{-i})]$, we have the inequality

$$|\tilde{u}_i(\theta_i, \theta_{-i}) - \tilde{u}_i^{\sigma}(\theta_i, \theta_{-i})| \leq \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, \sigma^2 I)}[|\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i})|]. \tag{S.7}$$

Next, we show that for fixed ε, $|\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i})| \leq ZL\|\varepsilon\|$. We compute

$$|\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i})| \leq \mathbb{E}_{v_i, o_i}\left[|\bar{u}_i(v_i, \pi_i(o_i, \theta_i + \varepsilon), \theta_{-i}) - \bar{u}_i(v_i, \pi_i(o_i, \theta_i), \theta_{-i})|\right].$$

Since by assumption, $\bar{u}_i$ is differentiable with respect to b_i and the differential is uniformly bounded by Z (equation (4) in Definition 1 of the main paper, we have for every ε

$$\begin{aligned} &|\bar{u}_i(v_i, \pi_i(o_i, \theta_i + \varepsilon), \theta_{-i}) - \bar{u}_i(v_i, \pi_i(o_i, \theta_i), \theta_{-i})| \\ \leq &\left\|\frac{\partial \bar{u}_i}{\partial b_i}\right\|_{\infty} \|\pi_i(o_i, \theta_i + \varepsilon) - \pi_i(o_i, \theta_i)\| \\ \leq &Z\|\pi_i(o_i, \theta_i + \varepsilon) - \pi_i(o_i, \theta_i)\|. \end{aligned}$$

Consequently, by Assumption S.4 in Definition 1 of an NPGA policy network,

$$|\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i})| \leq Z\,\mathbb{E}_{v_i, o_i}\left[\|\pi_i(o_i, \theta_i + \varepsilon) - \pi_i(o_i, \theta_i)\|\right] \leq ZL\|\varepsilon\|,$$

which implies by equation (S.7)

$$|\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i}) - \tilde{u}_i^{\sigma}(\theta_i, \theta_{-i})| \leq ZL\,\mathbb{E}_{\varepsilon \sim \mathcal{N}(0, \sigma^2 I)}[\|\varepsilon\|] \leq ZL\sqrt{d_i}\sigma.$$

This proves equation (S.6). Now let θ_i^* be a best response to θ_{-i} in the game Γ. Then

$$\begin{aligned} \tilde{\ell}_i^{\Gamma}(\theta_i, \theta_{-i}) &= \tilde{u}_i(\theta_i^*, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i}) \\ &= (\tilde{u}_i(\theta_i^*, \theta_{-i}) - \tilde{u}_i^{\sigma}(\theta_i^*, \theta_{-i})) + (\tilde{u}_i^{\sigma}(\theta_i^*, \theta_{-i}) - \tilde{u}_i^{\sigma}(\theta_i, \theta_{-i})) + (\tilde{u}_i^{\sigma}(\theta_i, \theta_{-i}) - \tilde{u}_i(\theta_i, \theta_{-i})) \\ &\leq ZL\sqrt{d_i}\sigma + \tilde{\ell}_i^{\sigma}(\theta_i, \theta_{-i}) + ZL\sqrt{d_i}\sigma \\ &= \tilde{\ell}_i^{\sigma}(\theta_i, \theta_{-i}) + 2ZL\sqrt{d_i}\sigma. \end{aligned}$$

Lemma 3. *Let the neural net π be sufficiently expressive, i.e. for any $\beta_i \in \Sigma_i$ one can find θ such that $\|\beta_i - \pi(\cdot,\theta)\|_{\Sigma_i} \leq \delta$. Then the loss of in G is bounded by that in Γ: $\tilde{\ell}(\theta) \leq \tilde{\ell}^{\Gamma}(\theta) + Z\delta$.*

Proof: The proof relies on boundedness of partial derivatives in the definition of interim smooth Bayesian games. With this regularity condition and universal approximation of the neural network, the derivation is straightforward. Let $\theta_{-i} \in \Theta_{-i}$ be an opponent strategy profile, θ_i^* be a best response to θ_{-i} in Γ, and β_i^* be a best response to $\pi_{-i}(\cdot,\theta_{-i})$ in G, and θ_i an arbitrary parameter vector for player i. Then

$$\begin{aligned}
\tilde{\ell}_i(\theta;\theta_{-i}) &= \tilde{u}_i(\beta_i^*,\theta_{-i}) - \tilde{u}_i(\theta_i,\theta_{-i}) \\
&= \tilde{u}_i(\beta_i^*,\theta_{-i}) - \tilde{u}_i(\theta_i^*,\theta_{-i}) + \tilde{u}_i(\theta_i^*,\theta_{-i}) - \tilde{u}_i(\theta_i,\theta_{-i}) \\
&= \left(\tilde{u}_i(\beta_i^*,\theta_{-i}) - \tilde{u}_i(\theta_i^*,\theta_{-i})\right) + \left(\tilde{u}_i^{\Gamma}(\theta_i^*,\theta_{-i}) - \tilde{u}_i^{\Gamma}(\theta_i,\theta_{-i})\right) \\
&= \mathbb{E}_{o_i}\left[\overline{u}_i(o_i,\beta_i^*(o_i),\theta_{-i}) - \overline{u}_i(v_i;\pi_i(o_i,\theta_i^*),\theta_{-i})\right] + \tilde{\ell}_i^{\Gamma}(\theta) \\
&\leq Z \cdot \mathbb{E}_{o_i}\left[\|\beta_i^*(o_i) - \pi_i(o_i,\theta_i^*)\|\right] + \tilde{\ell}_i^{\Gamma}(\theta) \\
&\leq Z\delta + \tilde{\ell}_i^{\Gamma}(\theta).
\end{aligned} \tag{S.8}$$

S.3. Sampling from Conditional Distributions

When faced with a one-dimensional distribution, sampling is easily done by evaluating the inverse CDF at uniformly sampled points. In the multivariate case, however, there exists no inverse CDF. The following procedure, called conditional distribution method [3, Chapter 11], effectively reduces the problem of sampling from multivariate distributions to multiple one-dimensional sampling tasks. Conditioned on the observation of agent i, we

1. sample the first opponent's observation conditioned on i's observation, $f(o_{-i,1}|o_i)$, by using $u_0 \sim \mathcal{U}[0,1]$ and setting

$$o_{-i,1} = F^{inv}_{o_{-i,1}|o_i}(u_0),$$

2. sample the second opponent's observation conditioned on all observations sampled so far, $f(o_{-i,2}|o_i,o_{-i,1})$, by using $u_1 \sim \mathcal{U}(0,1)$ and setting

$$o_{-i,2} = F^{inv}_{o_{-i,2}|o_i,o_{-i,1}}(u_1),$$

3. continue in this manner for all opposing agents $-i$ and agent i's own type $f(v_i|o)$.

Then the samples satisfy $(o_{-i},v_i) \sim f(o_{-i},v_i|o_i)$ by definition. For most settings in this work, all required functions are analytically known, making a precise sampling possible.

In the general case, it's not possible to state the conditional distribution explicitly, either because there is no access to the true distributions or because the integrals or inverse cumulative density functions are inaccessible.

Subsequently, all required distributions for sampling will be calculated. We will use f, F, and F^{inv} as the probability density function (PDF), the cumulative distribution function (CDF), and the inverse CDF (iCDF), respectively.

S.3.1. Derivation of Conditional Distributions in the Common Values Setting

Let us denote by the random variable $V \sim \mathcal{U}[0,1]$ the common type and by $O_i = V \cdot X_i$ agent i's observation with her unobserved private factor $X_i \sim \mathcal{U}[0,2]$. As X_j is conditionally independent of O_i, we observe that $(O_j|O_i{=}o_i) = (V|O_i{=}o_i) \cdot X_j$. Thus, access to samples of $V|O_i{=}o_i$ is sufficient to sample from $O_j|O_i = o_i$. In the following, we will derive the inverse cumulative distribution function (icdf) $F^{inv}_{V|o_i}$ which we can then use to transform samples from the standard uniform distribution into samples of $V|o_i$. We will rely on Bayes' theorem. To do so, let's first observe that the conditional $O_i|v$ is uniformly distributed on $[0, 2v]$ with pdf $f(o_i|v) = \frac{1}{2v}$ on that interval. The marginal pdf of O_i is then given by

$$f(o_i) = \int_v f(o|v)f(v)dv = \int_{o/2}^1 \frac{1}{2v} \cdot \frac{1}{1} dv = \frac{-\log\left(\frac{o}{2}\right)}{2}$$

on the interval $(0, 2]$ and 0 elsewhere. Given a realized observation o_i, we can then use Bayes' theorem to calculate the conditional pdf of $V|o_i$ on the interval $(\frac{o}{2}, 1]$ via

$$f(v|o_i) = f(o_i|v)f(v)\frac{1}{f(o_i)} = \frac{1}{2v} \cdot 1 \cdot \frac{-2}{\log\left(\frac{o_i}{2}\right)} = \frac{-1}{v \log\left(\frac{o_i}{2}\right)}$$

Integrating over v then yields the conditional cumulative distribution function

$$F(v|o_i) = \begin{cases} 0 & v < \frac{1}{2}o_i, \\ 1 - \frac{\log(v)}{\log\left(\frac{o_i}{2}\right)} & \frac{1}{2}o_i \le v < 1, \\ 1 & 1 < v. \end{cases}$$

Identifying the output with u and inverting, we then arrive at the icdf

$$F^{inv}_{V|o_i}(u|o_i) = \left(\frac{o_i}{2}\right)^{(1-u)}$$

Given a standard uniform RV $U \sim \mathcal{U}[0,1]$, when then have $V|o_i \sim F^{inv}_{V|o_i}(U|o_i)$.

S.3.2. Derivation of Conditional Distributions in the Affiliated Values Setting

For this game setting, we have $(O_j|O_i = o_i) = (U_j|O_i = o_i) + (T|O_i = o_i)$, where $(U_j|O_i = o_i) = U_j$ is independent of O_i. Using Bayes theorem, one has

$$(T|O_i = o_i) = o_i - U_i \sim \mathcal{U}(\max\{0, o_i - 1\}, \min\{1, o_i\})$$

and thus the observation of the opponent is the sum of the two uniform random variables U_j and $(T|O_i = o_i)$.

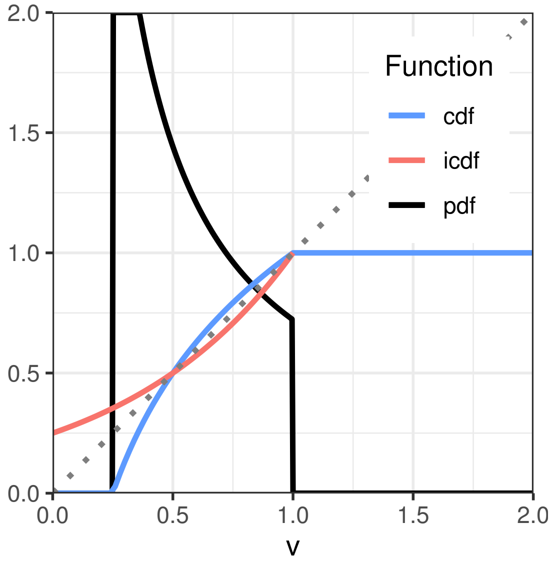


Figure S.5 **The probability functions in the common values setting for an example conditional value of $o_i = \frac{1}{2}$.**

S.3.3. Derivation of Conditional Distributions in the LLG Settings

In these settings, there are two groups of correlations: On one side, there is the global bidder whose prior is independent from all other bidders, and on the other side there are local bidders whose values depend on one another. In the Bernoulli weights model, the density of the local bidder j conditional on v_i is simply given as a uniform distribution on $[0, 1]$ with the addition that with a probability of γ the value will not be uniform but $v_j = v_i$.

In the constant weights model, the approach is similar to the one used for affiliated values above. We can directly derive the conditionals of player 1's individual component $\omega_1|v_1 \sim \mathcal{U}\{\max(0, \frac{v_1-w}{1-w}\}, \min\{1, \frac{v_1}{1-w}\}]$, and ω_2 is conditionally independent of v_1. Observe that $w\omega_4 = v_1 - (1-w)\omega_1$. We can thus sample $\omega_1|v_1$ and ω_2 and then calculate $v_2|v_1 = (1-w)\omega_2|v_1 + w \cdot \omega_4|v_1 = v_1 + (1-w)(\omega_2 - \omega_1|v_1)$, and vice versa for player 2.

References

[1] L. M. Ausubel, P. Cramton, M. Pycia, M. Rostek, and M. Weretka. Demand reduction and inefficiency in multi-unit auctions. *The Review of Economic Studies*, 81(4):1366–1400, 2014.

[2] J. Créemer and R. P. McLean. Optimal selling strategies under uncer0 tainty for a discriminating monopolist when demands are interdepen0 denty. *Econometrica*, 53:345–361, 1985.

[3] W. Hörmann, J. Leydold, and G. Derflinger. *Automatic nonuniform random variate generation*. Springer Science & Business Media, 2013.

[4] V. Krishna. *Auction Theory*. Academic press, 2009.

[5] P. R. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, pages 1089–1122, 1982.

[6] M. Perry and P. J. Reny. An efficient auction. *Econometrica*, 70(3):1199–1212, 2002.

[7] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *arXiv:1703.03864 [cs, stat]*, Mar. 2017.

[8] R. Wilson. *Competitive bidding with disparate information*. Number 114. Graduate School of Business, Stanford University, 1966.

Publication C: Learning Equilibria in Asymmetric Auction Games

Peer-Reviewed Journal Paper

Title: Learning Equilibria in Asymmetric Auction Games

Authors: M. Bichler, S. Heidekrüger, and N. Kohring

In: INFORMS Journal on Computing 35(3):523-542 (2023)

Abstract: Computing Bayesian Nash equilibrium strategies in auction games is a challenging problem that is not well understood. Such equilibria can be modeled as systems of nonlinear partial differential equations. It was recently shown that Neural Pseudogradient Ascent (NPGA), an implementation of simultaneous gradient ascent via neural networks, converges to a Bayesian Nash equilibrium for a wide variety of symmetric auction games. While symmetric auction models are widespread in the theoretical literature, in most auction markets in the field one can observe different classes of bidders having different valuation distributions and strategies. Asymmetry of this sort is almost always an issue in real-world multi-object auctions, where different bidders are interested in different packages of items. Such environments require a different implementation of NPGA with multiple interacting neural networks having multiple outputs for the different allocations the bidders are interested in. We analyze a wide variety of asymmetric auction models. Interestingly, our results show that we closely approximate Bayesian Nash equilibria in all models where the analytical Bayes-Nash equilibrium is known. Besides we analyze new and larger environments for which no analytical solution is known and verify that the solution found approximates equilibrium closely. The results provide a foundation for generic equilibrium solvers that can be used in a wide range of auction games.

Contribution of thesis author: design and implementation, empirical analysis, writing and revising the manuscript

References: Full Paper: Bichler et al. (2023a), Online Supplementary Material: Bichler et al. (2023b)

INFORMS JOURNAL ON COMPUTING
Vol. 35, No. 3, May 2023, 523–542
ISSN 1091-9856 | EISSN 1526-5528 | 23 | 202305

INFORMS
DOI 10.1287/ijoc.2023.1281

Learning Equilibria in Asymmetric Auction Games

Martin Bichler, Nils Kohring, Stefan Heidekrüger
Technical University of Munich, Department of Computer Science, 85748 Garching, Germany

Computing Bayesian Nash equilibrium strategies in auction games is a challenging problem that is not well understood. Such equilibria can be modeled as systems of nonlinear partial differential equations. It was recently shown that Neural Pseudogradient Ascent (NPGA), an implementation of simultaneous gradient ascent via neural networks, converges to a Bayesian Nash equilibrium for a wide variety of symmetric auction games. While symmetric auction models are widespread in the theoretical literature, in most auction markets in the field one can observe different classes of bidders having different valuation distributions and strategies. Asymmetry of this sort is almost always an issue in real-world multi-object auctions, where different bidders are interested in different packages of items. Such environments require a different implementation of NPGA with multiple interacting neural networks having multiple outputs for the different allocations the bidders are interested in. In this paper, we analyze a wide variety of asymmetric auction models. Interestingly, our results show that we closely approximate Bayesian Nash equilibria in all models where the analytical Bayes-Nash equilibrium is known. Additionally, we analyze new and larger environments for which no analytical solution is known and verify that the solution found approximates equilibrium closely. The results provide a foundation for generic equilibrium solvers that can be used in a wide range of auction games.

Key words: equilibrium learning, neural networks, Bayes-Nash equilibria

1. Introduction

Auction theory is arguably the best-known and practically most relevant application of Bayesian game theory, central to modern economic theory (Klemperer 2000) and with a multitude of applications in the field, ranging from industrial procurement to treasury auctions and spectrum sales (Krishna 2009, Milgrom 2017, Bichler and Goeree 2017). The derivation of Bayesian Nash equilibrium strategies (BNE) for the first-price and second-price sealed-bid auction led to a comprehensive theoretical framework for the analysis of single-item auctions by Nobel laureate William Vickrey, a landmark result of economic theory (Vickrey 1961). Also, the Nobel Prize in Economic Sciences 2020 to Paul Milgrom and Robert Wilson was awarded for contributions to auction theory. However, while single-item auctions are well understood and closed-form BNE strategies are known for a variety of auction formats, we only know equilibrium strategies for a few restricted multi-item auction environments with heterogeneous goods. Even for uniform or discriminatory multi-unit auctions with homogeneous goods and symmetric bidders, we can only characterize properties of the Bayes-Nash equilibrium but do not have a general closed-form solution (Krishna 2009). So, the realm of auction markets where we know a Bayes-Nash equilibrium is very limited.

Equilibrium computation is well-known to be hard even for simple, finite, complete-information games: Finding Nash equilibria in normal-form games is known to be in the complexity class PPAD[1] (Daskalakis et al. 2009). Mathematically, auctions are typically described as Bayesian games. Bidders' valuations are considered samples from some continuous and atomless prior valuation distribution and their strategies are represented by continuous bid functions mapping these valuations to bids. Vickrey (1961) have enabled a deep understanding of common single-item auction formats. However, there still remain many open questions for more involved multi-item auctions such as *combinatorial auctions*, in which players bid on *bundles* of multiple goods simultaneously. We also know little about the existence of Bayesian Nash equilibria in such auction games (Jackson and Swinkels 2005). Importantly, the computational complexity of computing BNE is hardly understood. Typically, for a fully specified setting, we can model the equilibrium problem as systems of nonlinear partial differential equations for which no exact solution theory is known (Klainerman 2010). Given the relevance of auctions, understanding their equilibria is crucial, and numerical methods for computing or approximating such steady states would be a significant step forward in the theory of auctions and also in their design and in applications.

This paper can be viewed in the context of equilibrium learning via gradient dynamics. Whether learning agents' strategies in repeated games converge to equilibria has been studied for complete-information normal-form games (Fudenberg and Levine 2009). In contrast, equilibrium learning in Bayesian auction games is largely unexplored (see Section 2). First, the ex-post utility function is non-differentiable in auctions, which makes it difficult to apply gradient dynamics. Secondly, it is a known fact that multi-agent gradient dynamics do not converge in general games: Convergence to Nash equilibria has only been

[1] The class of *Polynomial Parity Arguments on Directed graphs* (PPAD) problems is believed to be hard and is related to NP.

established for restricted classes of complete-information normal-form games. In summary, it is all but clear how gradient dynamics would be implemented in Bayesian auction games, and even if this was done, whether the algorithm would converge to a BNE in auction games. We draw on Bichler et al. (2021), who recently introduced Neural Pseudogradient Ascent (NPGA), an algorithm that relies on simultaneous gradient ascent of bidders with respect to their ex-ante utility functions. More specifically, NPGA models all players' bidding strategies as neural networks, and trains them via self play based on approximate ex-ante gradients computed from observations of the discontinuous ex-post utility function using evolutionary strategies. It can be applied to a wide range of Bayesian auction games, since it does not require any auction-specific sub-procedures beyond access to simulating auction outcomes. Likewise, its computational steps can exploit massive parallelization and GPU hardware acceleration.

The results by Bichler et al. (2021) focus on *symmetric* auction models, assuming symmetric prior distributions and symmetric equilibrium bidding strategies of the bidders. This allows them to train only a single neural network to provably find the symmetric equilibrium bidding strategy. While symmetric models cover some important auctions in the theoretical literature, many interesting environments include asymmetries. For example, asymmetric priors are a concern for single-object auctions with strong and weak bidders drawn from different distributions, but they are even more prevalent in multi-item auctions where it is unlikely that bidders are interested in the same items with their values drawn from the same distribution. It is also these more general market environments for which the literature on auction theory does not provide analytical equilibrium predictions.

1.1. Contributions

In this paper, we explore a number of challenging environments, models which clearly violate the symmetry assumption. Nothing is known about the convergence and speed of equilibrium learning in such environments where one needs to train multiple neural networks with multi-dimensional outputs modeling different actions of bidders. We show that the NPGA algorithm also converges with multiple neural networks which are required to model asymmetric environments. We explore a wide range of wicked models from the literature where the BNE is known analytically and find that NPGA computes a very close approximation of the Bayes-Nash equilibrium in all of them. In addition, we explore large environments where no analytical solution is known and we can verify empirically that a close approximation to a BNE is found.

We start with a single-object auction with two asymmetric priors (Plum 1992). Apart from this original model, we also analyze one that allows for (rational) overbidding and admits multiple equilibria (Kaplan and Zamir 2015). Here, we only know closed-form equilibrium strategies for uniform prior distributions, for which NPGA finds a BNE. However, we also discuss a specific model with two bidders competing for a single object where the valuations are drawn from a non-linear beta distribution. No equilibrium strategy is known, but we find bidding strategies with a very low estimated utility loss for all players. This indicates that the computed strategy profile is a close approximation of a BNE.

Second, we explore a specific type of multi-unit uniform-price auction of homogeneous goods with two classes of bidders, those with a high and with a low type. The environment is very large with up to 12 units and NPGA is able to compute a sufficiently close equilibrium in under five minutes. Such mechanisms are used in treasury bill auctions but also electricity markets. Demand reduction is a well-known phenomenon in such auctions and it is interesting to observe how it plays out under different model assumptions about the strength of the competitors.

Third, we analyze a combinatorial auction in the well-known local-local-global (LLG) model. The model has two items and three bidders and it has become a standard environment to discuss spectrum auction design and more generally combinatorial auctions (Bichler and Goeree 2017). Two local bidders want to win one item each and they compete against a global bidder interested in the package of both items. Bidders are assumed to only bid for the single item for which they have a strictly positive expected valuation. In this standard LLG model, the local bidders are assumed to have symmetric priors, and NPGA converges quickly to the BNE strategy (Bichler et al. 2021). In contrast to this standard setting, we analyze a variant where one of the local bidders is favored and bidders are not precluded a priori from bidding on bundles for which they do not have a strictly positive value. While a bidder would not actually be interested in *winning* such a bundle, it turns out that sometimes it may nevertheless be rational to submit a positive bid for it. Ott and Beck (2013) showed that, in fact, this version of the local-local-global model has an equilibrium where the second local bidder bids on the package of both items and even overbids—in spite of being interested in a single item only. Such equilibrium strategies are not obvious. Again, we find that NPGA recovers this analytical solution with high precision.

Fourth, we experiment with another reverse combinatorial auction model with two homogeneous objects and two bidders. This model is interesting because there are two pure BNE (Anton and Yao 1992, Kokott et al. 2019). Similar to the analysis of the asymmetric single-object environment (Kaplan and Zamir 2015), NPGA finds an equilibrium, which is also the efficient one.

Finally, we report the results for a large combinatorial auction model with six bidders, belonging to two symmetry classes, and eight items, which has recently been proposed as a challenging problem for equilibrium computation and which, to the authors' knowledge, is the largest combinatorial auction for which an approximate BNE has been computed numerically with a setting-specific algorithm (Bosshard et al. 2020). Going beyond the existing challenge model, we also study NPGA in an even larger extension by introducing an additional seventh bidder belonging to a new third symmetry class. In both these settings, strategy profiles learned by NPGA converge to approximate BNE. Such environments can already be considered very large and beyond what is typically analyzed in auction theory.

Overall, The empirical results we show in this paper provide evidence that gradient dynamics implemented in NPGA are significantly more powerful than expected and they converge in a much wider range of (asymmetric) auction games. This raises hope that gradient dynamics can be used to compute equilibria in a much broader variety of market models and that general auction equilibrium solvers are in reach.

1.2. Organization

In the next section, we discuss related literature. Section 3 introduces preliminaries and notation before we discuss gradient dynamics in the context of auctions in Section 4. Section 5 introduces metrics to evaluate the quality of our results before we report our results in Section 6. Finally, we provide a summary and conclusions in Section 7. The source code and configurations can be found at the repository (Bichler et al. 2023).

2. Related Literature

In what follows, we survey existing hardness results, approaches to equilibrium learning, and initial research on computing approximate Bayes-Nash equilibria.

2.1. Hardness of Equilibrium Computation

The computation of Nash equilibria received significant attention after the initial contribution by John Nash on the existence of such equilibria in complete-information normal-form games (Nash et al. 1950). However, it was shown that the problem is already PPAD-complete for two-agent normal-form games (Daskalakis et al. 2009) and it is hard to approximate (Rubinstein 2016). The computation of Nash equilibria for three or more agents is even FIXP-complete, i.e., complete for the class of search problems that can be cast as fixed-point computation problems (Etessami and Yannakakis 2007).

Determining whether a pure-strategy BNE exists in a finite Bayesian game is NP-complete and these hardness results also hold if there are only two agents and the game is symmetric (Conitzer and Sandholm 2008). Finding a mixed Bayesian equilibrium in a Bayesian game is, of course, PPAD-hard, but might be even harder; however, little is known in general. As indicated in the introduction, Cai and Papadimitriou (2014) show that finding a BNE in simultaneous single-item Vickrey auctions for which the bidders have combinatorial valuations is hard for the class PP (the decision version of ♯P), which is much harder than NP. Even certifying a BNE is PP-hard, which casts doubt on the question of whether BNE can be at all predictive in the field. Additionally, the authors show that it is even NP-hard to find an approximate BNE in the simultaneous Bayesian auction game. Note that environments with continuous action space are not finite games, and the existence result by Nash does not carry over. We are not aware of proof that a possibly mixed Bayesian equilibrium always exists in such games. Athey (2001) showed conditions for pure BNE to exist, Carbonell-Nicolau and McLean (2018) provided conditions that guarantee the existence of a BNE, while Ui (2016) characterized strong payoff-monotonicity as a sufficient condition for uniqueness of BNE in ex-post differentiable continuous-action Bayesian games.

2.2. Equilibrium Learning

Our research is best situated in the literature on equilibrium learning (Fudenberg and Levine 2009). Learning in complete-information normal form games has a long history and has been extensively studied in game theory and, more recently, multi-agent reinforcement learning. One class of methods is formed by *best response dynamics.* The earliest such method, published by Cournot in 1838, has agents play a pure strategy best response against other agents' strategies used in the previous iteration. In Fictitious Play (FP)

(Brown 1951), a best response is instead played against the strategy profile induced by opponents' empirical frequencies of play in all previous iterations. Whenever the *empirical frequencies* of FP converge, the limit constitutes a Nash equilibrium, but the actual (last-iteration) play only converges in special cases of normal form games such as potential games (Monderer and Shapley 1996).

Gradient dynamics constitute another class of equilibrium learning algorithms. Generalized infinitesimal gradient ascent (GIGA) (Zinkevich 2003) or GIGA-WoLF (Bowling 2005) are examples of gradient dynamics in normal form games, where in each iteration, for each agent we move a step along the direction of the utility gradient and then project the resulting point back to the set of feasible mixed strategies. If aggregating over the stages of the process, the agent's regret grows sublinearly, then there is "no regret" asymptotically. GIGA's total regret is $O(\sqrt{T})$, where T is the number of steps in a repeated strategic game. Hazan et al. (2007) have given an algorithm with a total regret of $O(\log(T))$. Complete-information games with continuous action spaces and smooth utility functions have also received some attention in the context of generative adversarial networks (Letcher et al. 2019, Balduzzi et al. 2018, Schäfer and Anandkumar 2019). A common observation in this line of research is that gradient-based learning does not necessarily converge to an equilibrium and may even exhibit cycling or chaotic behavior. However, it often achieves no-regret properties and thereby converges to a weaker form of equilibrium, so called coarse correlated equilibria (CCE). Similar conclusions were drawn for finite-type (and possibly continuous-action) Bayesian games. Here, no-regret learners were shown to converge to Bayesian CCEs (Hartline et al. 2015).

Gradient dynamics are only known to converge to a Nash equilibrium in certain types of normal-form games such as potential games, bilinear games (Singh et al. 2000), and convex games (Mertikopoulos and Zhou 2019). Letcher et al. (2019) explore gradient dynamics in complete-information continuous-action *differential games*. If ex-post payoffs are twice continuously-differentiable, they find properties such that gradient dynamics converge to at least *local equilibria*. Unfortunately, the ex-post utility in our auction games is not differentiable. More importantly, these techniques are defined for complete-information games with finite-dimensional action spaces while we search for strategies over a function space. Unfortunately, a thorough understanding of the convergence and limiting behaviors in general, continuous games is missing. Actually, the analysis of gradient dynamics, in general, can be arbitrarily complex (Andrade et al. 2021).

2.3. Algorithms for Computing Approximate BNE

Earlier approaches to compute approximate BNE in auctions either comprised solving the set of nonlinear differential equations resulting from the first-order conditions of simultaneous maximization of the bidders' payoffs (Marshall et al. 1994, Bajari 2001) or of restricting the action space, e.g., through discretization (Athey 2001). Then, however, one has no guarantees on the quality of the corresponding ϵ-BNE of the original auction game. Armantier et al. (2008) introduced a BNE-computation method that is based on expressing the Bayesian game as the limit of a sequence of complete-information games, but defining this sequence requires setting-specific analysis.

Numerical BNE in more complex combinatorial auctions were first computed by Bosshard et al. (2017, 2020) in two recent papers; in particular, they study the LLG and LLLLGG markets, both of which are also analyzed in this paper. Their algorithm computes point-wise best responses in a linearization of the strategy space via Monte Carlo integration. They prove an an upper bound ϵ on the interim utility loss achieved by their algorithm using a verification method that assumes identical independent priors ($F_{v_i|v_{-i}} = F_{v_i}$) and risk-neutral attitudes of all bidders. High worst-case interim precision comes at a computational cost for more complex environments with multi-minded bidders.

NPGA (Bichler et al. 2021) follows a different approach and is rooted in gradient dynamics rather than best response dynamics. It directly learns the bid functions expressed across the entire value space (as opposed to point-wise) by updating the parameters of the neural networks via ex-ante gradient ascent. NPGA neither requires discretization of the value or action space as in Athey (2001) nor does it rely on twice differentiable payoff or loss functions as required in the literature on differentiable games (Singh et al. 2000, Letcher et al. 2019). Further, it makes no assumptions about the risk attitude or independence of the bidders' valuations. For symmetric auctions, Bichler et al. (2021) show that NPGA converges (at least) to local BNE.

3. Problem Statement and Notation

We next introduce the necessary notation and concepts from Bayesian game theory relevant to our paper.

3.1. Auctions as Bayesian Games

A *Bayesian game* or game with *incomplete information* is defined by the quintuple $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$. The set of players is denoted by $\mathcal{I} = \{1, \ldots, n\}$, $\mathcal{A} \equiv \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ is the set of possible action profiles, where agent $i \in \mathcal{I}$ has access to the action set $\mathcal{A}_i$. $\mathcal{V} \equiv \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ we denote the set of *type profiles* and $F\colon \mathcal{V} \to [0,1]$ defines the joint probability distribution over $\mathcal{V}$ that is known to all players. Throughout this paper, F_X will denote the cumulative distribution function of a random variable X. For example, F_{v_i} will denote the marginal distribution of player i's type. In each game, a type profile $v \sim F$ is drawn and all agents i are privately informed of their own types v_i. Based on this private information, each player must then choose an action b_i from $\mathcal{A}_i$. After actions have been chosen, every player will observe their *ex-post* utility according to a function $u_i : \mathcal{A} \times \mathcal{V}_i \to \mathbb{R}$ that notably depends on all agents' actions but only on i's own type.

This paper considers not only sealed-bid auctions of a single object, but also multi-unit auctions and combinatorial auctions with m heterogeneous items, $\mathcal{M} = \{1, \dots, m\}$. In these auctions, each agent, also called *bidder*, is allocated a bundle $x_i \in \mathcal{K} \equiv 2^{\mathcal{M}}$ of items (possibly $x_i = \varnothing$). In the *private value* setting most commonly studied in auction theory, types $v_i \in \mathcal{V}_i$ can then be interpreted as a vector of *private valuations* that is composed of the valuations the bidder has for all possible bundles: $v_i \equiv (v_i(k))_{k \in \mathcal{K}}$. For a treatment beyond private values (e.g., interdependent bidder types) we refer the interested reader to Bichler et al. (2021). Bidders map these valuations to their individual bids $b_i = \beta_i(v_i)$ according to some *pure strategy* or *bid function* $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$. In line with most work in auction theory, we will focus on pure strategies that choose a specific action with certainty.

In an exclusive-OR (XOR) bid language, a bidder submits bids for every possible bundle but can only win one of the bids. This means that bids are generally in $\mathcal{A}_i \subseteq \mathbb{R}_+^{|\mathcal{K}|}$, and every player must thus submit a total of 2^m scalar bids.

By $\Sigma_i \subseteq \mathcal{A}_i^{\mathcal{V}_i}$ we denote the strategy space of bidder i and by $\Sigma \equiv \prod_i \Sigma_i$ the space of available joint strategies. Note that the spaces Σ_i are infinite-dimensional as a consequence of infinite $\mathcal{V}_i$.

The auctioneer then applies an *auction mechanism* which will determine an allocation x and a price vector p. The allocation determines the bundles of goods $x_i \in \mathcal{K}$ received by each bidder which must be disjoint: $x_i \cap x_j = \varnothing$. Payments $p \in \mathbb{R}^n$ determine a scalar amount of money that each payer will have to pay to the auctioneer in exchange for receiving the bundle x_i. We will rely on the standard environment in auction theory where bidders have a *quasi-linear* utility function given by $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$,

$$u_i(v_i, b_i, b_{-i}) = v_i(x_i) - p_i, \tag{3.1}$$

where the index $-i$ denotes a profile of types, actions, or strategies for all agents but agent i. That is, each bidder's utility is given by her valuation of her allocated bundle, minus the payment she has to make. Quasi-linear utilities correspond to risk-neutral bidders. Note that NPGA is not restricted to the risk-neutral setting. (See e.g. Bichler et al. (2021) or Ewert et al. (2022) for applications in the presence of risk-averse agents.) However, the environments discussed in this paper assume quasi-linear utility, which simplifies notation. We will differentiate between the *ex-ante*, *interim*, and *ex-post* states of the game, where bidders first know only F, then additionally their valuations $v_i \sim F_{v_i}$, and finally also the observed utility $u_i(v_i, b)$, respectively.

3.2. Bayes-Nash Equilibrium

The notion of Nash equilibria (NE) is the central equilibrium solution concept in noncooperative game theory. An action profile b^* is a pure-strategy NE of the complete-information game $G = (\mathcal{I}, \mathcal{A}, u)$ iff no player has any incentive to deviate unilaterally while other agents adhere to the equilibrium: $u_i(b_i^*, b_{-i}^*) \geq u_i(b_i, b_{-i}^*)$ for all $b_i \in \mathcal{A}_i$ and all $i \in \mathcal{I}$. Bayesian Nash equilibria (BNE) generalize this concept to incomplete-information games. To do so, we will need to consider the expected *interim utility* $\overline{u}_i$ of i of a given bid choice $b_i \in \mathcal{A}_i$ over the conditional distribution of opponent valuations v_{-i}, given i's observed type v_i and assuming opponents play fixed strategies $\beta_{-i} \in \Sigma_{-i}$:

$$\overline{u}_i(v_i, b_i, \beta_{-i}) \equiv \mathbb{E}_{v_{-i}|v_i}[u_i\left(v_i, b_i, \beta_{-i}(v_{-i})\right)], \tag{3.2}$$

In our analysis, we will also use the *interim utility loss* of action b_i that is incurred, in hindsight, by not playing a best response action. Given v_i and β_{-i} it is defined as

$$\overline{\ell}_i(b_i; v_i, \beta_{-i}) = \sup_{b_i' \in \mathcal{A}_i} \overline{u}_i(v_i, b_i', \beta_{-i}) - \overline{u}_i(v_i, b_i, \beta_{-i}). \tag{3.3}$$

Typically, $\overline{\ell}_i$ is not actually observable to any agent because it requires knowledge of (a) the opponents' strategies and (b) a corresponding best response.

An interim *ϵ-Bayesian Nash Equilibrium (ϵ-BNE)* is a strategy profile $\beta^* = (\beta_1^*, \ldots, \beta_n^*) \in \Sigma$ in which no deviation could yield an interim utility improvement of more than $\epsilon \geq 0$ for any player. Formally, an ϵ-BNE is described as follows:

$$\overline{\ell}_i\left(b_i; v_i, \beta_{-i}^*\right) \leq \epsilon \quad \text{for all } i \in \mathcal{I}, v_i \in \mathcal{V}_i, \text{ and } b_i \in \mathcal{A}_i. \tag{3.4}$$

In a true BNE, where $\epsilon = 0$, every bidder's strategy maximizes her expected interim utility everywhere on her type space $\mathcal{V}_i$ given the opponents' strategies. While this *interim* stage definition of BNE is most common in the literature, we will instead focus on *ex-ante* Bayesian equilibria as strategy profiles that concurrently maximize each player's *ex-ante* expected utility $\tilde{u}$, i.e., at the stage where only the priors F are known, but players have not yet learned their own private valuation. We thusly define $\tilde{u}$ and the *ex-ante utility losses* $\tilde{\ell}$ of a strategy profile $\beta \in \Sigma$ by

$$\begin{aligned}\tilde{u}_i(\beta_i, \beta_{-i}) &\equiv \mathbb{E}_v[u_i(v_i, \beta_i(v_i), \beta_{-i}(v_{-i}))] &(3.5)\\ &= \mathbb{E}_{v_i \sim F_{v_i}}[\overline{u}_i(v_i, b_i, \beta_{-i})], &(3.6)\end{aligned}$$

and

$$\tilde{\ell}_i(\beta_i, \beta_{-i}) \equiv \sup_{\beta'_i \in \Sigma_i} \tilde{u}_i(\beta'_i, \beta_{-i}) - \tilde{u}_i(\beta_i, \beta_{-i}). \tag{3.7}$$

Ex-ante BNE strategy profiles $\beta^* \in \Sigma$ can be characterized by the equations $\tilde{\ell}_i(\beta^*_i, \beta^*_{-i}) = 0$ for all $i \in \mathcal{I}$. Note that interim BNE also constitute an ex-ante equilibria and the reverse holds almost everywhere: every ex-ante equilibrium fulfills Equation 3.4, except possibly on a set of type profiles with measure 0 under F. In this paper, we concern ourselves with finding ex-ante equilibria of auction games.

4. Neural Pseudogradient Ascent

In this section, we introduce Neural Pseudogradient Ascent (NPGA), an algorithm that was recently introduced by Bichler et al. (2021) for Bayesian games with continuous type- and action-spaces. We briefly summarize the algorithm for the paper to be self-contained before we discuss issues around computational hardness and scalability.

4.1. The Algorithm

Intuitively, NPGA simply follows the ex-ante gradient dynamics of the game. However, computing these dynamics is not trivial for auctions, where the ex-post utility functions have discontinuities. Suppose that in each iteration of the learning algorithm players have access to a gradient-oracle $\nabla_{\beta_i}\tilde{u}_i(\beta_i, \beta_{-i})$ with respect to the current joint strategy profile β^t. Then the gradient dynamics would require that each player perform a projected gradient update:

$$\beta_i^t \equiv \mathcal{P}_{\Sigma_i}\left(\beta_i^{t-1} + \Delta_i^t\right) \qquad \text{where} \qquad \Delta_i^t \propto \nabla_{\beta_i}\tilde{u}_i(\beta_i, \beta_{-i}), \tag{4.1}$$

where $\mathcal{P}_{\Sigma_i}(\,\cdot\,)$ projects its argument onto the set of feasible strategies. Some nuances of Equation 4.1 deserve discussion: Importantly, the gradient dynamics are to be understood with respect to the *ex-ante* utility $\tilde{u}$, rather than interim or ex-post utilities. As such, any update iteration aims to marginally improve on the player i's expected utility across all possible type realizations of the game. Furthermore, when computing the gradient oracle $\nabla_\beta \tilde{u}$ via self-play, one may need to rely on access to other players' strategies, but evaluating each player's policy requires only on their own valuation. Finally, $\beta_i \in \Sigma_i$ are functions in an infinite-dimensional function space, so the gradient $\nabla_{\beta_i} \tilde{u}_i$ is itself a *functional* derivative. We formally consider this to be the Gateaux derivative, a generalization of the directional derivative in Euclidean spaces, over the Hilbert space Σ_i equipped with the inner product $\langle \psi, \beta_i \rangle = \mathbb{E}_{v_i \sim F_{v_i}}\left[\psi(v_i)^T \beta_i(v_i)\right]$. This choice of space specifies the projection operation in Equation 4.1 to $\mathcal{P}_{\Sigma_i}(\beta) \equiv \arg\min_{\sigma \in \Sigma_i} \langle \sigma - \beta, \sigma - \beta \rangle$.

To implement these gradient updates in practice, NPGA considers all bidders' strategies to be policy networks $\beta_i(v_i) \equiv \pi_i(v_i; \theta_i)$ specified by some neural network architecture and parameters $\theta_i \in \mathbb{R}^{d_i}$. Importantly, when a suitable neural network architecture is chosen, all relevant θ_i will yield feasible bids, and the projection operation in the update can be neglected as a result. In the empirical part of this study, we restrict ourselves to fully-connected feed-forward neural networks with SeLU activations in the hidden layers (Klambauer et al. 2017) and ReLU activations in the output layer. The latter guarantees satisfaction of nonnegativity of the bids – the only feasibility constraint in the auctions studied below. Note that in contrast to Bichler et al. (2021), we analyze more complex auction models with multi-minded bidders, such that the output layer includes multiple neurons defining bids for different packages of items in a multi-item auction. Importantly, we need to train multiple neural networks that compete rather than only a single one. As network sizes $d_i \in \mathbb{N}$ are finite, the problem of choosing an infinite-dimensional strategy is thus transformed into choosing a finite-dimensional parameter vector θ_i.

As auction allocations x are inherently discrete, the ex-post utilities $u_i(v_i, b_i, b_{-i})$ in auction games have discontinuities and, as a result, are not (sub)differentiable in b_i. While the set of discontinuities is typically a v-nullset, taking the analytical gradient elsewhere nevertheless would yield systematically misleading updates: As an example, consider a first-price sealed-bid auction of a single item where the winner i pays her bid $p_i = b_i$. Players' utility functions are then separated into two intervals: When bidding below the highest other bid, one will lose the auction, have a constant payoff of 0, with $\nabla_{b_i} u_i = 0$ on this interval. Thus there will be no usable learning feedback. When i's bid wins the auction, any further increase in b_i will marginally decrease u_i, $\nabla_{b_i} u_i = -1$. Analytical gradient updates via backpropagation on the ex-post utility will thus always send nonincreasing feedback, until all players finally bid a constant amount of zero no matter their type.

NPGA alleviates this feedback-breakdown of the ex-post gradients by instead estimating the effect of parameter changes on the *ex-ante* utility using finite differences, and computing gradient estimates $\nabla_\theta \tilde{u}$ using a natural evolution strategy (ES) approach Salimans et al. (2017). Given parameters $P \in \mathbb{N}$ and $\epsilon > 0$, we perturb the parameter vector P times $\theta_{i;p} \equiv \theta_i + \varepsilon_p$ using zero-mean Gaussian noise $\varepsilon_p \sim \mathcal{N}(0, \sigma^2)$. NPGA then calculates each perturbation's *fitness*, $\varphi_p \equiv \tilde{u}_i(\pi_i(v_i; \theta_{i;p}), \beta_{-i})$, via Monte Carlo integration, and estimates the gradients as the fitness-weighted perturbation noise $\nabla_\theta^{ES} \equiv \frac{1}{\sigma^2 P} \sum_p \varphi_p \varepsilon_p$. This results in an unbiased estimator of the ex-ante gradients $\nabla_\theta \tilde{u}$ even when the ex-post gradients $\nabla_b u$ are not well-defined. Pseudo-code of NPGA is given in Algorithm 1.

Unlike in Bichler et al. (2021), where the "symmetric" version of NPGA has been analyzed, here we focus on the asymmetric case where agents can differ (in their prior F_{v_i}, or in how the auctioneer treats their bids) and each agent must learn their own optimal bid function. As indicated earlier, this necessitates each bidder to train her own neural network, rather than allowing a simplification of a single *shared* network, which is essential to the theoretical convergence analysis in Bichler et al. (2021). Instead, in each iteration, we iterate over bidders who perform their own individual gradient updates.

In summary, NPGA "implements" Equation 4.1 by parametrizing strategies using neural networks and training them with ES-pseudogradients:

$$\beta_i^t \equiv \pi_i(\,\cdot\,; \theta_i^t) \quad \text{with} \quad \theta_i^t \equiv \theta_i^{t-1} + \Delta_i^t \quad \text{where} \quad \Delta_i^t \propto \nabla_{\theta_i^t}^{ES}. \tag{4.2}$$

The computation of these updates in each iteration only relies on values of the ex-ante utility $\tilde{u} = \mathbb{E}_{v \sim F}[u]$. No further information about the game is necessary. Thus, whenever the joint ex-post utility u can be calculated in a vectorized fashion, $\tilde{u}$ can leverage parallel computations to efficiently perform Monte Carlo integration over $\mathcal{V}$. In practice, this approach lends itself to accelerated computation using GPUs. We built custom vectorized implementations of many common auction mechanisms using the PyTorch framework (Paszke et al. 2017) that allow us to perform the Monte Carlo estimation multiple orders of magnitude faster than prior numerical work on auctions.

Algorithm 1 Neural Pseudogradient Ascent using Evolutionary Strategies

```
input: Initial policy, ES population size $P$, ES noise variance, learning rate, batch size
for $t = 1, 2, \ldots$ do
    Sample a batch of valuation profiles from prior
    Calculate joint utility of current strategy profile
    for each agent $i \in \mathcal{I}$ do
        for each $p \in \{1, \ldots, P\}$ do
            Perturb agent $i$'s current policy
            Evaluate fitness of perturbation $p$ by playing against current opponents
        end for
        Calculate ES pseudogradient as fitness-weighted perturbation noise
        Perform a gradient ascent update step on the current policy
    end for
end for
```

5. Evaluation

We will provide three metrics for evaluating the quality of the learned strategy profiles β. Whenever an analytical BNE β^* is known, we may simply check whether $\beta \to \beta^*$. To do so, we calculate the agents' utility losses $\mathcal{L}_i$ that result from playing the learned strategy β_i rather than the equilibrium strategy β_i^*.

The *relative utility loss* is then given by

$$\mathcal{L}(\beta_i) \equiv 1 - \hat{u}_i(\beta_i, \beta_{-i}^*)/\hat{u}_i(\beta^*). \tag{5.1}$$

Additionally, we will also measure the distance in strategy space, which tells us how close the learned strategy is to the analytical one:

$$L_2(\beta_i) \equiv \|\beta_i - \beta_i^*\|_{\Sigma_i}. \tag{5.2}$$

Both of these metrics use Monte Carlo integration over a large number of valuations $v \sim F$ to approximate $\hat{u} \approx \tilde{u}$.

When no equilibrium is available for comparison we will instead qualify β by considering the potential gains of deviating from β itself: $\hat{\ell}_i \approx \tilde{\ell}_i(\beta_i, \beta_{-i})$. We will also estimate the "true" epsilon of β, i.e., the smallest ϵ such that β forms an interim ϵ-BNE. This estimator will be denoted by $\hat{\epsilon}$. As we will see, these additional metrics in the absence of analytical solutions are costly: Calculating $\hat{\ell}$ and $\hat{\epsilon}$ relies on a grid $\{b_{i,w} | w = 1, \dots, n_{\text{grid}}\}$ of equidistant feasible bids for each player i, in order to cover the spaces $\mathcal{A}_i$. For given v_i and b_i, one can then approximate the interim loss $\overline{\ell}$ via

$$\hat{\lambda}_i(v_i; b_i, \beta_{-i}) \equiv \max_{w \in \{1,\dots,n_{\text{grid}}\}} \frac{1}{n_{\text{batch}}} \sum_{h=1}^{n_{\text{batch}}} u_i\left(v_i; b_{i,w}, \beta_{-i}(v_{h,-i})\right) - u_i\left(v_i; b_i, \beta_{-i}(v_{h,-i})\right). \tag{5.3}$$

Here the batch n_{batch} only runs across opponent valuations v_{-i}. Evaluating $\hat{\lambda}_i$ for a single valuation v_i therefore requires $(n_{\text{batch}} + 1) \cdot n_{\text{grid}}$ simulations of the auction. The ex-ante loss can then be estimated as $\hat{\ell} = \frac{1}{n_{\text{batch}}} \sum_h \hat{\lambda}_i(v_{h,i}; \beta_i(v_{h,i}), \beta_{-i})$.

The worst-case interim loss is then given by $\hat{\epsilon} = \max_h \hat{\lambda}_i(v_{h,i}; \beta_i(v_{h,i}), \beta_{-i})$. Bosshard et al. (2020) proofed that this estimator can be shown to be an upper bound under further assumptions on the mechanism and the strategies. They additionally provide empirical evidence of the approximation quality of the estimator which justifies its usage.

Both computations can use a shared state for the estimations of $\hat{\lambda}$ but nevertheless $\mathcal{O}(n \cdot n_{\text{grid}} \cdot n_{\text{batch}}^2)$ auction simulations are necessary to compute these metrics. In comparison, a learning update in NPGA needs $\mathcal{O}(n \cdot P \cdot n_{\text{batch}})$ simulations only, with the population size $P \ll n_{\text{grid}}$. Due to the high cost of these additional metrics on dense grids $b_{i,w}$, we evaluate the metrics $\hat{\ell}$ and $\hat{\epsilon}$ on smaller batch sizes than $\mathcal{L}$, and only once at the end of an experiment. Finally, to approximate the relative utility loss (Equation 5.1) in the absence of known BNE, we estimate the relative ex-ante utility loss incurred in hindsight by not playing a best response, given as

$$\hat{\mathcal{L}}(\beta_i) \equiv 1 - \frac{\hat{u}_i(\beta)}{\hat{u}_i(\beta) + \hat{\ell}_i(\beta)}. \tag{5.4}$$

We choose this as our main evaluation criterion as its calculation is feasible and its values are comparable across the variety of settings considered.

6. Results

In this section, we report the results of several challenging auction models that allow for various types of asymmetries among bidders and fairly general market environments. In many of these environments, we have analytical solutions which provide unambiguous baselines. Note that these environments already describe some of the most challenging equilibrium problems to solve analytically. For more complex models, closed-form solutions of Bayesian Nash equilibrium strategies are typically not available. We introduce these environments individually and report the results and the runtimes. As we will observe, NPGA converges to approximate equilibria in all presented settings.

We use common hyperparameters across almost all settings (except where noted otherwise): fully connected neural networks with two hidden layers of ten nodes each with SeLU activations (Klambauer et al. 2017), as well as ReLU activations in the output layer. The parameters θ_i are then given by the weights and biases of these networks. The resulting parameter dimensionality d_i for each bidder thus depends on the dimensionality of the input and output layers and ranges from $d_i = 141$, in the single-item settings, to $d_i = 372$ in the 12-item multi-unit setting. All experiments were performed on a single Nvidia GeForce 2080Ti with 11GB of RAM and batch sizes in Monte Carlo sampling were chosen to maximize GPU-RAM utilization: A learning batch size of 2^{18}; primary evaluation batch size (for $\mathcal{L}$, L_2) of 2^{22}; and secondary evaluation batch size $n_{\text{batch}} = 2^{12}$ and grid size $n_{\text{grid}} = 2^{10}$ (for $\hat{\ell}$, $\hat{\epsilon}$). Each experiment was repeated ten times with 2,000 learning iterations each. Section 1 in the online supplement (Bichler et al. 2023) gives insights on the influence of the batch size and the population size, arguably the most important hyperparameters of NPGA. In the single-item auctions, it takes approximately 0.3 seconds to compute each learning iteration, whereas the combinatorial LLG auction takes about 2.0 seconds due to the complexity of the auction mechanism. For the larger LLLLGG and LLLLRRG auctions under the first-price payment rule, the computation takes under one second per iteration. We present a thorough discussion of the factors which influence the computational cost and runtimes in Subsection 6.6.

6.1. Single-Item Auctions with Asymmetric Priors

Our initial analysis focuses on a standard *single-object* first-price sealed-bid (FPSB) auction with asymmetric priors, where bidder valuations are drawn from two different distributions. FPSB auctions have mostly been analyzed with symmetric priors and equilibrium bid functions. Asymmetric prior distributions are harder to analyze analytically compared to symmetric environments, but a few environments with analytical solutions are known. We analyze three different environments, one with two overlapping uniform distributions and a unique BNE, one with two disjunct uniform distributions and multiple BNE, and another one where the priors are non-linear beta functions.

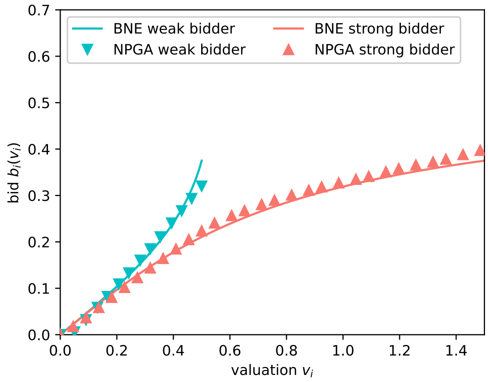


Figure 1 Equilibrium bid function and strategies learned by NPGA in the asymmetric single-item setting with overlapping valuations.

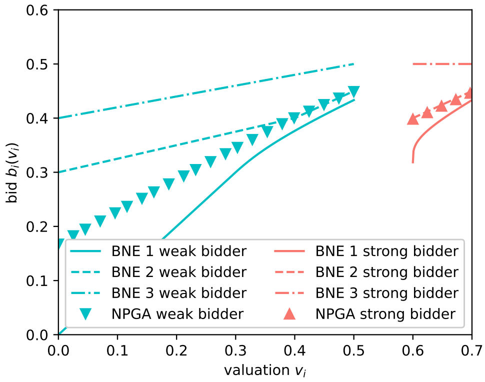


Figure 2 Equilibrium bid function and strategies learned by NPGA in the asymmetric single-item setting with non-overlapping valuations.

Table 1 Average losses achieved in the asymmetric first-price setting with overlapping valuations. Mean and standard deviation are aggregated over ten runs of 2,000 iterations each. The time per iteration is 0.2886 (0.0280) seconds.

bidder	$\mathcal{L}$	$\hat{\mathcal{L}}$	L_2
strong bidder	0.0024 (0.0026)	0.0178 (0.0037)	0.0104 (0.0055)
weak bidder	0.0074 (0.0031)	0.0524 (0.0134)	0.0128 (0.0029)

6.1.1. Asymmetries Induced by Priors with Different Domains.

We first analyze an environment with two bidders who have overlapping uniform prior distributions supported on $(0, 1/2)$ and $(0, 1)$ describing a weak and a strong bidder, respectively. The analysis goes back to Plum (1992). In the BNE, the weaker bidder bids more aggressively than the strong bidder. Figure 1 shows an example of the learned and the analytical BNE bid functions for both bidders. NPGA achieves a relative loss $\mathcal{L}$ below 1% for both types of bidders. Aggregated performance results over ten runs are displayed in Table 1.

This Bayes-Nash equilibrium is unique (Maskin and Riley 2000, Lebrun 2006) given the requirement that bidders may never bid above their observed valuation. Kaplan and Zamir (2015) relaxed this assumption. In their model, the prior distributions are non-overlapping, which results in additional equilibria. In particular, in BNE 1 and 2, which are also depicted in Figure 2, the weaker bidder has incentives to overbid. They conclude that the commonly used assumption of no overbidding, or more generally, the elimination of weakly dominated strategies, should be taken more carefully in asymmetric auctions.

Table 2 Average NPGA losses achieved in asymmetric first-price setting with non-overlapping valuations. Aggregated over ten runs of 2,000 iterations each and compared against the second equilibrium of Kaplan and Zamir (2015). The time per iteration is 0.2856 (0.0221) seconds.

bidder	$\mathcal{L}^{\text{BNE2}}$	$\hat{\mathcal{L}}$	L_2^{BNE2}
strong bidder	0.0080 (0.0097)	0.0104 (0.0012)	0.0109 (0.0085)
weak bidder	0.1687 (0.2310)	0.0229 (0.0140)	0.0544 (0.0161)

We analyzed NPGA in this setting with bidders that have non-overlapping uniform prior distributions, $\mathcal{V} = (0, 0.5) \times (0.6, 0.7)$ (see Table 2). In this model, there are three Bayesian Nash equilibria. Despite the equilibrium selection problem in this game, starting from truthfully initialized strategies, the bidding converges to BNE 2. The stronger bidder is able to decrease her relative utility loss below 1%. Only the weaker bidder has difficulties finding a particular strategy for low valuations because bids in this range are far from competitive for *any* opposing bids and rarely, if ever, win. This strategic disadvantage leads to sparse opportunities to learn in this specific setting, which in turn causes higher relative errors. In fact, only about 1/5 of the sampled data, i.e., the highest valuations of the weak bidder, are relevant for learning.

6.1.2. Asymmetries Induced by Different Prior Densities. For the single-item symmetric FPSB auction with two bidders and assuming uniform priors on $(0, 1)$, the equilibrium strategies and market outcomes are well understood analytically. Apart from the uniform distribution, we also want to analyze asymmetric environments with more complex non-linear prior distributions. Therefore, we analyzed an environment with two bidders whose values are drawn from a beta distribution B with parameters $\alpha, \beta > 0$. Note that for $\alpha = \beta = 1$ the beta distribution equals the uniform distribution. Except for this special case, no analytical equilibrium is known for the asymmetric case. Now we can analyze diverse market outcomes by running NPGA for various combinations of these parameters. As an example, we have selected a valuation prior of $B(0.8, 1.2)$ for the weak bidder and $B(1.2, 0.8)$ for the strong bidder's valuations prior. Note that NPGA has no access to the underlying distributions explicitly, but it learns the opponent's prior implicitly by observing frequencies of the played actions.

As a result of the change in the prior distributions, we already see the change in strategy in Figure 3 compared to the BNE of $\beta(v) = \frac{1}{2}v$ under common uniform priors. As expected,

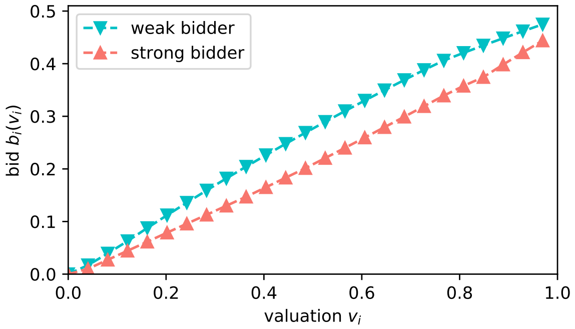


Figure 3 **Learned bid functions after 2,000 iterations for bidders with asymmetric priors:** $B(0.8, 1.2)$ **for the weak bidder with an expected valuation of** 0.4 **and** $B(1.2, 0.8)$ **for the strong bidder with an expected valuation of** 0.6**.**

because of its strategic disadvantage, the weaker bidder bids more aggressively, whereas the stronger bidder can lower its bids. As no analytical equilibrium is available to compare against, we report the approximated utility loss $\hat{\mathcal{L}}$ from Equation 5.4—the amount of a possible utility gain against the opponent—which decreases below 2.37% for the strong bidder and to 3.48% for the weaker bidder. The time per iteration of NPGA of 0.3131 (±0.0220) seconds is comparable with the previous single-item experiments.

6.2. Multi-Unit Auctions with Asymmetric Bidders

This section is concerned with a specific type of multi-unit uniform-price auction with two different classes of bidders for which a closed-form expression of the equilibrium is not available. Such mechanisms are used in treasury bill auctions and also in electricity markets. The environment is very large, with up to 12 units, and NPGA is able to compute a sufficiently close equilibrium in a few minutes.

Demand reduction is an important characteristic of equilibrium bidding strategies in uniform-price auctions (Krishna 2009): Bidders submit bids on fewer items in order to reduce competition, lower the price, and increase their payoffs. The phenomenon of demand reduction can be observed in all our experiments.

In our experiments, we consider two weak bidders with uniform, marginally decreasing valuations on $\mathcal{V}_i = \{v_i \in [0,1]^m : v_{i,1} \geq \cdots \geq v_{i,m}\}$ and one strong bidder with analogously distributed valuations on $[0, 2]$. Unlike in general CAs in multi-unit auctions it is sufficient to bid on individual items rather than bundles. Thus, the action space is given by $\mathcal{A}_i = \mathbb{R}^m_+$ and the neural network strategies take m inputs (the marginal valuations) and produce m outputs (the bids for each incremental unit received).

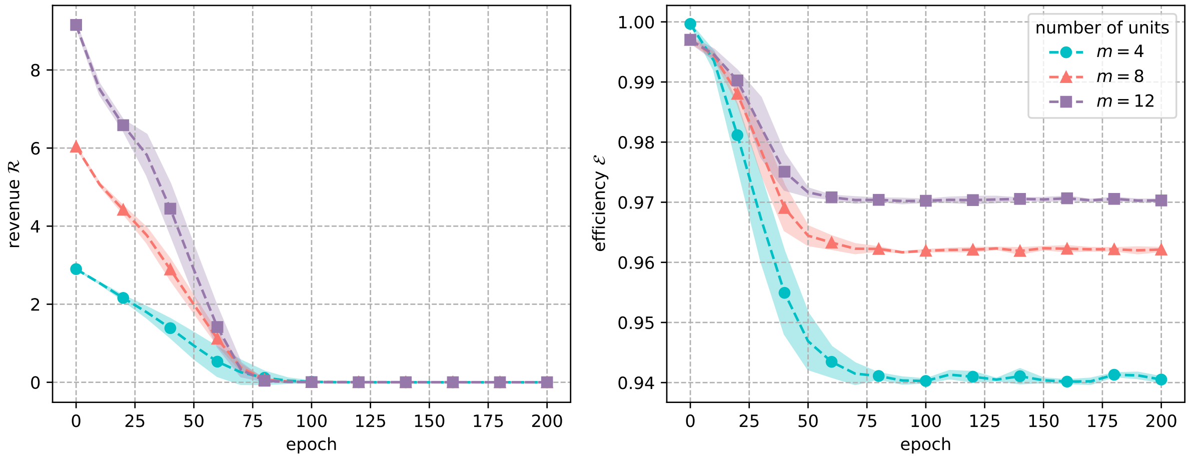


Figure 4 **Revenue $\mathcal{R}$ and efficiency $\mathcal{E}$ during self-play in different asymmetric multi-unit auctions. The means (opaque lines) and the standard deviations (shaded area) are depicted.**

Simulating market sizes with $m \in \{4, 8, 12\}$, one can observe that the agents reduce their demand by lowering bids for multiple goods to zero. For example, in the market with four goods, the strong bidder and weak bidders only bid on two items and one item, respectively. Thus, they learn to collaborate to maximize their payoff. For the remaining demand, the bids are approximately truthful. Similar observations can be made in the other markets for a corresponding higher demand. Figure 4 shows how the seller's revenue decreases to zero when bidders learn to reduce demand and how the efficiency decreases when initialized with truthful bidding strategies. We do not plot the exact bid functions learned due to space constraints. Note that demand reduction happens when the bidders' demand can be easily distributed among the available goods. In other experiments, where there are only very few items and many bidders, the prices stay high.

The approximate utility loss decreases consistently below 1% for all runs.[2] With the default batch size, the experiments with 4, 8, and 12 units took on average about 1.1557 (±0.022), 1.3056 (±0.0207), and 2.405 (±0.0259) seconds, respectively.

6.3. The Asymmetric LLG Model

Next, we focus on the LLG model with three single-minded bidders and two heterogeneous objects or items (Ausubel et al. 2006). This model has received significant attention

[2] Note that we have increased the grid size used for computing the utility loss for the 4, 8, and 12 item case to 2^{14}, 2^{16}, and 2^{22}, respectively. The resulting grid is not as dense as if it was applied in single-dimensional environments.

in the study of core-selecting combinatorial auctions, which are being used in spectrum auctions worldwide (Goeree and Lien 2016, Bichler and Goeree 2017). After the auctioneer has determined the welfare-maximizing allocation, she computes a minimum-revenue core-selecting payment, where winning bidders merely have to pay enough such that no coalition of bidders could potentially deviate together with the auctioneer.

The LLG model is small enough to allow for game-theoretical analysis. Here, the global bidder is interested only in winning the package of both objects while the two local bidders desire exactly one of the objects each. The local bidders thus only need to outbid the global bidder. Both local bidders have an incentive to free-ride on each other, reminiscent of public goods problems. In the original model, the local bidders' priors are symmetric and the global bidder has a simple dominant strategy to bid truthfully in any core-selecting auction. Analytical solutions for core-selecting combinatorial auctions with different payment rules exist (Goeree and Lien 2016, Ausubel and Baranov 2019). Gradient dynamics were shown to achieve very good results in this standard model and approximate the BNE of the local bidders closely (Bichler et al. 2021).

Ott and Beck (2013) introduced a version with asymmetry among the local bidders that causes overbidding by one of the local bidders, which may help explain the outcomes observed in several real-world spectrum auctions. Unlike in the original LLG model, Ott and Beck (2013) define bidder local 2 to be *favored*, meaning that she pays VCG prices[3] for every realization of bids and for every optimal assignment of the items. As a result, bidder local 1 has to pay a higher price. The authors derive an intriguing BNE in which bidder local 1 overbids while both other bidders report their valuations truthfully. More precisely, bidder local 1 places bids for two bundles: the bundle containing only her desired item, as well as the package of both items. Her bid for the package of both items always exceeds the bid for the single desired good, which implies positive demand for the second item even though it provides no additional value to the bidder. This results from an incentive of bidder local 1 to raise the other bidders' payments so that her payment decreases. Such overbidding can increase the prices for opponents, which might lead to high revenues and price differences among bidders. They characterize the exact BNE strategy which is depicted in Figure 5 below. This model is important as it shows that the assumption

[3] Vickrey-Clarke-Groves (VCG) payments are calculated such that each bidder pays for the harm they cause to other bidders by participating in the auction.

Table 3 **Results in the asymmetric LLG setting after 2,000 iterations and averaged over ten repetitions. The mean and standard deviation are shown.**

bidder	$\mathcal{L}$	$\hat{\mathcal{L}}$	L_2
local 1	0.0005 (0.0005)	0.0119 (0.0107)	0.0353 (0.0082)
local 2	0.0001 (0.0001)	0.0172 (0.0151)	0.1146 (0.0600)
global	0.0000 (0.0000)	0.0058 (0.0054)	0.0281 (0.0112)

that each player only needs to bid for her bundle of interest is, in fact, restrictive, even when the single-mindedness of bidders is common knowledge. Without this assumption, very different equilibrium behavior can emerge as was recently discussed by Bosshard and Seuken (2021).

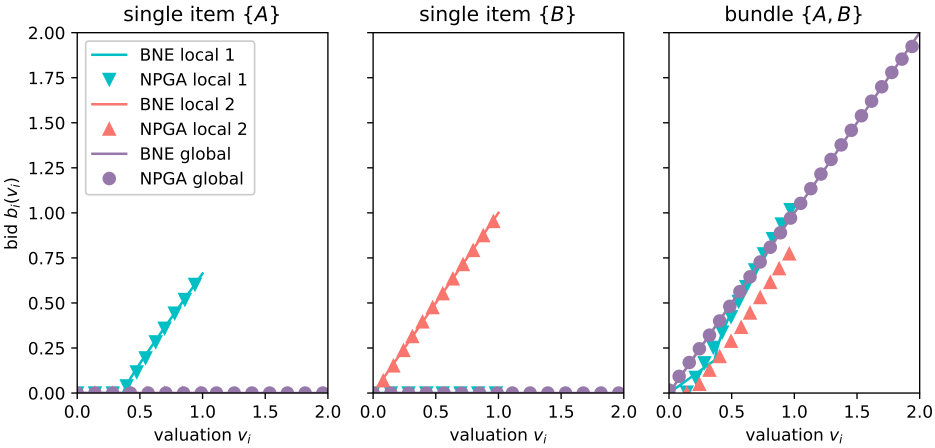


Figure 5 **Learned strategies in the asymmetric LLG setting. The left two subplots depict the bids on the individual items that must compete with the bundle bids in the rightmost plot. Bidders 1 and 3 learn to bid almost truthfully and bidder 1 indeed learns to overbid on the bundle as the theory suggests.**

Table 3 shows the performance of NPGA in this market. The resulting loss in equilibrium compared to adhering to the analytical BNE strategy $\mathcal{L}$ is well below 0.1% across all agents. Note that the bidders local 2 and global indeed learn to report their valuations truthfully for item B and the bundle, respectively. There is a small deviation from the analytical BNE for bidder local 2, who decreases her bundle bid slightly below her valuation. However, she would not have to bid on the bundle at all in equilibrium: Note that when bidding the same value on item B and the bundle, she would never be allocated the bundle in this auction. As such, the bid on the package of both items learned is irrelevant and the outcome of the auction under the learned NPGA strategies will always be identical in terms of prices and allocations to those in the analytical BNE. Importantly, the bundle bid for bidder local 1 indeed lies above the truthful bid for high valuations, which describes a non-obvious bidding strategy. Notably, NPGA can discover this incentive for overbidding. We point out that there is a minor difference in the NPGA strategy and the BNE in the bundle bid of bidder local 1 for low valuations. However, this difference has a negligible impact on the expected utility of any of the agents.

The time per NPGA update iteration averages at 1.9602 (±0.0358) seconds. Here, we clearly see the computational impact of allocating a bundle of goods, and computing the corresponding prices, as compared to auctions with a single good or multiple goods that are sold individually. The computational workload lies mainly in simulating the auction outcomes and not in learning and updating the strategies themselves, as we will discuss in Subsection 6.6.

6.4. The Split-Award Auction Model

Even in the asymmetric LLG model discussed above, each bidder is only interested in one package. An environment of a combinatorial auction with multi-minded bidders was analyzed in Anton and Yao (1992) and later in Kokott et al. (2019). This model is known to have multiple pure BNE, and it is interesting to understand how NPGA deals with the resulting equilibrium selection problem.

The model is a reverse auction and it is described by the bidders' type (or cost) distribution

$$\mathcal{V}_i = \{v_i \in \mathbb{R}^2 : v_{i,1} \sim F,\ v_{i,2} = C \cdot v_{i,1}\}, \quad i = 1, 2,$$

where $v_{i,1}$ corresponds to the cost of the 50% lot (or items) and the *efficiency parameter* C corresponds to the fraction of total costs for one of the lots. In our experiments we set parameters $F = \mathcal{U}(1.0, 1.4)$ with $C = 0.3$, being consistent to prior experimental work (Kokott et al. 2019). The environment describes diseconomies of scale in the production costs, which make the game strategically interesting.

There are two classes of Bayesian Nash equilibria in this game: First, there is a (single) so-called "winner-takes-all" equilibrium (WTA), which is economically inefficient and in which one bidder wins both items. The other class comprises a continuum of efficient "pooling equilibria" where both suppliers coordinate and reach a common price such that each

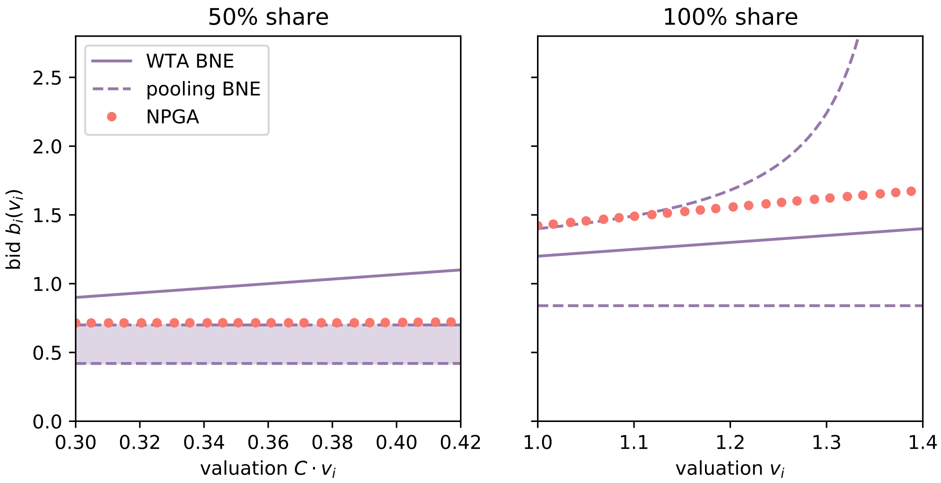


Figure 6 **The figure depicts the winner-takes-all equilibrium (solid line), the bounds for the range of efficient pooling equilibria (shaded), and the NPGA strategies (dotted line) for the first-price split-award auction. As the NPGA strategy is within the continuum of efficient pooling equilibria, two bidders playing according to this strategy always end up with a split contract for one lot each.**

bidder wins one of the goods (Anton and Yao 1992). In Figure 6, this class is represented by the shaded area. In such a pooling strategy, the two bidders select a price independent of their type or value. The payoff-dominant strategy for each bidder is achieved in the pooling equilibrium with the highest bids on a single lot. Apart from these two classes of pure-strategy Nash equilibria, hybrid equilibria are known to exist and there might also be mixed equilibria in nondeterministic strategies, which makes this setting strategically challenging.

Figure 6 depicts the analytically known pure-strategy BNE alongside a strategy learned via NPGA. Running NPGA multiple times, it always converges to a state close to the bidder-optimal pooling BNE: the bidders cooperate in the split equilibrium, where each one wins one lot a high price. NPGA reaches an average utility of 0.384 over ten runs compared to an expected utility of 0.34 in the analytical BNE. This outcome is notable as it requires coordination between the players which is strategically much more challenging than the simple competition to win both items at once, which resembles a single-item auction: To achieve a pooling equilibrium, players must not only submit a high bid for the single-lot, but also need to coordinate on a bid for the two-item bundle, such that deviating from the pooling strategy does not become profitable for the opponent.

Table 4 Results of NPGA after 5,000 iterations in the LLLLGG first-price auction. Results are averages over ten replications and the standard deviation is displayed in brackets.

bidder	$\tilde{u}$	$\hat{\epsilon}$	$\hat{\mathcal{L}}$
globals	0.2366 (0.0040)	0.0235 (0.0026)	0.0171 (0.0006)
locals	0.1793 (0.0012)	0.0241 (0.0024)	0.0230 (0.0006)

With NPGA, the agents learn to bid accordingly on the 100% share in this equilibrium, but this bid becomes subject to minor random changes as there is no “reward signal,” that is, the bid does not determine the price. In Figure 6 one can also see that bidding on the 50% lot is very close to the payoff-dominant (highest) pooling bid, whereas the bid on the 100% share lies within the continuum of possible equilibria. The distance in strategy space L_2 decreases to 0.0251, where we only measure the distance of the winning bid as the other bid falls within the continuum of possible BNE bids. The relative ex-ante utility loss $\mathcal{L}$ decreases to 0.0185 and $\hat{\mathcal{L}}$ also falls below 2%. The average time per iteration of 0.4627 (± 0.0154) seconds is again much lower than in the combinatorial LLG auction.

6.5. Large Combinatorial Auction Models

Finally, we analyze the LLLLGG model which was introduced by Bosshard et al. (2020) as a benchmark for equilibrium computation, as well as an extension, the LLLLRRG model. There is little hope for analytical solutions to such problems and the fact that the winner determination and payment rules involve NP-hard problems makes them challenging problems for equilibrium computation.

In the LLLLGG model six bidders compete for eight items: Inspired by geographical constraints, four of the bidders are “local” and are interested in two overlapping bundles of two items each. The other two bidders are “global”, and each aims to win one of two larger bundles comprising four items each. These bidder classes are asymmetric and no analytical BNE is known. Therefore, we again report the utility loss that we find after learning with NPGA.

As shown in Figure 7, the bidders’ utility converges quickly to around 0.24 (local bidders) and 0.18 (global bidders) and the utility losses drop quickly. Due to the computational requirements of this model, we reduced the number of experiments. Both bidders show a small relative ex-ante utility loss of $\hat{\mathcal{L}} < 1.8\%$ and 2.4% for the global and local bidders, respectively. Direct runtime comparisons to other state-of-the-art methods like

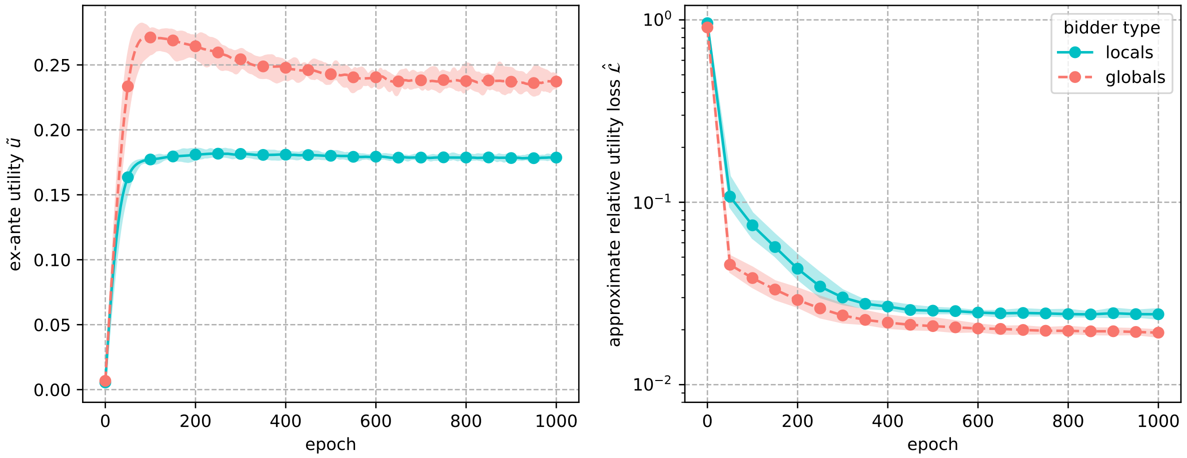


Figure 7 **Ex-ante utility $\tilde{u}$ and loss $\hat{\mathcal{L}}$ of in NPGA self-play in the LLLLGG first-price auction. The shaded area and line show mean and standard deviation over ten repetitions.**

Bosshard et al. (2017, 2020) are difficult due to differences of NPGA and their method in terms of goals (ex-ante vs. “stronger” ex-interim equilibria), implementation (generic vs. setting-specific), and hardware architecture (consumer-grade GPU vs. CPU-cluster). For the LLLLGG first-price auction, Bosshard et al. (2017) report an estimated absolute ex-interim 0.0037-BNE computed in 54,384 CPU-core hours. NPGA, on the other hand, finds an estimated (absolute) ex-ante 0.0042-BNE (absolute ex-interim 0.0241) in 38.3 minutes (corresponding to 0.4616 ($\pm$0.0010) seconds per iteration times 5,000 iterations) on a single GPU ($\approx$ 2,895 CUDA-core-hours).

We also explore a modified version, which we call LLLLRRG, which adds a third class of bidders interested in winning all eight items. Figure 4 in the online supplement (Bichler et al. 2023) depicts the valuation structure. This larger setting has not been explored in the literature previously and, to our knowledge, is the largest combinatorial auction for which a numerical BNE has been computed to date. Note that this environment is highly challenging for equilibrium computation because the auction mechanism needs to solve an NP-hard problem. One iteration of NPGA in this first-price auction takes on average 0.8097 ($\pm$0.0010) seconds on our machine. Table 5 shows the full results under the same hyperparameters as in the LLLLGG experiments. The relative utility loss decreases proportionally to the bidders’ strength: to values below 1% for the global bidder and to values below 3.9% for the local bidders.

Table 5 Results (mean and standard deviation) in the LLLLRRG setting after 5,000 iterations and averaged over three repetitions.

bidder	$\tilde{u}$	$\hat{\epsilon}$	$\hat{\mathcal{L}}$
locals	0.0939 (0.0009)	0.0194 (0.0013)	0.0382 (0.0016)
regionals	0.1069 (0.0024)	0.0192 (0.0004)	0.0288 (0.0001)
global	0.4396 (0.0044)	0.0356 (0.0031)	0.0093 (0.0009)

Table 6 Overview of all auction environments with the corresponding NPGA hyperparameters and the resulting number of simulations and runtimes. One NPGA iteration requires $n_{\text{models}} \cdot n_{\text{batch}} \cdot (P+1)$ auction evaluations, where batches are computed in parallel and model perturbations sequentially.

setting	n_{batch}	P	number of iterations	per iteration time (s)	per iteration auctions	total time (h:m:s)	total auctions
single-item uniform overlapping FPSB (6.1.1)	262,144	64	2,000	0.2886	34,078,720	0:09:37.20	68,157,440,000
single-item uniform non-overlapping FPSB (6.1.1)	262,144	64	2,000	0.2856	34,078,720	0:09:31.20	68,157,440,000
single-item beta asymmetric FPSB (6.1.2)	262,144	64	2,000	0.3131	34,078,720	0:10:26.20	68,157,440,000
multi-unit with 4 units (6.2)	262,144	64	2,000	1.1557	34,078,720	0:38:31.40	68,157,440,000
multi-unit with 8 units (6.2)	262,144	64	2,000	1.3056	34,078,720	0:43:31.20	68,157,440,000
multi-unit with 12 units (6.2)	262,144	64	2,000	2.4050	34,078,720	1:20:10.00	68,157,440,000
LLG, adapted VCG (6.3)	131,072	64	2,000	1.9602	25,559,040	1:05:20.40	51,118,080,000
split-award FPSB (6.4)	262,144	64	2,000	0.4627	17,039,360	0:15:25.40	34,078,720,000
LLLLGG FPSB (6.5)	262,144	64	5,000	0.4616	34,078,720	0:38:28.00	170,393,600,000
LLLLRRG FPSB (6.5)	262,144	64	5,000	0.8097	51,118,080	1:07:28.50	255,590,400,000

6.6. Scalability and Computational Costs

Let us now provide a summary of all experiments with their runtimes (Table 6) and a discussion of the computational cost. The runtimes range from a few minutes up to 80 minutes for the most complex scenarios with an NP-hard allocation problem and eight bidders. Let us put these empirical results into perspective. At first sight, it is surprising that we can solve such equilibrium problems at all. As discussed in the introduction, the computational complexity of computing BNE in auction games is, in general, an open problem. The analysis by Cai and Papadimitriou (2014) for a specific asymmetric multi-object auction model proves PP-hardness of exact BNE computation and this suggests that the class of problems is generally very hard. They also show that learning only approximate BNE cannot be polynomial in the number of items to be sold in the auction. As a result, no algorithm can be expected to efficiently compute approximate BNE in the general case. Note that the hardness of approximating a BNE also hinges on the observation that the number of strategies grows quickly in the number of items in their environment. In many auction models, the number of relevant strategies is small even with multiple items. For example, bidders might only be interested in a few out of many items in a combinatorial auction. Even in large combinatorial auctions with many bidders, one can typically limit attention to the strategic analysis of a few pivotal bidders.

In this paper, we have analyzed a number of challenging environments which are significantly more complex than models for which we can derive an equilibrium strategy analytically. For example, combinatorial auctions require solving an NP-hard winner determination problem. Yet, we can solve problems with eight items and seven bidders interested in multiple packages within 67 minutes.

Let us analyze the computational costs of NPGA in more detail. As a zeroth-order method, the vast majority of the computational cost required by NPGA results from calculating samples of the ex-post utilities u across the joint valuation space $\mathcal{V}$, in order to compute estimates of $\tilde{u}$ via Monte Carlo integration (i.e., lines 4 and 8 of Algorithm 1). Here, the main driver of computational cost is the auction mechanism itself, i.e., the cost of computing the winning allocation and the price vector. The role of the remaining computations in NPGA — namely sampling joint valuations $v \in \mathcal{V}$ and noise vectors ε_p, performing forward passes $b_i = \pi_i(v_i; \theta_i)$, aggregating auction sample results into the gradient estimates, and updating the parameters — is negligible in comparison. The cost of computing an approximate BNE using NPGA is thus determined, on the one hand, by the sample efficiency of the algorithm, that is, the number of auction simulations required, and, on the other hand, the computational cost of computing the individual auction samples. As we will see, both of these aspects vary significantly across different auction settings.

First, let us discuss the *computational cost of performing auction simulations* $u_i(v_i, b)$ for given joint valuations v and bids b according to the players' current or perturbed neural net strategies. This complexity varies significantly between auction settings and pricing rules. For example, in a single-item first-price auction, determining the allocation and prices only requires finding a (batch-wise) maximum, which is computable in $O(n)$, whereas computing core prices in a combinatorial auction requires solving a sequence of constrained quadratic problems, which themselves already constitute NP-hard problems (in the number of bidders and items) in general.

For all settings analyzed in this paper, we leverage custom implementations of the auction mechanisms that allow data-parallel simulation on GPUs. As a result, the time to compute auction samples is approximately constant in the batch size as long as an entire batch fits in GPU memory, and grows linearly with batch size thereafter. (The *utility loss estimator*, whose computation is independent of the learning algorithm NPGA, exhibits the same dynamic. Figure 3 in the online supplement (Bichler et al. 2023) depicts the constant-then-linear time complexity as a function of memory footprint.) For the experiments presented in this paper, performing a single iteration of NPGA, which involves computing $P+1$ batches of auctions for each player, takes between 0.3 and 2.4 seconds on a single Nvidia GeForce RTX 2080Ti GPU (see Table 6). While our implementation sequentially computes the utilities for each of the P model perturbations, these operations could easily be parallelized across larger or multiple GPUs.

The other important aspect is the *sample complexity of the algorithm*, which further breaks down into the number of samples needed for gradient estimation in each iteration, and the number of iterations needed to converge to an equilibrium. As discussed above, the asymmetric settings we study differ from those in Bichler et al. (2021) in that no theoretical convergence guarantee is available for simultaneous gradient methods (or any no-regret learner), even asymptotically. Consequently, it is difficult to characterize the number of gradient updates needed to converge to an approximate equilibrium, even if an exact oracle for the ex-ante gradient were available.

The gradient estimation in one iteration of NPGA requires $n_{\text{batch}} \cdot (P+1)$ auction simulations for each player (or class of identical players). Both higher batch sizes and higher population sizes will reduce the variance of the estimator at the expense of higher computational costs. An exemplary analysis of the impact of these hyperparameters on the learned equilibrium outcomes in the LLLLGG setting is presented in the online supplement (Bichler et al. 2023). As the estimation is performed via Monte Carlo integration, it is susceptible to the curse of dimensionality: Given a fixed batch size of samples, the variance of the estimator will increase with the dimensionality of the valuation space $\mathcal{V}$.

Furthermore, the specific prior distributions F may also affect the fidelity of the Monte Carlo estimator: For example, we observe that ceteris paribus, settings with uniform priors exhibit lower variance in the gradient estimator, compared to nonuniform priors. Intuitively, this is because the tails of the distribution, particularly for the highest valuations, play a significant role in the total achievable utility of a player. As a result, more samples are necessary to adequately calculate the utility contribution of these low-density regions. For example, NPGA would require more iterations to achieve the same performance in the asymmetric setting with Beta-distributed priors (Subsection 6.1.2). In practice, one may employ several variance-reduction techniques, such as importance-sampling or low-discrepancy sequences of quasi-random valuation samples to further improve the sample efficiency of Monte Carlo integration (Bosshard et al. 2020). While these methods are conceptually applicable to NPGA, they require setting-specific implementations and are not explored further in this work.

In summary, the key drivers for the total runtime of NPGA are the number of players, number of items, choice of prior distribution, and auction mechanism, as they influence the ability to efficiently calculate low-variance gradient estimates. Importantly, we demonstrate that NPGA, for the first time, finds close approximations of BNE in two of the largest settings to date, namely a 12-item, 3-bidder multi-unit auction (Subsection 6.2), and the 8-item, 7-bidder combinatorial auction with multi-minded bidders (“LLLLRRG”, Subsection 6.5).

7. Conclusion

Understanding the result of strategic interaction on markets is a fundamental problem and one that appears everywhere in economics and the management sciences. Equilibrium solution concepts are our primary approach to studying the outcome of games with multiple interacting agents. They help understand fundamental questions about the efficiency of markets, but equilibrium analysis can also provide tangible guidance for bidding in specific markets such as in procurement auctions or in high-stakes spectrum sales and for the design of specific auction mechanisms. Algorithms to compute equilibrium strategies in games would have a substantial impact on theory and practice. However, computing equilibrium in auction games with continuous action space and value distributions turned out very challenging. We know little about the existence of equilibrium in such auctions and do not have a mathematical solution theory for the underlying differential equations in more complex markets. Obviously, an equilibrium solution concept that is intractable is of little value and can hardly serve as a prediction for the outcome of a game. Equilibrium learning provides a reasonable behavioral model of agents in a market. While the implementation of equilibrium learning algorithms in auction games is challenging, we show that NPGA reliably finds equilibrium in a surprisingly wide array of complex auction models. The experimental results reported in this paper show that the gradient-based algorithm implemented in NPGA finds BNE even in asymmetric environments with multiple equilibria. Such asymmetric environments required us to train multiple neural networks with multiple outputs, where convergence to the bidder-optimal equilibrium is far from obvious.

An open question concerns a broader theoretical characterization of Bayesian games in which NPGA converges to a Bayesian Nash equilibrium. However, this is a very challenging theoretical endeavor that is beyond this article. Learning dynamics do not generally obtain a Nash equilibrium (Benaim and Hirsch 1999). A number of recent results on matrix games showed that gradient dynamics can either circle, diverge, or even be chaotic (Sanders et al. 2018). Actually, the study of gradient dynamics in games is akin to studying dynamical systems and characterizing environments, where gradient dynamics converge to a Nash equilibrium (if one exists), can be arbitrarily complex (Andrade et al. 2021).

However, even if we do not know a priori if an algorithm converges, we can verify an approximate BNE ex-post, if the algorithm converges. If we analyze many environments as in this article, we might be able to induce characteristics of auction models that can be learned via NPGA and those that cannot. In our experiments, we found that NPGA always converged to an approximate Bayes-Nash equilibrium in single- and multi-object auctions and we did not encounter cycling or chaotic behavior as was observed for finite games. As such, NPGA provides the foundation for widely applicable equilibrium solvers.

Acknowledgments

We are grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BI 1057/9.

References

Andrade GP, Frongillo R, Piliouras G (2021) Learning in matrix games can be arbitrarily complex. *arXiv preprint arXiv:2103.03405* .

Anton JJ, Yao DA (1992) Coordination in split award auctions. *The Quarterly Journal of Economics* 107(2):681–707.

Armantier O, Florens JP, Richard JF (2008) Approximation of Nash equilibria in Bayesian games. *Journal of Applied Econometrics* 23(7):965–981.

Athey S (2001) Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica* 69(4):861–889.

Ausubel LM, Baranov O (2019) Core-selecting auctions with incomplete information. *International Journal of Game Theory* ISSN 1432-1270, URL `http://dx.doi.org/10.1007/s00182-019-00691-3`.

Ausubel LM, Milgrom P, et al. (2006) The lovely but lonely Vickrey auction. *Combinatorial auctions* 17:22–26.

Bajari P (2001) Comparing competition and collusion: a numerical approach. *Economic Theory* 18(1):187–205.

Balduzzi D, Racaniere S, Martens J, Foerster J, Tuyls K, Graepel T (2018) The mechanics of n-player differentiable games. *International Conference on Machine Learning*, 354–363 (PMLR).

Benaim M, Hirsch MW (1999) Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games. *Games and Economic Behavior* 29:36–72, URL `https://escholarship.org/uc/item/4qj1335f`.

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

Bichler M, Goeree JK (2017) *Handbook of spectrum auction design* (Cambridge University Press).

Bichler M, Heidekrüger S, Kohring N (2023) Online supplement to ”learning equilibria in asymmetric auction games”, Version v2021.0151 URL `https://github.com/INFORMSJoC/2021.0151`.

Bosshard V, Bünz B, Lubin B, Seuken S (2017) Computing bayes-nash equilibria in combinatorial auctions with continuous value and action spaces. *IJCAI*, 119–127.

Bosshard V, Bünz B, Lubin B, Seuken S (2020) Computing bayes-nash equilibria in combinatorial auctions with verification. *Journal of Artificial Intelligence Research* 69:531–570.

Bosshard V, Seuken S (2021) The cost of simple bidding in combinatorial auctions. *2021 Conference on Economics and Computation* (New York, NY, USA).

Bowling M (2005) Convergence and no-regret in multiagent learning. *Advances in neural information processing systems*, 209–216.

Brown GW (1951) Iterative solution of games by fictitious play. *Activity analysis of production and allocation* 13(1):374–376.

Cai Y, Papadimitriou C (2014) Simultaneous bayesian auctions and computational complexity. *Proceedings of the Fifteenth ACM Conference on Economics and Computation - EC '14*, 895–910 (Palo Alto, California, USA: ACM Press), ISBN 978-1-4503-2565-3.

Carbonell-Nicolau O, McLean RP (2018) On the existence of nash equilibrium in bayesian games. *Mathematics of Operations Research* 43(1):100–129.

Conitzer V, Sandholm T (2008) New complexity results about nash equilibria. *Games and Economic Behavior* 63(2):621–641.

Cournot AA (1838) *Recherches sur les principes mathématiques de la théorie des richesses.* URL `https://gallica.bnf.fr/ark:/12148/bpt6k6117257c`.

Daskalakis C, Goldberg P, Papadimitriou C (2009) The Complexity of Computing a Nash Equilibrium. *SIAM Journal on Computing* 39(1):195–259, ISSN 0097-5397, URL `http://dx.doi.org/10.1137/070699652`.

Etessami K, Yannakakis M (2007) On the complexity of nash equilibria and other fixed points (extended abstract). *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 113–123, FOCS '07 (USA: IEEE Computer Society), ISBN 0769530109, URL `http://dx.doi.org/10.1109/FOCS.2007.48`.

Ewert M, Heidekrüger S, Bichler M (2022) Approaching the overbidding puzzle in all-pay auctions: Explaining human behavior through bayesian optimization and equilibrium learning. *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 1586–1588, AAMAS '22 (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), ISBN 9781450392136.

Fudenberg D, Levine DK (2009) Learning and equilibrium. *Annu. Rev. Econ.* 1(1):385–420.

Goeree JK, Lien Y (2016) On the impossibility of core-selecting auctions. *Theoretical Economics* 11(1):41–52, ISSN 1555-7561, URL `http://dx.doi.org/10.3982/TE1198`.

Hartline J, Syrgkanis V, Tardos E (2015) No-Regret Learning in Bayesian Games. Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, eds., *Advances in Neural Information Processing Systems 28*, 3061–3069 (Curran Associates, Inc.), URL `http://papers.nips.cc/paper/6016-no-regret-learning-in-bayesian-games.pdf`.

Hazan E, Agarwal A, Kale S (2007) Logarithmic regret algorithms for online convex optimization. *Machine Learning* 69(2-3):169–192.

Jackson MO, Swinkels JM (2005) Existence of equilibrium in single and double private value auctions 1. *Econometrica* 73(1):93–139.

Kaplan TR, Zamir S (2015) Multiple equilibria in asymmetric first-price auctions. *Economic Theory Bulletin* 3(1):65–77.

Klainerman S (2010) Pde as a unified subject. *Visions in Mathematics*, 279–315 (Springer).

Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self-normalizing neural networks. *Proceedings of the 31st international conference on neural information processing systems*, 972–981.

Klemperer P (2000) Why every economist should learn some auction theory. *Available at SSRN 241350* .

Kokott GM, Bichler M, Paulsen P (2019) The beauty of Dutch: Ex-post split-award auctions in procurement markets with diseconomies of scale. *European Journal of Operational Research* 278(1):202–210.

Krishna V (2009) *Auction Theory* (Academic press).

Lebrun B (2006) Uniqueness of the equilibrium in first-price auctions. *Games and Economic Behavior* 55(1):131–151.

Letcher A, Balduzzi D, Racanière S, Martens J, Foerster JN, Tuyls K, Graepel T (2019) Differentiable game mechanics. *Journal of Machine Learning Research* 20(84):1–40.

Marshall RC, Meurer MJ, Richard JF, Stromquist W (1994) Numerical analysis of asymmetric first price auctions. *Games and Economic Behavior* 7(2):193–220.

Maskin E, Riley J (2000) Asymmetric auctions. *The Review of Economic Studies* 67(3):413–438.

Mertikopoulos P, Zhou Z (2019) Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming* 173(1-2):465–507.

Milgrom P (2017) *Discovering prices: auction design in markets with complex constraints* (Columbia University Press).

Monderer D, Shapley LS (1996) Potential games. *Games and Economic Behavior* 14(1):124–143.

Nash JF, et al. (1950) Equilibrium points in n-person games. *Proceedings of the national academy of sciences* 36(1):48–49.

Ott M, Beck M (2013) Incentives for overbidding in minimum-revenue core-selecting auctions. Number F16-V3 in Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Auctions and Licensing (Kiel und Hamburg: ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft).

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in PyTorch. *NIPS-W*.

Plum M (1992) Characterization and computation of Nash-equilibria for auctions with incomplete information. *International Journal of Game Theory* 20(4):393–418.

Rubinstein A (2016) Settling the complexity of computing approximate two-player Nash equilibria. *arXiv:1606.04550 [cs]* URL `http://arxiv.org/abs/1606.04550`.

Salimans T, Ho J, Chen X, Sidor S, Sutskever I (2017) Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *arXiv:1703.03864 [cs, stat]* .

Sanders JB, Farmer JD, Galla T (2018) The prevalence of chaotic dynamics in games with many players. *Scientific reports* 8(1):1–13.

Schäfer F, Anandkumar A (2019) Competitive gradient descent. *Advances in Neural Information Processing Systems*, 7623–7633.

Singh SP, Kearns MJ, Mansour Y (2000) Nash convergence of gradient dynamics in general-sum games. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)*, 541–548.

Ui T (2016) Bayesian nash equilibrium and variational inequalities. *Journal of Mathematical Economics* 63:139–146.

Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance* 16(1):8–37.

Zinkevich M (2003) Online convex programming and generalized infinitesimal gradient ascent. *Proceedings of the 20th International Conference on Machine Learning (icml-03)*, 928–936.

List of Symbols

$\mathcal{A}$	The set of feasible *action profiles* in a Bayesian game, i.e., *bid profiles* in actions. Cross product of individual players' action sets: $\mathcal{A} \equiv \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$
β	A joint *strategy profile* in a Bayesian Game.
β^*	A strategy profile that constitutes a Bayesian Nash equilibrium.
β_i	A feasible *pure strategy* of player i: $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$.
b	An action/bid profile. $b \in \mathcal{A}$
$B(\alpha, \beta)$	The Beta-distribution with shape parameters α and β.
b_i	An action/bid for player i. $b_i \in \mathcal{A}_i$.
C	Efficiency parameter in split-award auction setting. See Subsection 6.4.
d_i	The dimension of the parameter vector θ_i of player i's neural network π_i.
ϵ	The approximation-bound in an approximate BNE, indicating that each player's incentive to deviate is less than $\epsilon \geq 0$.
$\hat{\epsilon}$	An ex-post estimator for the worst-case ex-interim loss. Does not require access to an analytical BNE. See Section 5.
ε_p	The Gaussian noise vector of perturbation p in NPGA gradient computation.
ϕ_p	The *fitness* of perturbation $\theta_{i;p}$ of player i's neural network in NPGA gradient computation.
F_v	The joint prior distribution over types, marginalized by F_{v_i}.

G A Bayesian Game $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$.

$-i$ Index identifying a *partial* action/bid/strategy profile for all bidders except player i.

$\mathcal{I}$ The set of players in a game. Indexed by i. Total number of players is n.

i Index identifying a particular player.

$\mathcal{K}$ The set of feasible bundles of items, generally the power set of $\mathcal{M}$.

$\hat{\ell}_i$ An estimator of i's ex-ante loss $\tilde{\ell}_i(\beta_i, \beta_{-i})$ computed ex-post from observational data. Does not require access to an analytical BNE. See Section 5.

$\hat{\lambda}_i$ Auxiliary quantity in the computation of $\hat{\ell}$ and $\hat{\epsilon}$. $\hat{\lambda}_i(v_i; b_i, \beta_{-i})$ constitutes an ex-post estimator for the interim utility loss $\bar{\ell}_i$ of playing b_i at valuation v_i. See Equation 5.3.

$\hat{\mathcal{L}}$ An ex-post estimator for the *relative utility loss* $\mathcal{L}(\beta_i)$, when no access to the analytical BNE is available. See Equation 5.4.

$\mathcal{L}$ The *relative utility loss* $\mathcal{L}(\beta_i)$ of strategy β_i compared to an analytical BNE β^*. See Equation 5.1.

$\bar{\ell}_i$ The *interim* utility loss of player i. See Equation 3.3.

$\tilde{\ell}_i$ The *ex-ante* utility loss of player i. See Equation 3.7.

L_2 The L2-loss $L_2(\beta_i)$ of a strategy β_i compared to BNE β^*, i.e., the distance of β_i and β_i^* in strategy space. See Equation 5.2.

$\mathcal{M}$ The set of items sold in an auction. Total number of items is m. Items can be homogenous or heterogeneous.

m Total number of items in an auction.

$\mathcal{N}(\mu, \sigma^2)$ Gaussian Distribution with mean μ and standard deviation σ.

$\mathbb{N}$ The set of natural numbers.

n The total number of players in a game.

n_{batch} Batch size used in sampling opponent behavior when computing ex-post estimators $\hat{\ell}$ and $\hat{\epsilon}$. See Section 5.

n_{grid} Size of the discrete grid of alternative bids evaluated to compute ex-post estimators $\hat{\ell}$ and $\hat{\epsilon}$. See Section 5.

$\mathcal{P}_{\Sigma_i}$ The projection function onto the set Σ_i.

π_i Neural network for player i, implementing i bidding strategy β_i via $\beta_i(v_i) := \pi_i(v_i; \theta_i)$, where $\theta_i \in \Theta_i = \mathbb{R}^{d_i}$ are the network's parameters.

P Hyperparameter in NPGA. The *population size*, or number of perturbations of θ_i considered for each iteration of gradient computation.

p Index used for permutations $1, \ldots, P$ in NPGA gradient computation.

p_i Price paid by bidder i to the auctioneer after receiving bundle x_i.

$\mathbb{R}$ The set of real numbers.

Σ The set of feasible joint strategy profiles.

σ Hyperparameter in NPGA. The standard deviation of the Gaussian noise used in permuting neural network parameters.

Σ_i The set of feasible (pure) *strategies* for player i. Generally an infinite-dimensionally Hilbert space.

θ_i The parameter vector of player i's neural network π_i.

t Time / iteration number.

$\hat{u}_i$ An estimator of ex-ante expected utility $\tilde{u}$, computed ex-post via Monte Carlo integration over a large batch of realizations of $v \sim F_v$.

$\mathcal{U}(l,h)$ Uniform distribution with lower bound l and upper bound h.

$\overline{u}_i$ The expected *interim* utility $\overline{u}_i(v_i, b_i, \beta_{-i})$ of player i. See Equation 3.2.

$\tilde{u}_i$ The expected *ex-ante* utility $\tilde{u}_i(\beta_i, \beta_{-i})$ of player i. See Equation 3.5.

u_i The ex-post utility function $u_i(v_i, b_i, b_{-i})$ of player i. Generally nondifferentiable.

$\mathcal{V}$ The set of possible *valuation profiles*, i.e., generally the support of F_v.

v The *private valuation* or *type* profile, $v \in \mathcal{V}$. Generally used to refer to the Random Variable, sometimes also used to refer to a realization of the RV.

v_i The private valuation of player i. Generally a random vector of length 2^m, indicating i's willingness to pay when allocated a certain bundle. We also write $v_i(x_i)$ for the entry of v_i corresponding to the (scalar) valuation of player i for bundle $x_i \in \mathcal{K}$.

x_i The bundle of items allocated to player i. $x_i \in \mathcal{K}$.

INFORMS Journal on Computing

Online Supplement “Learning Equilibria in Asymmetric Auction Games”

Martin Bichler*, Stefan Heidekrüger, Nils Kohring
Technical University of Munich, Department of Computer Science, 85748 Garching, Germany

bichler@in.tum.de

1. Performance and Runtime Analysis

In the following, we look at the influence of some of the hyperparameters on the performance and runtime of the computation. We will illustrate that based on the example of the FPSB LLLLGG auction format.

1.1. Influence of batch size

The batch size corresponds to the number of auction games that are simultaneously played under the current strategies. Strategy updates are solely based on any changes in the utility estimate averaged over these batches, thus making the batch size one of the most crucial parameters of NPGA. Except for the parameter of interest (the batch size here), we leave all other parameters unchanged from their default values from Section 6 in the main paper.

Results for different batch sizes can be seen in Figure 1 for the first 500 iterations. As expected, lower batch sizes dramatically increase the variance in the utility estimates (see left plot) and thus prevent quick learning (see right plot).

Furthermore, the average time per iteration increases slowly as to be expected by vectorized GPU implementation. The average time per iteration for batch sizes of 64, 1,024, and 262,144 are 0.3717 (±0.0878), 0.3741 (±0.0757), and 0.4681 (±0.0283) seconds, respectively. Only once the available memory and available cores run out, the computation would have to be done in a sequential manner, which would lead to a drastic increase in computation time.

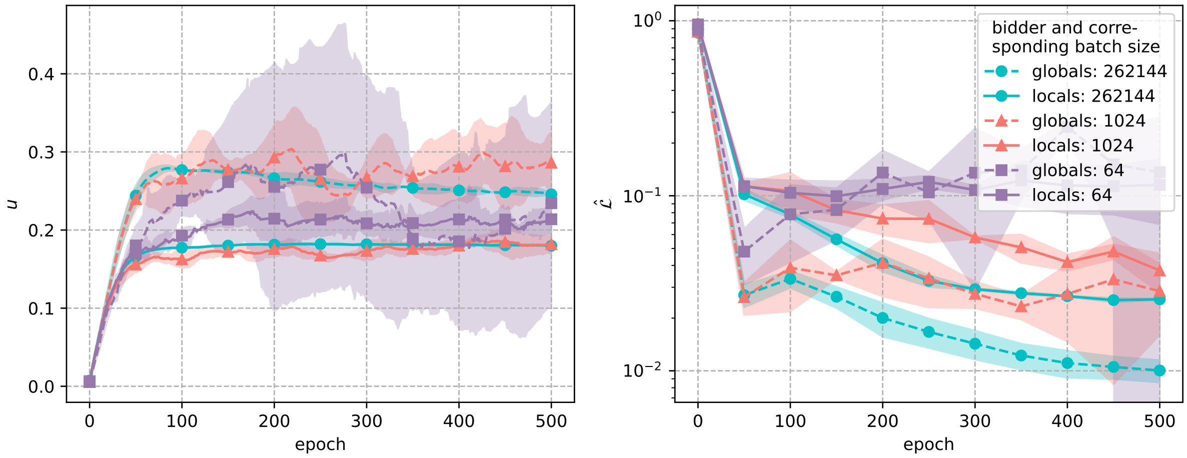


Figure 1 **Average utility of both bidder types (left) and relative utility loss (right) while learning via NPGA in LLLLGG first-price auction. Runs for three different population sizes, where each configuration was run three times (mean ± std), are depicted.**

1.2. Influence of population size

The population size refers to the number of sampled parameters that are considered during one estimation of the pseudogradient. Even though the needed utility estimates for each sample are independent of one another and could thus in principle be calculated in parallel, this is not implemented in favor of larger batch sizes that allow for better utility estimates. Therefore, it influences the running time linearly. Results for different population sizes can be seen in Figure 2 for the first 500 iterations.

As can be seen, the utility for larger population sizes changes quicker while the utility loss also decreases faster. The influence when increasing the population size is not as strong compared to changes in the batch size. This justifies using most of the GPU memory for sampling a large number of auction games for precise utility estimates.

Furthermore, the average time per iteration increases quickly as to be expected by the sequential computation of the fitness of individual population samples. The average time per iteration for population sizes of 16, 32, and 64 are 0.1264 (±0.0042), 0.2426 (±0.0162), and 0.4687 (±0.0283) seconds, respectively.

The observations extend qualitatively to other auction formats and payment rules except for an increased run time for core selecting payment rules.

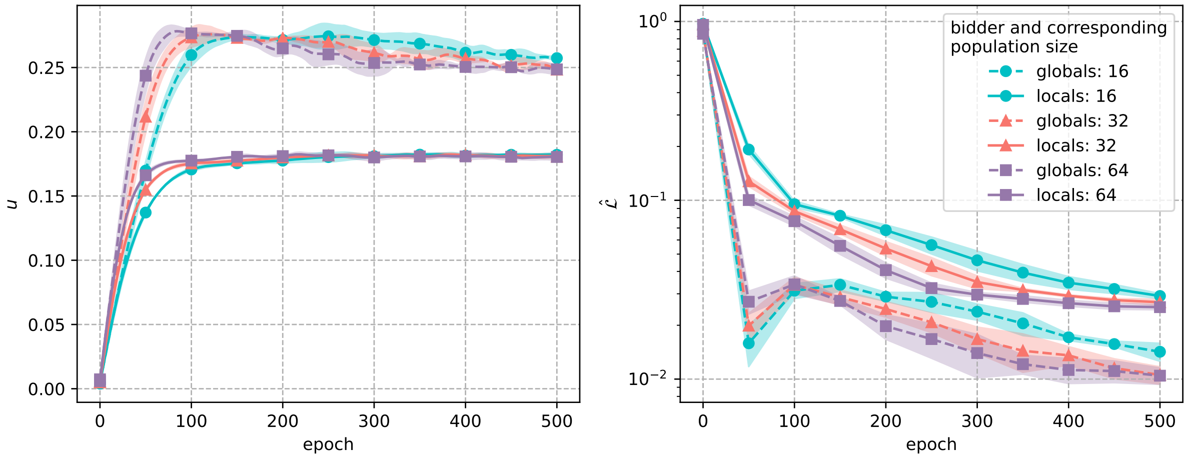


Figure 2 **Average utility of both bidder types (left) and relative utility loss (right) while learning via NPGA in LLLLGG first-price auction. Runs for three different batch sizes, where each configuration was run three times (mean ± std), are depicted.**

2. Approximation quality of utility loss

As we rely on the approximate utility loss $\hat{\mathcal{L}}$ from Equation 5.4, it is essential to have a reference of its estimation quality. Therefore, we have run the calculations for a variety of values for n_{batch} and n_{grid} in equilibrium for the weak bidder in the single-item auction with overlapping valuations. Of course, here the exact $\mathcal{L}$ is known to be zero. Figure 3 shows the results.

Let us emphasize two observations. First, counterintuitively, when increasing the grid size (comparing a finer grid of possible best responses), the utility loss increases. This is due to the bias introduced by choosing the best responses that maximize the utility of a particular sample, thus always selecting just that action that exploits the particular sample the most. One sees, however, that this effect decreases in severity with increasing grid sizes. Second, the right plot clearly shows the massive computational advantage of GPU computation as the calculations can be run in parallel, having nearly constant computational time for increasing batch sizes, as long as the tensor fits in memory.

In terms of scalability one has, on the one hand, the number of opposing agents that linearly affect the memory requirements, and on the other hand, the number of objects for sale that makes it exponentially harder to find a best response with the same level of accuracy.

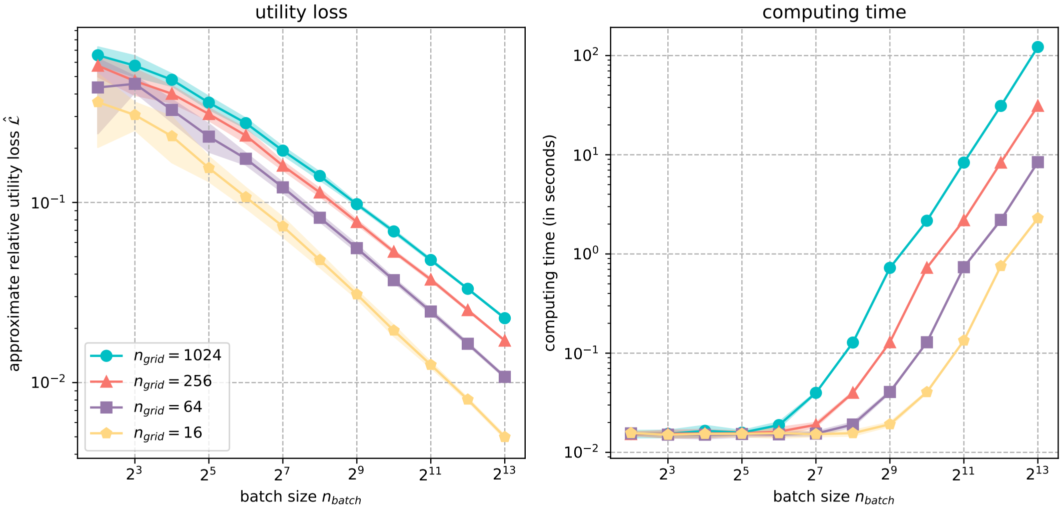


Figure 3 **Approximation quality of the utility loss (left) and computing time (right) for different batch sizes and different grid sizes averaged over ten runs each. The utility loss for the weak bidder in the single-item auction with overlapping valuations from Subsubsection 6.1.1 is depicted. Note that all axes are log-scaled.**

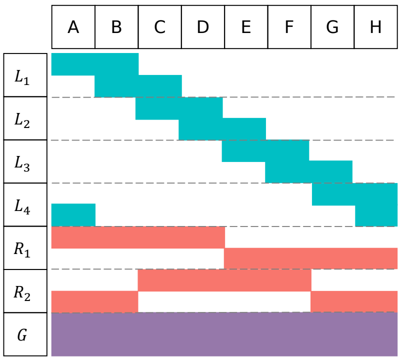


Figure 4 **Valuations in the LLLLRRG auction model. The columns depict items A through H and the rows correspond to the local, regional, and global bidders.**

3. Valuations in the LLLLRRG auction

As introduced in Subsection 6.5 of the main papaer, Figure 4 gives an overview of which agents are interested in which set of items (bundles) in the LLLLRRG auction.

Publication D: Approaching the Overbidding Puzzle in All-Pay Auctions

Peer-Reviewed Conference Paper

Title: Approaching the Overbidding Puzzle in All-Pay Auctions: Explaining Human Behavior through Bayesian Optimization and Equilibrium Learning

Authors: M. Ewert, S. Heidekrüger, M. Bichler

In: 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2022), Online, May 9–13, 2022. An extended abstract has been published in the conference proceedings.

Abstract: It is an established fact in behavioral economics that in lab experiments of auctions, human subjects do not adhere to the risk-neutral Bayesian Nash equilibria of such games. Several attempts at explaining this Overbidding Puzzle focus on the bidders' psychology and suggest they may have parametrized utility functions that differ from the risk-neutral payoff. However, analytical equilibria of the resulting modified games are generally not available. Consequently, it has been difficult to identify the specific parameters and assess the merits of these proposed modifications in explaining empirical observations.

With recent advances in equilibrium learning, it has become tractable to compute approximations of Bayesian Nash equilibria. Building on these advances and Bayesian optimization, we propose a novel regression framework to infer unobserved parameters of Bayesian games from behavioral data. We apply our method to two data sets of human bidding behavior in all-pay auctions. For the first time, this makes it possible to directly compare the goodness-of-fit of several proposed qualitative explanations of overbidding.

Contribution of thesis author: development and implementation, empirical analysis (auction settings), writing and revising the manuscript, project management

References: Full paper: Ewert et al. (2022a). Extended abstract: Ewert et al. (2022b)

Approaching the Overbidding Puzzle in All-Pay Auctions: Explaining Human Behavior through Bayesian Optimization and Equilibrium Learning

Markus Ewert, Stefan Heidekrüger, Martin Bichler
Technical University of Munich
Munich, Germany
{ewert,stefan.heidekrueger,bichler}@in.tum.de

ABSTRACT

It is an established fact in behavioral economics that in lab experiments of auctions, human subjects do not adhere to the risk-neutral Bayesian Nash equilibria of such games. Several attempts at explaining this *Overbidding Puzzle* focus on the bidders' psychology and suggest they may have parametrized utility functions that differ from the risk-neutral payoff. However, analytical equilibria of the resulting modified games are generally not available. Consequently, it has been difficult to identify the specific parameters and assess the merits of these proposed modifications in explaining empirical observations.

With recent advances in equilibrium learning, it has become tractable to compute approximations of Bayesian Nash equilibria. Building on these advances and Bayesian optimization, we propose a novel regression framework to infer unobserved parameters of Bayesian games from behavioral data. We apply our method to two data sets of human bidding behavior in all-pay auctions. For the first time, this makes it possible to directly compare the goodness-of-fit of several proposed qualitative explanations of overbidding.

KEYWORDS

Behavioural Game Theory, Equilibrium Learning, All-Pay Auctions

ACM Reference Format:
Markus Ewert, Stefan Heidekrüger, Martin Bichler. 2022. Approaching the Overbidding Puzzle in All-Pay Auctions: Explaining Human Behavior through Bayesian Optimization and Equilibrium Learning. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Auckland, New Zealand, May 9–13, 2022*, IFAAMAS, 8 pages.

1 INTRODUCTION

A standard assumption in economic theory is that market participants are utility-maximizing, rational agents, and thus, one would expect them to behave according to the market's equilibrium. However, studies in behavioral economics have repeatedly shown that human subjects do not conform to this assumption. A prominent example is the phenomenon of overbidding in auctions. Particularly in all-pay auctions, where all bidders have to pay their bid to the auctioneer, not just the winner, this behavior extends to a bimodal bidding pattern, where low-valued bidders underbid, and high-valued bidders overbid their optimal strategy expressed by the risk-neutral Bayesian Nash Equilibrium [28]. This type of auction has many practical applications like research and development activities in organizations or political campaigns [19]. Therefore, understanding the reasons for this bimodal behavior is essential and has been the subject of research in behavioral economics and psychology. This question has been dubbed the Overbidding Puzzle.

Previous research has approached this problem by suggesting that bidders either lack the cognitive capabilities to assess their winning probabilities [1] or questioning the validity of assuming risk-neutrality of the bidders [11, 18, 23]. A common threat in several of these, partly contradictory, approaches is that they model bidders having utility functions incorporating psychological effects which are quantified via some numerical parameters. However, to the authors' knowledge, there has been no unified approach for estimating such utility function parameters across several approaches and making their goodness-of-fit comparable on behavioral data. A key difficulty to this has been the computational complexity of computing the equilibria of the resulting parametrized auctions, which are commonly modeled as continuous-type-and-action Bayesian games. Analytical derivations of equilibria exist in some but not all settings, making parameter estimation intractable in those settings where they do not. Most recently, however, there has been progress in approximating equilibria of such Bayesian games via numerical techniques based on multi-agent learning.

In this study, we propose a novel estimation framework using Bayesian Optimization and equilibrium learning techniques that, for the first time, allows a quantitative comparison of the goodness-of-fit to experimental data of various behavioral explanation attempts to the Overbidding Puzzle. Here, we apply our method to overbidding in symmetric all-pay auctions and the assumptions of risk-averse and regretful bidders. Our empirical results coincide with established results in the empirical literature in aspects where quantitative results were previously available. However, our framework is not restricted to either a specific auction mechanism, type of utility function, or equilibrium learning method, such that future research may apply it to other problems in behavioral economics and behavioral psychology.

The remainder of this paper is structured as follows: In section 2, we will define the problem setup and discuss relevant literature in several overlapping fields. We formally introduce our estimation scheme in section 3 and our experimental setup in section 4, before discussing the results in section 5 and concluding the paper.

An extended abstract of this paper [17] appeared in the
Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Auckland, New Zealand. © 2022 Full paper copyright held by the owner/author(s).

2 PRELIMINARIES AND RELATED LITERATURE

2.1 Auctions and Equilibria

A Bayesian Game [21] is given by a tuple $G = (N, \mathcal{V}, \mathcal{A}, F, u)$. N players, indexed by $i = 1, \ldots, N$ participate in the game. $\mathcal{V} = \{\mathcal{V}_1, \ldots, \mathcal{V}_n\}$ is the set of possible *type profiles*, which describe private information available to the players when deciding their strategic behavior. These types are drawn from some prior probability distribution F that is assumed to be common knowledge. In the context of private-value auctions, we will refer to types as the *valuations* of the item(s) to be auctioned to the individual players. Given knowledge of these types, players i must then choose an action b_i from the set $\mathcal{A}_i$ of available actions. In auctions, these actions b_i are called *bids*. Finally, u is a vector of individual utility functions $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$ that describes the outcomes of the game. Crucially, for each player, these utilities depend only on their own type but on all players' chosen actions. In order to maximize her own utility, every player i, therefore, needs to decide on a strategy $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$ that will prescribe her action b_i for a given valuation input v_i[1]. A central solution concept in the study of Bayesian games is the *Bayesian Nash Equilibrium (BNE)*, which describes a strategy profile $\beta^\star = (\beta_1^\star, \ldots, \beta_N^\star)$ in which no agent can improve their expected utility by unilaterally deviating. We will write $\tilde{u}_i(\beta_i, \beta_{-i}) = \mathbb{E}_{v \sim F}[u_i(v_i, \beta_i(v_i), \beta_{-i}(v_{-i}))]$ for the (ex-ante) expected utility. Then, formally, $\beta^\star$ is an (ex-ante) BNE if for all players i and all possible strategies β_i it holds that $\tilde{u}_i(\beta_i^\star, \beta_{-i}^\star) \geq \tilde{u}_i(\beta_i, \beta_{-i}^\star)$.

In the following, we will concern ourselves with all-pay auctions in the independent private value model: Here, the players compete for a single indivisible good, and the private valuations $\{v_i\}_{i=1}^N$ are drawn independently from a continuous distribution function F. The uniform distribution is a common choice for F in many experiments. After drawing their valuations v_i, the bidders simultaneously submit their bids b_i and the one with the highest bid wins the prize. In contrast to winner-pay auctions, all bidders must pay their bid regardless of whether or not they won the good. This *payoff* leads to the following risk-neutral utility function (RNU):

$$u_i(v_i, b_i) = \begin{cases} v_i - b_i & \text{if } i \text{ wins, i.e. } b_i > \max\{b_{-i}\} \\ -b_i & \text{if } i \text{ loses} \end{cases}$$

Due to the symmetry of the bidder's valuations, the derivation of the *risk-neutral Bayes Nash equilibrium* (RNBNE) is straightforward in this setting. Assuming uniformly distributed valuations supported on the interval $[a, b] \subset \mathbb{R}^+$, its closed-form is given by the optimal bid function $b^\star(v_i) = \frac{N-1}{N(b-a)^{N-1}} v_i^N$ [29].

2.2 The Overbidding Puzzle

A common assumption in economical models is that in the long-term, market participants will follow the strategies prescribed by equilibria. However, in all-pay auctions, lab experiments have shown that real-world bidders do not follow this strategy but, in aggregate, submit higher bids than the RNBNE predicts [13]. From an individual perspective, they follow a pattern that Müller and Schotter [28] termed *bifurcation* - low-valued bidders tend to bid less than in the RNBNE, and high-valued bidders tend to overbid. This phenomenon has been dubbed the Overbidding Puzzle, and multiple explanations for this phenomenon have been proposed. Some arguments are based on a broader criticism of economic modeling, e.g. by suggesting that for several reasons, humans cannot be assumed to be rational utility-maximizing agents, or that markets generally can not be expected to reach equilibrium states due to hardness results in computational complexity [30]. Here, however, we will focus on explanation attempts *within the model* that suggest that psychological factors lead to human utility functions that differ from the RNU described above.

Although many authors investigated the deviation of the bidding behavior of lab subjects from the RNBNE due to psychological effects, it remains unclear what, if any, psychological factors are driving the overbidding puzzle. Moreover, there is empirical evidence for competing explanations. One of the observed phenomena is that bidders cannot estimate their winning probabilities correctly. For instance, Armantier and Treich [1] provided each bidder an independently drawn valuation and queried explicitly the bidders' estimation of their winning chance. They find that the subjects underestimate their probabilities leading to overbidding by playing their best responses given their biased estimations. Additionally, they conclude that competing explanations only play a minor role. In contrast to this finding, Goeree et al. [20] showed that this concept of biased probability estimations explains the behavior equally well as other phenomena, especially the idea of risk-aversion.

Risk Aversion. The theory of risk-averse bidders was one of the earliest to explain overbidding in auction mechanisms [9–11]. Its central idea is that subjects want to prevent losing the auction, and thus, bid more than the RNBNE predicts. Additionally, in the all-pay auction, they simultaneously prevent higher losses in low-valued settings by bidding too little. The usual approach to model this is by specifying a concave transformation of the RNU. Arrow [2] and Pratt [31] define the coefficients of absolute and relative risk measures that describe the risk attitude of a bidder through the curvature of the concave utility function, which builds the basis for deriving specific utility functions. While the absolute measure depends on the dimension of the analyzed unit, the relative one is dimensionless, and thus, we will focus on this measure to ensure the generalizability of our results.

The most used utility function in the experimental literature is the constant relative risk aversion (CRRA) [20]. It applies the concave transformation $U^C(u) = \frac{u^{1-\rho_i}}{1-\rho_i}$ to the quasilinear RNU, where $\rho_i \in [0, 1)$ depicts the risk attitude of bidder i. Another common family of utilities is the power functions that apply the transformation $U^P(u) = u_i^\rho$ to the RNU. Again, $\rho_i \in (0, 1]$ describes the risk attitude of bidder i. The special cases $\rho_i = 0$ in the CRRA model and $\rho_i = 1$ in the power model coincide with the risk-neutral setting. Although each bidder has an individual risk-attitude, we follow the experimental literature and assume that the parameter is equal to all bidders, and thus, $\rho_1 = \cdots = \rho_N$, to investigate the behavior of the average subject.

In the context of the all-pay auction, the analytical derivation of an optimal strategy results in an ordinary differential equation

[1] In this paper we will restrict ourselves to pure/deterministic strategies as common in the literature on auctions with continuous type and action spaces. However, in principle, randomizations or mixed strategies are possible.

that does not yield a unique solution assuming risk-averse bidders. Nevertheless, Fibich et al. [18] show that weakly risk-averse bidders follow the bifurcation pattern in equilibrium through a perturbation analysis. Hörisch and Kirchkamp [22] support this for a specific utility function. However, to the best of our knowledge there, does not exist a single empirical study that extensively tries to measure the risk attitude of lab subjects in the context of all-pay auctions. Additionally, the measurement of the risk-attitude of lab subjects still depicts a methodological challenge. Hence, even though these theoretical predictions coincide with the experimental data, the exact form of the optimal strategy is still unknown.

Regret. A separate stream of the literature adopts the idea of *post-auction regret* that bidders experience after observing auction outcomes [14, 16], which causes harm in the form of emotional suffering from having achieved suboptimal results in hindsight. One may now suggest that, when making a bidding decision, players anticipate and aim to avoid this emotional suffering[2]. In the context of the all-pay auction, Hyndman et al. [23] differentiate between the *all-pay loser regret* and the *general loser regret*. While the first type is generally relevant for all losers, and thus, always present, the latter type applies only to those bidders who would have an incentive to increase their bid because their valuation is greater than the winning bid b_w, which presupposes that the auctioneer announces the winning bid. The authors implement this context through the following utility function, where the coefficients α and γ determine the degree to which a bidder emotionally suffers from the all-pay regret, and respectively, the loser regret.

$$u_i^{Re}(v_i, b_i) = \begin{cases} v_i - b_i & \text{win, i.e. } b_i = b_w \\ -b_i - \alpha b_i - \gamma(v_i - b_w) & \text{lose and } b_w < v_i \\ -b_i - \alpha b_i & \text{lose and } v_i < b_w \end{cases}$$

Based on this utility function, Hyndman et al. calculate the analytical equilibrium strategy and conducted an experiment to empirically validate their theoretical findings. However, they were only able to provide evidence for the general loser-regret and the combination of both types, but not for the all-pay loser regret alone since it predicts that bidders bid, on average, less than in the RNBNE while the experiment showed overbidding. Furthermore, their study did not measure the regret coefficients. As those parameters are individual for each bidder, we follow the same strategy as in the risk settings and assume that all bidders share the same parameters to analyze the behavior of the average bidder.

2.3 Equilibrium Learning

The literature on learning Nash equilibria in Games goes back decades but has focussed primarily on finite complete-information games. Even in these settings, computing equilibria is known to be PPAD-complete [12]. For incomplete-information Bayesian games with continuous types and actions, i.e. infinite-dimensional strategy spaces, the exact worst-case complexity remains unknown but is likely significantly harder. Cai and Papadimitriou [8] analyze a specific setting for which the computation of BNE is PP-hard.

However, recent advances in numerical methods on auctions with continuous types and actions have shown many successes in approximating equilibria, which may suggest that the discouraging worst-case results may not be indicative of average-case difficulty, or that many Bayesian games belong to a special subclass of games, for which equilibrium computation is feasible. Bosshard et al. [6] compute high-fidelity approximations of pure-strategy equilibria in specific combinatorial auctions via smoothed best-response dynamics on locally linearized strategies. Bichler et al. [4] represent strategies via neural networks and learn pure-strategy BNE in a wide range of auctions via ex-ante gradient dynamics implemented via evolutionary strategy gradient approximation. Their method probably converges to *local BNE* in any symmetric auction, that is, a strategy profile where no player can improve her expected utility by making *small* deviations from her current strategy. In their empirical work, they indeed observe convergence to global equilibria, even in asymmetric auctions. Li and Wellman [26] also represent strategies via neural networks in symmetric games, and learn approximate equilibria by solving a meta game over mixtures of strategies in earlier iterations. However, the authors explicitly note that their method is inapplicable to all-pay auctions.

As described in section 3, our estimation framework to find optimal parameters will rely on computing equilibria of parametrized, symmetric all-pay auction games in its evaluation stage. To this end, we will rely on the method from Bichler et al. [4], called NPGA, unless a closed-form solution for the equilibria are known.

3 ESTIMATION FRAMEWORK

Given the fact that there are several competing notions of possible parametric utility functions that bidders might have to encode their psychological effects, we aim to determine which of these best fits the underlying data, assuming that bidders, on average, conform to their equilibrium strategy. However, even given fixed values for the parameters, determining these equilibrium strategies has not been possible, except in the cases where an analytical expression of the BNE strategies is known for the specific auction setting.

With recent advances in equilibrium learning described above, it has now become feasible to approximate BNE in all-pay auctions with arbitrary parametric utility functions with sufficient accuracy.

Building on these advances, we propose a Bayesian Optimization scheme to infer the unknown parameters of the utility functions.

Equilibrium Oracles. To that end, let $G^\theta = (N, \mathcal{V}, \mathcal{A}, F, u^\theta)$ be a Bayesian Game where the utility functions of individual players take some parametric form $u_i(v_i, b) \equiv u_\theta(v_i, b)$ —we restrict ourselves to symmetric utility functions with a shared parameter θ. Further, we assume the existence of an *equilibrium oracle* $EO : \theta \to \mathcal{A}^\mathcal{V}$ that, for a given parameter θ computes an estimate $\hat{\beta}$ of a BNE in G^θ. In our empirical analysis in sections 4 and 5, we will rely on NPGA [4] whenever a closed-form solution is unknown.

Behavioral loss function. The ultimate goal is determining the value of θ that maximizes the goodness of fit of the resulting equilibrium to some behavioral dataset $\mathcal{D} \equiv \{(v_k, b_k)\}_{k=1}^K$. Here, we define goodness-of-fit in terms of some regression loss function $\ell(b, \hat{b})$

[2] This notion of *regret* should not be confused with the abstract concept of the same name used in Multi-Agent Learning and Online Optimization to analyse algorithms, which describes the utility gain a player could have achieved, in hindsight, by changing her strategy. To avoid confusion, we will call the latter concept *utility loss* in this paper. Here, *regret*, refers to the phenomenon of players *anticipating* harm from the psychological effects of ex-post utility losses, *before* the game is played.

that compares the estimated equilibrium bids $\hat{b}$ prescribed under θ_t, to those observed in the behavioral data, b. We will consider two choices of loss functions in this study:

Since the trajectory of the bid function is quadratic under the assumption of risk-neutral and regretful bidders, it is reasonable to assume that the equilibria under the risk-averse utilities also have a quadratic shape in the context of the all-pay auction. Based on this fact, the first approach estimates a quadratic regression model using the experimental data and subsequently calculates the *root mean squared error* (RMSE) between the predictions of the resulting model and the estimated bid function of the NPGA algorithm.

As the proposed estimation framework does not depend on a single type of auction mechanism, the assumption of a specific form of the optimal bid function is not always possible or even feasible. For this purpose, the second approach predicts the corresponding bid $\hat{b}$ of the estimated bid function and compares those with the bids b in the data using a *coefficient of determination* $R^2 = 1 - \sum(b-\hat{b})^2 / \sum(b-\bar{b})^2$, where $\bar{b}$ indicates the average bid of all lab subjects. It quantifies to which degree the estimated bid function explains the variance of the data, and thus, the BO algorithm aims at finding the set of parameters that maximizes this metric.

Bayesian Optimization. Tasked with achieving this goal, the Bayesian Optimization scheme is comprised of two stages that are applied alternatingly at each time step t: In the *evaluation stage*, the goodness of fit of θ_t is evaluated via a call to the equilibrium oracle $\hat{\beta} = EO(\theta_t)$, computing the estimated equilibrium bids $\hat{b}_k = \hat{\beta}(v_k)$ that subjects should have bid in the experiment if they were following $\hat{\beta}$, and then evaluating the resulting loss $\ell(b, \hat{b})$.

In the *estimation stage*, the algorithm fits a stochastic model of the loss function over the entire domain Θ of the parameters to be inferred, relying on the history of all previously seen samples and their corresponding losses. It then uses this stochastical model to select a "promising" next sample θ_t. Ideally, the model specification must be sufficiently expressive, provide a measure of output uncertainty over its domain, be suitable for iterative refitting in the presence of new data points, and be inexpensive to evaluate. Keeping in line with the BO literature, a reasonable choice is given by a Gaussian Process model [5, 32]. Given this model, the next sample to be evaluated is chosen according to some acquisition criterion, that should strike a balance between exploitation (selecting a point that is likely close to the optimum based on past information) and exploration (selecting a point that will reduce model uncertainty in the next iteration). Here, we choose the *expected improvement* criterion, which is again a common choice in the BO literature [7, 32]. The interested reader is referred to [7] for a review of further choices in Bayesian Optimization. The complete pseudocode of this procedure is given in 1.

Algorithm 1 Bayesian Optimization scheme to infer parameters θ of a Game G^θ from behavioral data.

Input: Domain $\Theta\subseteq\mathbb{R}^d$, data $\mathcal{D} \equiv \{(v_k, b_k)\}_{k=1}^K$
Parameters: initial sample choice $\theta_0\in\Theta$, regression loss function ℓ, equilibrium oracle EO, stochastical model specification SM, sample acquisition function AF, maximal number of time steps T
Output: Vector of optimal parameters $\theta^\star \in \Theta$

```
$H := \emptyset$                          ▷ History of observed samples and losses.
$t := 0$.
for $t \in 0, 1, \ldots, T$ do
    // Evaluation stage
    $\hat{\beta} := EO(\theta_t)$                  ▷ Compute approx. BNE.
    for $k \in [K]$ do
        $\hat{b}_k := \hat{\beta}(v_k)$            ▷ Get bids prescribed by $\hat{\beta}$
    end for
    $\hat{\ell}_t := \ell(b, \hat{b})$                  ▷ Get loss of $\theta_t$.
    $H := H \cup \{(\theta_t, \hat{\ell}_t)\}$          ▷ Record sample, loss.
    // Estimation stage
    $SM = \text{fitSM}(H)$               ▷ Fit stochastic loss model.
    $\theta_{t+1} = AF(SM)$                 ▷ Acquire next sample.
end for
return best observed sample: $\theta^\star := \arg\min_H \hat{\ell}_t$
```

4 EXPERIMENT SETUP

Settings and Hyperparameters. To test the estimation framework empirically, we apply it to the three utility functions mentioned above on the example of the all-pay auction using real-world experimental data. Thus, each utility will be tested multiple times through the settings prescribed by the data sets mentioned in the next section. Additionally, we calculate the goodness of fits using the R^2 and the regression method to compare their estimations. It is noticeable that we apply a Tobit model for the regression evaluation because not all valuations have been observed during the experiments. Finally, we repeat these steps multiple times to ensure the statistical validity of our results. Overall, we conducted 90 experiments.

Since the regret parameters α and γ have no upper limits, it is reasonable to reduce the search space by restricting their domains, which improves the performance of the estimation process. Initial experiments showed that $\alpha, \gamma \in [0, 5]$ are promising choices for the parameter domains. Besides this, we stick to the standard scope for the risk parameters, such that $\rho \in [0, 1)$ in the CRRA function and $\rho \in (0, 1]$ in the power utility. To initialize the estimation of the parameters, we collect five random samples and conduct then twelve iterations in the BO algorithm. In each of these iterations, we query a parameter grid of the analyzed parameters that contains 2^9 reference values describing the entire domain of the variables through the acquisition function to determine the next promising parameters. In each iteration, we add a white noise term $\nu \sim N(0, \sigma^2_{\text{noise}})$ to the grid values to explore the entire search space, where we set $\sigma^2_{\text{noise}} = 0.005$ in the experiments.

We apply the NPGA method for approximating the bid functions in the evaluation stage of the framework, relying on the open-source library provided by Bichler et al. [4]. This requires the specification of hyperparameters; for an extensive description, see their paper: In each instance, we perform 500 iterations of supervised pre-training of the model to initialize the algorithm with approximately truthful bids, followed by 3500 NPGA iterations using a batch size of 2^{22}. Due to the symmetry of the analyzed settings, all players share a single neural network with two hidden layers á ten hidden nodes and SELU activation functions and a ReLU function at the output node of the model to learn their optimal strategies. To verify that the resulting bid functions are in equilibrium, we use a grid size of $n_{\text{grid}} = 2^{10}$

and batch size $n_{\text{batch}} = 2^{12}$ for calculating the relative utility loss to ensure the goodness of the approximated bid functions.

All parameters were chosen to best possible fit the available hardware, an Nvidia Geforce RTX 2080Ti GPU with 11GB of RAM. Under these conditions, a single NPGA evaluation stage instance took roughly ten minutes and an entire BO instance 170 minutes.

Data. We base the empirical analysis of the estimation framework on the three data sets described in Aycinena et al. [3] ("4-P FF") and Hyndman et al. [23] ("2-P FF", "2-P PP"), where FF and PF represent full-feedback and partial-feedback environments. Both data sets have been made available to us for this study by the respective authors. They investigate the behavior of human lab subjects in all-pay auctions with four and two agents per session, respectively. Since the equilibria depend on the number of players, we estimate the parameters of the utilities for each data set separately. As a data preprocessing step for both data sets, we remove bidders that submit bids, which are above their valuation, as these bids are strictly dominated under any utility function and indicate misunderstandings of the auction mechanism by the subjects. The analysis of the data sets revealed that most bidders commit errors in terms of bidding above their valuation in the first five periods. Hence, we consider those as a learning phase that is irrelevant for our analysis, in which bidders can commit wrong inputs and learn about their payoffs and winning probabilities, such that they will not be removed if they overbid their valuation only in this phase.

Aycinena et al. [3] analyze the behavior of subjects under three different value structures, where two of them assumed that the bidders share a common value for the prize, and one structure uses a private value setting. We restrict our evaluation of the estimation to the latter setting, which corresponds to the incomplete information all-pay auction: The valuations are composed additively of a common component v^c and private components v_i^p that are each drawn i.i.d. from $\mathcal{U}[0, 50]$: $v_i = v^c + v_i^p$ Thus, bidders observe (correlated) valuations in $[0, 100]$. The resulting restricted and preprocessed data set comprises 1260 observations describing the behavior of 36 subjects that participated in 35 sessions each, where the subjects were randomly matched into groups of four subjects.

We derive the other data sets from the experiments of Hyndman et al. [23], which consist of randomly matched sessions with independently drawn valuations on the same interval, $[0, 100]$, but without correlation between bidders. The sessions are split into those with *full feedback*, where the winning bid was announced after each round, and those with *partial feedback*, where bidders only observed whether they won the auction or not. Since the experimental literature found evidence that bidders behave differently in both settings [16, 23], we also separate these environments in our analysis. After preprocessing, the data sets contain 969 observations in the full-feedback setting and 1007 observations in the partial-feedback setting. It is important to note that we scaled the valuations and respective bids, wlog., to the intervals $[0, 1]$ in both data sets to enhance the empirical runtimes of our experiments.

5 RESULTS

Figure 1a shows that the distances between the best-estimated bid functions and the Tobit regression models are, on average, low. Hence, the estimated parameters describe the bidder's behavior in the data sets reasonably well. It is striking that the regret-based models perform worse in the two-player setting and better in the four-player setting than the risk-based models. This could imply that the idea of risk-averse bidders is a more suitable explanation in the two-player environment than in the four-player setting. However, it is necessary to further investigate this observation by applying the framework to more experiments in the future.

Table 1: Estimated risk measures under the CRRA utility for both evaluation techniques and all settings.

Eval	2P FF	2-P PF	4-P FF
Tobit	0.350	0.353	0.014
R^2	0.479	0.353	0.068

In addition to this, the plot indicates that the performances of the estimations are overall better in the two-player settings, which can be explained by the fact that we used the same set of hyperparameters in all runs. Due to the higher number of players, more iterations in the NPGA algorithm would enhance the performance in these settings. In fact, optimizing the sets of hyperparameters in all scenarios could improve the performances of all estimations further. However, this imposes the threat of overfitting, and thus, would limit the generalizability of the results. Comparing the performances within the settings shows that both risk models perform very similarly. This observation also holds for the regret models, although the differences are larger in these cases.

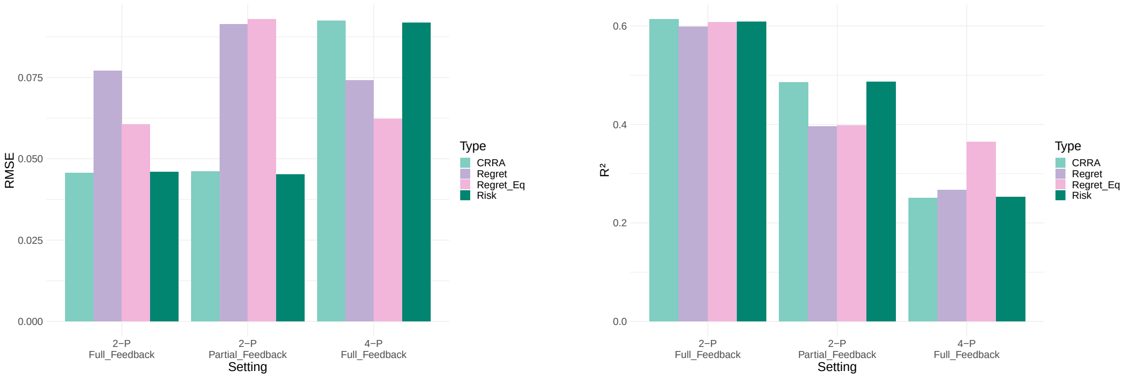
Besides the evaluation of the estimated bid functions via the comparison with the regression models, we find similar good performances in the evaluation via the R^2 metric, which is summarized in figure 1b. Again, the estimations achieve overall a better performance in the two-player settings. However, the approximated bid functions in the full-feedback environment yield greater R^2 values than in the partial-feedback scenario. Similar to the regression evaluation technique, the regret models outperform the risk models in the four-player setting, which supports the presumption that the concept of regretful-bidders explains the data in the four-player environment better than the idea of risk-averse bidders. Besides this, the plot indicates that the models perform very similarly for both psychological concepts.

Furthermore, the similarity of the regret models implies that the NPGA algorithm is a reasonable choice for an equilibrium learning algorithm in settings where no analytical solution is known.

Overall, the results of both evaluation methods show that our approach can generally estimate the parameters of utility functions reasonably well in settings where assumptions regarding the curvature of the optimal strategy are feasible and infeasible. This shows that our estimation framework builds a reasonable basis for comparing and measuring psychological assumptions regarding the utility functions of bidders. In the context of our experiments, it implies that the concept of risk-averse bidders is a better explanation for bifurcation than regretful bidders in the two-player setting.

5.1 Risk-Aversion

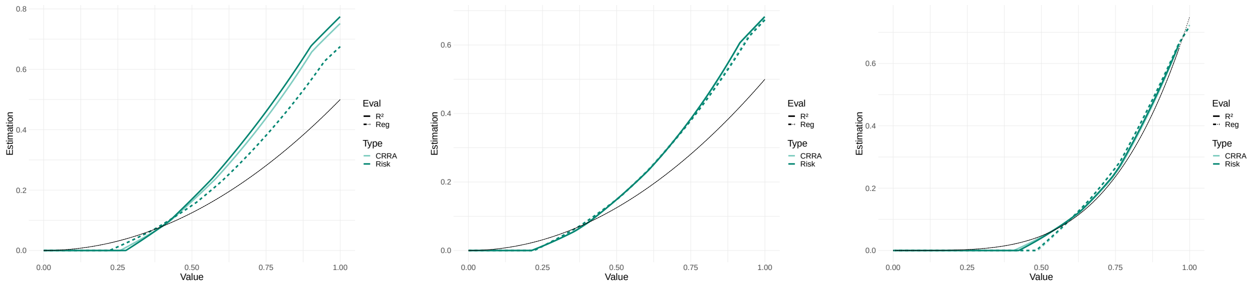
CRRA. Table 1 summarizes the estimated parameters of all settings per evaluation method for the CRRA utility. The values differ



(a) Results of the regression evaluation method

(b) Results of the R^2 evaluation method

Figure 1: Overview of the average model performances per used utility function and data setting for both evaluation techniques.

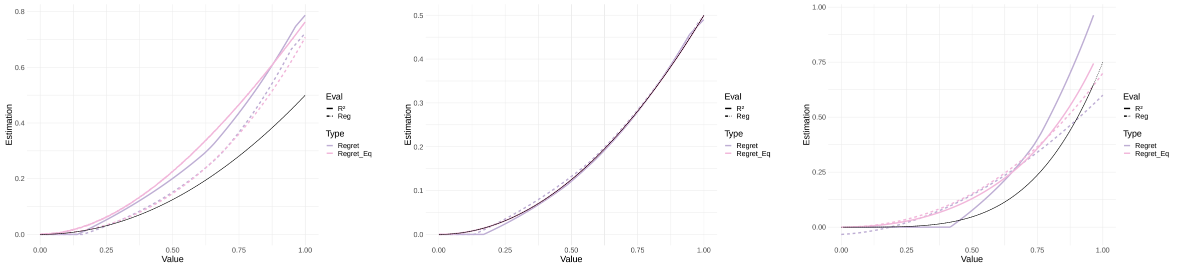


(a) Two-Player Full-Feedback

(b) Two-Player Partial-Feedback

(c) Four-Player Full-Feedback

Figure 2: Approximated bid functions for the risk utilities and the RNBNE in black. The dotted lines describes the result of the evaluation with the regression technique and the solid lines the evaluation with the R^2 metric.



(a) Two-Player Full-Feedback

(b) Two-Player Partial-Feedback

(c) Four-Player Full-Feedback

Figure 3: Approximated bid functions for the regret utility using NPGa and the closed-form of the equilibrium and the RNBNE in black. The dotted lines describes the result of the evaluation with the regression technique and the solid lines the evaluation with the R^2 metric.

between the evaluation methods in the sense that the models using the RMSE as its metric yield lower estimations than those based on the R^2 metric in the full-feedback environment. We followed the approach of Hyndman et al. [23] and did not apply any model selection or optimization technique for estimating the Tobit model. Consequently, these reference models could fail to summarize the variance of the corresponding data sets, which would explain the deviation between the evaluation techniques. Nonetheless, the estimations are all very similar within the settings.

The table also shows that the estimated values are in the regression approach almost identical in the two-player settings, which would imply that the type of feedback is generally irrelevant for

Table 2: Estimated risk measure under the power utility for both evaluation techniques and all settings.

Eval	2P FF	2-P PF	4-P FF
Tobit	0.650	0.653	0.995
R^2	0.511	0.647	0.926

shaping the risk-attitude of the bidders and hence also for explaining bifurcation. Conversely, the values differ with the other evaluation technique and imply that the announcement of the winning bid leads to more risk-averse bidders affecting their behavior, which results in their bimodal behavior. Since there exists empirical evidence in the literature for the latter observation [15, 23], it is reasonable to assume that the Tobit models require additional tuning effort to improve their explanatory power of the data. This shows that the estimation using the RMSE measure can only be as good as the estimation of the underlying regression model.

Power Utility. Table 2 summarizes the estimated parameters based on the power utility function. It is striking that the same observations hold as above. The estimations deviate between the Tobit and R^2 methods that the latter yields more risk-averse bidders in the two-player settings. Additionally, the values resulting from the regression technique imply that the type of feedback does not affect the bidder's risk-attitude crucially, while the other indicates an influence. Besides this, transforming the values in table 2 by $\tilde{r}^{\mathrm{P}} = 1 - r^{\mathrm{P}}$ shows that the estimations of both types of risk-averse utility functions result in very similar measures per setting and evaluation technique. Therefore, the approximated bid functions in figure 2 are very close to each other. The plots follow the analytical observations of Fibich et al. [18] by only deviating little from the RNBNE strategy and, combined with the estimated parameters, this indicates that the lab subjects were only weakly risk-averse.

The estimated parameters coincide with prior measurements in the literature since our results imply risk-averse bidders. For instance, Goeree et al. [20] conducted a two-player first-price auction with full feedback and observed a measure of $\hat{r} = 0.52$, which is similar to ours. Lu and Perrigne [27] found similar results for a two- ($\hat{r} = 0.5928$) and a three-player ($\hat{r} = 0.5994$) timber auction. Finally, Isaac et al. [24] estimated the risk parameter for auctions with a large but unknown number of bidders. They observed weakly risk-averse bidders ($\hat{r} = 0.375$) for experienced and risk-averse bidders ($\hat{r} = 0.455$) for new bidders in their analysis.

5.2 Post-Auction Regret

Table 3 provides an overview of the estimated regret coefficients α and γ for all settings and both evaluation techniques. It is striking that the majority of the estimated parameters is low, which implies that the lab subjects tend to be more risk-neutral than regretful in the given scenarios. The bid functions in figure 3 strenghten this view by indicating that the absolute distances between each of the curves and the RNBNE are low, and thus, there is no crucial deviation above or below the optimal risk-neutral strategy. This is especially true for the two-player partial-feedback environment, where the loser regret is almost zero in all estimations, and thus, the approximated bid function coincides with the risk-neutral case. Consequently, bidders act more risk-neutral in settings with less information about their environment, which corresponds to the observation under the assumption of risk-averse bidders.

Table 3: Estimated regret coefficients for both evaluation techniques using NPGA and the closed-form expression for determining the equilibrium strategy per setting.

		2-P FF		2-P PF		4-P FF	
Approx	Eval	α	γ	α	γ	α	γ
NPGA	Tobit	0.825	0.777	0.004		0.003	0.218
NPGA	R^2	0.412	1.24	0.004		0.493	1.47
Closed	Tobit	0.706	0.729	0.000		0.573	0.709
Closed	R^2	0.000	1.18	0.000		1.87	1.20

Besides this, it is noticeable that the estimated values deviate between the estimation models, although all of them achieved similar performance. Figure 3 supports this observation because the resulting graphs are very close to each other. One potential explanation is that multiple parameter constellations yield very similar bid functions, and thus, it is infeasible to determine the unique set that describes the bidders' behavior in the data best.

While the estimation of the risk-averse utilities implied that bidders act almost risk-neutral in the four-player setting, the plot and the table show the opposite for the regret scenarios. Since the performances of the latter models were better for both evaluation techniques, this observation implies that those models provide more sophisticated explanations for the behavior in this environment. Nonetheless, the estimations do not yield a unique parameter constellation, as mentioned before. Aside from that, the estimated bid functions in both figures 2a and 3a are also very similar, which corresponds to the similarity of performances of all models using the R^2 value as the loss function. This illustrates the complexity of the research area: Although our estimation framework enables to quantifying the performance of specific psychological concepts and their parametrization, given operationalizations of those concepts, it is necessary to develop more sophisticated loss functions in the future for improving the comparability of those concepts.

6 DISCUSSION AND CONCLUSION

In this study, we developed a regression framework based on BO and NPGA to infer unobserved parameters of Bayesian games and applied it to experiments in the context of symmetric all-pay auctions with incomplete information. Thereby, we implemented three utility functions that describe the idea of risk-aversion and regret to test reasonable psychological concepts for explaining overbidding in this type of auction. We showed that the approximated bid functions describe the variance of the data sufficiently, and thus, the estimated parameters depict a reasonable starting point for analyzing bidder's behavior in experimental auction settings. The resulting strategies coincide with the bifurcation pattern that is usually observed in the empirical literature, and thus, the analyzed utilities generally contribute to explaining the overbidding puzzle.

There exist differences in the explanatory power of the used utility functions, which is expressed by the loss function of our framework. The empirical analysis showed that the assumption of risk-averse lab subjects performs better than the concept of regretful bidders in the two-player environments. However, the differences are only marginal using both loss functions. Possible explanations are the lack of tuning the hyperparameters of the NPGA algorithm for each setting, which results in minor estimation errors, and the poorly estimated Tobit models that bias the overall estimation.

Our framework serves as a tool for measuring and comparing assumptions about the bidder's behavior that are not directly observable, as long as they can be expressed as utility functions. However, our experiments showed that it is necessary to develop more sophisticated loss functions for our framework in the future to clearly distinguish the performance of the analyzed behavioral concepts.

The derivation of psychological concepts and their operationalization, which could explain the overbidding puzzle is not trivial. However, our framework has the strength that it does not rely on information about a closed-form of the analyzed equilibrium because it applies NPGA to approximate it. Hence, future research should conduct further experiments to analyze more complex utilities comprising multiple behavioral concepts. Nonetheless, this is mainly limited by two reasons. First, bidders tend to behave differently between sessions, i.e. they behave very risk-averse in one session and almost neutral in another session [25]. This behavior makes a reasonable estimation of the individual risk-attitude or other concepts infeasible. Second, it is yet unclear which type of utility functions can be approximated by learning algorithms.

ACKNOWLEDGMENTS

If you wish to include any acknowledgments in your paper (e.g., to people or funding agencies), please do so using the 'acks' environment. Note that the text of your acknowledgments will be omitted if you compile your document with the 'anonymous' option.

REFERENCES

[1] Olivier Armantier and Nicolas Treich. 2009. Subjective probabilities in games: An application to the overbidding puzzle. *International Economic Review* 50, 4 (2009), 1079–1102.

[2] Kenneth Joseph Arrow. 1965. *Aspects of the theory of risk-bearing.* Helsinki: Yrjö Jahnsonian Sä iö.

[3] Diego Aycinena, Rimvydas Baltaduonis, and Lucas Rentschler. 2019. Valuation structure in incomplete information contests: experimental evidence. *Public Choice* 179, 3 (2019), 195–208.

[4] Martin Bichler, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring, and Paul Sutterer. 2021. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence* (2021), 1–9.

[5] Erik Bodin, Markus Kaiser, Ieva Kazlauskaite, Zhenwen Dai, Neill Campbell, and Carl Henrik Ek. 2020. Modulating Surrogates for Bayesian Optimization. In *International Conference on Machine Learning.* PMLR, 970–979.

[6] Vitor Bosshard, Benedikt Bünz, Benjamin Lubin, and Sven Seuken. 2020. Computing Bayes-Nash Equilibria in Combinatorial Auctions with Verification. *Journal of Artificial Intelligence Research* 69 (Oct. 2020), 531–570. https://doi.org/10.1613/jair.1.11525

[7] Eric Brochu, Vlad M Cora, and Nando De Freitas. 2010. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599* (2010).

[8] Yang Cai and Christos Papadimitriou. 2014. Simultaneous bayesian auctions and computational complexity. In *Proceedings of the fifteenth ACM conference on Economics and computation - EC '14*. ACM Press, Palo Alto, California, USA, 895–910. https://doi.org/10.1145/2600057.2602877 tex.ids: cai2014SimultaneousBayesianAuctionsa, cai2014SimultaneousBayesianAuctionsb, cai2014simultaneous.

[9] Kay-Yut Chen and Charles R Plott. 1998. Nonlinear behavior in sealed bid first price auctions. *Games and Economic Behavior* 25, 1 (1998), 34–78.

[10] James C. Cox, Vernon L. Smith, and James M. Walker. 1983. Tests of a heterogeneous bidders theory of first price auctions. *Economics Letters* 12, 3 (1983), 207–212. https://doi.org/10.1016/0165-1765(83)90039-3

[11] James C Cox, Vernon L Smith, and James M Walker. 1988. Theory and individual behavior of first-price auctions. *Journal of Risk and uncertainty* 1, 1 (1988), 61–99.

[12] C. Daskalakis, P. Goldberg, and C. Papadimitriou. 2009. The Complexity of Computing a Nash Equilibrium. *SIAM J. Comput.* 39, 1 (Jan. 2009), 195–259. https://doi.org/10.1137/070699652

[13] Emmanuel Dechenaux, Dan Kovenock, and Roman M Sheremeta. 2015. A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics* 18, 4 (2015), 609–669.

[14] Richard Engelbrecht-Wiggans. 1989. The effect of regret on optimal bidding in auctions. *Management Science* 35, 6 (1989), 685–692.

[15] Richard Engelbrecht-Wiggans and Elena Katok. 2007. Regret in auctions: Theory and evidence. *Economic Theory* 33, 1 (2007), 81–101.

[16] Richard Engelbrecht-Wiggans and Elena Katok. 2009. A direct test of risk aversion and regret in first price sealed-bid auctions. *Decision Analysis* 6, 2 (2009), 75–86.

[17] Markus Ewert, Stefan Heidekrüger, and Martin Bichler. 2022. Approaching the Overbidding Puzzle in All-Pay Auctions: Explaining Human Behavior through Bayesian Optimization and Equilibrium Learning. Extended Abstract.. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems* (Virtual Event, New Zealand) *(AAMAS '22)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1586–1588.

[18] Gadi Fibich, Arieh Gavious, and Aner Sela. 2006. All-pay auctions with risk-averse players. *International Journal of Game Theory* 34, 4 (2006), 583–599.

[19] Uri Gneezy and Rann Smorodinsky. 2006. All-pay auctions—an experimental study. *Journal of Economic Behavior & Organization* 61, 2 (2006), 255–275.

[20] Jacob K Goeree, Charles A Holt, and Thomas R Palfrey. 2002. Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory* 104, 1 (2002), 247–272.

[21] John C. Harsanyi. 1968. Games with Incomplete Information Played by "Bayesian" Players, I-III. Part II. Bayesian Equilibrium Points. *Management Science* 14, 5 (1968), 320–334. https://www.jstor.org/stable/2628673 Publisher: INFORMS.

[22] Hannah Hörisch and Oliver Kirchkamp. 2010. Less fighting than expected: Experiments with wars of attrition and all-pay auctions. *Public Choice* 144, 1/2 (2010), 347–367. http://www.jstor.org/stable/40661063

[23] Kyle Hyndman, Erkut Y Ozbay, and Pacharasut Sujarittanonta. 2012. Rent seeking with regretful agents: Theory and experiment. *Journal of Economic Behavior & Organization* 84, 3 (2012), 866–878.

[24] Mark Isaac, Svetlana Pevnitskaya, and Kurt S Schnier. 2012. Individual behavior and bidding heterogeneity in sealed bid auctions where the number of bidders is unknown. *Economic Inquiry* 50, 2 (2012), 516–533.

[25] John H Kagel. 2020. *7. Auctions: A Survey of Experimental Research.* Princeton University Press.

[26] Zun Li and Michael P Wellman. 2021. Evolution Strategies for Approximate Solution of Bayesian Games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5531–5540.

[27] Jingfeng Lu and Isabelle Perrigne. 2008. Estimating risk aversion from ascending and sealed-bid auctions: The case of timber auction data. *Journal of Applied Econometrics* 23, 7 (2008), 871–896.

[28] Wieland Müller and Andrew Schotter. 2010. Workaholics and dropouts in organizations. *Journal of the European Economic Association* 8, 4 (2010), 717–743.

[29] Charles Noussair and Jonathon Silver. 2006. Behavior in all-pay auctions with incomplete information. *Games and Economic Behavior* 55, 1 (2006), 189–206.

[30] Christos Papadimitriou and Georgios Piliouras. 2019. Game dynamics as the meaning of a game. *ACM SIGecom Exchanges* 16, 2 (May 2019), 53–63. https://doi.org/10.1145/3331041.3331048 tex.ids: papadimitriou2019GameDynamicsMeaning publisher: ACM New York, NY, USA.

[31] John W. Pratt. 1964. Risk Aversion in the Small and in the Large. *Econometrica* 32, 1/2 (1964), 122–136. http://www.jstor.org/stable/1913738

[32] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. 2012. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25 (2012).

3 Discussion and Conclusion

While the computation of Bayesian Nash equilibria is believed to be intractable in the general case, understanding the equilibrium outcomes of auction markets remains of utmost importance to economists and market design practitioners. In this dissertation, we studied an approach to computing such equilibria in auctions via an approximation of the ex-ante dynamics of the Bayesian game. The proposed algorithm, NPGA, approximates the Bayesian game as a complete information game on neural network parameters which map to functional strategies in the original Bayesian game. As auctions are ex-post non-differentiable, standard policy gradient methods like backpropagation cannot be applied. Therefore, to follow the gradient dynamics of this complete-information game, NPGA employs a non-standard gradient computation technique based on evolutionary strategies. The algorithm leverages massive parallelism to calculate these ex-ante gradients from ex-post data.

We proved two sufficient conditions for the convergence of NPGA to a close approximation of Bayesian Nash equilibria. These generalize existing results for differentiable complete-information games to ex-post non-differentiable Bayesian games: In Bayesian potential games, NPGA provably finds approximate local Nash equilibria (Publication B). In monotonic auction games, it provably converges to a close approximation of the unique (global) BNE (Heidekrüger et al., 2021c).

Beyond these theoretical results, a thorough empirical analysis of the method has been performed. In Publication A, the empirical behavior of NPGA in comparison to established best-response dynamics in normal-form games was analyzed. In Publication B, we studied symmetric environments in detail, where at least local convergence is guaranteed. In Publication C, we investigated asymmetric environments, where the established convergence guarantees no longer hold. We saw that NPGA converges to one of the (global) analytical BNE whenever the latter is known – even when the theoretical convergence criteria do not apply or cannot be numerically verified.

This is remarkable, as it seems to be in opposition to established hardness results (Cai and Papadimitriou, 2014; Daskalakis and Syrgkanis, 2016), and therefore it begs the question of how these positive results can be explained.

It may be possible that the kinds of sealed-bid auctions we study exhibit some structure that formally guarantees convergence, like monotonicity.[1] However, attempts to characterize when monotonicity holds beyond strategyproof settings in auction games have remained fruitless, even in simple two-player auctions. As such, understanding *why* ex-ante gradient dynamics work well in such a wide range of auctions remains an open problem. Further research may test the limits of global convergence in settings that are known to admit many socially undesirable local equilibria, like Blotto games or

[1] It should be noted that even then, NPGA is only known to converge when working with stylized *convex neural networks* that are unworkable in practice (Heidekrüger et al., 2021c; Bach, 2017). However, with general neural networks, the gap to guaranteed convergence is not game-theoretic in nature, but “only” consists of finding a global maximum in single-agent nonconvex optimization. While this is theoretically a computationally hard problem, this feat is routinely achieved in supervised machine learning with overparametrized neural networks, and we would expect to observe global convergence in our settings as well.

contests, or a continuous-type variant of the SiSPAs with XOS-bidding studied by Daskalakis and Syrgkanis (2016). In such games, NPGA as a local method may well get stuck in local equilibria, i.e. saddle points of the game where infinitesimal deviations of an agent do not yield utility improvements but step-changes may indeed do so.

Another obvious drawback of NPGA is the low sample efficiency of the ES-based gradient estimator in contrast to backpropagation. This hurts performance because, in each learning iteration, a large batch of auctions needs to be computed for the many perturbations of an agent's model, rather than just for the current model itself. A common effect visible in numerical equilibrium strategies of NPGA is that in low-dimensional settings, where explicit ex-interim methods (like those by Fichtl et al. (2022) or Bosshard et al. (2020)) are tractable, NPGA achieves somewhat lower accuracy than the latter. In particular, NPGA's learned strategies often differ somewhat in strategy space from the analytical solution, particularly in regions of the type space where an agent is unlikely to win the auction (compare Publication A). For such types, even the optimal bid achieves low expected utility compared to higher valuations. As a result, any subtle signal for optimal play in such regions is easily drowned out by aleatoric uncertainty in the ES-gradient estimates. This issue is likely to persist in higher dimensions where no comparisons are possible due to the lack of analytical results and the intractability of other learning methods. As such, improving the sample variance of NPGA would constitute a boon for scalability. One possible approach would be the reduction of aleatoric variance in utility and gradient estimation through the use of low-discrepancy (rather than pseudorandom) Monte-Carlo integration. However, while this approach may help, it is unlikely to be sufficient to fully alleviate the above effect (Belgacem, 2021). A separate approach informed by deep reinforcement learning may be more promising: The deterministic policy gradient theorem was originally proposed in the context of actor-critic models (Lillicrap et al., 2015)– which learn a (differentiable) Q-function of the environment and use it as a surrogate objective. In the context of auctions, one may introduce a target network $Q(v_i, b_i) \approx \overline{u}_i(v_i; b_i; \beta_{-i})$ and train it from ex-post samples of auction outcomes. Replacing the ex-post samples with observations from this Q-function would enable the use of standard backpropagation which may promise a significant improvement of the algorithm in sample efficiency, runtime, and memory consumption, provided that the Q network can be adequately adjusted to changing opponent behavior. Such an approach akin to Multi-Agent Deep Deterministic Policy Gradient (MA-DDPG) (Lowe et al., 2017) may be a promising future research direction. However, we want to stress that despite the drawbacks in efficiency, NPGA's ES-gradient computation has the advantage that it can be shown to be an unbiased estimator of the ex-ante gradient dynamics. Such statements are typically not possible to make in actor-critic methods. However, it is crucial for the theoretical analysis of NPGA, which is paramount because numerically computed results will not be trusted by the economics community without a solid theoretical underpinning of the computation method.

By design, the studies in this dissertation are limited to *sealed-bid* auctions in which every bidder only makes a single bidding decision and there is no time component. However, sequential decisions may be necessary for many real-world auction formats: On the one hand, auction formats themselves may be sequential in nature, such as in English, Dutch, or Japanese single-item auctions (Duetting et al., 2019), or Simultaneous Multi-Round combinatorial auctions (Bichler et al., 2014)). On the other hand, in many markets, it is common practice to hold multiple auctions sequentially to sell inventory (Jeitschko, 1998; Guerci et al., 2014). A common example is the market for computational display or search engine advertisement inventory, where millions of ad slots are sold in a virtual stream of auctions as new inventory arrives (Stange and Funk, 2014). In such settings, bidders are no longer interested in optimal behavior in a single auction, but rather aim to optimize their performance over a campaign, i.e. a sequence of auctions. As a result, following the BNE strategies of each single auction instance may no longer be optimal. Such sequential settings are beyond the reach of this dissertation. Modeling them as something akin to a Markov Game and studying multi-agent learning in such settings presents an exciting future research direction with close ties to MARL.

Beyond the further development of equilibrium learning methods themselves, the presented methods can already unlock novel lines of research in auction theory. As demonstrated in the included publications, NPGA allows the computation of equilibria in a wide range of settings where this was not previously possible. NPGA is applicable to any market environment that can be efficiently implemented in a simulator and requires no market-specific adjustments (although hyperparameter tuning may lead to a further increase in performance). For example, we have shown its ability to compute equilibria in all-pay auctions with arbitrary non-quasi-linear utility functions in Publication D, which has enabled performing statistical inference on parameters of these utility functions themselves. Moreover, NPGA has enabled the analysis of comparable statics in moderately-sized auctions with value interdependencies (Publication B, Publication C), and the computation in larger markets than ever before (Publication D). While the computational study of larger combinatorial auctions remains elusive, we hope that further development in numerical equilibrium computation may make such investigations approachable in the future.

Bibliography

O. Armantier, J.-P. Florens, and J.-F. Richard. Approximation of Nash equilibria in Bayesian games. *Journal of Applied Econometrics*, 23(7):965–981, Nov. 2008. ISSN 08837252, 10991255. doi: 10.1002/jae.1040. URL `http://doi.wiley.com/10.1002/jae.1040`.

S. Arora, E. Hazan, and S. Kale. The Multiplicative Weights Update Method: a Meta-Algorithm and Applications. *Theory of Computing*, 8:121–164, May 2012. doi: 10.4086/toc.2012.v008a006. URL `https://theoryofcomputing.org/articles/v008a006/`. Number: 6 Publisher: Theory of Computing.

I. Ashlagi, D. Monderer, and M. Tennenholtz. Simultaneous ad auctions. *Mathematics of Operations Research*, 36(1):1–13, 2011.

R. J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96, Mar. 1974. ISSN 0304-4068. doi: 10.1016/0304-4068(74)90037-8. URL `https://www.sciencedirect.com/science/article/pii/0304406874900378`.

L. M. Ausubel and O. Baranov. Core-selecting auctions with incomplete information. *International Journal of Game Theory*, July 2019. ISSN 1432-1270. doi: 10.1007/s00182-019-00691-3. URL `https://doi.org/10.1007/s00182-019-00691-3`.

L. M. Ausubel, P. Milgrom, et al. The lovely but lonely Vickrey auction. *Combinatorial auctions*, 17: 22–26, 2006.

F. Bach. Breaking the Curse of Dimensionality with Convex Neural Networks. *Journal of Machine Learning Research*, 18(19):1–53, 2017. ISSN 1533-7928. URL `http://jmlr.org/papers/v18/14-546.html`.

N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, I. Dunning, S. Mourad, H. Larochelle, M. G. Bellemare, and M. Bowling. The Hanabi Challenge: A New Frontier for AI Research. *Artificial Intelligence*, 280:103216, Mar. 2020. ISSN 00043702. doi: 10.1016/j.artint.2019.103216. URL `http://arxiv.org/abs/1902.00506`. arXiv:1902.00506 [cs, stat].

I. Belgacem. Improving Sample Efficiency in Multiagent Equilibrium learning settings via Advanced Monte Carlo Methods. *Student Project (unpublished)*, page 7, 2021.

M. Benaïm, J. Hofbauer, and S. Sorin. Perturbations of Set-Valued Dynamical Systems, with Applications to Game Theory. *Dynamic Games and Applications*, 2(2):195–205, June 2012. ISSN 2153-0793. doi: 10.1007/s13235-012-0040-0. URL `https://doi.org/10.1007/s13235-012-0040-0`.

M. Bichler. Market Design by Martin Bichler, Dec. 2017. URL /core/books/market-design/A946947368CC94047DFA0B4DEF236FEC.

M. Bichler and J. K. Goeree. *Handbook of Spectrum Auction Design*. Cambridge University Press, 2017.

M. Bichler, J. Goeree, S. Mayer, and P. Shabalin. Spectrum auction design: Simple auctions for complex sales. *Telecommunications Policy*, 38(7):613–622, 2014.

M. Bichler, M. Fichtl, S. Heidekrüger, N. Kohring, and P. Sutterer. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence*, 3(8):687–695, Aug. 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00365-4. URL https://www.nature.com/articles/s42256-021-00365-4.

M. Bichler, P. Milgrom, and G. Schwarz. Taming the Communication and Computation Complexity of Combinatorial Auctions: The FUEL Bid Language. *Management Science*, June 2022. ISSN 0025-1909. doi: 10.1287/mnsc.2022.4465. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2022.4465. Publisher: INFORMS.

M. Bichler, N. Kohring, and S. Heidekrüger. Learning equilibria in asymmetric auction games. *INFORMS Journal on Computing*, 35(3):523–542, May 2023a. doi: 10.1287/ijoc.2023.1281. URL https://doi.org/10.1287/ijoc.2023.1281.

M. Bichler, N. Kohring, and S. Heidekrüger. Online supplement to "learning equilibria in asymmetric auction games", Version v2021.0151. 2023b. URL https://github.com/INFORMSJoC/2021.0151.

A. Blum and Y. Mansour. Learning, Regret Minimization, and Equilibria. In N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, pages 79–102. Cambridge University Press, Cambridge, 2007. ISBN 978-0-511-80048-1. doi: 10.1017/CBO9780511800481.006. URL https://www.cambridge.org/core/product/identifier/CBO9780511800481A051/type/book_part.

A. L. Blum and R. L. Rivest. Training a 3-node neural network is NP-complete. *Neural Networks*, 5(1):117–127, Jan. 1992. ISSN 0893-6080. doi: 10.1016/S0893-6080(05)80010-3. URL https://www.sciencedirect.com/science/article/pii/S0893608005800103.

V. Bosshard, B. Bünz, B. Lubin, and S. Seuken. Computing Bayes-Nash Equilibria in Combinatorial Auctions with Continuous Value and Action Spaces. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 119–127, Melbourne, Australia, Aug. 2017. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-0-3. doi: 10.24963/ijcai.2017/18. URL https://www.ijcai.org/proceedings/2017/18.

V. Bosshard, B. Bünz, B. Lubin, and S. Seuken. Computing Bayes-Nash Equilibria in Combinatorial Auctions with Verification. *Journal of Artificial Inelligence Research*, 2020. URL http://arxiv.org/abs/1812.01955.

S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK ; New York, 2004. ISBN 978-0-521-83378-3.

G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. OpenAI Gym. *arXiv:1606.01540 [cs]*, June 2016. URL `http://arxiv.org/abs/1606.01540`.

G. W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

N. Brown and T. Sandholm. Solving Imperfect-Information Games via Discounted Regret Minimization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):1829–1836, July 2019a. ISSN 2374-3468. doi: 10.1609/aaai.v33i01.33011829. URL `https://ojs.aaai.org/index.php/AAAI/article/view/4007`. Number: 01.

N. Brown and T. Sandholm. Superhuman AI for multiplayer poker. *Science*, 365(6456):885–890, Aug. 2019b. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aay2400. URL `https://science.sciencemag.org/content/365/6456/885`.

Y. Cai and C. Papadimitriou. Simultaneous bayesian auctions and computational complexity. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation - EC '14*, pages 895–910, Palo Alto, California, USA, 2014. ACM Press. ISBN 978-1-4503-2565-3. doi: 10.1145/2600057.2602877. URL `http://dl.acm.org/citation.cfm?doid=2600057.2602877`.

V. Conitzer and T. Sandholm. New complexity results about Nash equilibria. *Games and Economic Behavior*, 63(2):621–641, July 2008. ISSN 0899-8256. doi: 10.1016/j.geb.2008.02.015. URL `http://www.sciencedirect.com/science/article/pii/S0899825608000936`.

A.-A. Cournot. *Recherches sur les principes mathématiques de la théorie des richesses*. 1838. URL `https://gallica.bnf.fr/ark:/12148/bpt6k6117257c`.

C. Daskalakis and V. Syrgkanis. Learning in Auctions: Regret is Hard, Envy is Easy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 219–228, Oct. 2016. doi: 10.1109/FOCS.2016.31.

C. Daskalakis, P. Goldberg, and C. Papadimitriou. The Complexity of Computing a Nash Equilibrium. *SIAM Journal on Computing*, 39(1):195–259, Jan. 2009. ISSN 0097-5397. doi: 10.1137/070699652. URL `https://epubs.siam.org/doi/abs/10.1137/070699652`.

R. Day and P. Milgrom. Core-selecting package auctions. *International Journal of Game Theory*, 36(3): 393–407, Mar. 2008. ISSN 1432-1270. doi: 10.1007/s00182-007-0100-7. URL `https://doi.org/10.1007/s00182-007-0100-7`.

R. W. Day and P. Cramton. Quadratic Core-Selecting Payment Rules for Combinatorial Auctions. *Operations Research*, 60(3):588–603, June 2012. ISSN 0030-364X. doi: 10.1287/opre.1110.1024. URL `https://pubsonline.informs.org/doi/abs/10.1287/opre.1110.1024`.

P. Duetting, Z. Feng, H. Narasimhan, D. Parkes, and S. S. Ravindranath. Optimal Auctions through Deep Learning. In *Proceedings of the 36th International Conference on Machine Learning*, pages 1706–1715. PMLR, May 2019. URL `https://proceedings.mlr.press/v97/duetting19a.html`. ISSN: 2640-3498.

M. Ewert, S. Heidekrüger, and M. Bichler. Approaching the Overbidding Puzzle in All-Pay Auctions: Explaining Human Behavior through Bayesian Optimization and Equilibrium Learning. *Author's Manuscript*, 2022a. Extended Abstract published in AAMAS '22, (See Ewert et al., 2022b).

M. Ewert, S. Heidekrüger, and M. Bichler. 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*, AAMAS '22, page 1586–1588, Richland, SC, 2022b. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450392136.

Farama Foundation. GitHub - Farama-Foundation/Gymnasium: A standard API for reinforcement learning and a diverse set of reference environments (formerly Gym), 2021. URL `https://github.com/Farama-Foundation/Gymnasium`.

M. Fichtl, M. Oberlechner, and M. Bichler. Computing Distributional Bayes Nash Equilibria in Auction Games via Gradient Dynamics. *AAAI-22 Workshop on Reinforcement Learning in Games*, 2022.

J. N. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. Learning with Opponent-Learning Awareness. *arXiv:1709.04326 [cs]*, Sept. 2017. URL `http://arxiv.org/abs/1709.04326`.

D. P. Foster and R. V. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21(1-2):40, 1997.

D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*, volume 2 of *MIT Press Series on Economic Learning and Social Evolution*. MIT Press, Cambridge, 2. edition, 1999. ISBN 978-0-262-06194-0.

D. Fudenberg and D. K. Levine. Learning and Equilibrium. *Annual Review of Economics*, 2009. ISSN 1941-1383. doi: 10.1146/annurev.economics.050708.142930. URL `https://dash.harvard.edu/handle/1/4382413`.

J. K. Goeree and Y. Lien. On the impossibility of core-selecting auctions. *Theoretical Economics*, 11(1): 41–52, 2016. ISSN 1555-7561. doi: 10.3982/TE1198. URL `https://onlinelibrary.wiley.com/doi/abs/10.3982/TE1198`.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

E. Guerci, A. Kirman, and S. Moulet. Learning to bid in sequential Dutch auctions. *Journal of Economic Dynamics & Control*, 48:20, 2014.

J. Hannan. Approximation to Bayes Risk in Repeated Play. In *Contributions to the Theory of Games (AM-39), Volume III*, pages 97–140. Princeton University Press, 1958. ISBN 978-1-4008-8215-1. doi: 10.1515/9781400882151-006. URL `http://www.degruyter.com/document/doi/10.1515/9781400882151-006/html?lang=en`.

J. C. Harsanyi. Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points. *Management Science*, 14(5):320–334, 1968.

J. Hartline, V. Syrgkanis, and E. Tardos. No-Regret Learning in Bayesian Games. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3061–3069. Curran Associates, Inc., 2015. URL `http://papers.nips.cc/paper/6016-no-regret-learning-in-bayesian-games.pdf`.

S. Heidekrüger. Equilibrium Learning in Auction Markets. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11):12882–12883, June 2022. ISSN 2374-3468. doi: 10.1609/aaai.v36i11.21578. URL `https://doi.org/10.1609/aaai.v36i11.21578`. Number: 11.

S. Heidekrüger, P. Sutterer, and M. Bichler. Computing Approximate Bayes-Nash Equilibria through Neural Self-Play. In *Workshop on Information Technology and Systems (WITS-19)*, Munich, Germany, Dec. 2019.

S. Heidekrüger, N. Kohring, and M. Bichler. Bnelearn: A Framework for Equilibrium Learning in Sealed-Bid Auctions. *Working Paper*, 2021a.

S. Heidekrüger, N. Kohring, P. Sutterer, and M. Bichler. Multiagent Learning for Equilibrium Computation in Auction Markets. *AAAI Spring Symposium on Challenges and Opportunities in Multiagent Reinforcement Learning (COMARL-21)*, Mar. 2021b. URL `https://sites.google.com/view/comarl-aaai-2021/accepted-papers`.

S. Heidekrüger, P. Sutterer, N. Kohring, M. Fichtl, and M. Bichler. Equilibrium Learning in Combinatorial Auctions: Computing Approximate Bayesian Nash Equilibria via Pseudogradient Dynamics. *AAAI Workshop on Reinforcement Learning in Games (AAAI-RLG-21), arXiv:2101.11946 [cs]*, Feb. 2021c. URL `http://arxiv.org/abs/2101.11946`.

J. Heinrich and D. Silver. Deep Reinforcement Learning from Self-Play in Imperfect-Information Games. *arXiv:1603.01121 [cs]*, Mar. 2016. URL `http://arxiv.org/abs/1603.01121`.

J. Heinrich, M. Lanctot, and D. Silver. Fictitious self-play in extensive-form games. In *International Conference on Machine Learning*, pages 805–813, 2015.

D. Hennes, D. Morrill, S. Omidshafiei, R. Munos, J. Perolat, M. Lanctot, A. Gruslys, J.-B. Lespiau, P. Parmas, E. Duenez-Guzman, and K. Tuyls. Neural Replicator Dynamics. *arXiv:1906.00190 [cs, stat]*, Feb. 2020. URL `http://arxiv.org/abs/1906.00190`.

T. D. Jeitschko. Learning in Sequential Auctions. *Southern Economic Journal*, 65(1):98–112, 1998. ISSN 0038-4038. doi: 10.2307/1061354. URL `https://www.jstor.org/stable/1061354`.

S. Kakutani. A generalization of brouwer's fixed point theorem. *Duke mathematical journal*, 8(3): 457–459, 1941.

T. R. Kaplan and S. Zamir. Asymmetric first-price auctions with uniform distributions: analytic solutions to the general case. *Economic Theory*, 50(2):269–302, 2012. ISSN 0938-2259. URL `https://www.jstor.org/stable/41486012`. Publisher: Springer.

T. R. Kaplan and S. Zamir. Multiple equilibria in asymmetric first-price auctions. *Economic Theory Bulletin*, 3(1):65–77, 2015.

N. Kohring, C. Frohlich, S. Heidekrüger, and M. Bichler. Equilibrium Computation for Auction Games via Multi-Swarm Optimization. *AAAI Workshop on Reinforcement Learning in Games (AAAI-RLG-22)*, page 9.

V. Krishna. *Auction Theory*. Academic press, 2009.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL `https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html`.

M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, k. Tuyls, J. Perolat, D. Silver, and T. Graepel. A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4190–4203. Curran Associates, Inc., 2017. URL `http://papers.nips.cc/paper/7007-a-unified-game-theoretic-approach-to-multiagent-reinforcement-learning.pdf`.

M. Lanctot, E. Lockhart, J.-B. Lespiau, V. Zambaldi, S. Upadhyay, J. Pérolat, S. Srinivasan, F. Timbers, K. Tuyls, S. Omidshafiei, D. Hennes, D. Morrill, P. Muller, T. Ewalds, R. Faulkner, J. Kramár, B. De Vylder, B. Saeta, J. Bradbury, D. Ding, S. Borgeaud, M. Lai, J. Schrittwieser, T. Anthony, E. Hughes, I. Danihelka, and J. Ryan-Davis. OpenSpiel: A Framework for Reinforcement Learning in Games. *arXiv:1908.09453 [cs]*, Aug. 2019. URL `http://arxiv.org/abs/1908.09453`.

A. Letcher, D. Balduzzi, S. Racaniere, J. Martens, J. Foerster, K. Tuyls, and T. Graepel. Differentiable Game Mechanics. *The Journal of Machine Learning Research*, 20(1):3032–3071, 2019.

Z. Li and M. P. Wellman. Evolution Strategies for Approximate Solution of Bayesian Games. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(6):5531–5540, May 2021. ISSN 2374-3468. URL `https://ojs.aaai.org/index.php/AAAI/article/view/16696`.

E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, J. Gonzalez, K. Goldberg, and I. Stoica. Ray RLlib: A Composable and Scalable Reinforcement Learning Library. page 18, 2017. URL `https://royf.org/pub/pdf/Liang2017Ray.pdf`.

T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *arXiv:1509.02971 [cs, stat]*, Sept. 2015. URL `http://arxiv.org/abs/1509.02971`.

R. Lowe, Y. WU, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6379–6390. Curran Associates, Inc., 2017. URL `http://papers.nips.cc/paper/7217-multi-agent-actor-critic-for-mixed-cooperative-competitive-environments.pdf`.

E. Mazumdar, L. J. Ratliff, and S. S. Sastry. On Gradient-Based Learning in Continuous Games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, Jan. 2020. doi: 10.1137/18M1231298. URL `https://epubs.siam.org/doi/abs/10.1137/18M1231298`.

R. D. McKelvey, A. M. McLennan, and T. L. Turocy. Gambit: Software Tools for Game Theory, 2016. URL `http://www.gambit-project.org/`.

B. McMahan. Follow-the-Regularized-Leader and Mirror Descent: Equivalence Theorems and L1 Regularization. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 525–533. JMLR Workshop and Conference Proceedings, June 2011. URL `https://proceedings.mlr.press/v15/mcmahan11b.html`. ISSN: 1938-7228.

P. Mertikopoulos and Z. Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1):465–507, Jan. 2019. ISSN 1436-4646. doi: 10.1007/s10107-018-1254-8. URL `https://doi.org/10.1007/s10107-018-1254-8`.

P. Milgrom. Auction Research Evolving: Theorems and Market Designs. *American Economic Review*, 111(5):1383–1405, May 2021. ISSN 0002-8282. doi: 10.1257/aer.111.5.1383. URL `https://pubs.aeaweb.org/doi/10.1257/aer.111.5.1383`.

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb. 2015. ISSN 1476-4687. doi: 10.1038/nature14236. URL `https://www.nature.com/articles/nature14236`. Number: 7540 Publisher: Nature Publishing Group.

D. Monderer and L. S. Shapley. Potential Games. *Games and Economic Behavior*, 14(1):124–143, May 1996. ISSN 0899-8256. doi: 10.1006/game.1996.0044. URL `http://www.sciencedirect.com/science/article/pii/S0899825696900445`.

R. B. Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

J. F. Nash. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust Stochastic Approximation Approach to Stochastic Programming. *SIAM Journal on Optimization*, 19(4):1574–1609, Jan. 2009. ISSN 1052-6234. doi: 10.1137/070704277. URL `https://epubs.siam.org/doi/abs/10.1137/070704277`. Publisher: Society for Industrial and Applied Mathematics.

Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical programming*, 120 (1):221–259, 2009.

A. Neyman. Correlated equilibrium and potential games. *International Journal of Game Theory*, 26(2):223–227, June 1997. ISSN 1432-1270. doi: 10.1007/BF01295851. URL `https://doi.org/10.1007/BF01295851`.

N. Nisan. Bidding and allocation in combinatorial auctions. In *Proceedings of the 2nd ACM conference on Electronic commerce*, EC '00, pages 1–12, New York, NY, USA, Oct. 2000. Association for Computing

Machinery. ISBN 978-1-58113-272-4. doi: 10.1145/352871.352872. URL https://doi.org/10.1145/352871.352872.

N. Nisan, editor. *Algorithmic Game Theory*. Cambridge University Press, Cambridge ; New York, 2007. ISBN 978-0-521-87282-9.

S. Omidshafiei, C. Papadimitriou, G. Piliouras, K. Tuyls, M. Rowland, J.-B. Lespiau, W. M. Czarnecki, M. Lanctot, J. Perolat, and R. Munos. α- Rank: Multi-Agent Evaluation by Evolution. *Scientific Reports*, 9(1):1–29, July 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-45619-9. URL https://www.nature.com/articles/s41598-019-45619-9.

C. Papadimitriou and G. Piliouras. Game dynamics as the meaning of a game. *ACM SIGecom Exchanges*, 16(2):53–63, May 2019. ISSN 15519031. doi: 10.1145/3331041.3331048. URL http://dl.acm.org/citation.cfm?doid=3331041.3331048.

A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in PyTorch. In *NIPS-W*, 2017.

M. Plum. Characterization and computation of Nash-equilibria for auctions with incomplete information. *International Journal of Game Theory*, 20(4):393–418, 1992.

D. Porter, S. Rassenti, A. Roopnarine, and V. Smith. Combinatorial auction design. *Proceedings of the National Academy of Sciences*, 100(19):11153–11157, Sept. 2003. doi: 10.1073/pnas.1633736100. URL https://www.pnas.org/doi/10.1073/pnas.1633736100. Publisher: Proceedings of the National Academy of Sciences.

J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.

M. H. Rothkopf. Thirteen reasons why the Vickrey-Clarke-Groves process is not practical. *Operations Research*, 55(2):191–197, 2007.

T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *arXiv:1703.03864 [cs, stat]*, Mar. 2017. URL http://arxiv.org/abs/1703.03864.

Y. Sato, E. Akiyama, and J. D. Farmer. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences*, 99(7):4748–4751, Apr. 2002. doi: 10.1073/pnas.032086299. URL https://www.pnas.org/doi/10.1073/pnas.032086299. Publisher: Proceedings of the National Academy of Sciences.

R. Savani and B. von Stengel. Game Theory Explorer - Software for the Applied Game Theorist. *Computational Management Science*, 12(1):5–33, Jan. 2015. ISSN 1619-697X, 1619-6988. doi: 10.1007/s10287-014-0206-x. URL http://arxiv.org/abs/1403.3969. arXiv:1403.3969 [cs].

J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. *Nature*, 588(7839):604–609, Dec. 2020. ISSN 0028-0836, 1476-4687.

doi: 10.1038/s41586-020-03051-4. URL http://arxiv.org/abs/1911.08265. arXiv:1911.08265 [cs, stat].

J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz. Trust Region Policy Optimization. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1889–1897. PMLR, June 2015. URL https://proceedings.mlr.press/v37/schulman15.html. ISSN: 1938-7228.

J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal Policy Optimization Algorithms, Aug. 2017. URL http://arxiv.org/abs/1707.06347. Number: arXiv:1707.06347 arXiv:1707.06347 [cs].

J. S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3):312–327, Mar. 2005. ISSN 0018-9286. doi: 10.1109/TAC.2005.843878.

Y. Shoham. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller. Deterministic Policy Gradient Algorithms. In *International Conference on Machine Learning*, pages 387–395, Jan. 2014. URL http://proceedings.mlr.press/v32/silver14.html.

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, Jan. 2016. ISSN 1476-4687. doi: 10.1038/nature16961. URL https://www.nature.com/articles/nature16961. Number: 7587 Publisher: Nature Publishing Group.

D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, Oct. 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature24270. URL http://www.nature.com/doifinder/10.1038/nature24270.

S. Singh, M. Kearns, and Y. Mansour. Nash Convergence of Gradient Dynamics in Iterated General-Sum Games. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)*, June 2000. URL http://arxiv.org/abs/1301.3892.

J. M. Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

M. Stange and B. Funk. Real-Time-Advertising. *Wirtschaftsinformatik*, 56(5):335–338, Oct. 2014. ISSN 1861-8936. doi: 10.1007/s11576-014-0435-1. URL https://doi.org/10.1007/s11576-014-0435-1.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge, Massachusetts, second edition edition, 2018. ISBN 978-0-262-03924-6.

J. K. Terry, B. Black, A. Hari, L. Santos, C. Dieffendahl, N. L. Williams, Y. Lokesh, C. Horsch, and P. Ravi. PettingZoo: Gym for Multi-Agent Reinforcement Learning. *arXiv:2009.14471 [cs, stat]*, Sept. 2020. URL `http://arxiv.org/abs/2009.14471`.

The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel. Improvements to auction theory and inventions of new auction formats. *Scientifc Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2020*, page 41, Oct. 2020.

K. Tuyls, J. Perolat, M. Lanctot, J. Z. Leibo, and T. Graepel. A Generalised Method for Empirical Game Theoretic Analysis. *arXiv:1803.06376 [cs]*, Mar. 2018. URL `http://arxiv.org/abs/1803.06376`.

T. Ui. Correlated equilibrium and concave games. *International Journal of Game Theory*, 37(1):1–13, Apr. 2008. ISSN 1432-1270. doi: 10.1007/s00182-007-0098-x. URL `https://doi.org/10.1007/s00182-007-0098-x`.

T. Ui. Bayesian Nash equilibrium and variational inequalities. *Journal of Mathematical Economics*, 63:139–146, Mar. 2016. ISSN 0304-4068. doi: 10.1016/j.jmateco.2016.02.004. URL `http://www.sciencedirect.com/science/article/pii/S0304406816000124`.

J. v. Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, Dec. 1928. ISSN 1432-1807. doi: 10.1007/BF01448847. URL `https://doi.org/10.1007/BF01448847`.

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16 (1):8–37, 1961.

O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, pages 1–5, Oct. 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1724-z. URL `https://www-nature-com.eaccess.ub.tum.de/articles/s41586-019-1724-z`.

J. Von Neumann and O. Morgenstern. *Theory of games and economic behavior*. Theory of games and economic behavior. Princeton University Press, Princeton, NJ, US, 1944. Pages: xviii, 625.

M. P. Wellman. Methods for Empirical Game-Theoretic Analysis. In *AAAI*, 2006.

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3):229–256, May 1992. ISSN 1573-0565. doi: 10.1007/BF00992696. URL `https://doi.org/10.1007/BF00992696`.

P. R. Wurman, S. Barrett, K. Kawamoto, J. MacGlashan, K. Subramanian, T. J. Walsh, R. Capobianco, A. Devlic, F. Eckert, F. Fuchs, L. Gilpin, P. Khandelwal, V. Kompella, H. Lin, P. MacAlpine, D. Oller, T. Seno, C. Sherstan, M. D. Thomure, H. Aghabozorgi, L. Barrett, R. Douglas, D. Whitehead, P. Dürr, P. Stone, M. Spranger, and H. Kitano. Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature*, 602(7896):223–228, Feb. 2022. ISSN 1476-4687. doi: 10.1038/

s41586-021-04357-7. URL `https://www.nature.com/articles/s41586-021-04357-7`. Number: 7896 Publisher: Nature Publishing Group.

S. Zheng, A. Trott, S. Srinivasa, N. Naik, M. Gruesbeck, D. C. Parkes, and R. Socher. The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies. *arXiv:2004.13332 [cs, econ, q-fin, stat]*, Apr. 2020. URL `http://arxiv.org/abs/2004.13332`.

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML'03, pages 928–935, Washington, DC, USA, Aug. 2003. AAAI Press. ISBN 978-1-57735-189-4.