Learning Equilibria in Symmetric Auction Games using Artificial Neural Networks

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

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 often turn out to be intractable systems of partial differential equations. Previous work has relied either on solving such PDEs explicitly, 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 based on gradient dynamics in self-play to provably learn local equilibria. Our empirical results show that these approximated Bayes-Nash equilibria 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 for winning or not winning.

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

1. Introduction

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, 2, 3]. While complete-information games have seen some progress, equilibrium learning for incomplete-information (aka. 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, 6, 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-price and second-price sealed-bid auction in the independent private values (IPV) model led to a comprehensive theoretical framework for the analysis of single-item auctions, a landmark result of economic theory [10, 11].

While 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 where 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 2020 that was awarded to Paul Milgrom and Robert B. Wilson highlighted in particular their contribution to auctions with interdependent values [14].

Numerical techniques to compute Bayes-Nash equilibria can be very valuable. Although there has been significant recent work on imperfect-information finite-dimensional extensive-form games such as Poker or other card games [15, 16, 17, 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 and quasi-linear utilities [19, 20, 21, 20, 22, 23, 24]. The motivation for such restrictions is the computational hardness of equilibrium computation.

For finite, complete-information games, we know of the existence of a mixed Nash equilibrium and that the computation is PPAD-hard [4]. For Bayesian games with continuous types and actions, we neither know whether (possibly mixed) Bayes-Nash equilibria exist in the general case, nor do we know how hard they are to find if they exist. Cai and Papadimitriou [25] 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 BNE in other multi-item auctions. Even approximating equilibria in these auction games is NP-hard [25].

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 [26]. 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 [27, 28, 29, 30] in their average history of play. Coarse correlated equilibria (CCE) encompass the set of correlated equilibria (CE). The latter is a nonempty convex polytope which in turn contains the convex hull of the game's Nash equilibria such that we get NE $\subset$ CE $\subset$ CCE. In contrast to correlated equilibria, coarse correlated equilibria may contain strictly dominated (pure) strategy profiles with positive probability. This means that while coarse correlated equilibria are learnable via no-regret algorithms, they are a rather weak solution concept [31]. Therefore, the question is when learning dynamics converge to a Nash equilibrium. A different relaxation of NE is given by local equilibria [32] which only 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 Bayes-Nash equilibria even in simple simultaneous single-item auctions in the worst case [25], it is far from obvious that no-regret dynamics can find a BNE in continuous-type and -action Bayesian games. There has been a recent literature using deep learning for auction design [33, 34, 35, 36] but this work does not attempt to find BNE in auctions. Challenges in computing Nash equilibria in general-sum games have also led to alternative solution concepts [37]. Apart from this, artificial intelligence and machine learning are increasingly used to predict strategic behavior of humans [38] or outcomes of auctions in the field [39].

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 (e.g., accommodating risk aversion). Neural networks are used to represent the players' bid functions, and the agents learn via self-play. Unfortunately, using neural self-play in this environment is not straightforward: While we assume players' *expected utility* (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 (i. e., 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 [25], it is not obvious that gradient ascent schemes such as ours would converge to BNE. We leverage the fact that the vast majority of auction games described in the literature assume symmetric bidders and equilibrium bid functions [11]. This restricted version of the game leads to a potential game, and gradient dynamics converge to local Nash equilibria in potential games. Whether we find global and not only local approximate equilibria can then be checked empirically. 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 while all the others bid constant zero, independent of their respective private valuations.

In our experiments in the main paper, we illustrate NPGA via a combinatorial auction in the local-local-global (LLG) model [40], which has received significant attention due to the use of core-selecting combinatorial auctions for spectrum sales world-wide [41]. 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, where correlation among bidder valuations has been investigated analytically with quasilinear utility functions, but this is not the case for risk aversion. Besides, 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 policy makers. 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 BNE that arise with different types of interdependencies, distributional assumptions, or different levels of risk aversion.

2. The Algorithm

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

2.1. Notation

An incomplete-information or *Bayesian game* is given by a sextuplet $G = (\mathcal{I}, \mathcal{V}, \mathcal{O}, \mathcal{A}, f, u)$. 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, f_{o_i}, etc.; its conditionals by $f_{v_i|o_i}$, etc.; 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 the observation 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 $u = (u_1, \ldots, u_n)$ 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 have observations nor valuations, only knowledge about f. In the *interim* stage, agents additionally observe o_i providing (possibly partial or noisy) information about their own valuations v_i. Full access to the outcomes $u(v, b)$ is given only after taking actions (*ex-post*). In our formulation, we do not assume explicit ex-post access any values (e.g., v_i, v_{-i}, 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 by $\Sigma_i \equiv \mathcal{A}_i^{\mathcal{O}_i}$ and by $\Sigma \equiv \prod_i \Sigma_i$, respectively, the resulting spaces of individual and joint pure strategies. 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 $\langle \cdot, \cdot \rangle_{\Sigma_i} : \Sigma_i \times \Sigma_i \to \mathbb{R}$, $(\alpha, \beta) \mapsto \mathbb{E}_{o \sim f_o} \left[\alpha(o)^T \beta(o)\right]$ and the norm $\|\beta\|_{\Sigma_i} \equiv \sqrt{\langle \beta, \beta \rangle_{\Sigma_i}}$ such that they form Hilbert spaces [42].

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 $|\mathcal{K}| = 2^m$. In the *private values* setting, we always have $o_i = v_i$; in the *common values* setting, there's some unobserved constant $v_c = v_1 = \cdots = v_n$ and the o_i can be considered noisy measurements of v_c. Mixed settings are likewise possible. In any case, based on bid profile b, an *auction mechanism* will determine two things: An allocation $x = x(b) = (x_1, \ldots x_n)$ which constitutes a partition of the m items, where bidder i is allocated the bundle x_i; and a price vector $p(b) \in \mathbb{R}^n$, where p_i is the monetary amount bidder i has to pay in order 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 the utilities u_i are then described by *quasilinear* payoff functions $u_i^{QL}(v_i, b) = (x_i(b) \cdot v_i - p_i(b))$, i.e. by how much a player values her allocated bundle minus the price she has to pay. An extension to this basic setting includes *risk-aversion*. Here, we model risk-aversion via utilities $u^{RA} = (u^{QL})^\rho$ where $\rho \in (0,1]$ is the risk attitude; $\rho = 1$ describes risk-neutrality, smaller values lead to strictly concave, risk-averse transformations of u^{QL}. Risk aversion is an established way to explain why in field studies of single-object first-price sealed-bid (FPSB) auctions, bidders bid higher than their risk-neutral counterparts in analytical BNE [43].

For fixed strategy profiles $\beta \in \Sigma$, 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: While other agents follow β, we define agent i's *interim utility* as the expected utility of choosing an action b_i conditioned on the observation o_i:

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

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

$$\overline{\ell}(o; b_i, \beta_{-i}) = \sup_{b_i' \in \mathcal{A}_i} \overline{u}_i(o_i, b_i', \beta_{-i}) - \overline{u}_i(o_i, b_i, \beta_{-i}). \tag{2}$$

Then, an *(interim) ϵ-Bayes-Nash Equilibrium (ϵ-BNE)* is a strategy profile $\beta^* = (\beta_1^*, \dots, \beta_n^*) \in \Sigma$ such that no agent can improve her own interim utility by more than $\epsilon \geq 0$ by unilaterally deviating from β^*. Thus, in an ϵ-BNE the following holds:

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

For $\epsilon = 0$, we will call the BNE *exact*, or simply drop the ϵ-prefix. Additionally, we will also need the ex-ante utility, defined as $\tilde{u}_i(\beta_i, \beta_{-i}) = \mathbb{E}_{o_i \sim f_{o_i}}[\overline{u}_i(o_i, \beta_i(o_i), \beta_{-i})]$, that can be interpreted as the expected utility over all of f for a particular strategy β_i against fixed opponents β_{-i}. Similarly, we will define *ex-ante loss* $\tilde{\ell}_i(\beta_i, \beta_{-i})$ and *ex-ante ϵ-BNE* 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 $\tilde{G} = (\mathcal{I}, \Sigma, \tilde{u})$ with an infinite-dimensional *action* space Σ that is identical to the *strategy* space of the Bayesian game. Clearly, every exact (interim) BNE also constitutes an exact ex-ante BNE. The reverse holds almost surely (a.s.), i.e. 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 $0 = \tilde{\ell}_i(\beta^*) = \mathbb{E}_{o_i}\left[\overline{\ell}_i(o_i; \beta_i^*(o_i), \beta_{-i}^*)\right]$ and the fact that $\overline{\ell}_i(o_i, \beta) \geq 0$ by definition. Importantly, this a.s. equivalence of ex-ante and (interim) BNE 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 NE and BNE, one is often interested in relaxations of equilibria that may be easier to find in some circumstances. For example, in *local BNE*, the loss requirement is relaxed to only consider best responses from a neighborhood of the equilibrium strategy profile: We call a strategy profile β^* a *local ex-ante BNE*, iff there exists an open set $\emptyset \neq W_i \subset \Sigma_i$ such that $\beta_i^* \in W_i$ and $\tilde{u}_i(\beta_i^*, \beta_{-i}^*) \geq \tilde{u}_i(\beta_i', \beta_{-i}^*)$ for all agents 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 [44] and no other local BNE.

Smoothness of the (ex-post) utilities are a standard assumption in the analysis of Bayesian games [44], but due to the discrete nature of allocations x this is commonly violated in auctions. Instead, let us introduce a weaker notion of smoothness at the interim stage that 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 (i) the interim utilities $\bar{u}_i(o_i, b_i, \beta_{-i})$ are continuously differentiable with respect to their second argument for each $i \in \mathcal{I}$ and any $o_i \in \mathcal{O}_i, \beta_{-i} \in \Sigma_{-i}$; (ii) all partial derivatives are uniformly bounded by a finite constant $Z < \infty$:

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

and (iii) the ex-post utilities are F-square-integrable:

$$\text{There exists } S < \infty, \text{ s.t. for all } i \in \mathcal{I}, \beta \in \Sigma: \quad \mathbb{E}_{v_i, o}\left[u_i\left(v_i, \beta_i(o_i), \beta_{-i}(o_{-i})\right)^2\right] \leq S. \tag{5}$$

To see why the assumption of interim differentiability is justified, consider that ex-post utilities in auctions are generally piecewise-smooth. Nondifferentiability only occurs at the bid profiles where the auctioneer is indifferent between multiple possible allocations 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 allocation x. The choice probabilities for these are given by $P(x|b_i, o_i, \beta_{-i})$, bidder i's Bayesian belief that x will be chosen if she bids b_i. If the β_{-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 $\nabla \bar{u}_i(o_i, b_i, \beta_{-i}) \equiv (\partial \bar{u}_i(o_i, b_i, \beta_{-i})/\partial b_{ik})_k$ and call it the *interim payoff gradient*. Furthermore, when G is interim-smooth, the ex-ante gradients $\nabla_{\beta_i} \tilde{u}_i(\beta_i, \beta_{-i}) \in \Sigma_i$ 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, i. e. $f_{v_i, o_i} = f_{v_j, o_j}$ (but not necessarily independent), as are their individual utilities (almost surely, up to tie-braking): $u_i(\beta_i, \beta_{-i}) = u_j(\beta_i; \beta_{-i})$ with probability 1. The literature primarily discusses equilibria which are likewise symmetric, i.e. $\beta^* = (\beta_1^*, \beta_1^*, \dots \beta_1^*)$ [11, Chapter 2.1]. We will refer to auctions that are booth symmetric and interim-smooth as *symmetric and smooth auction* games.

2.2. Neural Pseudogradient Ascent

Our numerical technique to learn Bayes-Nash equilibria called Neural Pseudogradient Ascent (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 the dynamics of the game, 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 Bayes-Nash equilibrium β^* for a continuum of observations o that bidders can draw. In other words, we search for a profile of equilibrium bid functions in infinite-dimensional spaces.

We start by taking the infinite-dimensional, complete information game interpretation $\tilde{G} = (\mathcal{I}, \Sigma, \tilde{u})$ 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 let's us define a finite-dimensional approximation of $\tilde{G}$, which we will call the *proxy game*:

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

Common neural network architectures have been shown to be able to approximate any sufficiently regular function arbitrarily well [45], so 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 thus 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 the utility $\tilde{u}_i$ and estimated via the empirical sample-mean of ex-post utilities from a large number H of auctions: After playing a batch of games, agents observe their utility, estimate its gradient with respect to θ_i, and apply an update to their policy network parameters 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 back-propagate gradients through the empirical objective. We solve this problem by leveraging an evolutionary strategy (ES) optimization technique that effectively smoothes the objective [46, 47]. This allows us to derive an adequate estimate of the ex-ante payoff gradients even under ex-post non-smoothness.

Algorithm 1: Neural Pseudogradient Ascent using Evolutionary Strategy gradients

Input: agents $i \in \mathcal{I}$ with initial policies $\beta_i^0 := \pi_i(\,\cdot\,; \theta_i^0)$ induced by initial parameters θ_i^0; ES population size P; ES noise standard deviation σ; learning rate η; batch size H

for $t := 1, 2, \ldots$ **do**

Sample a batch $(v_h, o_h)_{h=1,\ldots,H}$ of valuation and obseravtion profiles from the prior f

Calculate joint utility in current strategy profile:

$$\tilde{u}^{t-1} := \frac{1}{H} \sum_h \tilde{u}\left(v_h, \beta^{t-1}(o_h)\right)$$

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 $\varepsilon_p \sim \mathcal{N}(0, \sigma^2 I)$ i.i.d. for all $p \in \{1, \ldots, P\}$

For each p, evaluate the fitness of θ_p by playing against current opponents:

$$\varphi_p := \frac{1}{H} \sum_h u_i\left(v_{h,i}, \pi_{i;p}(o_{h,i}), \beta_{-i}^{t-1}(o_{h,-i})\right) - \underbrace{\tilde{u}_i^{t-1}}_{\text{baseline}}$$

Calculate ES pseudogradient as fitness-weighted perturbation noise:

$$\nabla^{ES} := \frac{1}{\sigma^2 P} \sum_p \varphi_p \varepsilon_p$$

Perform a gradient update step on the current policy:

$$\Delta\theta_i^t := \eta^t \nabla^{ES} \tilde{u}_i^{t-1}(\beta^{t-1}(o)), \qquad \theta_i^t := \theta_i^{t-1} + \Delta\theta_i^t, \qquad \beta_i^t := \pi_i(\,\cdot\,; \theta_i^t)$$

end

end

We provide the pseudo-code 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 her individual (ex-ante) payoff gradient at the current strategy profile. The noise is an artifact 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.

2.3. Convergence

In our experimental results below and in the supplementary material, we find that NPGA always converges very close to the global ϵ-BNE, which, at first, is surprising given the known results about non-convergence of gradient play to Nash equilibria in general [48] 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 [49].

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 Section 2.1). 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 $\beta^* = (\beta_1^*, \ldots, \beta_1^*)$ in the restricted game, limited to symmetric strategies.

Definition 3 (potential game [50]). A complete information game $\Gamma = (\mathcal{I}, \Theta, \tilde{u})$ is an *(exact) potential game* 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 $\theta_{-i} \in \Theta_{-i}$, it holds that

$$\tilde{u}_i(\theta_i, \theta_{-i}) - \tilde{u}_i(\theta_i', \theta_{-i}) \;=\; \phi(\theta_i, \theta_{-i}) - \phi(\theta_i', \theta_{-i}). \tag{6}$$

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's 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 [32]. 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 of the restricted game.*

A formal proof can be found in the Methods Section. If NPGA converges to a global approximate equilibrium o the game can be checked empirically, as we discuss in the next section.

3. Empirical Evaluation

We illustrate the versatility of NPGA in the context of combinatorial auctions in the well-known local-local-global (LLG) environment, which has been an important model for the discussion about spectrum auction formats [41, 51]. NPGA allows us to analyze 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 additional results for single-object auctions with different types of value interdependencies (including common values models), small and larger multi-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 first-price sealed-bid 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 minutes. However, the runtime depends on the specific model analyzed (e.g., the prior distribution, the number of bidders, and the auction format).

3.1. The Local-Local-Global Model

The LLG model consists of two objects $\{1,2\}$, two local bidders $i \in \{1,2\}$ and one global bidder $i = 3$, each being only interested in one specific bundle (of the single object i (locals) or both objects (global)) [40], and 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 licenses such as Australia or Canada, where local telecoms compete against operators who provide their services nation-wide, 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. While there are hardly any game-theoretical analyses of combinatorial auctions, this model is simple enough to allow for the derivation of analytical results [52]. It was shown that with independent private values and risk-neutral bidders, core-selecting payment rules lead to significant inefficiencies in equilibrium [40] in combinatorial auctions. Essentially, 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, it can happen that both local bidders bid too low in total and they fail to outbid the global bidder, even if their combined valuations are higher than the global bidder's. This results in an inefficient outcome. This fact has been used as an argument against core-selecting combinatorial auctions [41].

Now, 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. Recently, the model was analyzed with different types of correlation [52]. 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 analyzed, although it plays a role in the revenue ranking of single-object auctions. In 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, he might bid even lower as a result of this knowledge. The environment is not symmetric as there are two local and a global bidder. However, the global bidder has a simple dominant strategy to bid truthful and the two local bidders can indeed be considered symmetric whenever $f_{v_1} = f_{v_2}$.

Ausubel and Baranov [52] investigate two models of correlation among local bidders' private values and derive analytical BNE, which we will use as a baseline in our experiments. Let's define the joint prior f to be the five-dimensional uniform distribution of a latent random variable $\omega \sim \mathcal{U}[0,1]^5$. 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 a 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 [52] propose two different ways to choose w such that $\text{corr}(v_1, v_2) = \gamma$: 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), and nearest-bid (NB) rule in the Bernoulli weights model. These rules all choose the efficient allocation x (according to the submitted bids) but select different price vectors p from the set of core-stable outcomes. For example, the nearest-VCG rule picks the point in the core that minimizes the Euclidean distance to the (unique) Vickrey-Clarke-Groves payments. Similarly, the nearest-zero point takes the origin of the coordinate system as a reference point, while the nearest-bid rule minimizes the distance to the vector of submitted bids b. Only the nearest-VCG rule has been used in spectrum sales so far. Apart from these core-selecting payment rules, we will also report the results in first-price sealed-bid (FPSB) auctions, for which no analytical BNE are known, as these are used in some spectrum sales [41], and in the VCG mechanism, which is not core-stable but always prescribes truthful bidding as a BNE.

3.2. Evaluation Criteria

Let us discuss how we will evaluate any learned strategy profile β 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 β^* is known, we draw on it for direct comparison. In this case, we sample the *BNE-utility* of each player, $\hat{u}_i(\beta^*) \approx \tilde{u}_i(\beta^*)$, as well as the utility β_i played against the BNE, $\hat{u}_i(\beta_i, \beta^*_{-i}) \approx \tilde{u}_i(\beta_i, \beta^*_{-i})$, with a batch size of 2^{22}. We then report the resulting *relative utility loss*:

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

Additionally, we report the probability-weighted root mean squared error 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 $\tilde{\ell}_i(\beta_i, \beta_{-i}) = \sup_{\beta'_i \in \Sigma_i} \tilde{u}_i(\beta'_i, \beta_{-i}) - \tilde{u}_i(\beta_i, \beta_{-i})$. Evaluating this supremum in function space Σ_i exactly 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 an observation o_i and each of the alternative bids b_w we evaluate the interim utility, $\bar{u}_i(o_i, b_w, \beta_{-i})$, 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 o_i, see equation (1). For n_{batch} samples of o_i and n_{batch} samples of $v_i, o_{-i}|o_i$ for each of the o_i's, we then have

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

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

$$\lambda_i(o_i, b', \beta) = \frac{1}{n_{\text{batch}}} \sum_{v_i, o_{-i}|o_i} \left(u_i\left(v_i, b', \beta_{-i}(o_{-i})\right) - u_i\left(v_i, \beta_i(o_i), \beta_{-i}(o_{-i})\right)\right). \tag{13}$$

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 $v_i, o_{-i}|o_i$ 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 like the uniform or normal distribution. For more complicated multivariate conditional distributions, we use the conditional distribution method (for details, see the supplementary information, Section S.3). Based on these estimates, we can compute a relative ex-ante utility loss without access to the analytical BNE:

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

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 β^*, 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 $n_{\text{grid}} \cdot n_{\text{batch}}^2 = 2^{34}$ simulated auctions. To sample that many games efficiently, both NPGA and our evaluation procedures leverage parallelization on GPU hardware. Certification of BNE is a challenge in all computational approaches to equilibrium computation. A thorough discussion for environments with standard quasilinear utility functions and independent private values is provided by [23].

3.3. Results

Let us first provide the aggregate convergence results in Table 1, which almost perfectly reproduce the BNE found in [52]. 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.

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 the efficiency with the wide-spread VCG-nearest payment rule at a precision of $\pm 1\%$ of $\mathcal{E}$. For the highest level of risk aversion of $\rho = 0.1$, the efficiency rose to about 98%, from about 84% under risk neutrality. So, while higher correlation of valuations does not lead to higher efficiency, 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 [53], but the analysis could not yet be extended to the general sealed-bid case.

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

4. Discussion

Auction theory, and game theory in general, is often very sensitive to model assumptions. While 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 [54]. Value interdependencies and changes in the utility function can have substantial impact on the resulting equilibrium bidding strategies. While 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 behavioral assumptions, different priors, and value interdependencies. The supplementary information provides additional 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 Bayes-Nash equilibria 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 [55]. When gradient dynamics as implemented with NPGA converge to a global Bayes Nash equilibrium is an interesting open research question. Beyond the study of equilibria in games, our techniques can possibly contribute to automated and empirical mechanism design [56, 57].

References

[1] Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365**, 885–890 (2019). URL `https://science.sciencemag.org/content/365/6456/885`.

[2] Daskalakis, C., Ilyas, A., Syrgkanis, V. & Zeng, H. Training gans with optimism. *arXiv preprint arXiv: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). URL `http://www.sciencemag.org/lookup/doi/10.1126/science.aar6404`.

[4] Daskalakis, C., Goldberg, P. & Papadimitriou, C. The Complexity of Computing a Nash Equilibrium. *SIAM Journal on Computing* **39**, 195–259 (2009). URL `https://epubs.siam.org/doi/abs/10.1137/070699652`.

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

[6] Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, 928–936 (2003).

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

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

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

[10] Vickrey, W. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* **16**, 8–37 (1961).

[11] Krishna, V. *Auction Theory* (Academic press, 2009).

[12] Bergemann, D. & Morris, S. Robust implementation in direct mechanisms. *The Review of Economic Studies* **76**, 1175–1204 (2009).

[13] Campo, S., Perrigne, I. & Vuong, Q. Asymmetry in first-price auctions with affiliated private values. *Journal of Applied Econometrics* **18**, 179–207 (2003).

[14] Janssen, M. C. Reflections on the 2020 nobel memorial prize awarded to paul milgrom and robert wilson. *Erasmus Journal for Philosophy and Economics* **13**, 177–184 (2020).

[15] Heinrich, J. & Silver, D. Deep reinforcement learning from self-play in imperfect-information games. *arXiv preprint arXiv:1603.01121* (2016).

[16] Lanctot, M. *et al.* A unified game-theoretic approach to multiagent reinforcement learning. *arXiv preprint arXiv:1711.00832* (2017).

[17] Brown, N., Lerer, A., Gross, S. & Sandholm, T. Deep counterfactual regret minimization. In *International conference on machine learning*, 793–802 (PMLR, 2019).

[18] Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365**, 885–890 (2019). URL `https://science.sciencemag.org/content/365/6456/885`.

[19] Reeves, D. M. & Wellman, M. P. Computing equilibrium strategies in infinite games of incomplete information. In *Proc. of Game-theoretic and Decision-theoretic Workshop at AAMAS'03* (2003).

[20] Naroditskiy, V. & Greenwald, A. Using iterated best-response to find bayes-nash equilibria in auctions. In *AAAI*, 1894–1895 (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. *Artificial Intelligence* **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 *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 119–127 (2017).

[23] Bosshard, V., Bünz, B., Lubin, B. & Seuken, S. Computing Bayes-Nash Equilibria in Combinatorial Auctions with Verification. *Journal of Artificial Inelligence Research* (2020). `1812.01955`.

[24] Feng, Z., Guruganesh, G., Liaw, C., Mehta, A. & Sethi, A. Convergence analysis of no-regret bidding algorithms in repeated auctions. *arXiv preprint arXiv:2009.06136* (2020).

[25] Cai, Y. & Papadimitriou, C. Simultaneous Bayesian auctions and computational complexity. In *Proceedings of the fifteenth ACM conference on Economics and computation*, 895–910 (2014).

[26] Fudenberg, D. & Levine, D. K. Learning and equilibrium. *Annu. Rev. Econ.* **1**, 385–420 (2009).

[27] Jafari, A., Greenwald, A., Gondek, D. & Ercal, G. On no-regret learning, fictitious play, and nash equilibrium. In *ICML*, vol. 1, 226–233 (2001).

[28] Stoltz, G. & Lugosi, G. Learning correlated equilibria in games with compact sets of strategies. *Games and Economic Behavior* **59**, 187–208 (2007).

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

[30] 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 (2016).

[31] Viossat, Y. & Zapechelnyuk, A. No-regret dynamics and fictitious play. *Journal of Economic Theory* **148**, 825–842 (2013).

[32] Mazumdar, E., Ratliff, L. J. & Sastry, S. S. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science* **2**, 103–131 (2020).

[33] 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).

[34] Feng, Z., Narasimhan, H. & Parkes, D. C. Deep learning for revenue-optimal auctions with budgets. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, 354–362 (2018).

[35] Tacchetti, A., Strouse, D., Garnelo, M., Graepel, T. & Bachrach, Y. A neural architecture for designing truthful and efficient auctions. *arXiv preprint arXiv:1907.05181* (2019).

[36] Weissteiner, J. & Seuken, S. Deep learning—powered iterative combinatorial auctions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2284–2293 (2020).

[37] Morrill, D. *et al.* Hindsight and sequential rationality of correlated play. *arXiv preprint arXiv:2012.05874* (2020).

[38] Hartford, J. S. *Deep learning for predicting human strategic behavior*. Ph.D. thesis, University of British Columbia (2016).

[39] Ghani, R. & Simmons, H. Predicting the end-price of online auctions. In *International workshop on data mining and adaptive modelling methods for economics and management* (Citeseer, 2004).

[40] Goeree, J. K. & Lien, Y. On the impossibility of core-selecting auctions. *Theoretical Economics* **11**, 41–52 (2016).

[41] Bichler, M. & Goeree, J. K. *Handbook of spectrum auction design* (Cambridge University Press, 2017).

[42] Debnath, L., Mikusinski, P. *et al. Introduction to Hilbert spaces with applications* (Academic press, 2005).

[43] Bichler, M., Guler, K. & Mayer, S. Split-award procurement auctions—can bayesian equilibrium strategies predict human bidding behavior in multi-object auctions? *Production and Operations Management* **24**, 1012–1027 (2015).

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

[45] Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural networks* **4**, 251–257 (1991).

[46] Wierstra, D. *et al.* Natural Evolution Strategies. *Journal of Machine Learning Research* **15**, 949–980 (2014). URL `http://jmlr.org/papers/v15/wierstra14a.html`.

[47] Salimans, T., Ho, J., Chen, X., Sidor, S. & Sutskever, I. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *arXiv:1703.03864 [cs, stat]* (2017). URL `http://arxiv.org/abs/1703.03864`. `1703.03864`.

[48] Benaïm, M., Hofbauer, J. & Sorin, S. Perturbations of set-valued dynamical systems, with applications to game theory. *Dynamic Games and Applications* **2**, 195–205 (2012).

[49] Letcher, A. *et al.* Differentiable game mechanics. *Journal of Machine Learning Research* **20**, 1–40 (2019).

[50] Monderer, D. & Shapley, L. S. Potential games. *Games and economic behavior* **14**, 124–143 (1996).

[51] Bünz, B., Lubin, B. & Seuken, S. Designing core-selecting payment rules: A computational search approach. *ACM Conference on Electronic Commerce (EC)* (2018).

[52] Ausubel, L. M. & Baranov, O. Core-selecting auctions with incomplete information. *International Journal of Game Theory* (2019).

[53] Guler, K., Bichler, M. & Petrakis, I. Ascending combinatorial auctions with risk averse bidders. *Group Decision and Negotiation* **25**, 609–639 (2016).

[54] Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B. & Zame, W. R. The limits of ex post implementation. *Econometrica* **74**, 585–610 (2006).

[55] Daskalakis, C., Skoulakis, S. & Zampetakis, M. The complexity of constrained min-max optimization. *arXiv preprint arXiv:2009.09623* (2020).

[56] Vorobeychik, Y., Reeves, D. M. & Wellman, M. P. Constrained automated mechanism design for infinite games of incomplete information. In *UAI*, 400–407 (2007).

[57] Viqueira, E. A., Cousins, C., Mohammad, Y. & Greenwald, A. Empirical mechanism design: Designing mechanisms from data. In *Uncertainty in Artificial Intelligence*, 1094–1104 (PMLR, 2020).

[58] Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (2015). URL `http://arxiv.org/abs/1412.6980`. ArXiv: 1412.6980.

[59] Paszke, A. *et al.* Automatic differentiation in pytorch. In *NIPS-W* (2017).

Acknowledgements

We're grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BI 1057/1-9. We thank Vitor Bosshard, Ben Lubin, Panayotis Mertikopoulos, Sven Seuken, Takashi Ui, Felipe Maldonado, and participants of the NBER Market Design Workshop 2020 for valuable feedback on earlier versions.

Author Contributions Statement

Martin Bichler *(corresponding author)* initiated and supervised the project and contributed to the overall study design, theoretical analysis of NPGA, and writing the manuscript. Maximilian Fichtl contributed to the theoretical analysis of NPGA. Stefan Heidekrüger *(equal contribution)* contributed to the design, implementation and optimization of the algorithm and simulation framework, the theoretical analysis, and to the writing of the manuscript. Nils Kohring *(equal contribution)*

contributed to the optimization of the algorithm, design, implementation and optimization of the simulation framework, the theoretical and empirical analysis, and writing the manuscript. Paul Sutterer contributed to design and implementation of the algorithm and simulation framework, and the empirical analysis.

Competing Interests Statement

The authors declare no competing interests.

Figure Legends/Captions

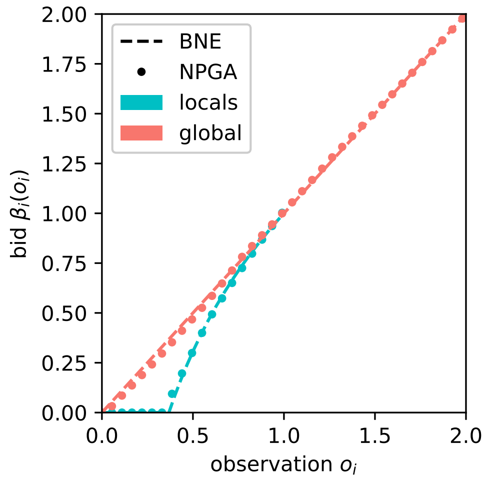

Figure 1 **Bid functions in the LLG auction with a nearest-zero core payment rule, where bidders are assumed independent and risk-neutral.**

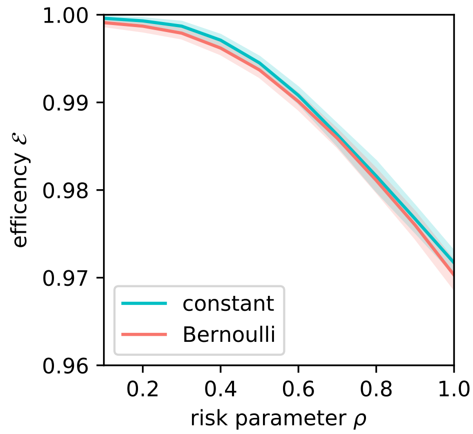

Figure 2 Comparison of the efficiency $\mathcal{E}$ of NPGA for different levels of risk attitudes ρ in the two correlation models in the LLG setting with the nearest-VCG payment rule. Depicted are mean and standard deviation of ten runs.

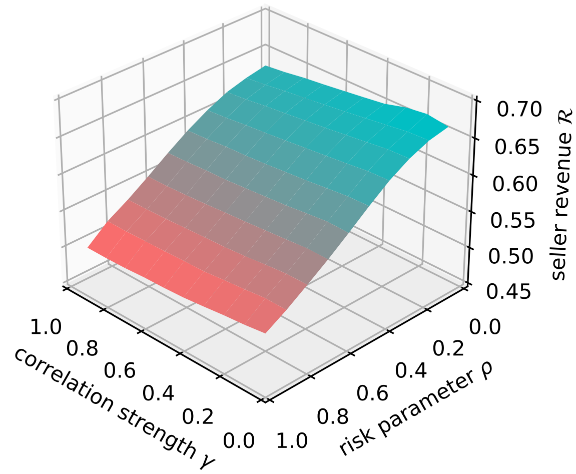

Figure 3 Comparison of the averaged seller revenue $\mathcal{R}$ for different levels of bidder correlation and risk attitudes. Approximate BNE is calculated via NPGA with the nearest-VCG payment rule in the LLG auction game.

Tables

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 standard deviation of experiments over ten runs.

Auction game	L_2	$\mathcal{L}$	$\hat{\mathcal{L}}$
LLG Bernoulli NZ	0.011 (0.005)	-0.000 (0.000)	0.007 (0.007)
LLG Bernoulli VCG	0.008 (0.003)	0.001 (0.000)	0.007 (0.005)
LLG Bernoulli NVCG	0.016 (0.016)	0.000 (0.000)	0.008 (0.007)
LLG Bernoulli NB	0.021 (0.021)	0.001 (0.000)	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)

Methods

Proof of Proposition 1

Proof: 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. Additionally, assume the auction mechanism to be anonymous: the identity and order of bidders almost surely have no influence on the allocation and payments (tie-braking on a nullset notwithstanding). Assume all players play the same strategy β_i. Then, the symmetric ex-ante utility function $\tilde{u}_i(\beta_i, \dots, \beta_i)$ is a potential function and $\tilde{G}$ is a potential game in this restricted environment. 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's 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 [32, Corollary 4.2]. It thus remains to be shown that (a) NPGA implements gradient-play in the proxy game Γ and thus finds a local Nash equilibrium θ^* of the proxy game, and (b) that this Nash equilibrium of the proxy game Γ—which restricts the strategy space to neural networks expressible by Θ— is also a BNE of the game G restricted to symmetric strategies. To show (a) and (b) below, we will rely on some auxiliary lemmata. The proofs of these lemmata are of a technical nature and can be found in Section S.2 of the supplementary information. 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.

To prove (a), 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's set $\tilde{u}_i^{\sigma}(\theta_i, \theta_{-i}) \equiv \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, \sigma^2 I)}[\tilde{u}_i(\theta_i + \varepsilon, \theta_{-i})]$ call $\Gamma^{\sigma} = (\mathcal{I}, \Theta, \tilde{u}^{\sigma})$ the *smoothed proxy game*, and define $\tilde{\ell}^{\sigma}$ analogously. Then Γ^{σ} is likewise a symmetric potential game, and we get the following lemmata:

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

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 by [32]. By Lemma 2, any Nash equilibrium of Γ^{σ} is an approximate Nash equilibrium of Γ.

For the latter (b), the universal approximation theorem [45] 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, i.e. 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 θ^* of Γ^σ which, by application of local versions of Lemma 2 and Lemma 3, retains a (local) ex-ante loss of at most $\kappa = Z(\delta + 2L\sqrt{d}\sigma)$, thus constituting a κ-BNE of G restricted to symmetric strategies. In practice, one may choose the parameters δ (via the NN architecture and size d) and σ sufficiently small such that the error vanishes.

Neural Network Architecture and Hyper-Parameters

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 [58] 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 by [47]). 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 (e. g., 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 [59] 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). The raw simulation artifacts underlying any figures and tables in the study will be made available on request.

Code availability

The source code of our simulation framework, 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.

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

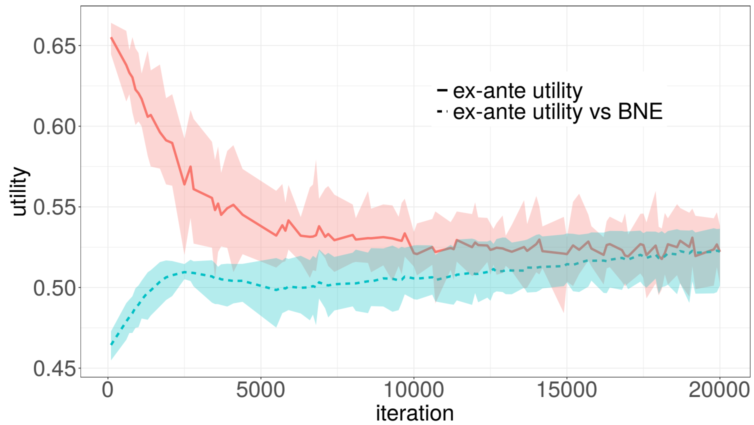

Figure S.1 **Learning curve of NPGA in a 10 player Gaussian first-price auction. Utility of learning opponents against each other (solid red line) and NPGA utility of learning opponents individually evaluated against the analytical BNE strategy (dashed blue line). Line and shaded area indicate mean,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

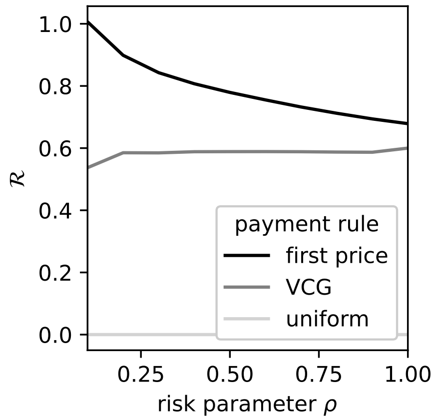

Figure S.2 **Comparison of the revenue $\mathcal{R}$ in a $n=m=2$ multi-unit auctions for different levels of risk aversion.**

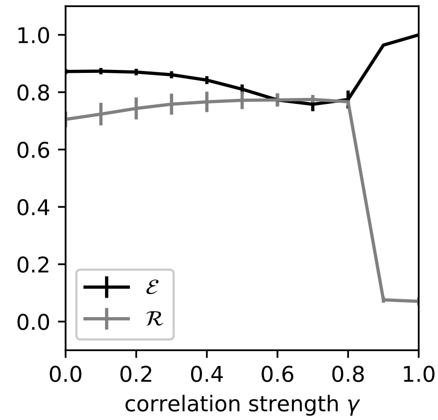

Figure S.3 **Impact of correlation on revenue and efficiency in a $n=m=2$ correlated multi-unit FPSB auction.**

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:

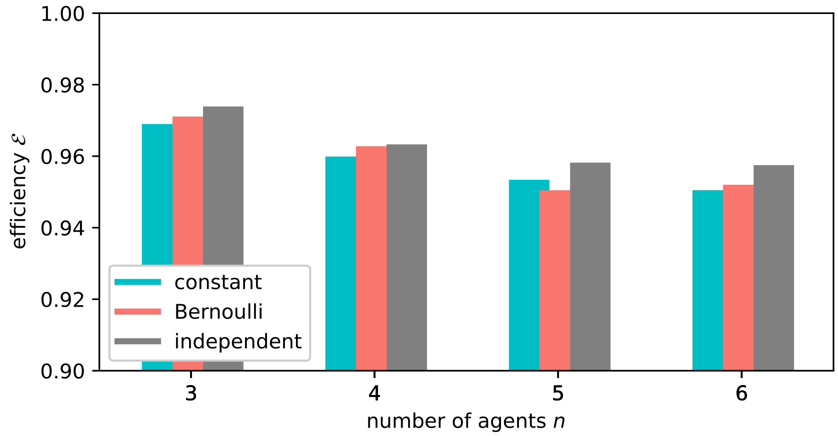

Figure S.4 **Efficiency in combinatorial auctions with increasing numbers of items and bidders. For the two correlated models, i. e. the constant weights model and the Bernoulli weights model, a correlation strength of $\gamma = 0.5$ is assumed and the selected pricing rule for all settings is nearest-VCG.**

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

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

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 $(T|O_i = o_i)$.

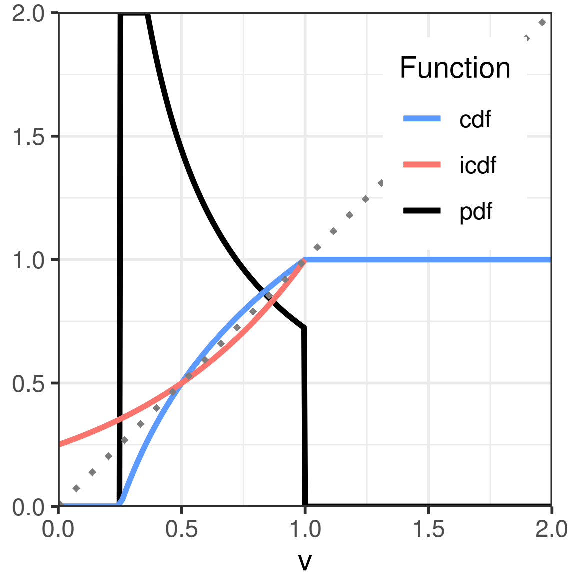

Figure S.5 **The probability functions in the common values setting for an example conditional value of $o_i = \frac{1}{2}$.**

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.