Revenue in First- and Second-Price Display Advertising Auctions: Understanding Markets with Learning Agents

Martin Bichler, Alok Gupta*, *MatthiasOberlechner*
Department of Computer Science, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany, m.bichler@tum.de
*Carlson School of Management, University of Minnesota, Minneapolis, MN 55455, USA, alok@umn.edu

---

Appendix A: Scope, Assumptions, and Limitations

This appendix summarizes what our model captures, what it abstracts away from, and how these choices bound the interpretation of our results. Throughout, we distinguish clearly between (i) the *strategic benchmark* we analyze analytically/numerically and (ii) the *engineering reality* of production systems.

A.1. Benchmark environment and equilibrium results

Our analytic results (incl. the FP/SP revenue comparison under ROI) are derived in the canonical *symmetric independent private values (SIPV)* benchmark:

- i.i.d. values $V_i \sim F$ on $[r, \bar{v}]$ with density $f > 0$ and reserve $r > 0$,
- single-slot, single-object auctions; no quality weights, no externalities,
- risk-neutral bidders; utilities are either quasilinear (QL) or ROI-based,
- symmetric, strictly increasing, C^1 equilibrium bid functions, $\beta(r) = r$.

These assumptions ensure (local) existence/uniqueness and monotonicity needed for our ODE-based characterization of symmetric FP equilibria and for the comparison argument establishing $\beta_{ROI}(v) \leq \beta_{QL}(v)$ (strictly for $v > r$ under $G'(v) > 0$), which in turn implies the SP > FP revenue ranking in our benchmark.

Limitations. (i) The SIPV abstraction rules out interdependent values/affiliation, bidder asymmetries, and common-value components. (ii) Our proof leverages monotonicity and regularity (G'/G well behaved on $[r, \bar{v}]$); irregular F may require separate analysis. (iii) A positive reserve avoids the division-by-zero issue in ROI ($p \geq r > 0$); how results adapt as $r \downarrow 0$ is nontrivial. (iv) In production systems, losing bids typically yield censored feedback rather than explicit zero rewards, and learning relies on counterfactual estimation or off-policy evaluation. Our model abstracts from these challenges to isolate strategic interaction under simplified feedback.

A.2. Learning experiments: design choices and caveats

We study ε-greedy, Exp3, and Thompson sampling in repeated play of the *fixed* Bayesian game.

• **On-policy, stationary benchmark.** The underlying game (players, payoff functions, value distribution) is fixed; policies evolve as agents learn. Our "convergence to equilibrium" claims refer to convergence of play to the analytically/numerically characterized BNE of this fixed game.

• **Reward definition.** Per round, ROI utility is $u_t = \frac{v_t - p_t}{p_t} \mathbf{1}\{\text{win at price } p_t\}$. Thus "reward $= 0$" on a loss is the *model's* period utility (no display, no spend, no return).

• *Censoring/estimation in practice.* In production, values are predicted and feedback when losing is censored. Addressing this setting would require an explicit value-estimation and off-policy evaluation layer, potentially using inverse-propensity weighting (IPW) or doubly robust (DR) correction. Our learning experiments do not incorporate such corrections. The robustness checks reported in Figure 2 and Appendix D instead refer to benchmark variations in prior distributions, bidder objectives, and utility specifications; they do not address censored-feedback learning or value-estimation bias.

Limitations. (i) We do not model nonstationary markets (entry/exit, drift in value distributions); extending to such dynamics requires different solution concepts. (ii) We omit contextual features; adding them turns the problem into a richer contextual-bandit problem that we leave for future work. (iii) Our three algorithms are representative but not exhaustive; guarantees for arbitrary learners are beyond scope.

A.3. Budgets and pacing: objectives versus constraints

ROI is a per-dollar *objective*; budgets/pacing are *constraints* in a dynamic, cross-auction problem. In practice, tight budgets induce a shadow price of spend, often implemented as multiplicative pacing $b_t^{\text{eff}} = \alpha_t b_t$ to track delivery; slack budgets do not force spend and a pure ROI maximizer may rationally abstain from low-ROI impressions.

Limitations. We do not model intertemporal budget constraints or platform-level pacing equilibria. Such models form a different game (stateful, potentially marketplace-level fixed points) and could change the model's counterfactual predictions. Our baseline result should therefore be read as isolating a distinct mechanism: ROI preferences alone tilt revenue toward second-price auction in the SIPV benchmark.

In our model, ROI maximization is treated as a per-impression objective that captures efficiency considerations, rather than as a pacing mechanism enforcing budget exhaustion. Conceptually, ROI maximization can be understood as the outcome of a Lagrangian relaxation of a constrained optimization problem in which a bidder faces a long-run budget constraint. In such a formulation, the shadow price of spending induces a focus on per-dollar efficiency. Importantly, however, ROI is a ratio objective and does not impose a hard constraint on total expenditure. When budgets are slack, they simply do not bind; bidders have no incentive to overbid to spend more, since doing so would reduce ROI. This ensures that equilibrium behavior is well defined even in the absence of explicit budget constraints.

A.4. Market thickness and heterogeneity

As N grows, top order statistics bunch and format differences compress; our ROI ranking persists but the gap shrinks with N. Conversely, at the impression level, effective competition may be thin (small N_{eff}) due to targeting/eligibility, reserves/brand-safety, geography/time segmentation, and private deals.

Limitation. We treat N as exogenous and identical across auctions; endogeneous participation and heterogeneous bidder pools are not modeled.

A.5. Summary

The paper provides a baseline, not a full-stack market model. We do not replicate the full DSP pipeline (supervised CTR/CVR prediction, creative selection, cross-campaign pacing, supply-path optimization, throttling/header bidding, deal IDs/walled gardens, ad fraud/measurement).

Our aim is a *clean strategic benchmark* to isolate how bidder objectives affect equilibrium bids and format revenues. Within SIPV, changing only the objective from QL to ROI yields SP > FP in expected revenue; this mechanism is clean and policy-relevant, but its quantitative impact in any given marketplace will depend on additional forces—budgets, asymmetries, information, and operations—that we document here and leave to future work.

Appendix B: Proofs

B.1. Proof of Theorem 1

Proof:

Step 0: Model and payoff. Fix a bidder with value $v \in [0,1]$. If she submits bid b and wins, she pays $p = b$ (first-price) and obtains ROI payoff

$$u(v,b) = \frac{v-b}{b}.$$

If she loses, she obtains payoff 0. The auction uses a reserve price $r > 0$, so a bid $b < r$ cannot win; bids $b \geq r$ are admissible.

We consider symmetric increasing equilibria on $[r,1]$ and write β for the common bidding function on $[r,1]$. Let β^{-1} denote its inverse on the image $\beta([r,1])$.

Step 1: Non-participation for $v < r$ and boundary condition at $v = r$. If $v < r$ and the bidder bids $b < r$, she cannot win and obtains payoff 0. If instead she bids $b \geq r > v$ and wins, her ROI payoff is $(v-b)/b < 0$, so her expected payoff is non-positive (strictly negative whenever winning occurs with positive probability). Hence every best response of types $v < r$ is to bid below r (e.g., bid 0), yielding expected payoff 0.

At $v = r$, bidding $b = r$ yields ROI payoff 0 upon winning, hence expected payoff 0 as well; thus $v = r$ can be taken as the participation threshold, and in any continuous symmetric equilibrium we must have the boundary condition

$$\beta(r) = r. \tag{1}$$

In what follows we derive β on $[r,1]$.

Step 2: Expected ROI from deviating to an arbitrary bid. Assume opponents use the increasing strategy $\beta(\cdot)$ on $[r,1]$ and do not bid above r when their value is below r. Consider a deviation bid $b \geq r$. Let $x = \beta^{-1}(b) \in [r,1]$ be the value type who bids b under β. Because values are i.i.d. uniform on $[0,1]$, the probability that a single opponent bids at most b is the probability that their value is at most x, i.e., $F(x) = x$. Therefore, the probability of winning is

$$\Pr(\text{win} \mid b) = \Pr(\max_{j \neq i} \beta(v_j) \leq b) = \Pr(\max_{j \neq i} v_j \leq x) = x^{N-1}.$$

Hence bidder i's expected ROI payoff from bidding $b = \beta(x)$ is

$$U(v,x) \;=\; x^{N-1}\,\frac{v-\beta(x)}{\beta(x)} \qquad \text{for } x \in [r,1]. \tag{2}$$

In a symmetric equilibrium, type v chooses $x = v$ (equivalently $b = \beta(v)$) to maximize $U(v,x)$ over $x \in [r,1]$.

Step 3: First-order condition and the equilibrium ODE. Fix $v \in (r,1]$. Since β will be increasing and differentiable on $(r,1]$ (verified below), the objective $x \mapsto U(v,x)$ is differentiable. In a symmetric interior optimum at $x = v$ we require

$$\frac{\partial}{\partial x} U(v,x)\Big|_{x=v} = 0.$$

Differentiate (2). Let

$$H_v(x) \;=\; \frac{v-\beta(x)}{\beta(x)} \;=\; \frac{v}{\beta(x)} - 1, \qquad \text{so} \qquad H_v'(x) \;=\; -\frac{v\,\beta'(x)}{\beta(x)^2}.$$

Then

$$\frac{\partial}{\partial x} U(v,x) = (N-1)x^{N-2} H_v(x) + x^{N-1} H_v'(x).$$

Evaluating at $x = v$ gives

$$0 = (N-1)v^{N-2}\Big(\frac{v-\beta(v)}{\beta(v)}\Big) + v^{N-1}\Big(-\frac{v\,\beta'(v)}{\beta(v)^2}\Big).$$

Multiply by $\beta(v)^2/v^{N-2}$ to obtain

$$0 = (N-1)\beta(v)\big(v-\beta(v)\big) - v^2\beta'(v).$$

Therefore, any differentiable symmetric equilibrium β on $(r,1]$ must satisfy the ODE

$$\beta'(v) \;=\; (N-1)\Big(\frac{\beta(v)}{v} - \frac{\beta(v)^2}{v^2}\Big) \;=\; (N-1)\frac{\beta(v)\big(v-\beta(v)\big)}{v^2}, \qquad v \in (r,1], \tag{3}$$

together with the boundary condition (1).

Step 4: Solving the ODE via the substitution $y(v) = \beta(v)/v$. Define $y(v) = \beta(v)/v$ for $v \in (r,1]$. Then $\beta(v) = v\,y(v)$ and

$$\beta'(v) = y(v) + v\,y'(v).$$

Substituting into (3) yields

$$y(v) + v\,y'(v) = (N-1)\big(y(v) - y(v)^2\big). \tag{4}$$

Rearranging gives

$$v\,y'(v) = (N-2)\,y(v) - (N-1)\,y(v)^2 = y(v)\Big((N-2) - (N-1)y(v)\Big). \tag{5}$$

The boundary condition $\beta(r) = r$ is equivalent to

$$y(r) = \frac{\beta(r)}{r} = 1. \tag{6}$$

We solve (5)–(6) separately for $N = 2$ and $N > 2$.

Case 1: $N = 2$**.** Then (5) becomes

$$v\,y'(v) = -y(v)^2.$$

Separate variables:

$$\frac{dy}{y^2} = -\frac{dv}{v}.$$

Integrate to obtain $-1/y = -\log v + C$, i.e.,

$$\frac{1}{y(v)} = \log v + C'.$$

Using $y(r) = 1$ gives $1 = \frac{1}{\log r + C'}$, hence $C' = 1 - \log r$. Therefore

$$y(v) = \frac{1}{1 - \log r + \log v}, \qquad \text{and hence} \qquad \beta(v) = v\, y(v) = \frac{v}{1 - \log r + \log v}.$$

Case 2: $N > 2$**.** Let $a = N - 2 > 0$ and $b = N - 1 > 0$. Then (5) is

$$v\, y'(v) = y(v)\big(a - b\, y(v)\big).$$

Separate variables:

$$\frac{dy}{y(a - by)} = \frac{dv}{v}.$$

Use partial fractions:

$$\frac{1}{y(a - by)} = \frac{1}{a} \cdot \frac{1}{y} + \frac{b}{a} \cdot \frac{1}{a - by}.$$

Integrate:

$$\frac{1}{a} \ln y - \frac{1}{a} \ln(a - by) = \ln v + C,$$

i.e.,

$$\ln\Big(\frac{y}{a - by}\Big) = a \ln v + C' \quad \Longrightarrow \quad \frac{y}{a - by} = K\, v^a$$

for some constant K. Solve for y:

$$y = K v^a (a - by) \implies y\big(1 + bKv^a\big) = aKv^a \implies y(v) = \frac{aKv^a}{1 + bKv^a}.$$

Impose $y(r) = 1$:

$$1 = \frac{aKr^a}{1 + bKr^a} \quad \Longrightarrow \quad 1 + bKr^a = aKr^a \quad \Longrightarrow \quad 1 = (a - b)Kr^a = (-1)\, Kr^a,$$

so $K = -r^{-a} = -r^{-(N-2)}$. Substitute back:

$$y(v) = \frac{a(-r^{-a})v^a}{1 + b(-r^{-a})v^a} = \frac{a\, v^a}{b\, v^a - r^a}.$$

Finally,

$$\beta(v) = v\, y(v) = \frac{a\, v^{a+1}}{b\, v^a - r^a} = \frac{(N-2)\, v^{N-1}}{(N-1)\, v^{N-2} - r^{N-2}},$$

as claimed.

Step 5: Verification (monotonicity and optimality). We verify that the obtained β constitutes a symmetric equilibrium on $[r, 1]$.

(i) Feasibility and nonnegative ROI on $[r, 1]$. For $v \geq r$, the formulas above satisfy $\beta(v) \geq r$ and $\beta(v) \leq v$. Indeed, from (3) we have $\beta'(v) \geq 0$ whenever $0 \leq \beta(v) \leq v$. At the boundary $\beta(r) = r \leq r$, and the explicit solutions yield $\beta(v) \leq v$ for all $v \in [r, 1]$, so $v - \beta(v) \geq 0$ and the ROI payoff upon winning is nonnegative.

(ii) Strict monotonicity. For $v > r$, since $\beta(v) \in (r, v)$, the right-hand side of (3) is strictly positive:

$$\beta'(v) = (N-1)\frac{\beta(v)\big(v-\beta(v)\big)}{v^2} > 0.$$

Thus β is strictly increasing on $(r, 1]$ and hence invertible there.

(iii) Best response condition. Given opponents play β, bidder type $v \in (r, 1]$ maximizes $U(v, x)$ in (2). We showed that $x = v$ satisfies the first-order condition. Moreover, under the derived β, the objective $x \mapsto U(v, x)$ is single-peaked on $[r, 1]$ (because the log-derivative condition in Step 3 induces a unique stationary point), so $x = v$ is the unique maximizer. Hence bidding $b = \beta(v)$ is a best response for each $v \in [r, 1]$. For $v < r$, Step 1 shows that any bid below r is optimal (yielding payoff 0).

Therefore, the strategy profile in which each bidder bids $\beta(v)$ for $v \in [r, 1]$ and bids below r for $v < r$ forms a symmetric Bayesian Nash equilibrium. This completes the proof.

B.2. Proof of Corollary 1

Proof: For the second-price auction, Proposition 1 implies truthful bidding, so the payment equals $\max\{r, V_{(N-1)}\}$ whenever $V_{(N)} \geq r$, and is 0 otherwise. This yields the first equality in (6). To compute it in closed form, note that

$$\mathrm{Rev}^{\mathrm{SP}}(r, N) = r\,\Pr(V_{(N-1)} < r < V_{(N)}) \;+\; \mathbb{E}[V_{(N-1)}\mathbf{1}\{V_{(N-1)} \geq r\}].$$

The first term corresponds to exactly one bidder exceeding the reserve: $\Pr(V_{(N-1)} < r < V_{(N)}) = N(1-r)r^{N-1}$, hence it contributes $Nr^N(1-r)$. For the second term, the density of $V_{(N-1)}$ under Unif$[0,1]$ is $f_{(N-1)}(x) = N(N-1)x^{N-2}(1-x)$, so

$$\mathbb{E}[V_{(N-1)}\mathbf{1}\{V_{(N-1)} \geq r\}] = \int_r^1 x\,N(N-1)x^{N-2}(1-x)\,dx = \frac{N-1}{N+1} - (N-1)r^N + \frac{N(N-1)}{N+1}r^{N+1}.$$

Adding the two contributions yields (6) after simplification.

For the first-price auction, if all bidders use the increasing equilibrium strategy β on $[r, 1]$ (Theorem 1), the winner is the highest-value bidder whenever $V_{(N)} \geq r$ and pays $\beta(V_{(N)})$. Since $V_{(N)}$ has density $f_{(N)}(v) = Nv^{N-1}$, we obtain (7). The identity (8) follows immediately. The strict inequality $\Delta(r, N) > 0$ is equivalent to Result 1 in the uniform benchmark and follows from the revenue comparison established there.

B.3. Proof of Theorem 2

Proof: First, we derive the ODE characterizing the BNE function of an ROI-maximizing bidder, analogously to Section 3.2. We assume that all opponents play according to a symmetric, increasing and differentiable strategy β with inverse β^{-1}. The ex-interim utility for bidder i submitting bid b with valuation v is given by

$$\bar{u}_i(b, \beta_{-i}, v) = G(\beta^{-1}(b)) \cdot \frac{v-b}{b}$$

where $G(v) = F(v)^{N-1}$ is the probability that all $N-1$ opponents have a value less than v. To find the optimal bid $b = \beta(v)$, take the derivative of $\bar{u}_i(b, \beta_i, v)$ with respect to b and set it to zero:

$$G'(\beta^{-1}(b))\beta^{-1\prime}(b)\frac{v-b}{b} + G(\beta^{-1}(b))\frac{-v}{b^2} = 0$$

Assuming symmetry of all bidders, i.e., $b = \beta(v)$ and using the inverse function theorem we get

$$G'(v)\frac{1}{\beta'(v)}\frac{v-\beta(v)}{\beta(v)} - G(v)\frac{v}{\beta(v)^2} = 0.$$

Solving for $\beta'(v)$, we obtain the ODE for ROI-maximizing agents:

$$\beta'(v) = \frac{G'(v)}{G(v)} \cdot \frac{\beta(v)}{v} \cdot (v - \beta(v)) =: \Psi(v, \beta(v)). \tag{7}$$

The well-known ODE for quasi-linear bidders agents can be derived analogously and is given by:

$$\beta'(v) = \frac{G'(v)}{G(v)} \cdot (v - \beta(v)) =: \Phi(v, \beta(v)). \tag{8}$$

The corresponding equilibria can be computed by solving the initial value problem (IVP) given by the ODEs and the initial condition $\beta(r) = r$, where $r > 0$ is the reserve price.

These differential equations are describing the learning rule or adjustment path of a bidding agent as their valuation increases. We want to understand how fast each bidder type increases their bid as their value increases. We define the change rate of the bid functions on the rectangle $\mathcal{R} := \{(v, \beta) \in [r, \bar{v}] \times [r, \bar{v}]\}$ using the right-hand sides of the ODEs $\Psi(v, \beta)$ and $\Phi(v, \beta)$.

Because $r > 0$, G, G' are continuous with $G > 0$ on $(r, \bar{v}]$, and the right-hand sides are locally Lipschitz in β on $\mathcal{R}$, each IVP admits a unique C^1 solution on $[r, \bar{v}]$, based on the Picard-Lindelöf theorem. Moreover, $\mathcal{R}$ is invariant for both flows, i.e. if the solution starts inside $\mathcal{R}$, then it stays in $\mathcal{R}$ for all $v \in [r, \bar{v}]$: if $\beta = v$ then $\Phi(v, \beta) = \Psi(v, \beta) = 0$, and if $\beta < v$ then $\Phi(v, \beta) \geq 0$ and $\Psi(v, \beta) \geq 0$, so solutions starting at (r, r) remain in $\mathcal{R}$ (in particular, no overbidding). For all $(v, \beta) \in \mathcal{R}$,

$$\Psi(v, \beta) - \Phi(v, \beta) = \frac{G'(v)}{G(v)}(v - \beta)\left(1 - \frac{\beta}{v}\right) = \frac{G'(v)}{G(v)}\frac{(v-\beta)^2}{v} \;\geq\; 0,$$

with strict inequality whenever $G'(v) > 0$ and $\beta < v$. Hence, on $\mathcal{R}$, $\Phi \leq \Psi$ pointwise. In words, if two functions start at the same point, and one always increases faster than the other, then it will always stay ahead.

Equal initial data $\beta_{ROI}(r) = \beta_{QL}(r)$ and the dominance $\Phi \leq \Psi$ imply $\beta_{ROI}(v) \leq \beta_{QL}(v)$ for all $v \in [r, \bar{v}]$. Uniqueness plus the strict inequality of the drifts whenever $G'(v) > 0$ and $\beta < v$ yields strict separation $\beta_{ROI}(v) < \beta_{QL}(v)$ for all $v > r$ with $G'(v) > 0$.

Finally, in a symmetric FP auction, expected revenue equals the expected winning bid $\mathbb{E}[\beta(V_{(1)})]$. Pointwise inequality with strict separation on a set of positive measure yields $\mathbb{E}[\beta_{ROI}(V_{(1)})] < \mathbb{E}[\beta_{QL}(V_{(1)})]$.

Appendix C: ROI- vs. ROS-Maximizing Agents

Apart from ROI, some practitioners also discusses ROS and we want to provide a robustness check with respect to this utility model.

$$\mathbb{E}[ROS] = \mathbb{E}[\tfrac{v}{p}]$$

ROS can be motivated from agents that aim to maximize total value of the impressions subject to payments being less than a budget constraint B (Tunuguntla and Hoban 2021). In an offline auction where all M impressions are auctioned off at the same time, the utility of the bidder could be described as

$$\max \sum_{m=1}^{M} \mathbb{E}(x_m v_m) \quad s.t. \sum_{m=1}^{M} x_m p_m \leq B,$$

where x_m is a binary variable that equals one upon winning and zero otherwise, p_m is the payment, and v_m the value per impression. Tunuguntla and Hoban (2021) only require the budget constraint for a campaign to hold in expectation and derive the Lagrangian of the resulting optimization problem. Based on the first-order condition for this Lagrangian, they derive a bidding strategy $b_m = v_m/\lambda^*$, where b_m is the bid, and λ^* defines a threshold. In a first-price auction $b_m = p_m$. Tunuguntla and Hoban (2021) devise an algorithm to learn an optimal λ^*. With an optimal λ^*, the advertiser wins the subset of auctions for which the value per expenditure is greatest. By rearranging terms, the optimal shading factor $\lambda = v_m/p_m$ is the ratio of the value for an impression and its payment, similar to ROS. This is reminiscent of widespread heuristics for the knapsack problem, which rank-order items by the ratio between value and weight of objects. $\lambda \in [0,1]$ is also described as *multiplicative shading factor* in recent publications on real-time bidding (Balseiro and Gur 2019, Fikioris and Tardos 2022). Looking at the practitioner literature, ROI or ROS seem to be dominant in the practice of real-time bidding.[1]

There is an important difference between ROI and ROS: While the expected utility for ROI becomes negative once the price is higher than the value, ROS never becomes negative, independent of how high the payment is. In other words, with a monotone allocation rule, if a bidder is losing, they can always increase their bid without ever having a negative payoff. This leads to a situation where agents can place the highest possible bid, regardless of how high this value is and which payment rule is being used.

Proposition 1 *In the first- and second-price sealed-bid single-object auction where agents are ROS maximizers, ties are broken randomly, and the action space is bounded from above, submitting the maximal bid is an equilibrium.*

Proof. Assume all agents bid according to $\beta(v) = \bar{b}$, where $\bar{b} > 0$ is the maximal bid the agents can submit. Then, the ex-interim utility is $\bar{u}(b, \beta_{-i}, v_i) = \frac{1}{N}\frac{v_i}{\bar{b}} > 0$ if $b = \bar{b}$, and 0 if $b < \bar{b}$. Since agents can only deviate from $\beta(v)$ by bidding less, this would strictly decrease their payoff. Therefore, $\beta(v) = \bar{b}$ is indeed an equilibrium strategy. □

This equilibrium strategy is also found by equilibrium learning algorithms. Obviously, this is unrealistic, because it ignores that bidders always have only a finite budget available. In this article, we consider ROS-maximizing bidders with a per-auction budget (ROSB), which we simply model using a log barrier function. This ROSB utility is given by

$$u_i^{ROSB}(b, v_i) = x_i(b)\left(\frac{v_i}{p_i(b)} + \log(B - p_i(b))\right), \tag{ROSB}$$

where B is the budget. It is difficult to derive a BNE analytically for such a utility function. The first-order condition of the ex-interim utility $\bar{u}_i$ for the first-price payment rule, under the assumption of symmetric bidders, leads to

$$\frac{d}{dv}\beta(v) = (B - \beta(v))\frac{G'(v)}{G(v)}\frac{v\beta(v) + \beta(v)^2\log(B - \beta(v))}{v(B - \beta(v)) - \beta(v)^2}. \tag{9}$$

[1] see `https://www.indeed.com/career-advice/career-development/roas-vs-roi`, `https://instapage.com/blog/roas-vs-roi-which-metric-should-you-use/`, `https://hawksem.com/blog/roi-vs-roas-similarities-differences/`

Similar to the first-order condition for ROI-maximizing bidders, this is a general non-linear ODE. But even for a uniform prior, we cannot solve this ODE analytically, highlighting the need for numerical approximation methods.

Appendix D: Additional Results for Numerical Equilibrium Analysis

D.1. SODA for Numerical Equilibrium Computation

Let us briefly describe a gradient-based method, the SODA learning algorithm (Bichler et al. 2025). This method provides us with an equilibrium strategy to compare against in situations where no analytical solution is available.

The main idea of SODA is to use distributional strategies on a discretized version of the auction game $\mathcal{G} = (\mathcal{I}, \mathcal{V}, \mathcal{A}, F, u)$. The discretized game is obtained by discretizing the type and action space, i.e., $\mathcal{V}^d = \{v_1, \ldots, v_n\} \subset \mathcal{V}$ and $\mathcal{A}^d = \{b_1, \ldots, b_m\} \subset \mathcal{A}$, and the prior distribution F, which leads to the discrete auction game $\mathcal{G}^d = (\mathcal{I}, \mathcal{V}^d, \mathcal{A}^d, F, u)$. Distributional strategies are an extension of mixed strategies in complete-information games to settings with incomplete information (Milgrom and Weber 1985). For the game $\mathcal{G}^d$, they can be represented by matrices $s \in \mathbb{R}^{n \times m}$, where each entry s_{ij} of this matrix represents the probability of a valuation-action pair $(v_i, b_j) \in \mathcal{V}^d \times \mathcal{A}^d$. To be consistent with the auction game, the distributional strategy has to satisfy the marginal condition $\sum_{j=1}^{m} s_{ij} = F^d(v_i)$, i.e., the probability of a value over all actions has to be equal to the probability of the value given by the prior. We denote the set of such distributional strategies by $\mathcal{S}^d$. Given a profile of such discrete distributional strategies, one can compute the expected utility $\tilde{u}_i$ as sum over all outcomes weighted by the probabilities induced by $s = (s_1, \ldots, s_N)$. If we consider an example with $N = 2$ agents, we get an expected utility for agent 1 by

$$\tilde{u}_1(s_1, s_2) = \sum_{i_1, j_1 = 1}^{n,m} (s_1)_{i_1 j_1} \sum_{i_2, j_2 = 1}^{n,m} (s_2)_{i_2 j_2} \, u_1(b_{j_1}, b_{j_2}, v_{i_1}), \tag{10}$$

where $u_1(b_1, b_2, v_1)$ is the ex-post utility of agent 1. The expected utility, together with the set of distributional strategies, allows us to interpret the auction game as a complete-information game $\Gamma = (\mathcal{I}, \mathcal{S}, \tilde{u})$. Note that the set of discrete distributional strategies is a compact and convex subset of $\mathbb{R}^{n \times m}$ and the expected utility is differentiable and linear in the bidder's own strategy. This allows us to rely on standard tools from online convex optimization to compute the NE of Γ, which corresponds to a BNE of $\mathcal{G}^d$. In our experiments, we use dual averaging (Nesterov 2009) with an entropic regularization term (known as exponentiated gradient ascent), mirror ascent (Nemirovskij and Yudin 1983) with an Euclidean mirror map (standard projected gradient ascent), and the Frank-Wolfe algorithm (Frank and Wolfe 1956). Given the respective method, all agents simultaneously compute their individual gradients and perform the corresponding update step, as described in Algorithm 1 for dual averaging. Specifically for dual averaging, one can show that if this procedure converges to a single point $s \in \mathcal{S}^d$, then this strategy profile is a NE in the approximation game, and thereby a BNE for the discretized auction game (Bichler et al. 2025, Corollary 1). Moreover, for some single-object auction formats, such as first- or second-price sealed-bid and all-pay auctions with quasilinear utility functions, it is shown that if SODA finds an approximate equilibrium of the discretized auction game, this is also an approximate equilibrium of the continuous game (Bichler et al. 2025, Theorem 1). Therefore, SODA provides an ex-post certificate.

ALGORITHM 1: Simultaneous Online Dual Averaging (SODA)

Input: Initial distributional strategy s_1, sequence of step sizes $\{\eta_t\}_{t=1}^{T}$

Initialize $y_{i,t} = 0$ for all $i \in \mathcal{I}$ **for** $t = 1, 2, \ldots, T$ **do**
  for *each agent* $i \in \mathcal{I}$ **do**
    observe gradient $c_{i,t} = \nabla_{s_i} \tilde{u}_i(s_{i,t}, s_{-i,t})$;
    update dual variable $y_{i,t+1} = y_{i,t} + \eta_t c_{i,t}$;
    update strategy $s_{i,t+1} = \arg\min_{s \in \mathcal{S}^d} \|s - y_{i,t+1}\|_2^2$;
  end
end

Note that SODA serves only as a baseline as it allows us to approximate the BNE quickly and with high accuracy. It is not meant as an algorithm to simulate display ad auctions. While bandit algorithms such as Exp3 only require information about the price and whether they won or lost in an auction, in SODA the auctioneer would need all agents' mixed strategies to provide the respective feedback for the agents.

D.2. Approximation of Equilibria for Payoff- and ROI-Maximizing Agents

To verify the computed strategies for payoff- and ROI-maximizing agents, we compare them to the analytical solutions derived for symmetric settings with uniform prior (cf. Section 3). To this end, we report the following metrics:

Discrete Relative Utility Loss (DRUL) ℓ To measure the quality of computed strategies, we use the relative utility loss ℓ in the discretized game, which also serves as stopping criterion for the learning algorithm. Given a distributional strategy s_i, ℓ measures the relative improvement of the expected utility we can get by best responding to the other agents strategies s_{-i}, i.e.,

$$\ell(s_i, s_{-i}) = 1 - \frac{\tilde{u}_i(s_i^{br}, s_{-i})}{\tilde{u}_i(s_i, s_{-i})}. \tag{11}$$

Note that in the discretized game, the best response s_i^{br} is the solution of a simple LP. If the absolute utility loss $\ell \cdot \tilde{u}_i(s_i, s_{-i}) < \varepsilon$ for all agents, the computed strategy profile is a ε-BNE in the discretized auction game. Note that this metric does not depend on the knowledge of the analytical BNE and can be used to check if an approximate BNE in the discrete setting is found.

Relative Utility Loss $\mathcal{L}$ We approximate the expected utility using the sample-mean of the ex-post utilities, i.e., $\tilde{u}_i(\beta_i, \beta^*_{-i}) \approx \hat{u}_i(\beta_i, \beta^*_{-i}) := \frac{1}{n_v} \sum_v u_i(\beta_i(v_i), \beta_{-i}((v_{-i})))$. The relative loss in the expected utility the agent receives when playing the computed strategy s_i instead of the analytical equilibrium β_i^*

$$\mathcal{L}_i(\beta_i) \;=\; 1 - \frac{\hat{u}_i(\beta_i, \beta^*_{-i})}{\hat{u}_i(\beta^*_i, \beta^*_{-i})}, \tag{12}$$

while all other agents play the equilibrium strategy β^*_{-i}.

L_2 Distance The probability-weighted root mean squared error of β_i and β_i^* in the action space, which approximates the L_2 distance of these two functions:

$$L_2(\beta_i) \;=\; \left(\frac{1}{n_{\text{batch}}} \sum_{v_i} \left(\beta_i(v_i) - \beta_i^*(v_i)\right)^2 \right)^{\frac{1}{2}}. \tag{13}$$

This way we ensure that our computed strategy not only performs similarly in terms of utility but also approximates the actual known BNE.

To approximate the utility loss $\mathcal{L}$ and the L_2 distance, we sample valuations (batch size $n_v = 2^{22}$), identify for each valuation the nearest discrete valuation $v \in \mathcal{V}^d$, and then sample a discrete bid $b \in \mathcal{A}^d$ from the computed strategy. The sampled bids are denoted by $\beta_i(v)$ in the definitions above. The same simulation procedure is used to approximate the expected revenue. As Table 1 shows, the computed strategies of the discretized game closely approximate the analytical BNE in the original continuous setting.

Table 1 Evaluation of SODA against Analytical Equilibrium Strategies.

Utility Model	Pricing Rule	Bidders	ℓ	$\mathcal{L}$	L_2
QL	First-Price	2	0.000 (0.000)	0.002 (0.000)	0.010 (0.001)
		3	0.000 (0.000)	0.001 (0.000)	0.010 (0.001)
		5	0.000 (0.000)	0.002 (0.000)	0.023 (0.001)
		10	0.001 (0.000)	0.010 (0.001)	0.084 (0.001)
	Second-Price	2	0.000 (0.000)	0.000 (0.000)	0.013 (0.000)
		3	0.000 (0.000)	0.001 (0.000)	0.014 (0.001)
		5	0.000 (0.000)	0.002 (0.000)	0.028 (0.001)
		10	0.000 (0.000)	0.011 (0.001)	0.092 (0.001)
ROI	First-Price	2	0.000 (0.000)	0.014 (0.001)	0.013 (0.000)
		3	0.000 (0.000)	0.004 (0.000)	0.011 (0.000)
		5	0.000 (0.000)	0.002 (0.000)	0.014 (0.000)
		10	0.001 (0.001)	0.005 (0.001)	0.070 (0.001)
	Second-Price	2	0.000 (0.000)	0.000 (0.000)	0.017 (0.000)
		3	0.000 (0.000)	0.001 (0.000)	0.013 (0.000)
		5	0.000 (0.000)	0.002 (0.000)	0.020 (0.001)
		10	0.000 (0.000)	0.009 (0.000)	0.078 (0.001)

The mean (and standard deviation) of the relative utility loss in the discretized game ℓ and the approximated utility loss $\mathcal{L}$ and the L_2 norm with respect to the analytical BNE are reported for different utility models (payoff-maximizing, ROI-maximizing), payment rule (first- and second-price) and different number of agents ($N \in \{2, 3, 5, 10\}$). In all settings, we assume a uniform prior.

D.3. Additional Analysis for ROS-Maximizing Agents

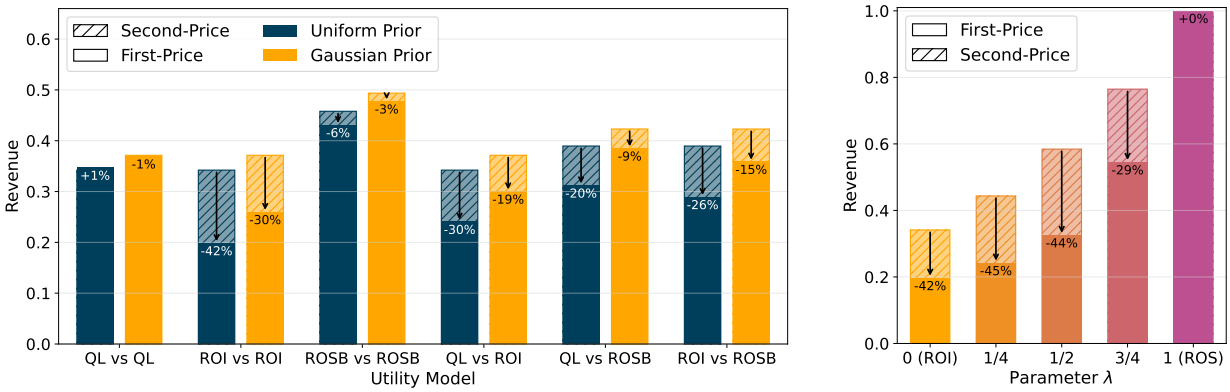
To check for robustness of our results, i.e., if revenue in first-price auctions is also lower compared to the second-price auctions for other utility models agents might have, we again use SODA to compute the equilibrium strategies. We extend our analysis to ROS-maximizing agents (see Appendix C) and consider different variations and asymmetric settings.

First, we consider different combinations of bidders that maximize payoff-, ROI, or ROSB (with budget $B = 1.01$). We observe that while the effects are strongest for symmetric ROI-maximizing agents, a decrease in revenue when switching from first- to second-price auctions can be seeon over all combinations of utility models for uniform and truncated Gaussian prior ($\mu = 0.5, \sigma = 0.3$) distributions (cf. Figure 1a).

In another experiment, we considered a generalized version of the ROI and ROS utility model, where we consider convex combinations of both, i.e.,

$$u_i^{ROIS}(b, v_i) := (1 - \lambda) u_i^{ROI}(b, v_i) + \lambda u_i^{ROS}(b, v_i), \tag{ROIS}$$

Figure 1 Expected Revenue in the FPSB and SPSB Auction under Mixed Utility Models.



(a) Asymmetric Utility Models: QL, ROI and ROSB.

(b) Symmetric Utility Models: ROIS.

Note. We use the computed strategies from SODA to compare the expected revenue between the first- and second-price auction for two agents with different utility models. In the first plot, we consider two asymmetric agents with i.i.d. prior and different combinations of utility models, namely QL, ROI, and ROSB. The budget for ROSB is $B = 1.01$. In the second plot, we consider two symmetric agents with uniform prior (i.i.d.) and utility functions that are convex combinations of ROI and ROS with different parameters.

with some parameter $\lambda \in [0, 1]$. SODA converges in all settings with two symmetric agents and a uniform prior, achieving a utility loss of at most $\ell \leq 10^{-4}$ in the discrete setting. Using the computed equilibria to approximate revenue, we observe trends consistent with our other results: revenue in first-price auctions is lower than in their second-price counterparts. However, this gap decreases as the utility function approaches ROS (Figure 1b). This aligns with our theoretical analysis, as for ROS, bidding the maximum bid constitutes an equilibrium under both first- and second-price payment rules (Proposition 1), resulting in identical revenue.

Appendix E: Auto-bidding Algorithms

As described in Section 4, methods from online optimization, in particular bandit algorithms, can be used as auto-bidding algorithms. We focus on a few well-known bandit learners that handle the exploration-exploitation tradeoff differently (see for instance Sutton and Barto (2018), Lattimore and Szepesvári (2020)).

E.1. Bandit Algorithms

ε-Greedy This policy compares the current average reward of each action a, which is defined by

$$\hat{R}(a) = \frac{1}{N(a_t)} \sum_{s=1}^{t} \chi_a(a_t) r_s(a_t) \tag{14}$$

with $\chi_a(a_t) = 1$ if $a_t = a$ and 0 else. $N_t(a)$ denotes the number an action a has been played up to iteration t. The ε-Greedy selects the action with the highest average reward with probability $1 - \varepsilon_t$ and some random action with probability ε_t ($\varepsilon_t = 0.05$ in our experiments).

Exp3 Similar to the Hedge algorithm (Freund and Schapire 1997), the algorithm maintains a mixed strategy x_t over the action space $\mathcal{A}^d$ and updates the probabilities in a multiplicative manner. The update rule is given by

$$x_{t+1}(b) = \frac{x_t(b) \exp(\alpha R_t(b))}{\sum_{b' \in \mathcal{A}^d} x_t(b') \exp(\alpha R_t(b'))}, \quad \forall b \in \mathcal{A}^d . \tag{15}$$

Since the reward is not available for all actions in the bandit feedback setting, the algorithm uses an importance-weighted estimator defined by

$$R_t(b) = \frac{r_t}{x_t(b)} \text{ for } b = b_t \text{ and } R_t(b) = 0 \text{ for all } b \in \mathcal{A}^d \setminus \{b_t\}, \tag{16}$$

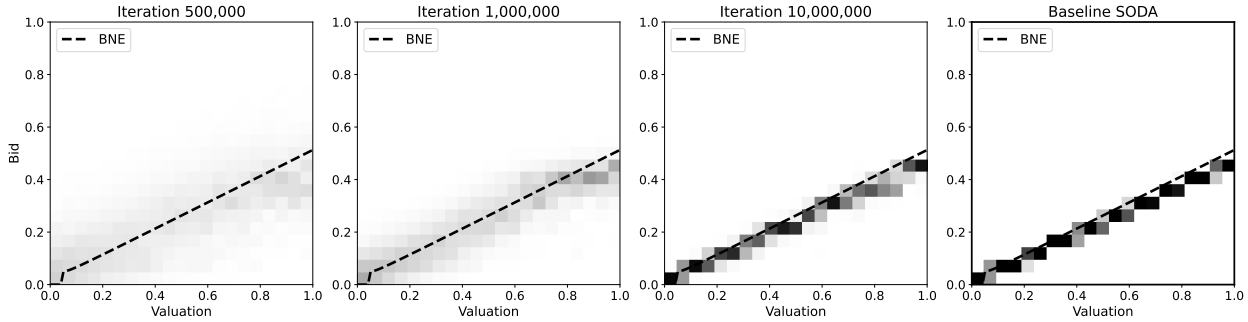
where r_t is the reward of the played action b_t. We use a specific version of Exp3 (Braverman et al. 2017) with an additional exploration parameter. This means, in each round the action b_t is sampled from the probability $P_t := (1-\varepsilon)x_t + \varepsilon \frac{1}{m} \mathbf{1}_m$. This version can also be found in the bandit version of the online stochastic mirror descent from Lattimore and Szepesvári (2020, Sec. 31). It is well-known that the Exp3 algorithm is a no-regret learner for $\varepsilon = \alpha \sim T^{-\beta}$ with $\beta \in (0,1)$ (Auer et al. 2002). For our experiments, we use a fixed exploration rate of $\varepsilon = 0.01$ and a fixed learning rate of $\alpha = 0.01$ for QL and $\alpha = 0.0005$ for ROI (as the magnitude of the utilities is larger for ROI).

Thompson-Sampling Finally, we also use a Bayesian method first introduced in (Thompson 1933). Thompson sampling assumes a distribution $\mathcal{P}_r(\theta_a)$ (with some parameter θ_a) over the rewards of each action. In each iteration t, the learners sample parameters $\theta_a^t \sim \mathcal{P}_\theta$ from some parameter distribution for each action a, play the action a^* that maximizes the expected reward given the distribution $\mathcal{P}_r(\theta_a^t)$, and update the parameter distribution $\mathcal{P}_\theta$ for action a^* given the observed reward $r_t(a^*)$. In our implementation, we assume that the rewards for each action follow a Gaussian distribution (with known variance σ) and thereby use a Gaussian prior for the parameter $\theta_a = \mu_a$ (conjugate prior).

E.2. Results of Simulations

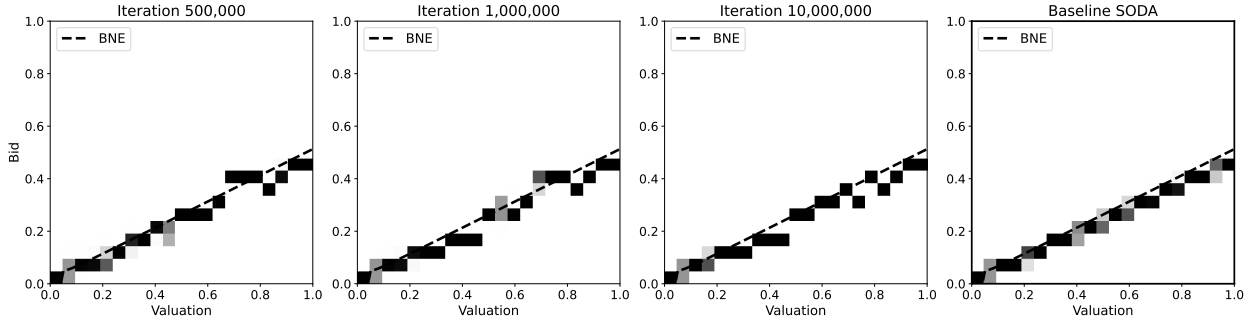
In this section, the simulation results from Section 4, are presented in more detail. In addition to the revenue, we report the relative utility loss, which allows us to evaluate if the played actions are close to an approximate Bayes-Nash equilibrium. Specifically, after 10^7 learning iterations, we examine the final 50 000 iterations, recording each action along with its respective valuation to construct a distributional strategy (see Appendix D.1). The empirical distributional strategies for different learning algorithms after $5 \cdot 10^5$, 10^6, and 10^7 iteratoins are visualized in Figures 2 and 3. Both algorithms adopt strategies that closely align with the equilibrium computed by SODA, though the convergence speed may vary. Notably, Thompson Sampling approaches the BNE almost immediately, reaching near-equilibrium performance after just a few hundred thousand iterations.

Figure 2 Empirical Strategies for ROI-Maximizing Agents using Exp3 in a FPSB Auction.



Note. We run Exp3 for all three ROI-maximizing agent in a FPSB auction for 10 million iterations and visualize the frequency of the last 50 000 bids of for the respective valuations, i.e., the induced distributional strategies after $5 \cdot 10^5$, 10^6, and 10^7 iterations of one agent. On the last plot, we show the distributional equilibrium strategy computed with SODA. The colored lines denote the BNE in the continuous setting.

Figure 3 Empirical Strategies for ROI-Maximizing Agents using Thompson Sampling in a FPSB Auction.



Note. We run Thompson Sampling for all three ROI-maximizing agent in a FPSB auction for 10 million iterations and visualize the frequency of the last 50 000 bids of for the respective valuations, i.e., the induced distributional strategies after $5 \cdot 10^5$, 10^6, and 10^7 iterations of one agent. On the last plot, we show the distributional equilibrium strategy computed with SODA. The lines denote the BNE in the continuous setting.

Based on the constructed distributional strategy profiles, we can also compute the discrete relative utility loss ℓ, representing the relative performance improvement achievable if the agent were to best respond to the opponents' strategies. We compare the results with the approximated equilibrium strategies using SODA (with the standard projected gradient update step due to better performance). The setting is identical to the one described in Section 4. The results for symmetric agents, i.e., all three agents use the same bandit algorithm, can be found in Table 2, while the results for asymmetric agents, where different bandit algorithms compete against each other, cen be found in Table 3.

Table 2 Evaluation of Symmetric Bandit Algorithms

Utility Model	Pricing Rule	Algorithm	DRUL ℓ	Revenue
QL	First-Price	SODA	< 0.001	0.496
		ε-Greedy	0.132 (0.008)	0.509 (0.006)
		Exp3	0.027 (0.002)	0.497 (0.003)
		Thompson Sampl.	0.003 (0.000)	0.495 (0.001)
	Second-Price	SODA	< 0.001	0.497
		ε-Greedy	0.053 (0.003)	0.496 (0.004)
		Exp3	0.020 (0.004)	0.496 (0.002)
		Thompson Sampl.	0.007 (0.001)	0.497 (0.002)
ROI	First-Price	SODA	< 0.001	0.344
		ε-Greedy	0.116 (0.034)	0.355 (0.016)
		Exp3	0.026 (0.005)	0.350 (0.003)
		Thompson Sampl.	0.067 (0.025)	0.349 (0.009)
	Second-Price	SODA	< 0.001	0.502
		ε-Greedy	0.032 (0.004)	0.499 (0.006)
		Exp3	0.030 (0.006)	0.486 (0.008)
		Thompson Sampl.	0.015 (0.008)	0.489 (0.009)

The mean (and standard deviation) of the discrete relative utility loss (DRUL) ℓ and the average revenue for the agents in the discrete auction game are reported for settings with $N = 3$ agents using the same algorithm, uniform prior, different utility models (QL and ROI), and different payment rules (first- and second-price). We report the mean and standard deviation over all agents and 10 runs for each experiment

For a better interpretation of the values: If agents with quasilinear utility function would collude and bid only half of the amount they would bid in equilibrium, i.e., $\frac{1}{3}v$, the relative utility loss would be at $\ell = 0.20$ with expected revenue of 0.25. The relatively high numbers for DRUL for ε-Greedy are due to the constant exploration factor ε. Other than that, we observe that bandit algorithms such as Exp3 and Thompson-Sampling perform very well and learn to play an approximate Bayes-Nash equilibrium with payoff- and ROI-maximizing agents.

Table 3 Evaluation of Asymmetric Bandit Algorithms

Utility Model	Pricing Rule	Agent	Algorithm	DRUL ℓ	Revenue
QL	First-Price	-	SODA	< 0.001	0.496
		1	ε-Greedy	0.121 (0.004)	0.500 (0.002)
		2	Exp3	0.027 (0.003)	-
		3	Thompson Sampl.	0.006 (0.002)	-
	Second-Price	-	SODA	< 0.001	0.497
		1	ε-Greedy	0.055 (0.002)	0.497 (0.002)
		2	Exp3	0.020 (0.004)	-
		3	Thompson Sampl.	0.007 (0.001)	-
ROI	First-Price	-	SODA	< 0.001	0.344
		1	ε-Greedy	0.103 (0.023)	0.363 (0.011)
		2	Exp3	0.028 (0.008)	-
		3	Thompson Sampl.	0.184 (0.079)	-
	Second-Price	-	SODA	< 0.001	0.502
		1	ε-Greedy	0.033 (0.002)	0.490 (0.006)
		2	Exp3	0.035 (0.013)	-
		3	Thompson Sampl.	0.016 (0.006)	-

The mean (and standard deviation) of the discrete relative utility loss (DRUL) ℓ and the average revenue for the agents in the discrete auction game are reported for settings with with $N = 3$ asymmetric agents, i.e., different algorithms, uniform prior, different utility models (QL and ROI), and different payment rules (first- and second-price). We report the mean and standard deviation over 10 runs for each experiment

References

Auer P, Cesa-Bianchi N, Freund Y, Schapire RE (2002) The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing* 32(1):48–77, URL http://dx.doi.org/10.1137/S0097539701398375.

Balseiro SR, Gur Y (2019) Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science* 65(9):3952–3968.

Bichler M, Fichtl M, Oberlechner M (2025) Computing bayes–nash equilibrium strategies in auction games via simultaneous online dual averaging. *Operations Research* 73(2):1102–1127, URL http://dx.doi.org/10.1287/opre.2022.0287.

Braverman M, Mao J, Schneider J, Weinberg SM (2017) Selling to a no-regret buyer.

Fikioris G, Tardos E (2022) Liquid welfare guarantees for no-regret learning in sequential budgeted auctions. URL http://dx.doi.org/10.48550/ARXIV.2210.07502.

Frank M, Wolfe P (1956) An algorithm for quadratic programming. *Naval Research Logistics Quarterly* 3(1-2):95–110, ISSN 1931-9193.

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1):119–139.

Lattimore T, Szepesvári C (2020) *Bandit algorithms* (Cambridge University Press).

Milgrom PR, Weber RJ (1985) Distributional strategies for games with incomplete information. *Mathematics of Operations Research* 10(4):619–632, ISSN 0364765X, 15265471.

Nemirovskij AS, Yudin DB (1983) Problem complexity and method efficiency in optimization .

Nesterov Y (2009) Primal-dual subgradient methods for convex problems. *Mathematical programming* 120(1):221–259.

Sutton RS, Barto AG (2018) *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning Series (Cambridge, Massachusetts: The MIT Press), second edition edition, ISBN 978-0-262-03924-6.

Thompson WR (1933) On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25(3/4):285–294, ISSN 00063444.

Tunuguntla S, Hoban PR (2021) A near-optimal bidding strategy for real-time display advertising auctions. *Journal of Marketing Research* 58(1):1–21.