

Behavioral Study of Dashboard Mechanisms^{*}

Paula Kayongo¹, Jessica Hullman¹, and Jason Hartline¹

Department of Computer Science, Northwestern University, 1856 Sheridan Rd,
Evanston, IL 60201, USA
paulakayongo2023@u.northwestern.edu, jhullman@northwestern.edu,
hartline@northwestern.edu

Abstract. Visualization dashboards are increasingly used in strategic settings like auctions to enhance decision-making and reduce strategic confusion. This paper presents behavioral experiments evaluating how different dashboard designs affect bid optimization in reverse first-price auctions. Additionally, we assess how dashboard designs impact the auction designer’s ability to accurately infer bidders’ preferences within the dashboard mechanism framework. We compare visualizations of the bid allocation rule—commonly deployed in practice—to alternatives that display expected utility. We find that utility-based visualizations significantly improve bidding by reducing cognitive demands on bidders. However, even with improved dashboards, bidders systematically under-shade their bids, driven by an implicit preference for certain wins in uncertain settings. This limitation bears directly on the feasibility of dashboard-based mechanisms, which rely on rational response assumptions for inference and efficient allocation. Given persistent undershading, dashboard mechanisms that assume fully rational or risk-neutral bidder responses to dashboards can produce significant estimation errors when inferring private preferences, which may lead to suboptimal allocations in practice. Explicitly modeling agents’ behavioral responses to dashboards substantially improves inference accuracy, highlighting the need to align visualization design and econometric inference assumptions in practice.

1 Introduction

A key challenge to obtaining good outcomes in strategic decision-making settings, e.g., online auctions [4] and school matching markets [32, 29, 10] arises from the fact that decision-makers (*henceforth agents*) often struggle to identify effective decision-making strategies. To encourage more efficient strategic behavior that leads to improved resource allocation, market designers (*henceforth principals*) commonly deploy *visualization dashboards* that communicate payoff-relevant market information to the agents. For instance, the Google AdWords auction dashboard visualizes a line chart that depicts predicted click volume and cost per click as a function of advertiser bids. This paper conducts behavioral

^{*} Full version of this paper, which includes an appendix, is available at <https://arxiv.org/abs/2507.20985>.

studies to experimentally evaluate different dashboard designs and their efficacy at inducing more optimal bidding in first-price single-item winner-pays-bid auctions.

Systematic deviations from optimal bidding (i.e., the exact right amount of bid shading) in first-price auctions have been extensively documented in the literature. A classic finding by Cox et al. [4] is that bidders in first-price auctions often undershade, submitting bids that exceed those expected under risk-neutral Bayes-Nash equilibrium. Undershading in first-price auctions has been attributed to unusual mental models [11, 6, 27], computational constraints [9, 24], and unusual preferences [5]. In first-price auctions, the bidding process involves several steps, including understanding the competitors’ value distribution, deducing how it leads to a specific bidding pattern, and how this pattern determines the allocation rule, and finally, optimizing one’s bid based on this rule. Compared to this extensive behavioral literature on bidding in auctions, our study removes the problem of reasoning about how other bidders’ strategize and asks whether bidders can respond optimally given fixed competition and what format of the visual depictions of competition should take to best aid in good bidding.

A recent line of work [15] demonstrates that good mechanisms can be designed around the idea that bidders respond optimally to dashboards. Specifically, the *dashboard mechanism* of Hartline et al. [15] and Deng et al. [8], incentivize bidders to rationally optimize according to the dashboard and efficiently allocate resources when they do. The dashboard mechanism approach can be viewed as a reduction from a many-agent allocation problem to the problem of each agent bidding against a single opponent with a known bid distribution, which is visualized in the dashboard. A critical step in dashboard mechanisms is the use of econometric inference to infer the bidders’ private values given their bids. Hartline et al. [15] and Deng et al. [8] assume rational agents that perfectly optimize their bids against the dashboard when conducting this econometric inference, and therefore, the bidders’ values can be exactly inferred. One goal of this paper is to understand how behavioral bidders deviate from the rational bidding assumption and the extent to which these deviations affect bid inference in dashboard mechanism design.

To evaluate how dashboard design influences bidding behavior and the practical feasibility of dashboard mechanisms, we contribute the results of a large-scale reverse first-price experimental auction. The reverse auction differs from the forward auction in that, while in forward auctions, buyers with private values bid to acquire items, in reverse auctions, sellers with private costs bid to provide goods.¹ In each auction round, we provide participants with payoff-relevant visualizations, solicit bids, determine outcomes, and provide feedback. We vary three visualization types between subjects, chosen to vary in the extent to which they support direct reasoning about bids conditional on endowed costs (see Figure 1):

¹ In auction experiments like the one we conduct, to establish normative behavior, participants need to be endowed with values or costs. We endow costs (of sellers in a reverse auction) because it is more natural to convey to study participants.

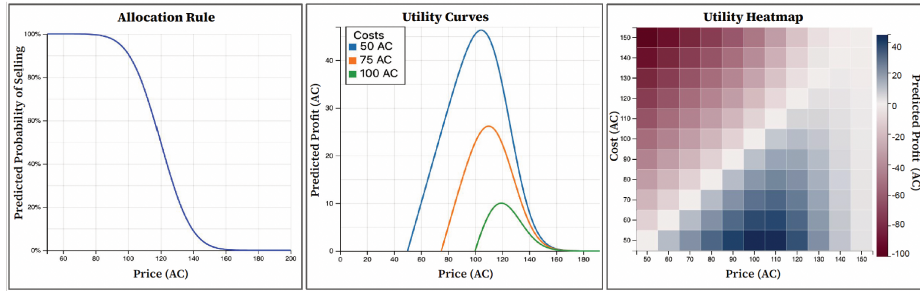


Fig. 1. Visualization designs evaluated in behavioral experiments

Allocation Rule depicts the probability of selling an item as a function of the seller’s bid. This visualization is similar to the one deployed in practice in dashboards such as Google Adwords.

Utility Curves depicts the expected payoff for three hypothetical sellers with different costs as a function of the seller’s bid.

Utility Heatmap depicts the expected payoff on an eleven-by-eleven grid of hypothetical seller costs and bids.

We hypothesize that visualizations presenting bid-relevant information in payoff space (Utility Curves and Utility Heatmap) will improve bid optimality for boundedly rational agents by reducing the cognitive effort needed to identify the best response. We also examine how two types of feedback influence bidding: (1) feedback on the bidder’s payoff, and (2) feedback on inferred costs assuming utility-maximizing behavior. We expect that inferred cost feedback will further improve bidding by helping participants recognize when they have over- or underbid.

We observe statistically significant variation in bid optimization across visualization types despite persistent undershading. Participants using the Allocation Rule visualization achieved the lowest bid optimization accuracy (ratio between the participant’s utility given their bid and the optimal utility), while the Utility Heatmap performed best. Utility Curves yielded slightly lower bid optimization accuracy than Heatmaps, likely due to fewer reference points in payoff space, though the difference is not significant. Even when most of the computational complexities in determining optimal bidding strategies are removed, participants consistently undershade,²

Further analysis reveals that many participants submitted bids that were effectively insensitive to the competitive environment. Across experimental conditions, their bids on average clustered near their private cost, with little adjustment in response to differences in the implied win probability or expected payoff. This pattern suggests that participants relied on simple heuristics rather than engaging with the information in the dashboard. Participants’ self-reported

² In the reverse auction we study they underbid this is consistent with prior findings in forward auctions showing persistent overbidding.

strategies further support this interpretation; several participants reported an implicit preference for wins, which were guaranteed by bidding closer to their costs.

To better understand whether fixed bidding reflects difficulty interpreting the dashboards or a deeper behavioral tendency, we conducted a follow-up study that introduced training in a deterministic environment where outcomes could be verified. If undershading was the result of limited visualization literacy, then training effects should carry over to the setting with stochastic payoffs. If, however, undershading reflects a preference for win certainty, then gains from training should diminish once uncertainty is reintroduced. We find support for the latter: training improves bid optimization when payoffs are deterministic, but performance reverts to baseline when outcomes are stochastic.

These behavioral findings motivate our contributions to the accurate inference of agent preferences within dashboard mechanisms. We show that the accuracy of inferring bidders' preferences is enhanced by employing a behavioral model that accommodates imprecise optimization and undershading. Specifically, the quantal response model, which relaxes the perfect rationality assumption, typically addresses the former, while the risk aversion model, which accounts for bidder preference for more certain outcomes, often accounts for the latter, reducing mean squared error in estimated costs by an order of magnitude. We do not have a stance on whether these models are behaviorally correct; instead, we focus on the improvements to inference that they enable.

2 Related Work

2.1 Behavioral Studies of Auctions

Research in auction design has traditionally focused on identifying mechanisms that achieve desirable outcomes when bidders employ optimal strategies. However, there is limited research on how to assist decision-makers in the real world in determining effective bidding strategies. Behavioral studies of first-price auctions show that humans experience difficulty choosing actions that optimize their utility and consistently overbid [4, 10, 31, 19, 16]. To simplify strategic reasoning under uncertainty, dashboard mechanisms propose visualizing the bid allocation rule to improve bidding behavior in auctions that require complex reasoning under uncertainty. If agents respond rationally, dashboard mechanisms offer provable allocation and welfare guarantees. This motivates our investigation into how providing bid-relevant visualizations influences bidders' behavior and inference accuracy in strategic mechanisms. While prior work has shown that boundedly rational agents continue to struggle with bid optimization when presented with probabilistic information [19], whether improved dashboard designs can elicit more optimal bidding remains an open question.

2.2 Visualizations for Decisions and Inference

Visualization research on decision-making under uncertainty has explored ways of visualizing distributional information that make variation and uncertainty

more salient to users, e.g., through the use of frequency representations like hypothetical outcome plots [17, 34] or quantile dot plots [12, 22]. However, Hullman et al. [18] argue that even when individuals accurately perceive visualizations of probabilistic prediction, like those depicting the allocation rule, they do not necessarily make optimal decisions. While many evaluations have tended to focus on lower-level perceptual tasks, i.e., probability perception, rather than decision quality in an incentivized task [18], more recent work has studied the role of visualization in aiding decision-making using incentivized decision-making [e.g, 12, 21, 1] and strategic-making decision tasks [e.g, 35, 23] tasks. Our work goes a step further, studying a case where visualization-aided decisions are used for downstream inference in a mechanism.

3 Preliminaries

We study a single-item reverse auction. In each auction round, the participant endowed with a private cost and a computer opponent simultaneously bid to sell a single item. The participant submits a bid from a set of feasible bids ($b \in B$), while the opponent submits a bid randomly drawn from a normal distribution $x \sim \mathcal{N}(\mu, \sigma^2)$. The seller with the lowest bid wins and is paid their bid. By fixing the computerized agents' bid distribution, we isolate potential errors in bidding to the bid optimization of the agent. This allows us to exclude other sources of bidding error, such as errors in estimating the strategies of other agents, which have been extensively studied in the literature on strategic bidding in auctions (see Crawford et al. [7] for a summary).

3.1 Best Response Inference

The participant best responds by submitting a bid that maximizes their utility given by:

$$\arg \max_{b \in \mathbf{B}} \pi(b) = (b - c) \cdot Pr(win|bid = b) \quad (1)$$

where c is the sellers participants cost and b is their bid. To infer a participant's cost, we perform a computational grid search that inverts their submitted bid. We construct the computation grid by discretizing the bid and cost space at a resolution of 0.01 and compute the best response bid corresponding to each cost. The inferred cost is then identified as the value whose best response most closely matches the participant's submitted bid.

3.2 Quantal Response Inference

To improve cost inference, we use the logit-based quantal response model, where bidders choose bids probabilistically based on expected utility. Given a seller with cost c and allocation rule a , the probability of choosing bid b is:

$$Pr(b|s) = \frac{e^{\lambda\pi(b|a)}}{\sum_{\alpha \in B} e^{\lambda\pi(\alpha|a)}} \quad (2)$$

where B is the set of feasible bids, $\pi(b|a)$ is the expected utility, and λ is a precision parameter. As $\lambda \rightarrow \infty$, bids converge to best responses; as $\lambda \rightarrow 0$, bids are random. We jointly estimate λ and cost via maximum likelihood estimation.

Quantal Response with Constant Relative Risk Aversion We extend the quantal response model to incorporate risk aversion using a constant relative risk aversion (CRRA) utility function:

$$\pi_{\text{CRRA}} = \frac{(b - c)^{1-r}}{1 - r} \cdot Pr(\text{win}|\text{bid} = b) \quad (3)$$

where r is the Arrow Pratt coefficient typically estimated to be between 0.3 to 0.7 across numerous auction experiments [13, 3, 5, 14]. To avoid overfitting, we estimate a single r across all participants rather than allowing individual variation³. We estimate the quantal precision $\hat{\lambda}$ and cost \hat{c} via maximum likelihood [25], minimizing the squared error between observed and predicted bid probabilities estimated by sampling from a model fitted to agent bids.⁴

4 Experiment 1: Baseline Dashboard Evaluation

4.1 Design

We conduct a mixed-design repeated-measures experimental auction to assess how dashboard visualization and feedback design influence bid optimization and inference. Participants, each endowed with the same private cost (85 AC), bid against a computer opponent across multiple rounds. In each round, they view a bid-relevant visualization, submit a bid and receive feedback. We vary three visualization types (Fig. 1) between subjects, two feedback conditions within subjects, and ten allocation rules within subjects. To guarantee that the ten allocation rules were invertible, we constructed them by sampling competitor bids from normal distributions with parameters that varied across rules.

Visualization We vary three visualization conditions between subjects: Allocation Rules, Utility Curves, and Utility Heatmaps (Fig. 1).

Allocation Rule visualizes the $Pr(\text{win}|\text{bid} = b)$ based on the computer opponent’s Cumulative Distribution Function ($F_B(b)$). Identifying the optimal bid given the allocation rule requires participants to compute their utility for each bid, given the endowed cost, and select the bid that maximizes their utility

³ Estimating individual r values yields only marginal improvements, likely due to widespread underbidding and limited data per participant.

⁴ Parameters are estimated using the L-BFGS-B algorithm in R.

in expectation. A task that may be cognitively demanding task for behavioral agents.

Utility Curves visualize the expected utility for hypothetical sellers with different costs ($\{50, 75, 100\}$) at given bids (Fig. 1). Participants endowed with a cost of 85 AC can use the visualization to approximate an optimal bid between the utility-maximizing bids for hypothetical sellers with costs of 75 AC and 100 AC.

Utility Heatmap visualizes the expected utility for hypothetical sellers with different costs at given bids using color encoding in a data matrix. To generate the heatmap, we discretize the continuous bid and cost space into fixed intervals $\{50, 60, 70, \dots, 150\}$. The smaller interval size in the Utility Heatmap visualizes the expected utility at a higher level of granularity than the Utility Curves while not directly depicting the optimal bid for the participant’s endowed cost. Ideally, given the heatmap, participants with an endowed cost of 85 AC submit a bid between the utility optimal bids of sellers with costs of 80 AC and 90 AC (Fig. 1).

We hypothesize that, relative to CDFs, the Utility Curves and Utility Heatmap visualizations will improve bid optimization by presenting bid-relevant information in utility space, reducing the cognitive demands of bid optimization to a simpler approximation task. Additionally, we hypothesize that relative to the Utility Curves, the Utility Heatmap will enable participants to approximate the best response bid more accurately since it provides closer reference points to their endowed cost.

Feedback Feedback has been shown to have a significant behavioral effect on bidding in First Price Auctions (FPAs), especially when it clarifies foregone payoffs [28, 26]. Prior work also shows that presenting probabilities alone does not improve bidding without informative feedback [30]. Based on these findings, we test two feedback conditions within subjects. In one condition, participants receive feedback on whether they won and their payoff if they win. In the second condition, they also receive feedback on their inferred cost, assuming a best response. The role of the second condition is to determine whether participants’ ability to best respond improves when they receive feedback on the optimality of their bids. An inferred cost higher (resp. lower) than the endowed cost implies overbidding (resp. underbidding).

Experimental Procedure Participants were randomly assigned to one of the three visualization conditions and completed a tutorial on interpreting their assigned dashboard, followed by four short training tasks. They then completed two blocks of ten randomized auction rounds, differing only in feedback condition, with block order counterbalanced across participants. Each round presented a dashboard visualizing bid-relevant information and allowed free entry of bids in auction coins (AC). On the final trial of each block, participants briefly explained their bidding rationale.

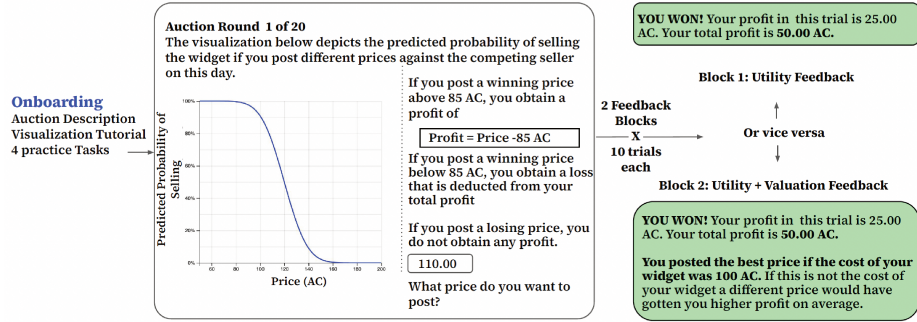


Fig. 2. Experimental Auction Study Overview and Interface

Participants were randomly assigned to a visualization condition and introduced to the auction scenario and bonus structure⁵. They then completed two blocks of ten randomized auction rounds, with feedback conditions varied within the blocks and block order counterbalanced across participants to control for order effects. Each auction round contains a bidding dashboard with a visualization of bid-relevant information. The data visualized for participants is framed as a prediction of the expected outcome based on past observations of the opponent’s bidding behavior. Participants can submit bids in auction coin (AC) units with no minimum⁶. On the last trial in each block, we ask participants to provide a rationale for their submitted bid.

4.2 Analysis

We quantify the effect of dashboard design on bid optimality using the bid optimization ratio:

$$\text{Bid Optimization Ratio} = \frac{\pi_{bid}}{\pi_{OPT}}$$

where π_{bid} is the participant’s expected utility given their bid, and π_{OPT} is the expected utility of the optimal (best response) bid for their endowed cost.

⁵ The bonus was selected based on Wu et al. [33] that proposes that the incentive to consult a visualization in experimental decision-making tasks should be quantified as the difference between the optimal utility obtained given access to visualization $\pi(b_{opt})$ and the baseline utility obtained by a Bayesian rational agent who only has prior information on the distribution of the allocation rules $\pi(b_{prior})$. We incentivize agents to use the visualization by setting payments such that the rational optimal behavior pays 35% more than the baseline behavior.

⁶ Although bids below cost were allowed to measure task misunderstanding, we exclude bids yielding negative utility from the analysis. Replicating findings from previous first-price auctions studies [20] bid resulting in negative utility occur in roughly 10% of trials.

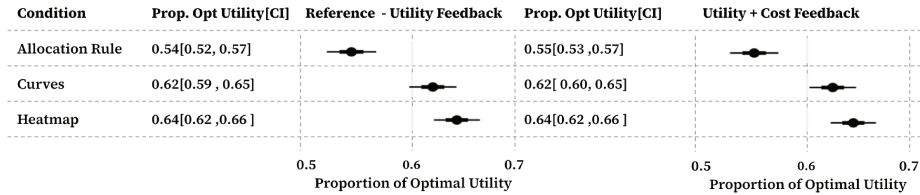


Fig. 3. Model estimates of the proportion of optimal utility by visualization and feedback condition (Utility: left; Utility + Cost: right). Dots denote means; lines indicate 95% credible intervals.

We fit the following preregistered⁷ Bayesian hierarchical model to estimate the effects of experimental manipulations on bid accuracy, quantified through the bid optimization ratio:

$$\begin{aligned}
 \text{Bid Optimization Ratio} = & \text{vis} \cdot \text{feedback} + \text{best response} + \text{block order} + \text{trialnum} \\
 & + (1 + \text{trialnum} \mid \text{participant})
 \end{aligned}
 \tag{4}$$

where *vis* is the visualization condition, *feedback* is an indicator variable indicating the feedback type, *best response* is the optimal bid given the allocation rule in the auction round, *block order* is an indicator variable for the order of the feedback blocks, *trialnum* is an index of the trial number which accounts for learning effects, and *participant* is a unique participant id for modeling individual random effects. We include an interaction term (*vis* * *feedback*) to measure if the bid optimization ratio varies as a joint function of the visualization and the feedback condition. We use a zero-one-inflated beta (ZOIB) prior [2] as the bid optimization ratio is bounded between 0 and 1 and frequently takes on exact values at both extremes. Effects are evaluated via posterior credible intervals (CIs); non-overlapping CIs between conditions indicate reliable differences. We aimed to recruit ~ 700 participants to ensure 80% power to detect a 3% difference in bid optimization between visualization conditions.

4.3 Results

Data Preliminaries We recruited 704 participants from Prolific⁸. The median completion time was approximately 13 minutes, with an average base payment of \$18.77 per hour (excluding bonuses).

Bid Optimization We report the conditional means derived from the model estimates in Figure 3. These conditional means communicate the average bid optimization ratio for visualizations and feedback conditions. Relative to the baseline Allocation rule visualization, the Utility Heatmap and Utility Curves led

⁷ Preregistration https://aspredicted.org/8WZ_JBT

⁸ <https://www.prolific.co>

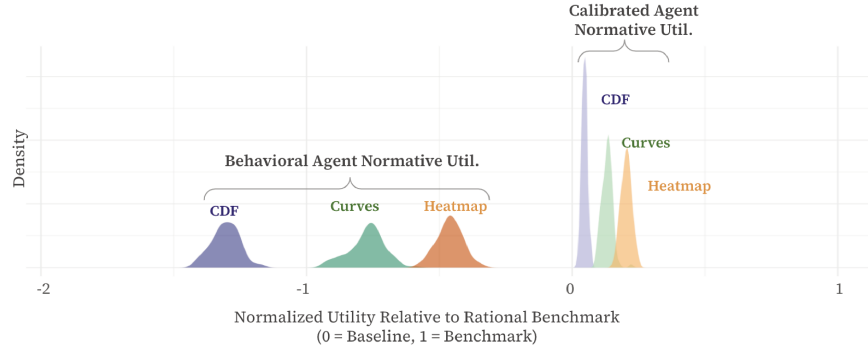


Fig. 4. Normalized utility scores by visualization condition; 0: baseline (prior-only) agent, 1: benchmark agent.

to significantly higher average bid optimization ratios. However, the 95% Confidence Intervals for the Utility Curves and the Utility Heatmap visualization conditions overlap, indicating that the improvement to bid optimization of the Heatmap over Utility Curves is not statistically significant. Although we hypothesized that providing feedback on inferred costs would improve bidding, it did not result in a statistically significant increase in the bid optimization ratio when compared to the baseline utility feedback condition. The absence of an inferred cost feedback effect suggests that the information it provides was dominated by the dashboard’s provision of this information. Despite our initial hypothesis that this feedback condition might address overbidding, typical in first-price auctions, we find that it does not. Additionally, even with the best performing dashboard, participants only obtain $\sim 64\%$ of the optimal utility suggesting that dashboard design choices do not fully eliminate suboptimal behavior, which may reflect not only cognitive limitations but also bidder preferences. Participants’ self-reported strategies suggest that some undershade—submit lower than optimal bids—intentionally to increase their chances of winning. Notably, a subset of participants using the Utility Curves and Utility Heatmap visualizations reported recognizing the optimal bid but choosing to submit lower bids anyway, indicating an awareness of their own deviation from optimal behavior.

Post Hoc Analysis: Bayesian Rational Agent Analysis To evaluate participant performance, we apply the *Bayesian rational agent framework* introduced by Wu et al. [33], which defines a set of normative and diagnostic reference points for decision-making with visualization under uncertainty. The framework compares observed behavior against a family of hypothetical agents who differ in what information they access and how optimally they respond. These comparisons help disentangle whether suboptimal behavior arises from poor information use or imprecise decision-making.

Specifically, we compare participants’ behavioral outcomes against three hypothetical rational agents: the *baseline rational agent*, the *benchmark rational agent*, and the *calibrated rational agent*. The baseline agent optimizes bids using only the prior distribution over competitor bids. This agent serves as a normative lower bound on rational, utility-maximizing behavior without dashboards. The benchmark rational agent represents an idealized upper bound: a Bayesian agent who observes the dashboard visualization, correctly interprets its informational content, and selects the bid that maximizes expected utility given the visualized distribution over competitor bids. The calibrated rational agent optimally responds not to the true state but to the belief distribution implicitly revealed by participants’ observed bids. In other words, we treat each participant’s bid as if it reflects a belief about the underlying competitive environment (e.g., win probability), and we ask: “What payoff would a rational agent achieve if they held that belief and bid optimally?”. The calibrated agents’ payoffs always lie between the performance of the baseline agent and the benchmark agent. The calibrated agent allows us to assess whether participants are meaningfully using the information provided by the dashboard. If behavioral performance is poor but calibrated performance is high, it suggests participants extracted useful information but failed to optimize. If both are low, it indicates a failure to incorporate or interpret the information altogether. In Wu et al. [33], the performance of the benchmark rational agent is normalized to 1, the performance of the baseline agent is normalized to 0, and the performance of all other agents (behavioral, calibrated) is expressed relative to that benchmark.

Figure 4 depicts the normalized score for behavioral and calibrated agents across visualization conditions. We find that participants underperform relative to the baseline Bayesian rational agent (score normalized to 0) in all visualization conditions. Moreover, since participants bid in a manner that is insensitive to the signal, the scores achieved by the calibrated agents—who optimally act on beliefs implicit in participant bids—only marginally exceed those of the baseline agent. This suggests that participants often fail to extract or utilize meaningful information from the dashboard. Together, these results raise doubts about the effectiveness of dashboards alone in improving strategic behavior.

Inference Figure 5 depicts the distribution of inferred costs for different visualizations and models and the corresponding Mean Square Error (MSE) point estimates. Under the best response assumption, inferred costs systematically underestimate the true endowed cost (85 AC) and exhibit high variance, resulting in the highest MSE values. The risk-neutral quantal response model, which relaxes the assumption of perfectly optimal bidding, reduces estimation variance and achieves lower MSE relative to the best response model. However, it still underestimates true costs, as it does not fully account for systematic underbidding, which is observed in approximately 80% of trials. Incorporating risk aversion into the quantal response model yields substantially more accurate cost estimates with lower variance, reducing MSE by an order of magnitude relative to best response inference. Nonetheless, even the risk-averse quantal response

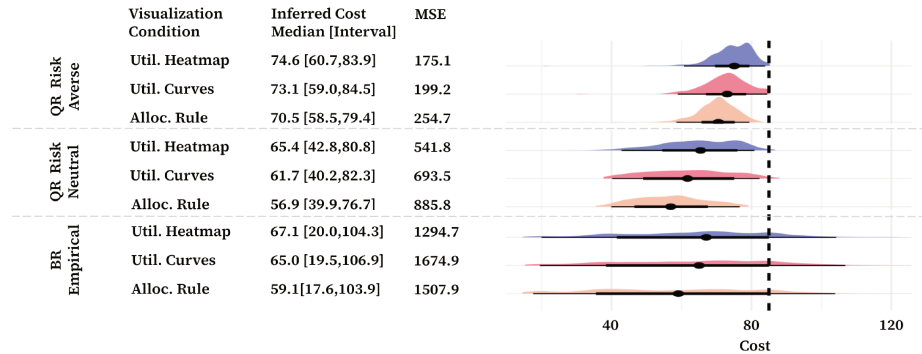


Fig. 5. Distribution of inferred costs across behavioral models (BR: Best Response; QR: Quantal Response) relative true endowed cost (dotted line).

model continues to underestimate costs, albeit to a lesser extent than both the best response and risk-neutral quantal response models. Across all models, visualizations that supported more accurate bidding, namely, the Utility Curves and Utility Heatmap, consistently produced the lowest inference error, underscoring the complementary roles of effective visualization design and behaviorally informed inference models.

5 Experiment 2: Deterministic Training and Stochastic Transfer

To isolate the sources of suboptimal bidding observed in Experiment 1, we run a follow-up experiment that tests whether training with verifiable dashboards in a deterministic environment improves bidding under uncertainty. To allow comparison with Experiment 1, participants again complete 20 reverse auction trials (two blocks of 10) using the same allocation rules and fixed cost (85 AC). However, three key changes are introduced: (1) feedback is held constant—all participants receive outcome and payoff feedback; (2) the first block features training trials with deterministic payoffs, allowing participants to verify the dashboard’s accuracy, while the second block reintroduces stochastic payoffs to assess transfer; and (3) visualization type is manipulated only during training. Participants were randomized to one of three training conditions:

Deterministic True-Cost Utility Curve: visualized a utility curve for the participant’s cost (85 AC) in all 10 training trials.

Deterministic Utility Curves: visualized utility curves for hypothetical sellers with different costs as a function of the seller’s bid (as in Experiment 1) in all 10 training trials.

Deterministic Combined True-Cost Curve and Hypothetical Curves: visualized the True-Cost Curve for the first 5 trials and the Hypothetical Seller Utility Curves for the remaining 5.

In all trials with stochastic payoffs, participants view utility curves for hypothetical bidders, which improved bidding in Experiment 1 and generalizes naturally from the curve for the participants’ true cost used in the training.

Beyond testing whether training with dashboards in deterministic settings improves bidding under uncertainty, Experiment 2 isolated additional sources of suboptimality. Persistent undershading in the deterministic setting after exposure to the True-Cost Utility Curve may indicate failures in learning or attention, while undershading with Approximate Utility Curves in deterministic settings suggests difficulty identifying an optimal bid from inexact stimuli. Finally, differences in performance between deterministic and stochastic blocks may indicate the impact of uncertainty on bidder behavior.

We fit a Bayesian hierarchical model to estimate the effects of training on bid accuracy, measured by the bid optimization ratio. The model includes fixed effects for the interaction between training phase and visualization type, with a random intercept for participants. The model is specified as follows:

$$\text{Bid Optimization Ratio} \sim \text{training} * \text{vis} + (1|\text{participant}) \quad (5)$$

where *training* is the training condition, *vis* is the visualization type and *participant* is a unique participant id for modeling individual random effects.

We report model-based conditional means of the bid optimization ratio by training condition and visualization type. These estimates reflect the average proportion of the optimal utility that participants obtained for a given condition.

5.1 Results

Data Preliminaries We recruited 240 participants from Prolific. The median completion time was approximately 17 minutes, with an average base payment of \$14.02 per hour (excluding bonuses).

Bid Optimization Training with deterministic payoffs improved bidding, but these gains did not transfer to settings with stochastic payoffs. As shown in Fig. 6, participants who trained exclusively on their true-cost curve, as well as those in the combined condition (training first on true-cost, then on approximate utility curves), demonstrated improved bid optimization during training, achieving approximately 71% of the utility-optimal welfare. However, across all training conditions, participants’ performance when using Utility Curves in the stochastic test trials remained similar to that observed in Experiment 1.) The persistence of sub-optimal bidding under stochastic payoffs could reflect individual risk preferences, attentional limits, or difficulties generalizing strategies learned under deterministic conditions, rather than a lack of understanding of

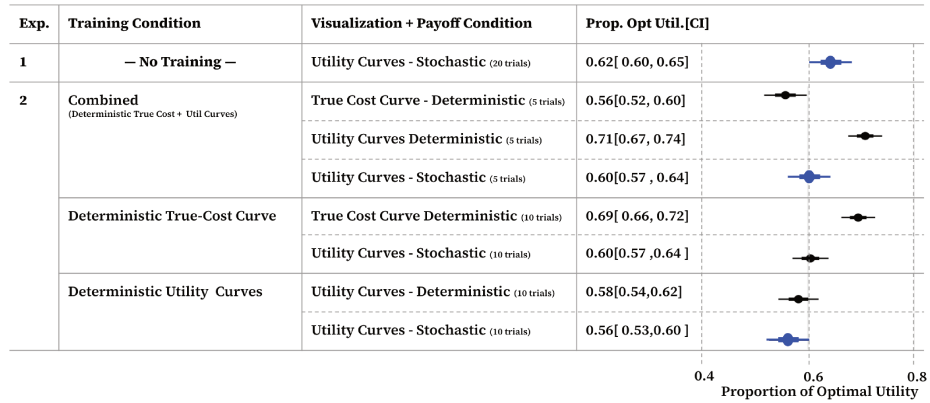


Fig. 6. Bayesian hierarchical model posterior mean and 95% credibility intervals estimates of the proportion of optimal utility achieved with Utility Curve visualizations in Experiment 1 and across all training and payoff conditions in Experiment 2.

the dashboard itself. Moreover, participants’ inability to achieve optimal payoffs even when viewing cost-specific curves in deterministic trials may indicate that factors beyond uncertainty, such as inattention or imperfect mental models of auction dynamics, limited performance improvements. Together with the results from the first experiment, these findings suggest that visualization dashboards, while helpful for improving bid behavior, may be insufficient for promoting optimal bidding.

6 Discussion and Future Work

Our research findings carry implications for strategic environments that employ visualization dashboards to aid agent decision-making. While dashboards can improve decision-making, especially for boundedly rational agents, they do not eliminate suboptimal behavior. In our reverse auction study, visualizations such as Utility Curves and Utility Heatmaps significantly improved bid optimization compared to normative Allocation Rule dashboards. Yet even with improved visualizations dashboards and training on how to use them in verifiable contexts, bounded rationality and a preference for wins led to persistent undershading. A key performance benchmark in our study is the bid behavior of a Bayesian rational agent with access to the prior distribution over competitor bids but no dashboard Wu et al. [33]. This agent represents a lower bound on rational behavior in the absence of information offered by the dashboard. Surprisingly, despite having access to visualized payoff information, participants consistently earned less utility on average than the Bayesian rational agent with only the prior. This underperformance persisted even with improved dashboard design, training, and deterministic payoffs. These findings reveal a practical limitation of dashboard mechanisms [15, 8]: merely publishing a transparent dashboard

is insufficient to ensure near-optimal agent behavior or accurate inference. In particular, assuming agents respond optimally (i.e., best response) can lead to significant inference errors. Incorporating behavioral models, such as quantal response with risk aversion, yields more accurate inference and should be the default approach when designing dashboard mechanisms.

Future work should examine whether these findings generalize to more strategically sophisticated agents or higher-stakes settings. Our study focused on novice participants and modest incentives, offering insight into boundedly rational behavior with dashboards. However, conducting large-scale, well-powered experiments with expert participants presents practical challenges and remains an open area for future exploration. Additionally, while our study focused on dashboards that truthfully implement the published allocation rule, in real-world platforms, such consistency may not always hold due to auction dynamics. An important direction for future work is to empirically evaluate how agents respond to imperfect dashboards. Prior theoretical work [15, 8] proposes methods to mitigate such deviations, but empirical validation remains an open question. Future work could explore alternative decision models (e.g., prospect theory, loss aversion), adaptive dashboards that personalize to user behavior, and new experimental paradigms to test dashboard mechanisms under more realistic, dynamic market conditions. Additionally, given the weak responsiveness to varied visualized signals observed in our study, models that assume fixed bid shading or weakly signal-dependent shading—such as constant or linear shading in signal space—may better capture observed deviations from optimal bidding and improve the accuracy of value inference in behavioral settings.

Bibliography

- [1] Melanie Bancilhon, Zhengliang Liu, and Alvitta Ottley. Let’s gamble: How a poor visualization can elicit risky behavior. In *2020 IEEE Visualization Conference (VIS)*, pages 196–200. IEEE, 2020.
- [2] Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017. <https://doi.org/10.18637/jss.v080.i01>.
- [3] Kay-Yut Chen and Charles R Plott. Nonlinear behavior in sealed bid first price auctions. *Games and Economic Behavior*, 25(1):34–78, 1998.
- [4] James C Cox, Vernon L Smith, and James M Walker. A test that discriminates between two models of the dutch-first auction non-isomorphism. *Journal of Economic Behavior & Organization*, 4(2-3):205–219, 1983.
- [5] James C Cox, Vernon L Smith, and James M Walker. Experimental development of sealed-bid auction theory; calibrating controls for risk aversion. *The American Economic Review*, 75(2):160–165, 1985.
- [6] Vincent P Crawford and Nagore Iriberri. Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *Econometrica*, 75(6):1721–1770, 2007.
- [7] Vincent P Crawford, Miguel A Costa-Gomes, and Nagore Iriberri. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62, 2013.
- [8] Yuan Deng, Jason Hartline, Jieming Mao, and Balasubramanian Sivan. Welfare-maximizing guaranteed dashboard mechanisms. *Available at SSRN 3858104*, 2021.
- [9] Robert Dorsey and Laura Razzolini. Explaining overbidding in first price auctions using controlled lotteries. *Experimental Economics*, 6:123–140, 2003.
- [10] Umut Dur, Robert G Hammond, and Thayer Morrill. Identifying the harm of manipulable school-choice mechanisms. *American Economic Journal: Economic Policy*, 10(1):187–213, 2018.
- [11] Erik Eyster and Matthew Rabin. Cursed equilibrium. *Econometrica*, 73(5):1623–1672, 2005.
- [12] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.
- [13] Jacob K Goeree, Charles A Holt, and Thomas R Palfrey. Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory*, 104(1):247–272, 2002.
- [14] Glenn W Harrison. Risk attitudes in first-price auction experiments: A bayesian analysis. *The Review of Economics and Statistics*, pages 541–546, 1990.

- [15] Jason D Hartline, Aleck Johnsen, Denis Nekipelov, and Onno Zoeter. Dashboard mechanisms for online marketplaces. In *Proceedings of the 2019 acm conference on economics and computation*, pages 591–592, 2019.
- [16] Avinatan Hassidim, Assaf Romm, and Ran I Shorrer. ” strategic” behavior in a strategy-proof environment. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 763–764, 2016.
- [17] Jessica Hullman, Paul Resnick, and Eytan Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one*, 10(11):e0142444, 2015.
- [18] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE transactions on visualization and computer graphics*, 25(1):903–913, 2018.
- [19] Pooya Jalaly, Denis Nekipelov, and Éva Tardos. Learning and trust in auction markets. *arXiv preprint arXiv:1703.10672*, 2017.
- [20] John H Kagel, Ronald M Harstad, and Dan Levin. Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica: Journal of the Econometric Society*, pages 1275–1304, 1987.
- [21] Alex Kale, Matthew Kay, and Jessica Hullman. Visual reasoning strategies for effect size judgments and decisions. *IEEE transactions on visualization and computer graphics*, 27(2):272–282, 2020.
- [22] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 chi conference on human factors in computing systems*, pages 5092–5103, 2016.
- [23] Paula Kayongo, Glenn Sun, Jason Hartline, and Jessica Hullman. Visualization equilibrium. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):465–474, 2021.
- [24] Dan Levin, James Peck, and Asen Ivanov. Separating bayesian updating from non-probabilistic reasoning: An experimental investigation. *American Economic Journal: Microeconomics*, 8(2):39–60, 2016.
- [25] Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.
- [26] Tibor Neugebauer and Reinhard Selten. Individual behavior of first-price auctions: The importance of information feedback in computerized experimental markets. *Games and Economic Behavior*, 54(1):183–204, 2006.
- [27] Muriel Niederle and Emanuel Vespa. Cognitive limitations: Failures of contingent thinking. *Annual Review of Economics*, 15(1):307–328, 2023.
- [28] Axel Ockenfels and Reinhard Selten. Impulse balance equilibrium and feedback in first price auctions. *Games and Economic Behavior*, 51(1):155–170, 2005.
- [29] Parag A Pathak and Tayfun Sönmez. Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–1652, 2008.

- [30] Anmol Ratan. Does displaying probabilities affect bidding in first-price auctions? *Economics Letters*, 126:119–121, 2015.
- [31] Alex Rees-Jones. Mistaken play in the deferred acceptance algorithm: Implications for positive assortative matching. *American Economic Review*, 107(5):225–229, 2017.
- [32] Alex Rees-Jones. Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games and Economic Behavior*, 108: 317–330, 2018.
- [33] Yifan Wu, Ziyang Guo, Michails Mamakos, Jason Hartline, and Jessica Hullman. The rational agent benchmark for data visualization. *arXiv preprint arXiv:2304.03432*, 2023.
- [34] Dongping Zhang, Eytan Adar, and Jessica Hullman. Visualizing uncertainty in probabilistic graphs with network hypothetical outcome plots (nethops). *IEEE Transactions on Visualization and Computer Graphics*, 28(1):443–453, 2021.
- [35] Dongping Zhang, Jason Hartline, and Jessica Hullman. Designing shared information displays for agents of varying strategic sophistication. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1):1–34, 2024.