Problem Set 2

Peilin Wu

September 21, 2025

1 Introduction

Auctions have long served as a central mechanism for resource allocation, from petroleum leases (Capen et al., 1971) to spectrum rights and online advertising markets. A key challenge in common-value settings is the *winner's curse*, the phenomenon where the auction winner tends to systematically overpay due to selection bias. Early empirical evidence from oil lease bidding showed that firms frequently incurred losses by failing to account for this bias, while behavioral studies have confirmed similar systematic deviations among human subjects (Thaler, 1988; Kagel and Levin, 2002). Auction theory formalized this problem, showing how rational bidders should optimally shade their bids in the presence of noisy signals and competition (Milgrom and Weber, 1982).

While the winner's curse has been extensively studied among human participants, far less is known about whether large language models (LLMs), which is increasingly deployed as autonomous agents, are subject to similar biases. Recent advances suggest that LLMs can perform strategic reasoning in structured tasks, but their susceptibility to systematic errors remains an open question. If LLMs behave like human novices, they may ignore adverse selection and fall victim to the winner's curse. Conversely, if they approximate rational bidding behavior, they may correctly adjust for informational noise and competition intensity.

This paper investigates whether LLMs exhibit systematic exposure to the winner's curse in controlled auction simulations. We design a series of experiments across three conditions: (i) a one-shot common-value purchase game (Control), (ii) auctions with private noisy signals of varying precision (T1), and (iii) a first-price sealed-bid common-value auction with multiple bidders (T2). Four representative LLMs (GPT-5, Qwen3-Max-Preview, DeepSeek, and Hunyuan) were tested. We compare their bidding strategies against theoretical benchmarks and behavioral findings from the literature.

Our contribution is twofold. First, we extend the literature on the winner's curse by introducing LLMs as experimental subjects in mechanism design, bridging economic theory, behavioral experimentation, and computational AI. Second, by comparing LLM outcomes with human predictions and equilibrium benchmarks, we evaluate whether LLMs resemble rational economic agents, boundedly rational humans, or a hybrid of both. The findings offer implications for the design of AI-augmented markets and the broader study of machine rationality.

2 Mechanism Design: Auction Game Selection and Variations

Baseline Game (Control): One-shot Common-Value Purchase

We begin with the painting-purchase game as the control condition. The **true common value** of the painting is denoted by V, which is uniformly distributed:

$$V \sim \text{Uniform}[0, 1000],$$

meaning that every value between 0 and 1000 is equally likely Capen et al. (1971).

The buyer submits a single bid $b \in [0, 1000]$. The trader accepts if and only if the bid is weakly greater than the true value:

$$\text{Accept if } b \geq V.$$

If accepted, the buyer can resell the painting for $1.5V$, yielding a realized profit:

$$\pi = \begin{cases} 1.5V - b, & \text{if } V \leq b, \\ 0, & \text{if } V > b. \end{cases}$$

Here, π denotes the buyer's profit, V is the true value, and b is the bid.

The expected profit given bid b is

$$E[\pi \mid b] = \int_0^b (1.5v - b)\,\frac{1}{1000}\,dv = -\,\frac{b^2}{4000} \le 0.$$

This shows that any positive bid has a non-positive expected return, and the optimal bid is $b^* = 0$. Intuitively, if the buyer wins, it is more likely that V is lower than the bid, leading to losses—a direct form of the *winner's curse* Thaler (1988).

Treatment 1 (T1): Private Noisy Signals

In this variation, the buyer observes a **private signal** s about the true value before bidding:

$$s = V + \varepsilon,$$

where s is the observed signal, V is the true value, and ε is a random noise term with mean zero and variance σ^2. A smaller σ means more precise information, while a larger σ implies less reliable information.

We consider two cases:

- Low-noise case: $\sigma = \sigma_{\text{low}}$ (signal is informative).
- High-noise case: $\sigma = \sigma_{\text{high}}$ (signal is less informative).

With high noise, the conditional expectation of V given ac4ceptance is biased upward, increasing the likelihood of losses. Thus, the winner's curse is more severe when signals are less precise Milgrom and Weber (1982); Kagel and Levin (2002).

Treatment 2 (T2): Competition Intensity

We extend the game to $N \in \{2, 4\}$ symmetric bidders in a first-price sealed-bid common-value auction. Each bidder i observes a private signal

$$s_i = V + \varepsilon_i,$$

where s_i is bidder i's signal, V is the common value, and ε_i is independent noise. Each bidder submits a bid $b_i = b(s_i)$, and the highest bidder wins, paying her own bid.

In this setting, the event of winning is itself informative: the winner is most likely the bidder who received the highest (most optimistic) signal. As the number of bidders N increases, the gap between the winner's signal and the true value tends to widen, amplifying the winner's curse Milgrom and Weber (1982); Kagel and Levin (2002).

Hypotheses (Winner's Curse Incidence)

H1 (Control). In the baseline game, the optimal bid is $b^* = 0$. Any positive bid produces non-positive expected returns, so naïve bidding directly induces a winner's-curse loss Capen et al. (1971); Thaler (1988).

H2 (Information Precision). In T1, the winner's curse is more severe when σ is larger, because the observed signal is less informative and conditional selection bias is stronger Milgrom and Weber (1982).

H3 (Competition Intensity). In T2, the winner's curse worsens as the number of bidders N increases: the winner is increasingly likely to be the one with the most optimistic signal, exacerbating overpayment and losses Kagel and Levin (2002).

3 AI Agent Experiments

3.1 Experiment Setup

To examine whether large language models (LLMs) exhibit systematic exposure to the winner's curse, we conducted controlled auction simulations under three experimental conditions. Four representative LLMs were selected: **GPT-5** (OpenAI), **Qwen3-Max-Preview** (Alibaba), **DeepSeek** (Chinese open-source), and **Hunyuan** (Tencent). Each model was instructed to act as an autonomous auction participant, constrained to output a single line of valid JSON containing its bid, role, and optional reasoning.

The experimental conditions replicated the designs in Section 2:

1. **CONTROL**: A one-shot common-value purchase game with resale payoff $1.5V$, where $V \sim U[0, 1000]$ and the seller accepts iff $b \geq V$.

2. **T1**: A private noisy signal condition, where the agent observes $s = V + \varepsilon$ with $s = 620.37$ and $\sigma = 180$.

3. **T2**: A first-price sealed-bid common-value auction with $N = 4$ symmetric bidders, each observing $s_i = V + \varepsilon_i$, with $s_{you} = 540.22$ and $\sigma = 120$.

In all cases, the prompt explicitly emphasized the rules of the game, profit structure, and the need to account for selection bias (winner's curse). Each run was executed independently, producing structured decision logs later transcribed into Tables 1–3.

3.2 Results

Control condition (one-shot purchase). Theoretically, the optimal bid is $b^* = 0$, since any positive bid yields a non-positive expected return. Contrary to this benchmark, the LLMs consistently placed strictly positive bids. GPT-5 proposed values between 375 and 750, Qwen produced both 0 and 375, DeepSeek offered bids at 75 and 375, while Hunyuan submitted midrange bids (50, 500). These outcomes reveal that none of the models fully internalized the adverse selection implied by the acceptance rule, instead behaving as though positive expected gains were achievable.

Treatment 1 (private noisy signal). Given $s = 620.37$ and high noise $\sigma = 180$, rational bidders should shade downward from the observed signal. All models exhibited such shading. GPT-5 bid 600, only slightly below the signal, while Qwen and Hunyuan placed more conservative bids at 465 and 400 respectively. DeepSeek also shaded its bid to 465, aligning more closely with Qwen than GPT-5. This variation suggests differential sensitivity to informational noise across models: some agents (e.g., Hunyuan, DeepSeek) appeared highly cautious, while others (e.g., GPT-5) retained optimism despite uncertainty.

Treatment 2 (competition with $N = 4$). Increased competition is expected to exacerbate the winner's curse, inducing stronger bid shading. The models conformed to this prediction: GPT-5 averaged ≈ 442.5 across runs, Qwen bid 390, DeepSeek bid 405, and Hunyuan bid 380. All of these values are noticeably lower than their corresponding T1 bids. This behavior aligns with theoretical expectations that the event of winning in multi-bidder settings is correlated with having the most optimistic signal, necessitating deeper shading.

Table 1: Bids in CONTROL condition

Model	Bids	Optimum
GPT-5	375, 750	0
Qwen3	0, 375	0
DeepSeek	75, 375	0
Hunyuan	50, 500	0

Table 2: Bids in T1 (signal $s = 620.37$, $\sigma = 180$)

Model	Bid	Signal
GPT-5	600	620.37
Qwen3	465	620.37
DeepSeek	465	620.37
Hunyuan	400	620.37

Table 3: Bids in T2 (first-price, $N = 4$, $s = 540.22$)

Model	Bid	vs. T1
GPT-5	∼442	Lower
Qwen3	390	Lower
DeepSeek	405	Lower
Hunyuan	380	Lower

Table 4: Summary of LLM behavior vs. theory

Condition	Theory	LLM outcome
CONTROL	$b^* = 0$	All > 0 (50–750)
T1	Shade from s	GPT-5: 600; Qwen/DeepSeek: 465; Hunyuan: 400
T2	More shading	GPT-5: ∼442; Qwen: 390; DeepSeek: 405; Hunyuan: 380

3.3 Analysis

H1 (Control). Our first hypothesis posited that rational bidders would converge to $b = 0$ in the baseline purchase game. Theoretically, the expected profit of bidding b under $V \sim U[0, 1000]$ is

$$E[\pi(b)] = \int_0^b (1.5v - b)\,\frac{1}{1000}\,dv \;=\; -\frac{b^2}{4000},$$

implying $b^* = 0$ is the unique equilibrium bid (Milgrom and Weber, 1982). The LLM outcomes, however, contradict this prediction: every model submitted strictly positive bids, some substantially above the true optimum. This reflects the same systematic bias documented in early studies of the winner's curse in petroleum lease auctions (Capen et al., 1971) and behavioral experiments (Thaler, 1988). Like human novices, LLMs appear to anchor on average resale value and neglect adverse selection.

H2 (Information precision). The second hypothesis suggested that less precise signals ($\sigma = 180$) would intensify the winner's curse, leading to stronger bid shading. Bayesian updating implies

$$b(s) < E[V \mid s],$$

since winning conveys negative information about V. Empirically, Hunyuan (400), Qwen (465), and DeepSeek (465) shaded their bids considerably below $s = 620.37$, while GPT-5 (600) remained close to the signal. This heterogeneity suggests that some models incorporated conditional expectation reasoning more effectively than others, echoing experimental findings that humans often under-correct for noisy signals (Kagel and Levin, 2002). Thus, H2 is only partially supported: directionally correct, but quantitatively inconsistent.

H3 (Competition intensity). Our final hypothesis predicted that as the number of bidders increases, the winner's curse becomes more severe. With $N = 4$ participants, the event of winning is correlated with holding the most optimistic signal, i.e.,

$$E[V \mid s_i, i = \arg\max_j b(s_j)] < E[V \mid s_i],$$

which requires deeper shading (Milgrom and Weber, 1982). This was borne out in the data: all models reduced bids in T2 relative to T1 (GPT-5: 600 → 442.5, Qwen: 465 → 390, DeepSeek: 465 → 405, Hunyuan: 400 → 380). This pattern matches both theoretical predictions and laboratory evidence on competitive common-value auctions (Kagel and Levin, 2002).

3.4 Discussion

Overall, the results reveal a **mixed rationality profile**. On the one hand, LLMs consistently failed to adopt the equilibrium strategy in CONTROL, exposing them to systematic losses—an outcome well aligned with the classic descriptions of the winner's curse in resource auctions (Capen et al., 1971; Thaler, 1988). On the other hand, when exposed to informational noise or competitive intensity, the models shaded bids in the theoretically correct direction, demonstrating bounded rationality comparable to human experimental subjects (Kagel and Levin, 2002).

The divergence from H1 highlights a fundamental limitation: LLMs appear to optimize via heuristic anchoring rather than fully internalizing conditional expectations. Yet their relative responsiveness under T1 and T2 suggests emergent rationality in comparative statics, consistent with the equilibrium adjustments proposed in auction theory (Milgrom and Weber, 1982). These findings position LLMs between purely rational agents and behavioral human bidders, making them promising proxies for mechanism design experiments where both theory and behavioral anomalies matter.

References

Capen, E. C., Clapp, R. V., and Campbell, W. M. (1971). Competitive bidding in high-risk situations. *Journal of petroleum technology*, 23(06):641–653.

Kagel, J. H. and Levin, D. (2002). Bidding in common-value auctions: A survey of experimental research. *Common value auctions and the winner's curse*, 1:1–84.

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

Thaler, R. H. (1988). Anomalies: The winner's curse. *Journal of economic perspectives*, 2(1):191–202.

Appendix A: LLM Decision Logs (JSON Summary)

CONTROLT1T2GPT-5, Qwen3, DeepSeek, Hunyuan `avg_bid` `shading`

$$\text{shading} = \frac{s-b}{s} \times 100\%.$$


```
{
  "experiments": {
    "CONTROL": {
      "theory_optimum": 0,
      "models": {
        "GPT-5": { "bids": [375, 750], "avg_bid": 562.5 },
        "Qwen3": { "bids": [0, 375], "avg_bid": 187.5 },
        "DeepSeek": { "bids": [75, 375], "avg_bid": 225 },
        "Hunyuan": { "bids": [50, 500], "avg_bid": 275 }
      }
    },
    "T1": {
      "signal": 620.37,
      "sigma": 180,
      "models": {
        "GPT-5":   { "bid": 600,  "shading": 3.3 },
        "Qwen3":   { "bid": 465,  "shading": 25.0 },
        "DeepSeek":{ "bid": 465,  "shading": 25.0 },
        "Hunyuan": { "bid": 400,  "shading": 35.5 }
      }
    },
    "T2": {
```

```
      "signal": 540.22,
      "sigma": 120,
      "n_bidders": 4,
      "models": {
        "GPT-5":   { "bid": 442.5, "shading": 18.0 },
        "Qwen3":   { "bid": 390,   "shading": 27.8 },
        "DeepSeek":{ "bid": 405,   "shading": 25.0 },
        "Hunyuan": { "bid": 380,   "shading": 29.7 }
      }
    }
  }
}
```