# Final Project Report

Sherry Zhuge, Sahil Parikh, Grace Lin
szhuge, snparikh, glin2

April 2026

## 1 Introduction

Combinatorial auctions allow an auctioneer to allocate multiple distinct items to bidders whose values may depend on entire bundles rather than on individual items alone. This is useful because bidders can express complementarities and substitutes: for example, two items together may be much more valuable than either item separately. However, this flexibility creates a major information problem. With $n$ items, there are $2^n - 1$ nonempty bundles, so asking every bidder to fully reveal their valuation function is often infeasible.

Preference elicitation studies how an auctioneer can ask targeted queries to bidders and collect only the information needed to determine a good, or ideally welfare-maximizing, allocation. This report focuses on two complementary perspectives on this problem. The first is the query-learning view of Blum et al., which treats each bidder's valuation function as a target function and asks how many queries are necessary to find the optimal allocation without fully learning every bidder's preferences. Their main result shows that even for relatively simple monotone DNF preferences, preference elicitation can require exponentially many queries in the worst case.

The second perspective comes from Sandholm and Boutilier's chapter on preference elicitation in combinatorial auctions. We focus especially on the Efficient Best-First (EBF) algorithm, which uses a rank-lattice structure to search through candidate allocations while eliciting only partial information from bidders. Unlike Blum et al.'s lower-bound result, EBF gives a constructive strategy for deciding which information to request next.

Together, these works highlight the central tension in preference elicitation. From a theoretical perspective, worst-case hardness shows that limited revelation cannot always avoid exponential complexity. From an algorithmic perspective, structured elicitation methods such as EBF show that partial revelation can still be useful in practice. From the game-theory perspective, the setting connects to the mechanism design portion of the course because the goal is to maximize social welfare while using VCG-style payments to make truthful bidding compatible with the elicitation process. These frameworks are useful in real-world settings such as spectrum auctions, procurement auctions, cloudresource allocation, and logistics markets, where bidders often care about combinations of items and cannot feasibly evaluate or reveal every possible bundle.

## 2 Formal Setting and Definitions

Let $S$ be a set of $n$ indivisible items to be allocated among $k$ bidders. A *bundle* is any subset $A \subseteq S$. Each bidder $i$ has a private valuation function

$$f_i : 2^S \to R_{\geq 0},$$

where $f_i(A)$ is bidder $i$'s value for bundle $A$. We assume valuations are monotone, also known as the *free disposal* assumption: if $A \subseteq B$, then $f_i(A) \leq f_i(B)$.

An allocation $(S_1, \ldots, S_k)$ is feasible if the assigned bundles are disjoint, meaning no item is given to more than one bidder. The goal of preference elicitation is to find a feasible allocation maximizing social welfare,

$$\sum_{i=1}^{k} f_i(S_i),$$

while asking bidders only limited questions about their valuation functions. In the two-bidder case studied by Blum et al., we write the bidders' valuations as $f$ and $g$. An allocation can be represented by $x \in \{0,1\}^n$, where $x$ is the bundle given to the first bidder and $\bar{x}$ is given to the second. The goal is to maximize

$$f(x) + g(\bar{x}).$$

**Query Types.** A *value query* asks a bidder for the exact value of a bundle, $f_i(A)$. A *demand query* gives the bidder item prices $w \in R^n_{\geq 0}$ and asks which bundle maximizes value minus price:

$$\arg\max_{A \subseteq S} \left( f_i(A) - \sum_{j \in A} w_j \right).$$

A *rank query*, used in the EBF framework, asks a bidder for their $r$th most preferred bundle.

**DNF Preferences.** A DNF preference function is specified by desired terms $T = \{T_1, \ldots, T_\ell\}$ with associated values $v_1, \ldots, v_\ell$. The value of a bundle $A \subseteq S$ is

$$f_{T,v}(A) = \max_{T_j \subseteq A} v_j.$$

In the Boolean case, all $v_j = 1$, so the bidder is satisfied exactly when $A$ contains at least one desired term.

**Rank-Lattice Notation.** For EBF, let $b_i(r_i)$ denote bidder $i$'s bundle at rank $r_i$, where rank 1 is most preferred. A rank vector

$$r = [r_1, \ldots, r_k]$$

represents the collection where bidder $i$ receives $b_i(r_i)$. One rank vector $r$ dominates another $r'$ if $r_i \leq r'_i$ for every bidder $i$. This domination relation is what allows EBF to prune parts of the rank lattice during search.

# 3 Blum et al.: Preference Elicitation as Query Learning

Blum et al. study preference elicitation through the lens of query learning. In this view, each bidder's valuation function is treated like an unknown target function. A bundle of items corresponds to an input $x \in \{0,1\}^n$, and a value query asks for the bidder's value on that bundle, much like a membership query in learning theory.

The important difference is that preference elicitation does not require learning the full valuation functions. In the two-bidder case, the auctioneer only needs to find a bundle $x$ for one bidder and the complementary bundle $\bar{x}$ for the other bidder that maximizes

$$f(x) + g(\bar{x}).$$

Thus, the goal is not complete learning, but finding enough information to certify the welfare-maximizing allocation.

This distinction makes preference elicitation potentially easier than exact learning, but Blum et al. show that it can still be very hard. Their main lower bound focuses on monotone DNF preferences, which are closely related to XOR-style bidding languages in combinatorial auctions. Even though these preferences have a simple form, the optimal allocation may be hidden among exponentially many possibilities. Theorem 1 formalizes this by constructing two bidders with row and column preferences over an $m \times m$ grid of items, plus one hidden random term for each bidder. Although each bidder has only $O(\sqrt{n})$ terms, any value-query elicitor is unlikely to discover the hidden allocation quickly. This yields the $2^{\Omega(\sqrt{n})}$ worst-case lower bound analyzed below.

# 4 Main Result: Theorem 1 Proof Sketch

The central hardness result of the paper is:

**Theorem 1** (Blum et al., 2004)**.** *Preference elicitation of monotone DNF formulas requires $2^{\Omega(\sqrt{n})}$ value queries in the worst case, even when each bidder's preference function has only $O(\sqrt{n})$ terms.*

This result is surprising because the preference functions involved are *small* - each bidder has only $O(\sqrt{n})$ terms - yet elicitation remains exponentially hard.

We start the proof by constructing a family of hard instances and showing that no query strategy can find the optimal allocation efficiently.

## 4.1 The Construction

Let $n = m^2$. Arrange items in an $m \times m$ matrix, labeling the item in row $i$, column $j$ as $x_{ij}$. Define two bidders:

- **Row bidder** $f_R$: satisfied by any complete row, $f_R^{\text{base}} = \bigvee_{i=1}^{m} \bigwedge_{j=1}^{m} x_{ij}$
- **Column bidder** $f_C$: satisfied by any complete column, $f_C^{\text{base}} = \bigvee_{j=1}^{m} \bigwedge_{i=1}^{m} x_{ij}$

Now flip a fair coin for each item independently, partitioning $S$ into disjoint sets $H$ (heads) and $T$ (tails). We append one hidden term to each bidder: $H$ to $f_R$ and $T$ to $f_C$. The optimal allocation is then to give $H$ to $f_R$ and $T$ to $f_C$, satisfying both - but $H$ and $T$ are unknown to the elicitor. Each bidder ends up with only $m + 1 = O(\sqrt{n})$ terms total.

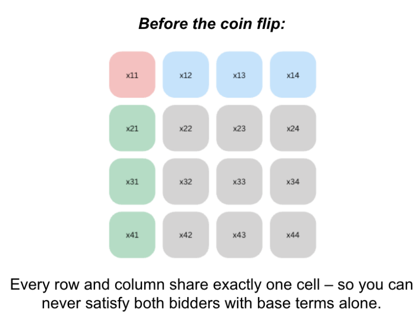


Figure 1: Before coin flip

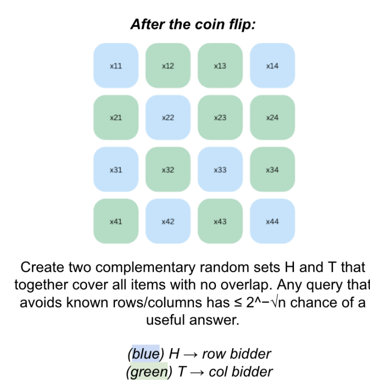


Figure 2: After coin flip

## 4.2 Why This Instance Is Hard

Any informative query to $f_R$ must be missing at least one item per row (otherwise a full row is contained and the answer is trivially "yes"). Such a query only returns a positive response if *all* missing items land in $T$ — which, by independence of coin flips, occurs with probability at most $2^{-m} = 2^{-\sqrt{n}}$. The same holds symmetrically for $f_C$.

After $q$ queries, the probability of receiving *any* informative positive response is at most $q \cdot 2^{-\sqrt{n}}$ by a union bound. For this to reach $\frac{1}{2}$, we need $q \geq \frac{1}{2} \cdot 2^{\sqrt{n}}$, giving the desired lower bound.

### 4.3 The Formal Bound

Suppose the elicitor issues $q$ value queries before attempting to name an allocation. Each query returns a useful positive response with probability at most $2^{-\sqrt{n}}$. By a union bound, the probability that *any* of the $q$ queries returns "yes" is at most:

$$\Pr[\text{at least one positive response in } q \text{ queries}] \ \leq \ q \cdot 2^{-\sqrt{n}}$$

Without a positive response, the elicitor has received no information about $H$ or $T$, and so cannot identify the optimal allocation. For this probability to be at least $\frac{1}{2}$ (i.e., for the elicitor to have a reasonable chance of succeeding), we need:

$$q \ \geq \ \frac{1}{2} \cdot 2^{\sqrt{n}}$$

Therefore, the expected number of value queries required by *any* elicitation algorithm is at least $\frac{1}{2} \cdot 2^{\sqrt{n}}$, establishing the $2^{\Omega(\sqrt{n})}$ lower bound.

### 4.4 The Key Intuition

The base terms (rows and columns) are fully transparent — the elicitor gains nothing from querying bundles that obviously contain them. All difficulty is concentrated in the *one hidden term per bidder*, and finding it is essentially equivalent to guessing a random subset without any useful signal. This also explains why the result sits strictly between approximate and exact learning for monotone DNF:

| **Goal** | **Query complexity for monotone DNF** |
|---|---|
| Approximate learning | Polynomial |
| Preference elicitation | $2^{\Omega(\sqrt{n})}$ (this theorem) |
| Exact learning | $2^{\Omega(n)}$ |

## 5 Rank-Lattice Elicitation and the Efficient Best-First Algorithm

### 5.1 The Rank-Lattice Framework

Sandholm introduces the topic of rank-lattice elicitors where this framework organizes bidder allocations by ranking bundles for each bidder from most to least preferred. A rank vector, $r = [r_1, r_2, \ldots, r_n]$, represents an allocation where each bidder $i$ receives the bundle at rank $r_i$. The set of these vectors, combined with a domination relation where one vector dominates another if all its ranks are numerically less than or equal to the other, defines the rank lattice. This topological structure allows the elicitor to identify Pareto efficient allocations as those feasible nodes not dominated by any other allocation within said lattice.

### 5.2 How EBF Uses Partial Information

The Efficient Best-First algorithm (EBF) is a top-down search strategy designed to find a social welfare maximizing allocation by traversing the rank lattice starting from the root node. Unlike standard search methods, EBF must operate with partial information because bundle values and bidder identities may not be fully known (initially). The algorithm expands the fringe node with the highest current value while using constraint network inference to propagate value bounds and prune off those provably dominated nodes. EBF specifically utilizes rank and value queries to hone its model of bidder valuations until the first feasible node reached is verified as the optimal allocation.

### 5.3 Strengths and Limitations of EBF

EBF is significant because it provides a concrete elicitation algorithm that identifies an optimal allocation rather than simply proving the theoretical hardness of the problem at hand. While it is designed to be as effective as possible within its set of elicitors, its practicality is heavily dependent on the number of items and variables involved. Maintaining an explicit rank lattice becomes computationally expensive as the scale of the problem goes up, often leading to an elicitation ratio $\approx$1 in larger auctions. A unique advantage of this approach is that no additional information beyond what EBF collects is necessary to determine VCG payments, meaning they are essentially gotten for free (no cost) once the algorithm terminates.

## 6 Compare: Worst-Case Barriers and Practical Elicitation

Blum et al. and Sandholm and Boutilier examine preference elicitation from two complementary perspectives. Blum et al. focus on the information-theoretic limits of elicitation: even if the bidders' preferences have a compact monotone DNF structure, finding the welfare-maximizing allocation can still require exponentially many value queries. Sandholm and Boutilier, through the Efficient Best-First (EBF) algorithm, focus on how an elicitor can strategically use partial information to search for an optimal allocation without asking every bidder for every bundle value.

### 6.1 Lower Bounds Versus Search Procedures

Theorem 1 from Blum et al. shows that the difficulty of preference elicitation is not just caused by a poor choice of algorithm. In their construction, the row and column preferences are easy to understand, but the optimal allocation depends on hidden random terms. A value query is very unlikely to reveal these hidden terms, so any elicitor needs $2^{\Omega(\sqrt{n})}$ queries in the worst case. The key point is that the hardness is built into the structure of the information problem itself.

EBF provides a practical counterpoint to this lower-bound view. Rather than trying to learn full valuation functions, EBF organizes candidate allocations in a rank lattice and traverses it top-down. It uses rank and value queries to expand the highest-value fringe node, prune dominated allocations, and eventually identify a welfare-maximizing feasible allocation. This makes EBF a concrete example of how partial revelation can guide the elicitation process more intelligently than naive full revelation.

### 6.2 Why EBF Does Not Eliminate the Hardness

Although EBF is algorithmically useful, it does not contradict the lower-bound result from Blum et al. The Sandholm and Boutilier chapter notes that rank-lattice methods can still have severe worst-case behavior. In particular, Hudson and Sandholm show that EBF may need to call the valuation routine on up to

$$\frac{2^{mn} - n^m}{2} + 1$$

collections in the worst case. Thus, even a clever traversal strategy can be forced to examine a very large portion of the search space.

This connects directly to the intuition behind Theorem 1. In the hard instance from Blum et al., the useful information is hidden in a way that ordinary queries almost never uncover. EBF can prune allocations when rank and value information reveal domination, but if the relevant hidden structure has not been exposed, there may be little useful pruning to do. The lower bound therefore explains why practical elicitation algorithms can still face exponential barriers on adversarial instances.

Overall, the two works should be read together rather than as opposing claims. EBF demonstrates what is achievable with a structured elicitation strategy: partial information can be used to avoid unnecessary revelation and guide the search for an optimum. Blum et al. show the limits of this hope: in the worst case, the information needed to certify the optimum may itself be exponentially hard to find. Together, they capture the central tension of preference elicitation in combinatorial auctions.

## 7 Key Reflections and Further Direction

The most interesting point of Theorem 1 is how it establishes a powerful lower bound, and paper *Preference Elicitation and Query Learning* provides a crucial counter-perspective by identifying specific conditions that make the problem more interpretable. This helps pinpoint exactly which structural properties (i.e. prior knowledge) drive that exponential complexity.

From here it is also observed that if the auctioneer knows the valuation function of even one bidder, the allocation task becomes significantly easier. The known bidder's preferences act as a set of constraints, allowing the auctioneer to use targeted queries to find the "missing piece" of the social welfare problem based on the remaining bidders. This establishes a coarse-to-fine intuition, which is an efficient strategy by starting with known information to prune the $2^n$ search space before moving to "finer" queries. The hardness in Theorem 1 stems directly from the lack of such an anchorage point because neither bidder's hidden term is known such that there are no initial constraints to narrow the search down with.

Ultimately Blum et al. provided a rigorous foundation for communication bounds, though we remain curious about whether poly-size DNFs can be elicited in $2^{O(\sqrt{n})}$ queries. While Theorem 1 provides a lower bound of $2^{\Omega(\sqrt{n})}$, is this gap truly tight? We also want to explore how this scales beyond two bidder auctions since we suspect that as $k$ increases, overlapping hidden terms between players will compound the elicitation bottleneck. By moving beyond worst-case theory, we can further explore how these auctions behave in more realistic settings. For example, learning more about if "structured" DNFs (such as those following natural orderings) are practically easier to elicit than the random instances used in the paper's proof, and how/if extending the model to $k$ bidders would effect query count blow-up under increased competition.

# 8 Contributions

All three team members read and discussed the primary sources together, including Blum et al. and Sandholm and Boutilier. As a group, we worked through the main technical ideas behind preference elicitation in combinatorial auctions, with particular attention to how Blum et al.'s lower-bound result connects to Sandholm and Boutilier's algorithmic treatment of elicitation through EBF.

Sherry focused primarily on the Blum et al. paper. She wrote the section introducing Blum et al.'s view of preference elicitation as a query-learning problem and explained how this perspective sets up the main lower-bound result. She was also responsible for the detailed discussion of Theorem 1, including the row-column construction, the hidden random terms, and the intuition behind the $2^{\Omega(\sqrt{n})}$ lower bound. In addition, she helped connect Theorem 1 to EBF by explaining how the lower bound illustrates the limits of practical elicitation algorithms.

Sahil was primarily responsible for the introduction and formal setup of the report. He wrote the motivation for studying preference elicitation in combinatorial auctions and helped define the core objects used throughout the paper, including bidders, bundles, valuation functions, welfare maximization, query types, and DNF preference representations. He also worked on the high-level comparison between Blum et al. and Sandholm and Boutilier, focusing on the contrast between worst-case query-complexity barriers and constructive elicitation strategies.

Grace focused primarily on the Sandholm and Boutilier chapter. She introduced the rank-lattice framework and explained how the Efficient Best-First

(EBF) algorithm uses rank and value queries to search for a welfare-maximizing allocation without full preference revelation. She also worked on the discussion and conclusion sections, including the broader interpretation of what these papers show about theory versus practice in preference elicitation. In addition, she helped connect the material to topics from class, especially mechanism design, welfare maximization, incentives, and learning from limited feedback.

## References


[1] Blum, A., Jackson, J., Sandholm, T., and Zinkevich, M. Preference elicitation and query learning. *Journal of Machine Learning Research*, 5:649–667, 2004.

[2] Sandholm, T. and Boutilier, C. Preference elicitation in combinatorial auctions. In Cramton, P., Shoham, Y., and Steinberg, R. (Eds.), *Combinatorial Auctions*. MIT Press, 2006.