Is Selling Complete Information (Approximately) Optimal?

DIRK BERGEMANN*, Yale University, USA

YANG CAI†, Yale University, USA

GRIGORIS VELEGKAS‡, Yale University, USA

MINGFEI ZHAO§, Google Research, USA

We study the problem of selling information to a data-buyer who faces a decision problem under uncertainty. We consider the classic Bayesian decision-theoretic model pioneered by Blackwell [9, 10]. Initially, the data buyer has only partial information about the payoff-relevant state of the world. A data seller offers additional information about the state of the world. The information is revealed through signaling schemes, also referred to as *experiments*. In the single-agent setting, any mechanism can be represented as a menu of experiments. A recent paper by Bergemann et al. [8] present a complete characterization of the revenue-optimal mechanism in a binary state and binary action environment. By contrast, no characterization is known for the case with more actions. In this paper, we consider more general environments and study arguably the simplest mechanism, which only sells the *fully informative experiment*. In the environment with binary state and $m \geq 3$ actions, we provide an $O(m)$-approximation to the optimal revenue by selling only the fully informative experiment and show that the approximation ratio is tight up to an absolute constant factor. An important corollary of our lower bound is that the size of the optimal menu must grow at least linearly in the number of available actions, so no universal upper bound exists for the size of the optimal menu in the general single-dimensional setting. We also provide a sufficient condition under which selling only the fully informative experiment achieves the optimal revenue.

For multi-dimensional environments, we prove that even in arguably the simplest matching utility environment with 3 states and 3 actions, the ratio between the optimal revenue and the revenue by selling only the fully informative experiment can grow immediately to a polynomial of the number of agent types. Nonetheless, if the distribution is uniform, we show that selling only the fully informative experiment is indeed the optimal mechanism.

CCS Concepts: • **Theory of computation** → **Algorithmic mechanism design**; **Algorithmic game theory**.

Additional Key Words and Phrases: mechanism design, selling information, revenue maximization, fully informative experiment

1 INTRODUCTION

As large amounts of data become available and can be communicated more easily and processed more effectively, information has come to play a central role for economic activity and welfare in our age. In turn, markets for information have become more prominent and significant in terms of trading volume. Bergemann and Bonatti [7] provide a recent introduction to markets for information with a particular emphasis on data markets and data intermediation in e-commerce.

Information is a valuable commodity for any decision-makers under uncertainty. By acquiring more information, the decision-maker can refine his initial estimates about the true state of the world and consequently improve the expected utility of his decision. In many economically significant situations, a decision-maker can acquire additional information from a seller who either already has the relevant information or can generate the relevant information at little or no cost. Beyond digital and financial markets, the value of information is particularly important for health economics where it appears as the expected value of sample information, see [1] or more recently, [23] and [21] with applications to COVID-19 testing. The value of additional information is also central in the literature on A/B testing, see [3].

For the seller of information, the question is therefore how much information to sell and at what price to sell it. A natural starting point is to make all the information available and sell the information at a price that maximizes the revenue of the seller. The problem of selling information at an optimal price is therefore closely related to the classic problem of the monopolist selling a single unit of an object at an optimal price. Just as in the classic optimal monopoly problem, the seller faces a trade-off in the choice of the price. A higher price generates a higher revenue for every additional sale, but there might be few buyers who value the object higher than the price asked. A lower price generates a larger volume of sales but with low marginal revenue. The optimal policy finds the balance between the marginal revenue and the inframarginal revenue considerations. The sale of complete information faces similar trade-offs. A high price for the complete information will be acceptable for decision-makers with diffuse prior information, thus those who value the additional information most, but may not be acceptable for those buyers who already have some information.

We analyze these issues in the classic Bayesian decision-theoretic model pioneered by Blackwell [9, 10]. Here we interpret the decision-theoretic model as one where a data buyer faces a decision problem under uncertainty. A data seller owns a database containing information about a "state" variable that is relevant to the buyer's decision. Initially, the data buyer has only partial information about the state. This information is private to the data buyer and unknown to the data seller. The precision of the buyer's private information determines his willingness to pay for any supplemental information. Thus, from the perspective of the data seller, there are many possible types of data buyer.

A recent contribution by Bergemann et al. [8] analyzes the optimal selling policy with the tools of mechanism design. Their analysis is mostly focused on the canonical decision-theoretic setting with a binary state and a binary action space. In this setting, the type space is naturally one-dimensional and given by the prior probability of one state, the probability of the other state being simply the complementary probability.

As in [8], we investigate the revenue-maximizing information policy, i.e., how much information the data seller should provide and how she should price access to the data. In order to screen the heterogeneous data buyer types, the seller offers a menu of information products. In the present context, these products are statistical experiments— signals that reveal information about the payoff-relevant state. Only the information product itself is assumed to be contractible. By contrast, payments cannot be made contingent on either the buyer's action or the realized state and signal.

Consequently, the value of an experiment to a buyer is determined by the buyer's private belief and can be computed independently of the price of the experiment. The seller's problem is then to design and price different versions of experiments, that is, different information products from the same underlying database.

Indeed, Bergemann et al. [8] find that in the binary action and state setting it is often optimal to simply sell the complete information to the buyers and identify the optimal monopoly price. This is for example the case when the distribution of the prior beliefs, the types of the buyers, is symmetric around the uniform prior. Yet, they find that sometimes it can be beneficial to offer two information goods for sale, one that indeed offers complete information, but also an additional one that offers only partial information. Thus, a menu of information goods sometimes dominates the sale of a single object, namely the complete information.

Given the proximity of the problem of selling information to the problem of selling a divisible good with a unit capacity, the superiority of a menu over a single choice may appear to be surprising. After all, in the single good problem, Riley and Zeckhauser [35] have famously established that selling the entire unit at an optimal price is always an optimal policy. When we consider the problem of selling information, we find that the utility of the buyer is also linear in the posterior probability, just as it is linear in the quantity in the aforementioned problem. Thus the emergence of a menu rather than a single item appears puzzling. The difference is however that the utility of the buyer is only piecewise linear. In particular, in the binary state binary action setting, it displays exactly one kink in the interior of the prior probability. The kink emerges at an interior point of the prior belief where the decision-maker is indifferent between the two actions that are at his disposal. This particular prior indeed identifies the type of the decision-maker who has the highest willingness-to-pay for complete information. Moving away from this point, the utility is then linear in the prior probability. The kink is foremost an expression of the value of information for the decision-maker. But following the utility descent away from the kink, the value of information is decreasing linearly, and thus the seller faces not one, but two endpoints at which the participation constraints of the buyer have to be satisfied.

Naturally, with complete information about binary states, there will be two actions, one for each state which will lead to the highest utility of the decision-maker. Thus, we might expect that the cardinality of the optimal menu remains at most binary when we allow the agents to have a larger choice or action set but stay with binary state space that represents the uncertainty. In this paper we pursue this question and find that *the cardinality of the optimal menu increases at least linearly with the number of available actions*. Thus, selling information forces us to consider larger menus even when the space of uncertainty remains binary, and thus the utility as a function of the one-dimensional prior remains piecewise linear everywhere. Moreover, the cardinality of a set of optimal actions will always remain at two. Thus, the nature of selling information is high-dimensional even when the underlying state and ex-post optimal action space remains binary and thus small.

The main idea behind the revenue-maximizing mechanism for the information seller is akin to offering "damaged goods" to low-value buyers. However, when selling information goods (see [38]), product versioning allows for richer and more profitable distortions than with physical goods. This is due to a peculiar property of information products: Because buyers value different dimensions (i.e., information about specific state realizations), the buyers with the lowest willingness to pay also have very specific preferences. For example, in the context of credit markets, very aggressive lenders are interested in very negative information only, and are willing to grant a loan otherwise. The seller can thus leverage the key insight that information is only valuable if it changes optimal actions—to screen the buyer's private information.

1.1 Our Results and Techniques

We first study the environment when the state is binary, investigating the ratio between the optimal revenue (denoted as OPT) and the revenue that can be attained with the sale of complete information (denoted as FRev). [8] showed that in the environment with a binary state and a binary action space, $\frac{\text{OPT}}{\text{FRev}} \leq 2$. What happens when there are more actions? Since there are only two states, the agent has a single dimensional preference. Conventional wisdom from the monopoly pricing problem suggests that the ratio should be no more than a fixed constant. Nonetheless, in the paper we show that the ratio $\frac{\text{OPT}}{\text{FRev}}$ is $\Theta(m)$, which is neither a fixed constant nor a function that scales with the number of agent types. Here m is the number of actions.

Our first main result, Theorem 1, shows that the revenue obtainable with the sale of complete information is only a fraction $1/\Omega(m)$ of the optimal revenue. Using Theorem 1 as a springboard, our second main result, Theorem 2, shows that the cardinality of the optimal menu is at least $\Omega(m)$. The reason is that any menu with ℓ experiments can obtain no more than ℓ times the revenue attained with the sale of complete information (Lemma 3). To prove Theorem 1, we explicitly construct an environment with m actions, where selling the complete information yields low revenue. However, characterizing the optimal menu appears to be challenging for this environment, and even providing any incentive compatible mechanism with high revenue seems to be non-trivial. We take an indirect approach by first constructing an approximately incentive compatible mechanism that has a revenue at least $\Omega(m)$ times larger than the revenue attained with the sale of complete information, then converting the mechanism to a menu with negligible revenue loss.

In Theorem 3, we show that selling the complete information can always obtain at least a fraction $1/O(m)$ of the optimal revenue, matching the lower bound in Theorem 1 up to a constant factor. This result is established by considering a relaxed problem for the optimal revenue maximization, where we only keep a subset of the original incentive constraints. We show that the cardinality of the optimal menu in the relaxed problem is always $O(m)$, thus Theorem 3 follows from Lemma 3. Note that we only show that the relaxed problem under which the optimal revenue can be obtained with $O(m)$ experiments. It remains an open question whether the cardinality of the optimal menu in the original problem grows linearly in the number of available actions, or perhaps exceeds linear growth in the feasible actions.

These results show that the optimal mechanism to sell information or data are likely high-dimensional even when the underlying decision problem is low-dimensional. By tailoring the information to different decision problems the seller can extract substantially more revenue from the decision-maker as if he were to rely on a simple mechanism. More specifically, for any constant $c > 0$ and any fixed finite menu, there exists a binary-state environment such that the revenue attained by the menu is no more than a c-fraction of the optimal revenue (Corollary 1). Nonetheless, we provide in Theorem 4 conditions under which it is indeed optimal to only sell the largest possible amount of information, akin to the single unit monopoly problem. In Examples 1, 2, and 3 we apply this condition to various parametrized classes of information problems.

Our results above provide a complete understanding of $\frac{\text{OPT}}{\text{FRev}}$ in the binary-state environments. What happens if there are more than two states? In fact, we show that the ratio becomes substantially larger when the agent has a multi-dimensional preference. We consider arguably the simplest multi-dimensional environment, where there are 3 states and 3 actions, and the buyer receives payoff 1 if they match their action j with the state i ($j = i$) and receives payoff 0 otherwise. We refer to this as the three state *matching utility* environment. We show in Theorem 5 that in this multi-dimensional environment, the ratio $\frac{\text{OPT}}{\text{FRev}}$ can scale polynomially with the number of agent types N. In particular, $\frac{\text{OPT}}{\text{FRev}} = \Omega(N^{1/7})$. With Lemma 3, the result also implies that the optimal menu contains at least $\Omega(N^{1/7})$ experiments. The proof is adapted from the approach in [29], which was originally used in

the multi-item auction problem. Their approach does not directly apply to our problem due to both the non-linearity of the buyer's value function and the more demanding incentive compatibility constraints in our setting. In the proof, we construct a sequence of types placed on a sequence of concentric thin circular sectors. [29] placed all types in a sequence of complete circles. We place the types in these carefully picked circular sectors, so that the constructed mechanism satisfies the more demanding incentive compatibility constraints. Then we construct a discrete distribution on those types, so that the ratio $\frac{\text{OPT}}{\text{FREV}}$ is large.

An alternative interpretation of the result is that: There does not exist a universal finite upper bound on either the cardinality of the optimal menu or the ratio $\frac{\text{OPT}}{\text{FREV}}$ that holds for all possible type distributions. To complement this result, in Theorem 7, we show that in the same environment – matching utility with 3 states/actions, selling the complete information is indeed optimal if the distribution is uniform. To prove the result, we propose another relaxation of the problem optimizing over the agent's utility functions. We then construct a dual problem in the form of optimal transportation [18, 36, 40]. To verify the optimality of selling complete information, we construct a feasible dual that satisfies the complementary slackness conditions. Although Theorem 7 focuses on a special matching utility environment with 3 states, our relaxation provides a general approach to certify the optimality of any specific menu.

The paper is organized as follows. In Section 2, we introduce the model and notations used in the paper. In Section 3 we prove lower bounds for $\frac{\text{OPT}}{\text{FREV}}$ and size of the optimal menu in the binary-state case. In Section 4 we upper bound $\frac{\text{OPT}}{\text{FREV}}$, matching the lower bound proved in the previous section. Then we provide a sufficient condition under which selling complete information is optimal. In Section 5 we consider the environment with $n \geq 3$ states. In Section 6 we conclude the paper.

1.2 Additional Related Work

Unlike selling a single unit of item, the monopolist pricing problem becomes much more involved in multi-dimensional settings. Complete characterizations are known only in several special cases [18–20, 24–27]. A recent line of work provides simple and approximately-optimal mechanisms [4, 11, 13–15, 22, 28, 30, 41]. Our results (Theorem 1 and Theorem 3) provide matching upper and lower bounds for the performance of selling the complete information, arguably the simplest mechanism for selling information, in single-dimensional settings. In sharp contrast to the monopolist pricing problem, our Theorem 2 indicates that the optimal solution for selling information is unlikely to be simple even in single-dimensional settings.

Babaioff et al. [5] studies a problem that is related to ours. They consider a model where a seller knows the state of the world ω and sells the information to a buyer who has a private type. However, there are two major differences between their model and ours. Firstly, our analysis considers the direct sale of information. Here contracting takes place at the ex ante stage: The buyer purchases an information structure (i.e., a Blackwell experiment), as opposed to paying for specific realizations of the seller's informative signals. By contrast, Babaioff et al. [5] studies a model of data lists (i.e., pricing conditional on signal realizations). Secondly, their paper considers a different model of uncertainty and information. In particular, the ex-post payoff function of the buyer depends on two random variables, namely ω, which is the state of the world about which the seller has information, and θ which is payoff type that the buyer has information about. This two-dimensional representation of the uncertainty differs from our setting. In our setting, the payoff function of the buyer depends only on the state of the world, and both the buyer and the seller have some, partial or complete, information about the state. In our view this is the canonical model of decision making under uncertainty where the buyer seeks to complement (or augment) his information by additional information from the seller. Finally, our model assumes that the buyer's private type

θ is his private prior belief about the underlying state ω. Babaioff et al. [5] shows that when the correlation between θ and ω is complex enough, the optimal mechanism can extract all surplus, as in Crémer and McLean [17]. However, their result does not seem to apply to the specific type of correlation between ω and θ in our model. To sum up, the models of [5] and ours are not nested, and neither is a special case of the other. Chen et al. [16] study a similar setting to Babaioff et al. [5] where the buyer has budget constraints. Bergemann and Bonatti [7] consider the trade of information bits ("cookies") that are an input to a decision problem.

Moreover, there are some recent works studying the revenue-optimal mechanism in different information-selling models. Liu et al. [32] characterizes the revenue-optimal mechanism when the buyer has a linear value function on the scalar-valued state and action. A recent paper by Li [31] studies the case where the agent has endogenous information, meaning that she can perform her own experiment at a certain cost after receiving the signal. Finally, Cai and Velegkas [12] study the same problem as ours but focus on efficient algorithms to compute the optimal menu for discrete type distributions. We consider general type distributions and study the cardinality of the optimal menu as well as the performance of selling the complete information.

2 PRELIMINARIES

Model and Notation. A data buyer (also referred to as the agent) faces a decision problem under uncertainty. The state of the world ω is drawn from a state space $\Omega = \{\omega_1, \ldots, \omega_n\}$. For each $i \in [n]$, we refer to state ω_i as state i for simplicity.[1] The buyer chooses an action a from a finite action space A. We use m to denote the size of A and let $A = [m]$. For every $i \in [n]$, $j \in [m]$, the buyer's payoff for choosing action j under state i is defined to be u_{ij}. Denote $\mathcal{U}$ the $n \times m$ payoff matrix that contains all u_{ij}'s. The buyer has *matching utility* payoff if $n = m$ and $\mathcal{U}$ is an identity matrix ($u_{ij} = 1$ if $i = j$ and 0 otherwise).

The buyer has some prior information about the state of the world, which is captured by a distribution that represents the probability that the buyer assigns to each of the states. We call this piece of prior information the *type* of the buyer, denoted by $\theta = (\theta_1, \ldots, \theta_{n-1})$, where θ_i represents the probability that the buyer assigns to the state of the world ω_i for each $i \in [n-1]$ (and a probability of $1 - \sum_{i=1}^{n-1} \theta_i$ for state ω_n). Denote $\Theta = \{\theta = (\theta_1, \ldots, \theta_{n-1}) \in [0,1]^{n-1} \mid \sum_{i=1}^{n-1} \theta_i \leq 1\}$ the space of θ. When the state space is binary, θ is a scalar in $[0,1]$. For ease of notation, denote $\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$ throughout the paper.

The type θ is distributed according to some known probability distribution function F. We use $\text{Supp}(F)$ to denote the support of distribution F. When F is a continuous probability distribution (or discrete probability distribution), we use f to denote its probability density function (or its probability mass function). Apart from the buyer, there is also a *seller* who observes the state of the world and is willing to sell supplemental information to the buyer.[2] We refer to the buyer as he and to the seller as she.

Experiment. The seller provides supplemental information to the buyer via a *signaling scheme* which we call *experiment*. A signaling scheme is a commitment to n probability distributions over a finite set of signals S, such that when the state of the world is realized, the seller draws a signal from the corresponding distribution and sends it to the buyer. According to [8], we can without loss of generality restrict our attention to the experiments whose signal set is the same as the action space $[m]$ (see Lemma 14 in Appendix A). One can think of every signal j as the seller recommending the buyer to choose action j. We denote such an experiment E by a $n \times m$ matrix.

---

[1]Throughout the paper, we denote $[k] = \{1, 2, \ldots, k\}$ for any integer $k \geq 1$.
[2]It is not crucial to assume that the seller knows the state of the world, and it suffices to assume that the seller can send signals that are correlated with the state of the world.

For every $i \in [n], j \in [m]$, the (i, j)-th entry, denoted as $\pi_{ij}(E)$ or π_{ij} if the experiment is clear from context, is the probability that experiment E sends signal j when the state of the world is ω_i. It satisfies that $\sum_{j \in [m]} \pi_{ij}(E) = 1, \forall i \in [n]$.

We call an experiment E *fully informative* if the seller completely reveals the state of the world by sending signals according to E. This means that for every state of the world the experiment recommends the action that yields the highest payoff under this state, i.e. for every $i \in [n]$, $\pi_{ij}(E) = 1$ if $j = \arg \max_k u_{ik}$ and 0 otherwise. [3]

The Value of an Experiment. To understand the behavior of the buyer, we first explain how the buyer evaluates an experiment. Without receiving any additional information from the seller, the buyer's best action is $a(\theta) \in \arg \max_{j \in [m]} \{\sum_{i \in [n]} \theta_i u_{ij}\}$, and his maximum expected payoff is $u(\theta) = \max_{j \in [m]} \{\sum_{i \in [n]} \theta_i u_{ij}\}$. If he receives extra information from the seller, he updates his beliefs and may choose a new action that induces higher expected payoff. After receiving signal $k \in [m]$ from experiment E his posterior belief about the state of the world is: $\Pr[\omega_i | k, \theta] = \frac{\theta_i \pi_{ik}}{\sum_{\ell \in [n]} \theta_\ell \pi_{\ell k}}, \forall i \in [n]$. The best action is $a(k|\theta) = \arg \max_{j \in [m]} \left\{ \frac{\sum_{i \in [n]} \theta_i \pi_{ik} u_{ij}}{\sum_{i \in [n]} \theta_i \pi_{ik}} \right\}$, which yields maximum expected payoff $u(k|\theta) = \max_{j \in [m]} \left\{ \frac{\sum_{i \in [n]} \theta_i \pi_{ik} u_{ij}}{\sum_{i \in [n]} \theta_i \pi_{ik}} \right\}$. Taking the expectation over the signal the buyer will receive, we define the value of the experiment E for type θ to be $V_\theta(E) = \sum_{k \in [m]} \max_{j \in [m]} \left\{ \sum_{i \in [n]} \theta_i \pi_{ik} u_{ij} \right\}$.

Mechanism. Any mechanism $\mathcal{M}$ can be described as $\{(E(\theta), t(\theta))\}_{\theta \in \Theta}$, where $E(\theta)$ is the experiment type θ purchases and $t(\theta)$ is the payment. The interaction between the seller and the buyer in any mechanism works as follows:

(1) The seller commits to a mechanism $\mathcal{M} = \{E(\theta), t(\theta)\}_{\theta \in \Theta}$.
(2) The state of the world ω_i $(i \in [n])$ and the type of the buyer θ are realized.
(3) The buyer reports his type θ to the mechanism.
(4) The seller sends the buyer a signal $k \in [m]$ with probability $\pi_{ik}(E(\theta))$.
(5) The buyer chooses an action $j \in [m]$, based on his type θ and the signal k. He receives payoff u_{ij}, and pays $t(\theta)$ to the seller.

In subsequent sections, sometimes we abuse the notation and denote the experiment $E(\theta)$ as $\mathcal{M}(\theta)$. We assume that the buyer is quasilinear, i.e. he wants to maximize his utility – the maximum expected payoff minus the payment.

Incentive Compatibility. A mechanism is *Incentive Compatible* (IC) if reporting his type θ truthfully maximizes his expected utility: $V_\theta(E(\theta)) - t(\theta) \geq V_\theta(E(\theta')) - t(\theta'), \forall \theta, \theta' \in \Theta$. For any $\varepsilon > 0$, a mechanism is ε-IC if the inequality is violated by at most ε. Given any mapping $\sigma : [m] \to [m]$, denote $V_\theta^{(\sigma)}(E)$ the value of the experiment E, if the buyer chooses action $\sigma(j)$ whenever he receives signal j. Formally, $V_\theta^{(\sigma)}(E) = \sum_{i \in [n]} \sum_{j \in [m]} \theta_i \pi_{ij}(E) \cdot u_{i,\sigma(j)}$. Hence we have $V_\theta(E) = \max_\sigma V_\theta^{(\sigma)}(E)$. Denote $V_\theta^*(E)$ the value of the experiment E, if the buyer follows the recommendation of the seller, i.e., σ is an identity mapping. According to [8], we can without loss of generality restrict our attention to the mechanisms such that if the buyer reports truthfully, following the recommendation from the seller maximizes his expected payoff (see Lemma 14). In those mechanisms, $V_\theta(E(\theta)) = V_\theta^*(E(\theta))$. Then IC constraints are equivalent to:

$$V_\theta^*(E(\theta)) - t(\theta) \geq V_\theta^{(\sigma)}(E(\theta')) - t(\theta'), \forall \theta, \theta' \in \Theta, \sigma : [m] \to [m] \qquad (1)$$

---

[3]We break ties lexicographically.

Individually Rationality. A mechanism is *Individual Rational* (IR) if reporting his type θ truthfully induces expected utility at least $u(\theta)$: $V_\theta(E(\theta)) - t(\theta) \geq u(\theta), \forall \theta \in \Theta$. We remark that the agent has expected payoff $u(\theta)$ before receiving any additional information. Thus IR constraints guarantee that the agent has a non-negative utility surplus by participating in the mechanism. Any IC and IR mechanism can be described as a menu $\mathcal{M} = \{(E, t(E))\}_{E \in \mathcal{E}}$. The buyer with any type θ chooses the experiment E that maximizes $V_\theta(E) - t(E)$. The pair $(E, t(E))$ consisting of an experiment E and its price $t(E)$ is called an *option*. The option enables the data-buyer to improve his information and consequently improve his decision. A menu is called fully informative if it only contains the fully informative experiment. We also refer to it as "selling complete information".

Revenue. An *environment* is a particular choice of parameters of the model (i.e. the payoff matrix and type distribution). Fix an environment, we denote by REV($\mathcal{M}$) the revenue that mechanism $\mathcal{M}$ generates and by OPT the optimal revenue among all IC, IR mechanisms. We denote by FREV the maximum revenue achievable by any fully informative menu.

Payoff Matrix in Binary States. When the state space is binary, we can without loss of generality make some assumptions on the payoff matrix. For any action j, we say the action is *redundant*, if there exists a set of actions $S \subseteq [m] \backslash \{j\}$ and a distribution $d = \{d_k\}_{k \in S}$ such that: $u_{ij} \leq \sum_{k \in S} d_k u_{ik}, \forall i \in \{1, 2\}$. In this scenario, the buyer will never choose action j as it is dominated by choosing an action according to distribution d regardless of the underlying state. Without loss of generality, we assume that none of the actions is redundant, which clearly implies that there does not exist $k \neq j$ such that $u_{1k} \geq u_{1j}, u_{2k} \geq u_{2j}$. Thus we can assume that

$$u_{11} > u_{12} > ... > u_{1m} = 0 \quad \text{and} \quad 0 = u_{21} < u_{22} < ... < u_{2m},$$

otherwise we can change the action indexes, or modify every u_{1j} (or u_{2j}) by the same amount. We further assume $u_{11} \geq u_{2m} = 1$, otherwise we can scale all payoffs or swap the states. Under the assumption above, the fully informative experiment E satisfies that $\pi_{11}(E) = \pi_{2m}(E) = 1$.

(a) Payoff Matrix						(b) Experiment					
u	1	$\cdots$	j	$\cdots$	m	E	1	$\cdots$	j	$\cdots$	m
ω_1	u_{11}	$\cdots$	u_{1j}	$\cdots$	0	ω_1	π_{11}	$\cdots$	π_{1j}	$\cdots$	π_{1m}
ω_2	0	$\cdots$	u_{2j}	$\cdots$	$u_{2m} = 1$	ω_2	π_{21}	$\cdots$	π_{2j}	$\cdots$	π_{2m}

3 BINARY STATE: A LOWER BOUND FOR SELLING COMPLETE INFORMATION

We focus on the binary-state case ($n = 2$) in Section 3 and Section 4. In this section, we show that for any m, there exists some environment with 2 states and m actions such that selling the complete information is only an $\frac{1}{\Omega(m)}$-fraction of the optimal revenue OPT. Proofs in this section are postponed to Appendix B.

In fact, we prove a stronger statement that in this environment, FREV is an $\frac{1}{\Omega(m)}$-fraction of the maximum revenue among a special class of menus called *semi-informative menu*. In particular, an experiment E is semi-informative if it satisfies: (i) It only recommends the fully informative actions (Action 1 and m), i.e., $\pi_{ij}(E) = 0$, for all $i \in \{1, 2\}, 2 \leq j \leq m - 1$; (ii) Either $\pi_{11}(E) = 1$ or $\pi_{2m}(E) = 1$. Since $\sum_{j \in [m]} \pi_{1j}(E) = \sum_{j \in [m]} \pi_{2j}(E) = 1$, any experiment E that satisfies both of the above properties has one of the following patterns:

A mechanism (or menu) is called semi-informative if every experiment in the mechanism (or menu) is semi-informative. Denote SIREV the optimal revenue achieved by any semi-informative menu. Clearly OPT $\geq$ SIREV $\geq$ FREV since the fully informative experiment is also semi-informative. The class of semi-informative menus is also useful in Section 4.

(a) Pattern 1: $\pi_{2m}(E) = 1$

E	1	$\cdots$	j	$\cdots$	m
ω_1	π_{11}	$\cdots$	0	$\cdots$	$1 - \pi_{11}$
ω_2	0	$\cdots$	0	$\cdots$	1

(b) Pattern 2: $\pi_{11}(E) = 1$

E	1	$\cdots$	j	$\cdots$	m
ω_1	1	$\cdots$	0	$\cdots$	0
ω_2	$1 - \pi_{2m}$	$\cdots$	0	$\cdots$	π_{2m}

Table 2. Two Specific Patterns of the Semi-informative Experiment

THEOREM 1. *For every m, there exists a payoff matrix with 2 states and m actions, together with a type distribution F of the agent, such that SIREV = $\Omega(m) \cdot$ FREV, which implies that OPT = $\Omega(m) \cdot$ FREV.*

To prove the theorem, we first introduce the concept of IR curve in the binary-state case.

IR Curve. In binary-state case, $u(\cdot)$ can be viewed as a function in $[0, 1]$, which we refer to as the *IR curve*. By definition, $u(\cdot)$ is a maximum over m linear functions. Thus, it is a continuous, convex and piecewise linear function. Since none of the actions is redundant, the IR curve $u(\cdot)$ contains exactly m pieces. We show in Table 3 an example with 4 actions and in Figure 1 the corresponding IR curve.

u	1	2	3	4
ω_1	1	0.8	0.6	0
ω_2	0	0.5	0.8	1

Table 3. The payoff matrix of an example with 2 states and 4 actions
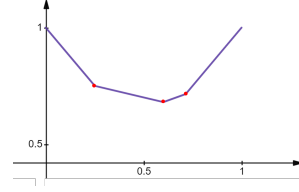
Fig. 1. The IR curve $u(\cdot)$. The slopes of the 4 pieces are $-1, -0.2, 0.3$ and 1 respectively.

LEMMA 1. *Given any $a, b \in (0, 1)$ such that $a < b$, let $h : [a, b] \to (0, 1)$ be any piecewise linear function that is strictly decreasing and convex. Consider any environment with IR curve h. [4] Then there exists a continuous distribution F over support $[a, b]$, such that FREV = $1 - h(a)$, and*

$$\frac{\int_a^b f(\theta) \cdot (1 - h(\theta)) d\theta}{FREV} = \log\left(\frac{1 - h(b)}{1 - h(a)}\right),$$

where f is the pdf of F. When $u_{11} = u_{2m} = 1$, the highest expected payoff an agent can achieve is 1, so $\int_a^b f(\theta) \cdot (1 - h(\theta)) d\theta$ is also the expected full surplus under distribution F.

We provide a sketch of the proof of Theorem 1 here. In the first step, we construct an environment with 2 states and m actions by constructing an IR curve then apply Lemma 1 to create a distribution F over the agent types to show that the ratio between the full expected surplus and FREV is $\Omega(m)$. In the second step, we indirectly construct a semi-informative menu whose revenue is a constant fraction of the total expected surplus, which implies that SIREV = $\Omega(m) \cdot$ FREV. In our construction, the payoffs satisfy $u_{11} = u_{2m} = 1$.

Construction of the IR curve: Given $m - 1$ types $0 < \theta_1 < \theta_2 < ... < \theta_{m-1} < 1$ which will be determined later, we consider the following IR curve $u(\cdot)$ on $[\theta_1, \theta_{m-1}]$. The IR curve is a piecewise

---

[4]When the type distribution has support $[a, b]$, we can restrict our attention to the IR curve at range $[a, b]$. In our construction, we construct a payoff matrix that implements the IR curve.

linear function with $m - 2$ pieces. For simplicity, for every $1 \le i \le m - 2$, we will refer to piece L_i as the piece for the corresponding action $m - i$.[5] Fix any ε less than $\frac{1}{2^m}$.

Let $d_0 := \varepsilon^{2^m}$. For every $1 \le i \le m - 2$, let $d_i := 2^i \cdot \varepsilon^{2^m - 2^i}$. For every $1 \le i \le m - 1$, let $\theta_i := \sum_{j=0}^{i-1} d_j$. Then when $\varepsilon < \frac{1}{2^m}$,

$$\theta_i = \varepsilon^{2^m} + \sum_{j=1}^{i-1} 2^j \cdot \varepsilon^{2^m - 2^j} < \varepsilon^{2^m - 2^{i-1} - 1} \tag{2}$$

For every $1 \le i \le m - 2$, piece L_i is in the range $\theta \in [\theta_i, \theta_{i+1}]$. The slope of piece L_i is $l_i := -\varepsilon^{2^i}$. Thus the height of piece L_i is $h_i := u(\theta_{i+1}) - u(\theta_i) = -l_i d_i = 2^i \cdot \varepsilon^{2^m}$. Let $u(\theta_1) := 1 - \theta_1$ and $h_0 := \theta_1 = d_0 = \varepsilon^{2^m}$. Then for every $1 \le i \le m - 1$, we have $u(\theta_i) = 1 - \sum_{j=0}^{i-1} h_j$. Since $\varepsilon < \frac{1}{2^m}$, $u(\theta_i) > 0, \forall i$. See Figure 2 for an illustration of notations.

Now we compute the payoff matrix according to the constructed IR curve $u(\cdot)$. $u_{11} = u_{2m} = 1$, $u_{1m} = u_{21} = 0$. For every $1 \le i \le m - 2$, before receiving any additional information, choosing action $m - i$ induces expected payoff $\theta \cdot u_{1,m-i} + (1 - \theta) \cdot u_{2,m-i}$. By the definition of the IR curve, it must go through the two points $(\theta_i, u(\theta_i))$ and $(\theta_{i+1}, u(\theta_{i+1}))$. Thus we have $\theta_i \cdot u_{1,m-i} + (1 - \theta_i) \cdot u_{2,m-i} = 1 - \sum_{j=0}^{i-1} h_j$ and $\theta_{i+1} \cdot u_{1,m-i} + (1 - \theta_{i+1}) \cdot u_{2,m-i} = 1 - \sum_{j=0}^{i} h_j$. Thus $u_{2,m-i} = 1 - \sum_{j=0}^{i-1} h_j - \theta_i l_i$, $u_{1,m-i} = 1 - \sum_{j=0}^{i-1} h_j - \theta_i l_i + l_i$.

By Lemma 1, there exists a distribution F under support $[\theta_1, \theta_{m-1}]$, such that FREV $= 1 - u(\theta_1) = \theta_1 = \varepsilon^{2^m}$, and

$$\int_{\theta_1}^{\theta_{m-1}} f(\theta) \cdot (1 - u(\theta)) d\theta \ge \log\left(\frac{1 - u(\theta_{m-1})}{1 - u(\theta_1)}\right) \cdot \text{FREV} = \text{FREV} \cdot \log\left(\frac{\sum_{i=0}^{m-2} h_i}{h_0}\right)$$

$$= \text{FREV} \cdot \log\left(\frac{\sum_{i=0}^{m-2} 2^i \cdot \varepsilon^{2^m}}{\varepsilon^{2^m}}\right) = \text{FREV} \cdot \log\left(2^{m-1} - 1\right) \ge \text{FREV} \cdot \log(2) \cdot (m - 2) \tag{3}$$

Next, we present a semi-informative mechanism $\mathcal{M}$ whose revenue is comparable to the integral $\int_{\theta_1}^{\theta_{m-1}} f(\theta) \cdot (1 - u(\theta)) d\theta$.

Construction of the mechanism: Consider the following $m - 2$ semi-informative experiments. For every $1 \le i \le m - 2$, we design experiment E_i and its price p_i so that the buyer's utility after purchasing experiment E_i coincides with the piece L_i of $u(\cdot)$ for $\theta \in [\theta_i, \theta_{i+1}]$, when the buyer follows the recommendation. Formally, experiment E_i is as follows:

Fig. 2. Notations in the construction of IR curve
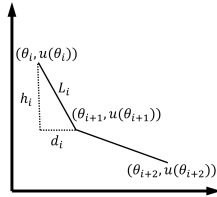
E_i	1	$\cdots$	m
ω_1	$1 + l_i$	0	$-l_i$
ω_2	0	0	1

Table 4. Experiment E_i

The price for experiment E_i is $p_i = \sum_{j=0}^{i-1} h_j + \theta_i l_i$. By definition of h_j and l_j, $p_i = \sum_{j=0}^{i-1} d_j \cdot (\varepsilon^{2^j} - \varepsilon^{2^i}) > 0$. It is not hard to verify that the buyer's utility for buying experiment E_i is $u(\theta_i)$ at θ_i and $u(\theta_{i+1})$ at θ_{i+1}.

---

[5]We choose the subscript this way so that the IR curve goes from a piece with smaller index to a piece with larger index as θ increases.

Consider the following mechanism $\mathcal{M}$. For every $\theta \in [\theta_1, \theta_{m-1}]$, let i be the unique number such that $\theta_i < \theta \leq \theta_{i+1}$.[6] The outcome and payment of $\mathcal{M}$ under input θ are defined as follows: It computes the buyer's utilities for (E_i, p_i), (E_{i-1}, p_{i-1}) (when $i \geq 2$), and (E_{i-2}, p_{i-2}) (when $i \geq 3$) under type θ, and chooses the option with the highest utility as the experiment and payment.

We prove in Lemma 2 that $\mathcal{M}$ is IR and δ-IC for some $\delta = o(\varepsilon^{2^m})$. Moreover, $\text{Rev}(\mathcal{M})$ is comparable to the integral $\int_{\theta_1}^{\theta_{m-1}} f(\theta) \cdot (1 - u(\theta)) d\theta$. $\mathcal{M}$ can then be converted to a semi-informative menu by losing no more than a constant fraction of the total revenue (see Lemma 16 in Appendix B). Since $\text{FRev} = \varepsilon^{2^m}$, Theorem 1 then follows from Inequality (3). The proofs of Lemma 2 and Theorem 1 are postponed to Appendix B.

LEMMA 2. *For any $\varepsilon \in (0, 2^{-m})$, $\mathcal{M}$ is IR and δ-IC where $\delta = 7 \cdot \varepsilon^{2^m+1}$. Moreover, $\text{Rev}(\mathcal{M}) \geq$* $\frac{1}{9} \cdot \int_{\theta_1}^{\theta_{m-1}} f(\theta) \cdot (1 - u(\theta)) d\theta$.

An important application of Theorem 1 is that, there is an $\Omega(m)$ lower bound of the size of the optimal menu. The main takeaway of this theorem is that, in our model where there is a single agent whose type is single-dimensional, the optimal menu, however, can be complex.

THEOREM 2. *For every m, there exists a payoff matrix with 2 states and m actions, together with a type distribution F, such that any optimal menu $\mathcal{M}^*$ consists of at least $\Omega(m)$ different experiments.*

To prove Theorem 2, it suffices to show that: Given any menu $\mathcal{M}$ with ℓ experiments, we can sell the complete information at an appropriate price to achieve revenue at least $\text{Rev}(\mathcal{M})/\ell$. In Lemma 3, we prove a more general result that applies to any IR (not necessarily IC) mechanism, which is useful in Section 4. A similar lemma also appears in [29] for the bundling mechanism in multi-item auctions.

DEFINITION 1. *For any finite integer $\ell > 0$, we say that a mechanism $\mathcal{M} = \{(E(\theta), t(\theta))\}_{\theta \in [0,1]}$ (not necessarily IC nor IR) has option size ℓ, if there exists ℓ different options $\{(E_j, t_j)\}_{j \in [\ell]}$ such that: For every $\theta \in [0, 1]$, there exists some $j \in [\ell]$ that satisfies $E(\theta) = E_j$ and $t(\theta) = t_j$.*

LEMMA 3. *For any positive integer ℓ, let $\mathcal{M}$ be any IR mechanism of option size ℓ that generates revenue $\text{Rev}(\mathcal{M})$. Then there exists a menu $\mathcal{M}'$ that contains only the fully informative experiment and generates revenue at least $\frac{\text{Rev}(\mathcal{M})}{\ell}$.*

Another important implication of Theorem 1 is shown in Corollary 1. It states that there is no menu with finite cardinality that achieves any finite approximation to the optimal revenue for all single-dimensional environments. The proof directly follows from Theorem 1 and Lemma 3.

COROLLARY 1. *For any finite $\ell > 0$, any menu $\mathcal{M}$ with size at most ℓ, and any $c > 0$, there exists an finite integer m, a payoff matrix with 2 states and m actions, together with a distribution F, such that $\text{Rev}(\mathcal{M})$ is at most a c-fraction of the optimal revenue in this environment.*

4 BINARY STATE: AN UPPER BOUND FOR SELLING COMPLETE INFORMATION

In this section, we provide upper bounds for the gap between the optimal revenue OPT and the revenue by selling complete information FRev. In Section 4.1, we prove an $O(m)$-approximation to the optimal menu using only the fully informative experiment, where m is the number of actions. In Section 4.2, we provide conditions under which selling complete information is the optimal menu. Proofs in this section are postponed to Appendix C and Appendix D.

---

[6] Choose $i = 1$ when $\theta = \theta_1$. It's a point with 0 measure, so it won't affect the revenue of the mechanism.

4.1 An Upper Bound for Selling Complete Information

The main result of this section is stated in Theorem 3. Together with Theorem 1, we know that the $O(m)$ approximation ratio is tight up to an absolute constant factor.

THEOREM 3. *For any environment with 2 states and m actions, there is a menu that contains only the fully informative experiment, whose revenue is at least $\Omega(\frac{OPT}{m})$. In other words, OPT = $O(m) \cdot$ FREV.*

To prove Theorem 3, a natural idea is to first show that the optimal menu contains $O(m)$ experiments, then use Lemma 3 to argue that we can sell the full information at an appropriate price to achieve revenue at least $\Omega\left(\frac{\text{REV}(\mathcal{M})}{m}\right)$. Unfortunately, we are not aware of such an upper bound on the size of the optimal menu, and it is not clear if the size of the optimal menu is indeed linear in m. Instead, we drop some of the IC constraints and consider the maximum revenue of a relaxed problem.

We first introduce the concept of responsive-IC and σ-IC constraints. Recall the IC constraint: $V_\theta^*(E(\theta)) - t(\theta) \geq V_\theta^{(\sigma)}(E(\theta')) - t(\theta'), \forall \theta, \theta' \in \Theta, \sigma : [m] \rightarrow [m]$ where $V_\theta^*(E(\theta))$ is the agent's utility for receiving experiment $E(\theta)$ and following the recommendation. We distinguish the IC constraints by the mapping σ. When σ is the identity mapping, we refer to the constraint as the *responsive-IC* constraint, and when σ is any non-identity mapping, we refer to the constraint as the σ-IC constraint. A mechanism is *responsive-IC* if it satisfies all responsive-IC constraints.

As Lemma 3 applies to any IR mechanism, we drop the σ-IC constraints and bound the number of experiments offered by the optimal responsive-IC and IR mechanism by $O(m)$, which suffices to prove Theorem 3. An important component of our proof is Lemma 4. Denote C_2' the set of all semi-informative, responsive-IC and IR mechanisms. We prove in Lemma 4 that the maximum revenue achievable by any responsive-IC and IR mechanism (denoted as OPT*) can be achieved by a semi-informative mechanism.

LEMMA 4. *There exists $\mathcal{M}^* \in C_2'$ such that REV($\mathcal{M}^*$) = OPT*.*

By [8], any semi-informative experiment is determined by a single-dimensional variable $q(E) = \pi_{11} \cdot u_{11} - \pi_{2m} \cdot u_{2m}$ (see also Observation 3). Given a mechanism $\mathcal{M}$ where every experiment is semi-informative, for any $\theta \in [0, 1]$, we slightly abuse the notation and let $q(\theta) := q(\mathcal{M}(\theta))$. $\mathcal{M}$ can also be described as the tuple ($\mathbf{q} = \{q(\theta)\}_{\theta \in [0,1]}, \mathbf{t} = \{t(\theta)\}_{\theta \in [0,1]}$). In Lemma 5, we present a characterization of all $\{q(\theta)\}_{\theta \in [0,1]}$ that can be implemented with a responsive-IC and IR mechanism. The proof uses the payment identity and it is similar to Lemma 1 in [8], where there are only two actions (Action 1 and m) in their setting. The main difference is that in our lemma, the IR constraints are not implied by the monotonicity of $q(\cdot)$.

LEMMA 5. *Given $\mathbf{q} = \{q(\theta)\}_{\theta \in [0,1]}$, there exists non-negative payment rule $\mathbf{t}$ such that $\mathcal{M} = (\mathbf{q}, \mathbf{t})$ is a responsive-IC and IR mechanism if and only if*

(1) $q(\theta) \in [-u_{2m}, u_{11}]$ is non-decreasing in θ.
(2) $\int_0^1 q(x)dx = u_{11} - u_{2m}$.
(3) For every $\theta \in [0, 1]$, $u_{2m} + \int_0^\theta q(x)dx \geq u(\theta)$. Recall that $u(\theta)$ is the value of the agent with type θ without receiving any experiment.

Moreover, the payment rule $\mathbf{t}$ must satisfy that for every θ,

$$t(\theta) = \theta \cdot q(\theta) + \min\{u_{11} - u_{2m} - q(\theta), 0\} - \int_0^\theta q(x)dx \qquad (4)$$

By Lemma 5, for any $\mathbf{q}$ that can be implemented with some responsive-IC and IR mechanism $\mathcal{M}$, the revenue of $\mathcal{M}$ can be written as an integral of $q(\cdot)$ using Equation (4) (see the objective of the

program in Figure 3). We notice that for continuous distribution F, Property 3 of Lemma 5 corresponds to uncountably many inequality constraints. To bound the size of the optimal responsive-IC and IR mechanism, another important component is to show that all (IR) constraints in Property 3 can be captured by $O(m)$ constraints (Constraint (4) in Figure 3). By Lemma 5, the optimal semi-informative, responsive-IC and IR mechanism is captured by the optimization problem in Figure 3. A formal argument is shown in Lemma 6.

$$\textbf{sup} \int_0^1 [(\theta f(\theta) + F(\theta))q(\theta) + \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}]d\theta$$

s.t. (1) $q(\theta)$ is non-decreasing in $\theta \in [0, 1]$

(2) $q(0) \geq -u_{2m}, \quad q(1) \leq u_{11}$

(3) $\int_0^1 q(\theta)d\theta = u_{11} - u_{2m}$

(4) $\int_0^1 (q(x) - \ell_k) \cdot \mathbb{1}[q(x) \leq \ell_k]dx \geq u_{2,m+1-k} - u_{2m}, \quad \forall k \in \{2, 3, ..., m - 1\}$

Fig. 3. Maximizing Revenue over Responsive-IC and IR Mechanisms

LEMMA 6. *For any optimal solution q^* to the program in Figure 3, the mechanism $\mathcal{M}^*$ that implements q^* (Lemma 5) achieves the maximum revenue among all responsive-IC and IR mechanisms.* [7]

In Lemma 7, we show that there exists an optimal responsive-IC and IR mechanism whose option size is $O(m)$. By Lemma 6, it's equivalent to prove that there is an optimal solution q^* that takes only $O(m)$ different values. When F is a discrete distribution, we turn the program into a collection of LPs (see Figure 8 in Appendix C.4), so that the highest optimum among the collection of LPs corresponds to the optimum of the program in Figure 3. Each LP has variables that represent the difference of the q-value between two adjacent θs (as the types are discrete). The LP has only $O(m)$ constraints. By the Fundamental Theorem of linear programming, each LP has an optimal solution where at most $O(m)$ variables are strictly positive, which corresponds to a $\mathbf{q}$ that takes only $O(m)$ different values. For continuous distribution F, we prove the claim by approximating the continuous program with an infinite sequence of discrete programs. Theorem 3 then follows from Lemmas 3, 4 and 7.

LEMMA 7. *There exists a semi-informative, responsive-IC and IR mechanism that has option size at most $3m - 1$ (Definition 1) and obtains revenue $\text{Rev}(\mathcal{M}^*) = \text{OPT}^*$.*

4.2 When is Selling Complete Information Optimal?

We have proved a tight approximation ratio of $\Theta(m)$ for selling complete information that applies to all binary-state environments. A natural follow-up question is whether the approximation ratio becomes significantly smaller for special environments. In this section, we provide a sufficient condition for the environment, under which selling complete information achieves revenue equal to OPT^*, the maximum revenue achievable by any responsive-IC and IR mechanisms.[8] Note that it immediately implies that selling complete information is the optimal menu in this environment.

---

[7]There is a feasible solution that achieves the supremum of the program. See Claim 3.

[8]In the paper, we indeed prove a necessary and sufficient condition under which selling complete information is optimal among all responsive-IC and IR mechanisms (Theorem 9). However, the conditions are in abstract terms and requires further definitions. For the purpose of presentation, we state here the sufficient condition that is easy to verify.

Throughout this section, we consider continuous distributions and assume that the pdf $f(\cdot)$ of the agent's type distribution is strictly positive on $[0, 1]$ and differentiable on $(0, 1)$.

THEOREM 4. *For every $\theta \in [0, 1]$, let $\varphi^-(\theta) = \theta f(\theta) + F(\theta)$ and $\varphi^+(\theta) = (\theta - 1)f(\theta) + F(\theta)$. Suppose $\varphi^-(\cdot)$ and $\varphi^+(\cdot)$ are both monotonically non-decreasing. Suppose the payoff matrix satisfies that $\varphi^-(\frac{\widehat{p}}{u_{11}}) \geq \varphi^+(1 - \frac{\widehat{p}}{u_{2m}})$, where $\widehat{p} = \min\left\{\frac{(u_{11}-u_{12})u_{2m}}{u_{11}-u_{12}+u_{22}}, \frac{(u_{2m}-u_{2,m-1})u_{11}}{u_{2m}+u_{1,m-1}-u_{2,m-1}}\right\}$. Then selling complete information at any price p such that $\varphi^-(\frac{p}{u_{11}}) = \varphi^+(1 - \frac{p}{u_{2m}})$ achieves the maximum revenue achievable by any responsive-IC and IR mechanisms.*

Here is an interpretation of Theorem 4: Both $\varphi^-(\cdot)$ and $\varphi^+(\cdot)$ are also considered in [8]. Intuitively, they can be viewed as the agent's "virtual value" when $q(\theta) \leq u_{11} - u_{2m}$ and when $q(\theta) > u_{11} - u_{2m}$ respectively, in a semi-informative mechanism. Both virtual value functions being non-decreasing is a standard regularity condition on the type distribution. The theorem applies to stardard distributions such as uniform, exponential and Gaussian distributions. See Appendix D.2 for several examples.

To understand the condition on the payoff matrix, we point out that if selling complete information at some price p is the optimal responsive-IC and IR mechanism, then the buyer's utility function must intersect with the IR curve at the first and last piece (∗). Otherwise we can add some extra experiments to strictly increase the revenue, while maintaining the same utility function and the responsive-IC and IR property, contradicting with the optimality (see Lemma 20 in Appendix D.2 for details). By some simple calculation, $\widehat{p}$ is exactly the largest price that satisfies (∗). The condition guarantees the existence of such a price $p \leq \widehat{p}$.

Note that although the theorem applies to arbitrary number of actions, the condition itself only depends on the payoffs of the the first two and the last two actions. Thus if the condition is satisfied, selling complete information is always optimal for any choice of the payoffs for other actions.

To prove the theorem, we provide an exact characterization of the optimal semi-informative, responsive-IC and IR mechanism, i.e., the optimal solution $\mathbf{q}^* = \{q^*(\theta)\}_{\theta \in [0,1]}$ of the program in Figure 3, by Lagrangian duality (Theorem 10). It is a generalization of the characterization by Bergemann et al. [8] to $m \geq 3$ actions.

5 SELLING COMPLETE INFORMATION: MORE THAN TWO STATES

In this section, we consider the environment with $n \geq 3$ states. In Section 5.1, we prove that even in arguably the simplest environment – matching utility environment with 3 states and 3 actions, the optimal revenue OPT and the revenue by selling complete information FRev can have an arbitrarily large gap, i.e. there is no universal finite upper bound of the gap that holds for all type distributions. Nonetheless, we prove in Section 5.2 that in the same environment, if the distribution is uniform, selling complete information is in fact the optimal mechanism. Proofs in this section are postponed to Appendix E.

5.1 Lower Bound Example for Matching Utilities

To avoid ambiguity, throughout this section we denote Rev($\mathcal{M}, D$) the revenue of $\mathcal{M}$ with respect to D, for any (not necessarily IC or IR) mechanism $\mathcal{M}$, and any distribution D. Denote OPT(D) and FRev(D) the optimal revenue and the maximum revenue by selling complete information, respectively. The main result of this section shows that in the matching utility environment with $n = 3$ states and $m = 3$ actions, the ratio $\frac{\text{OPT}(D)}{\text{FRev}(D)}$ arbitrarily large for some distribution D. Recall that in a matching utility environment, $n = m$ and the payoffs satisfy $u_{ij} = \mathbb{1}[i = j], \forall i, j$.

THEOREM 5. *Consider the matching utility environment with 3 states and 3 actions. For any integer N, there exists a distribution D with support size N such that $\frac{OPT(D)}{FRev(D)} = \Omega(N^{1/7})$.*

The following corollary directly follows from Theorem 5 and Lemma 3.

COROLLARY 2. *Consider the matching utility environment with 3 states and 3 actions. For any integer N, there exists a distribution D with support size N such that any optimal menu has option size $\Omega(N^{1/7})$.*

The proof of Theorem 5 is adapted from the approach in [29], which was originally used in problem of multi-item auctions. Their approach does not directly apply to our problem due to the non-linearity of the buyer's value function and the existence of the extra σ-IC and more demanding IR constraints in our problem.

For any mechanism $\mathcal{M}$, denote RATIO($\mathcal{M}$) the largest ratio between REV($\mathcal{M}, D$) and FREV(D) among all distributions, i.e.,

$$\text{RATIO}(\mathcal{M}) = \sup_D \frac{\text{REV}(\mathcal{M}, D)}{\text{FREV}(D)}$$

For every $\theta \in \Theta$, denote $U(\theta)$ the gain in value of a buyer with type θ, after receiving the fully informative experiment. Formally, $U(\theta) = \sum_{i=1}^n \theta_i \cdot \max_j u_{ij} - \max_j \{\sum_i \theta_i u_{ij}\}$. Here $\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$. In Theorem 6, we prove that RATIO($\mathcal{M}$) can be written as a concrete term using $U(\cdot)$ and the payment $t(\cdot)$ of $\mathcal{M}$.

THEOREM 6. *(Adapted from Theorem 5.1 of [29]) For any mechanism $\mathcal{M} = \{E(\theta), t(\theta)\}_{\theta \in \Theta}$ (not necessarily IC or IR), we have $\text{RATIO}(\mathcal{M}) = \int_0^\infty \frac{1}{r(x)} dx$, where $r(x) = \inf\{U(\theta) : \theta \in \Theta \wedge t(\theta) \geq x\}, \forall x \geq 0$.*

Here is a proof sketch of Theorem 5. We first show in Lemma 8 that: Given any sequence $\{y_k\}_{k=1}^N$ of vectors in Θ that satisfy all properties in the statement, we can construct a distribution D with support size N, together with an IC, IR mechanism $\mathcal{M}$ such that the ratio $\frac{\text{REV}(\mathcal{M}, D)}{\text{FREV}(D)}$ has a lower bound that only depends on the sequence $\{y_k\}_{k=1}^N$. Properties 2 and 3 in the statement are carefully designed to guarantee that the constructed mechanism is IC and IR. Property 1 ensures the mechanism to obtain enough revenue. With Lemma 8, we then complete the proof of Theorem 5 by constructing a valid sequence $\{y_k\}_{k=1}^N$ which induces a lower bound of $\Omega(N^{1/7})$.

LEMMA 8. *Consider the matching utility environment with 3 states and 3 actions. Given any integer N and a sequence of types $\{y_k = (y_{k,1}, y_{k,2})\}_{k=1}^N$ in Θ such that*

(1) $gap_k = \min_{0 \leq j < k} \{(y_{k,1} - y_{j,1}) \cdot y_{k,1} + (y_{k,2} - y_{j,2}) \cdot y_{k,2}\} \in (0, 0.09)$. *Here $y_0 = (0, 0, 1)$.*

(2) $\frac{y_{k,1}}{y_{k,2}} \in [\frac{9}{10}, \frac{10}{9}]$.

(3) $\|y_k\|_2 = \sqrt{y_{k,1}^2 + y_{k,2}^2} \in [0.3, 0.4]$.

Then for any $\varepsilon > 0$, there exists a distribution D with support size N such that

$$\frac{OPT(D)}{FREV(D)} \geq (1 - \varepsilon) \cdot \sum_{k=1}^N \frac{gap_k}{y_{k,1} + y_{k,2}} \geq \frac{3}{2}(1 - \varepsilon) \cdot \sum_{k=1}^N gap_k$$

To complete the proof of Theorem 5, for any integer N, we construct a sequence of types $\{y_k\}_{k=1}^N$ that satisfies all three properties in the statement of Lemma 8, and $gap_k = \Theta(k^{-6/7})$. Then by Lemma 8, there exists a distribution D with support size N such that (by choosing $\varepsilon = \frac{1}{2}$) $\frac{OPT(D)}{FREV(D)} \geq \sum_{k=1}^N gap_k = \Omega(\sum_{k=1}^N k^{-6/7}) = \Omega(N^{1/7})$. The construction is adapted from Proposition 7.5 of [29]: All points $\{y_k\}_{k=1}^N$ are placed in a sequence of shells centered at $(0, 0)$ with radius within the range of $[0.3, 0.4]$ (for Property 3). All points are placed in a thin circular sector close to the $45°$ angle so that Property 2 is satisfied. See Appendix E.1 for the complete construction and proof.

5.2 Matching Utility Environment with Uniform Distribution

As the main result of this section, we complement the result in Section 5.1, by showing that in the matching utility environment with 3 states and 3 actions, selling complete information is indeed optimal if the type distribution is uniform.

THEOREM 7. *Consider the matching utility environment with $n = 3$ states and 3 actions, where the buyer has uniform type distribution. Then selling only the fully informative experiment at price $p = \frac{1}{3}$ achieves the maximum revenue among all IC, IR mechanisms.*

The remaining of this section is dedicated to the proof of Theorem 7. We first provide a high-level plan of the proof. In Section 4, we considered a relaxation of our problem, which finds the optimal responsive-IC and IR mechanism. Here we propose another relaxed problem of finding the optimal menu for any matching utility environment (Figure 4). Then we construct a dual problem (Figure 5) in the form of optimal transportation [18, 36, 40]. This primal-dual framework provides a general approach to prove the optimality of any menu $\mathcal{M}$: If we can construct a feasible dual that satisfies the complementary slackness conditions, then $\mathcal{M}$ is the optimal primal solution to the relaxed problem, which implies that $\mathcal{M}$ is the optimal menu. We apply this framework to the matching utility environment with 3 states and uniform type distribution, proving that selling complete information at $p = \frac{1}{3}$ is optimal.

5.2.1 Construction of the Primal and Dual Problem. By Lemma 14, we will focus on responsive mechanisms $\mathcal{M} = \{(E(\theta), t(\theta))\}_{\theta \in \Theta}$. For any $\theta \in \Theta$, denote $\pi_{ij}(\theta)$ the (i, j)-entry of experiment $E(\theta)$. We also use $\pi_i(\theta)$ to represent $\pi_{ii}(\theta)$. For every measurable set $S \subseteq \Theta$, denote $\text{VOL}(S) = \int_S 1d\theta$ the volume of S. We first prove the following lemma that applies to any matching utility environment.

LEMMA 9. *In any matching utility environment, there is an optimal responsive mechanism $\mathcal{M}$ such that: for every $\theta \in \Theta$, there exists $i \in [n]$ such that $\pi_i(\theta) = 1$.*

By Lemma 9, we focus on all mechanisms that are responsive and satisfy: for every $\theta \in \Theta$, there exists $i \in [n]$ such that $\pi_i(\theta) = 1$. Recall that $V_\theta(E)$ (or $V_\theta^*(E)$) is the buyer's value of experiment E (or the value if she follows the recommendation) at type θ. For any θ, denote $G(\theta) = V_\theta(E(\theta)) - t(\theta)$ the buyer's utility at type θ. Since the mechanism is responsive, we have

$$G(\theta) = V_\theta^*(E(\theta)) - t(\theta) = \sum_{i=1}^{n-1} \pi_i(\theta) \cdot \theta_i + \pi_n(\theta) \cdot \left(1 - \sum_{i=1}^{n-1} \theta_i\right) - t(\theta)$$

We prove the following observation using the IC constraints.

OBSERVATION 1. *For any $i \in [n-1]$, $\frac{\partial G(\theta)}{\partial \theta_i} = \pi_i(\theta) - \pi_n(\theta)$.*

Let $\nabla G(\theta) = \left(\frac{\partial G(\theta)}{\partial \theta_1}, \ldots, \frac{\partial G(\theta)}{\partial \theta_{n-1}}\right)$. By Observation 1 and the fact that $\max_{i \in [n]}\{\pi_i(\theta)\} = 1, \forall \theta \in \Theta$, we have

$$t(\theta) = -G(\theta) + \nabla G(\theta) \cdot \theta + \pi_n(\theta) = -G(\theta) + \nabla G(\theta) \cdot \theta + 1 - \max\left\{\frac{\partial G(\theta)}{\partial \theta_1}, \ldots, \frac{\partial G(\theta)}{\partial \theta_{n-1}}, 0\right\}$$

In Lemma 10, we prove a necessary condition under which a function $G(\cdot)$ is derived by some IC, IR and responsive mechanism.

DEFINITION 2. *For every type $\theta, \theta' \in \Theta$, define $c(\theta, \theta') = \max_E \{V_\theta(E) - V_{\theta'}(E)\}$, where the maximum is taken over all possible experiments with n states and n actions.*

LEMMA 10. *For any IC, IR and responsive mechanism $\mathcal{M}$, let $G(\cdot)$ be the buyer's utility function. Then $G(\cdot)$ is convex and satisfies $G(\theta) - G(\theta') \leq c(\theta, \theta'), \forall \theta, \theta' \in \Theta$. Moreover, $G(\mathbf{0}) = 1$.*

We notice that given any convex function G that satisfies $G(\theta) - G(\theta') \leq c(\theta, \theta'), \forall \theta, \theta'$. The function $\widehat{G}(\theta) = G(\theta) - G(\mathbf{0}) + 1$ satisfies all properties in Lemma 10. Thus a relaxation of the problem of finding the revenue-optimal menu can be written as the optimization problem in Figure 4. Here we replace $G(\theta)$ by $G(\theta) - G(\mathbf{0}) + 1$ and remove the constraint $G(\mathbf{0}) = 1$.

$$\sup_{\substack{G \text{ is convex} \\ G(\theta) - G(\theta') \leq c(\theta, \theta'), \forall \theta, \theta'}} \int_{\Theta} \left[-G(\theta) + \nabla G(\theta) \cdot \theta - \max\left\{ \frac{\partial G(\theta)}{\partial \theta_1}, \ldots, \frac{\partial G(\theta)}{\partial \theta_{n-1}}, 0 \right\} + G(\mathbf{0}) \right] f(\theta) d\theta$$

Fig. 4. The Relaxed Problem of the Optimal Menu

In the next step, we construct a dual problem in the form of optimal transportation. We first introduce some useful notations in the measure theory.

- $\Gamma(\Theta)$ and $\Gamma_+(\Theta)$ denote the signed and unsigned (Radon) measures on Θ.
- Given any unsigned measure $\gamma \in \Gamma_+(\Theta \times \Theta)$, denote γ_1, γ_2 the two marginals of γ. Formally, $\gamma_1(A) = \gamma(A \times \Theta)$ and $\gamma_2(A) = \gamma(\Theta \times A)$ for any measurable set $A \subseteq \Theta$.
- Given any signed measure $\mu \in \Gamma(\Theta)$, denote μ_+, μ_- the positive and negative parts of μ respectively, i.e. $\mu = \mu_+ - \mu_-$.
- Given a measure $\mu \in \Gamma(\Theta)$. A *mean-preserving spread* operation in set $A \subseteq \Theta$ is a sequence of the following operation: picking a positive point mass on $\theta \in A$, splitting it into several pieces, and sending these pieces to multiple points in A while preserving the center of mass.

We define *strongly convex dominance* similar to the notion of convex dominance in [18, 37]. The main difference here is that only the mean-preserving spread operation is allowed during the transformation between two measures. [9]

DEFINITION 3. *Given any $\mu, \mu' \in \Gamma(\Theta)$, we say that μ strongly convex dominates μ' (denoted as $\mu \succeq_{cux} \mu'$) if μ' can be transformed to μ by performing a mean-preserving spread. By Jensen's inequality, if $\mu \succeq_{cux} \mu'$, then $\int_{\Theta} g d\mu \geq \int_{\Theta} g d\mu'$ for any continuous, convex function g on Θ.*

In the following definition, we define a measure μ^P for any partition $P = (\Theta_1, \ldots, \Theta_n)$ of Θ such that every component Θ_i is compact. Denote $\mathcal{P}$ the set of all partitions.

DEFINITION 4. *Given any partition $P = (\Theta_1, \ldots, \Theta_n)$ of Θ such that every Θ_i is compact. [10] We define a measure $\mu^P \in \Gamma(\Theta)$ as follows:*

$$\mu^P(A) = \mathbb{1}_A(\mathbf{0}) + \int_{\partial\Theta} \mathbb{1}_A(\theta) f(\theta)(\theta \cdot \mathbf{n}) d\theta - n \cdot \int_{\Theta} \mathbb{1}_A(\theta) f(\theta) d\theta - \sum_{i=1}^{n-1} \int_{\partial\Theta_i} \mathbb{1}_A(\theta) f(\theta)(\mathbf{e}_i \cdot \mathbf{n}_i) d\theta$$

$$- \int_{\Theta} \mathbb{1}_A(\theta) \cdot (\nabla f(\theta) \cdot \theta) d\theta + \sum_{i=1}^{n-1} \int_{\Theta_i} \mathbb{1}_A(\theta) \cdot (\mathbf{e}_i \cdot \nabla f(\theta)) d\theta$$

for any measurable set A. Here $\partial\Theta$ (or $\partial\Theta_i$) is the boundary of Θ (or Θ_i). $\mathbf{n}$ (or $\mathbf{n}_i$) is the outward pointing unit normal at each point on the boundary $\partial\Theta$ (or $\partial\Theta_i$). $\mathbf{e}_i$ is the i-th unit vector of dimension $n - 1$. $\mathbb{1}_A(\theta) = \mathbb{1}[\theta \in A]$.

---

[9]In [18, 37], the notion of convex dominance is less restricted: μ convex dominates μ' if $\int_{\Theta} g d\mu \geq \int_{\Theta} g d\mu'$ for any continuous, **non-decreasing**, convex function g on Θ. Under this notion, μ' can be transformed to μ by either performing the mean-preserving spread, or sending a positive mass to coordinatewise larger points. In our notion, the inequality holds for any convex function that is not necessarily non-decreasing. Thus the second operation may make $\int g d\mu'$ smaller and is not allowed during the transformation.

[10]Here we assume each Θ_i to be compact such that a intergral on Θ_i is well-defined. Two Θ_is may overlap on their boundaries. We still refer to $(\Theta_i, \ldots, \Theta_n)$ a partition of Θ as the set of overlapping points has 0 measure w.r.t. f.

We prove the following lemma by applying the divergence theorem.

LEMMA 11. *For any differentiable function G, we have*

$$\int_\Theta G(\theta)d\mu^P = \int_\Theta \left[-G(\theta) + \nabla G(\theta) \cdot \theta + G(\boldsymbol{0})\right] f(\theta)d\theta - \sum_{i=1}^{n-1} \int_{\Theta_i} \frac{\partial G(\theta)}{\partial \theta_i} f(\theta)d\theta$$

Choosing $G(\theta) = 1, \forall \theta \in \Theta$ implies that $\mu^P(\Theta) = 0$.

Now we are ready to define the dual problem (Figure 5).

$$\inf_{\substack{P \in \mathcal{P}, \gamma \in \Gamma_+(\Theta \times \Theta) \\ \gamma_1 - \gamma_2 \succeq_{cux} \mu^P}} \int_{\Theta \times \Theta} c(\theta, \theta')d\gamma(\theta, \theta')$$

Fig. 5. The Dual Problem in Form of Optimal Transportation

To have a better understanding of the dual problem, we prove the following lemma which shows that the weak duality holds.

LEMMA 12. *For any feasible solution G to the primal (Figure 4) and any feasible solution ($P = (\Theta_1, \ldots, \Theta_n), \gamma$) to the dual (Figure 5),*

$$\int_\Theta \left[-G(\theta) + \nabla G(\theta) \cdot \theta - \max\left\{\frac{\partial G(\theta)}{\partial \theta_1}, \ldots, \frac{\partial G(\theta)}{\partial \theta_{n-1}}, 0\right\} + G(\boldsymbol{0})\right] f(\theta)d\theta \le \int_{\Theta \times \Theta} c(\theta, \theta')d\gamma(\theta, \theta')$$

The inequality achieves equality if and only if all of the following conditions are satisfied:

(1) $\frac{\partial G(\theta)}{\partial \theta_i} = \max\left\{\frac{\partial G(\theta)}{\partial \theta_1}, \ldots, \frac{\partial G(\theta)}{\partial \theta_{n-1}}, 0\right\}$ *for every $i \in [n-1], \theta \in \Theta_i$; $\max_{i \in [n-1]} \frac{\partial G(\theta)}{\partial \theta_i} \le 0$ for every $\theta \in \Theta_n$.*

(2) $\gamma(\theta, \theta') > 0 \implies G(\theta) - G(\theta') = c(\theta, \theta')$.

(3) $\int_\Theta G(\theta)d(\gamma_1 - \gamma_2) = \int_\Theta G(\theta)d\mu^P$. *In other words, μ^P (in Definition 4) can be transformed to $\gamma_1 - \gamma_2$ by performing a mean-preserving spread operated in a region where G is linear.*

An important application of Lemma 12 is that: Given a menu $\mathcal{M}$ in any matching utility environment, we can certify the optimality of $\mathcal{M}$ by constructing a feasible dual (P, γ) that satisfies all conditions in Lemma 12 with respect to the utility function $G(\cdot)$ induced by $\mathcal{M}$. This is because $G(\cdot)$ being an optimal solution to the primal problem in Figure 4 immediately implies that $\mathcal{M}$ is the optimal menu since the primal problem is a relaxation of our problem.

THEOREM 8. *Given any matching utility environment and any menu $\mathcal{M}$. Let $G(\cdot)$ be the buyer's utility function induced by $\mathcal{M}$. If there exists a feasible dual solution (P, γ) to the problem in Figure 5 such that G and (P, γ) satisfies all conditions in the statement of Lemma 12, then $\mathcal{M}$ is the optimal menu in this environment.*

5.2.2 Proof of Theorem 7. Now we focus on the special case $n = 3$ and give a proof of Theorem 7 using Theorem 8. The type space $\Theta = \{(\theta_1, \theta_2) \in [0,1]^2 \mid \theta_1 + \theta_2 \le 1\}$ is a triangle. Let $\mathcal{M}^* = \{E^*(\theta), t^*(\theta)\}_{\theta \in \Theta}$ be the mechanism that only sells the full information at price $p = \frac{1}{3}$. Let $G^*(\theta) = V_\theta^*(E^*(\theta)) - t^*(\theta)$ be the buyer's utility function in $\mathcal{M}^*$. By Theorem 8, it suffices to construct a feasible dual (P^*, γ^*) that satisfies all conditions in Lemma 12.

Denote $\pi_i^*(\theta)$ the (i, i)-entry of $E^*(\theta)$. We consider the following partition $P^* = (\Theta_1^*, \Theta_2^*, \Theta_3^*)$ of Θ: $\Theta_i^* = \{\theta \in \Theta \mid \theta_i \ge \theta_j, \forall j \in \{1, 2, 3\}\}$, here $\theta_3 = 1 - \theta_1 - \theta_2$. See Figure 10a for an illustration. In uniform distribution, the density $f(\theta) = 2$ is a constant for all $\theta \in \Theta$. Thus we can simplify the description of μ^{P^*} as in Observation 2.

OBSERVATION 2. *By Definition 4, the measure μ^{P^*} with respect to the partition P^* is the sum of:*

- *A point mass of +1 at **0**.*
- *For each $(i, j) = (1, 2), (1, 3), (2, 3)$, a mass of $\frac{2}{3}$ uniformly distributed on the line segment $S_{ij} = \{\theta \in \Theta \mid \theta_i = \theta_j \geq \theta_k\}$. Here k is the index other than i, j. $\theta_3 = 1 - \theta_1 - \theta_2$.*
- *A mass of -3 uniformly distributed through out Θ.*

Now we construct the measure $\gamma^* \in \Gamma_+(\Theta \times \Theta)$ as follows: Let $\Omega_1 = \{\theta \in \Theta : 1 - \max_{i \in [n]} \theta_i \geq p\}$ be the set of types where the buyer purchases the full information in $\mathcal{M}^*$. Let $\Omega_2 = \Theta \setminus \Omega_1$. For any measurable set $A, B \subseteq \Theta$, define $\gamma(A \times B) = \mathbb{1}_A(\mathbf{0}) \cdot \int_{\Omega_2} \mathbb{1}_B(\theta) d\theta / \text{Vol}(\Omega_2)$. See Figure 10b for an illustration of the notations. To verify that G^* and (P^*, γ^*) are the optimal primal and dual solution respectively, it suffices to prove the following lemma.

LEMMA 13. *G^* and (P^*, γ^*) are feasible solutions to the primal (Figure 4) and dual problem (Figure 5) respectively. Moreover, they satisfy all the conditions in the statement of Lemma 12.*

The primal feasibility and the first two conditions are relatively easy to verify. To prove the dual feasibility and Condition 3, denote γ_1^*, γ_2^* the marginals of γ. By definition of γ, $\mu^{P^*} - (\gamma_1^* - \gamma_2^*)$ contains only a mass of -2 uniformly distributed throughout out the hexagon Ω_1, and a mass of $\frac{2}{3}$ uniformly distributed in each line segment S_{12}, S_{13}, S_{23}. Due to the special geometric shape of the hexagon, we can transform the positive mass on the line segments to the whole region Ω_1 via mean-preserving spread, to "zero-out" the negative mass. This proves both the dual feasibility and Condition 3 since $G^*(\theta) = \frac{2}{3}$ is constant throughout Ω_1.

6 CONCLUSION

We considered the problem of selling information to a data-buyer with private information. The (approximately) optimal mechanism typically offers menu of options at different prices to the data-buyer. We showed that this revenue maximization problem shares some features with the problem of selling multiple items to a single buyer. Yet, the problem of finding the optimal menu of information structure is richer in two important aspects. First, every item on the menu is information structure, thus matrix of signals given the true state, and second, the individual rationality constraint varies with the private information of the agent. Thus, the choice set as well as the set of constraints is richer than the standard multi-item. Our analysis thus focused on establishing lower and upper bounds for the optimality of the complete information structure, which is a natural information product and in a sense the equivalence of the grand bundle in the mutli-item problem.

We established a lower bound on the cardinality of the optimal menu in the binary-state environment. An interesting future direction is to show an upper bound of the cardinality of the optimal menu. Although we showed an $O(m)$ approximation ratio for selling complete information in this single-dimensional environment, no finite upper bound (in a function of m) of the cardinality that holds for all type distribution is known.

In general environments, we offered a primal-dual approach to prove the optimality of a given menu. We used the approach to show the optimality of selling complete information in a special environment. We believe that this approach should be productive to further characterize the optimal menu in other environments.

REFERENCES

[1] AE Ades, G Lu, and K Claxton. 2004. Expected value of sample information calculations in medical decision modeling. *Medical decision making* 24, 2 (2004), 207–227.

[2] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. 2000. *Functions of bounded variation and free discontinuity problems.* Courier Corporation.

[3] Eduardo M Azevedo, Alex Deng, José Luis Montiel Olea, Justin Rao, and E Glen Weyl. 2020. A/b testing with fat tails. *Journal of Political Economy* 128, 12 (2020), 4614–000.

[4] Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. 2014. A Simple and Approximately Optimal Mechanism for an Additive Buyer. In *the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[5] Moshe Babaioff, Robert Kleinberg, and Renato Paes Leme. 2012. Optimal mechanisms for selling information. In *Proceedings of the 13th ACM Conference on Electronic Commerce*. 92–109.

[6] Viorel Barbu and Teodor Precupanu. 2012. *Convexity and optimization in Banach spaces*. Springer Science & Business Media.

[7] Dirk Bergemann and Alessandro Bonatti. 2015. Selling cookies. *American Economic Journal: Microeconomics* 7, 3 (2015), 259–94.

[8] Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. 2018. The design and price of information. *American Economic Review* 108, 1 (2018), 1–48.

[9] David Blackwell. 1951. Comparison of Experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (1951), 99–102.

[10] David Blackwell. 1953. Equivalent comparisons of experiments. *The annals of mathematical statistics* (1953), 265–272.

[11] Yang Cai, Nikhil R. Devanur, and S. Matthew Weinberg. 2016. A Duality Based Unified Approach to Bayesian Mechanism Design. In *the 48th Annual ACM Symposium on Theory of Computing (STOC)*.

[12] Yang Cai and Grigoris Velegkas. 2021. How to Sell Information Optimally: An Algorithmic Study. In *12th Innovations in Theoretical Computer Science Conference, ITCS 2021, January 6-8, 2021, Virtual Conference (LIPIcs)*, Vol. 185. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 81:1–81:20. https://doi.org/10.4230/LIPIcs.ITCS.2021.81

[13] Yang Cai and Mingfei Zhao. 2017. Simple mechanisms for subadditive buyers via duality. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*. 170–183. https://doi.org/10.1145/3055399.3055465

[14] Shuchi Chawla, Jason D. Hartline, and Robert D. Kleinberg. 2007. Algorithmic Pricing via Virtual Valuations. In *the 8th ACM Conference on Electronic Commerce (EC)*.

[15] Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. 2010. Multi-Parameter Mechanism Design and Sequential Posted Pricing. In *the 42nd ACM Symposium on Theory of Computing (STOC)*.

[16] Yiling Chen, Haifeng Xu, and Shuran Zheng. 2020. Selling information through consulting. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2412–2431.

[17] Jacques Cremer and Richard P. McLean. 1988. Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions. *Econometrica* 56, 6 (1988), 1247–1257.

[18] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2017. Strong Duality for a Multiple-Good Monopolist. *Econometrica* 85, 3 (2017), 735–767.

[19] Nikhil Devanur, Kira Goldner, Raghuvansh Saxena, Ariel Schvartzman, and S Matthew Weinberg. 2020. Optimal Mechanism Design for Single-Minded Agents. *arXiv preprint arXiv:2002.06329* (2020).

[20] Nikhil R Devanur and S Matthew Weinberg. 2017. The optimal mechanism for selling to a budget constrained buyer: The general case. In *Proceedings of the 2017 ACM Conference on Economics and Computation*. 39–40.

[21] Kimon Drakopoulos and Ramandeep S Randhawa. 2020. Why perfect tests may not be worth waiting for: Information as a commodity. *Available at SSRN 3565245* (2020).

[22] Paul Dütting, Thomas Kesselheim, and Brendan Lucier. 2020. An O(log log m) Prophet Inequality for Subadditive Combinatorial Auctions. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*. IEEE, 306–317. https://doi.org/10.1109/FOCS46700.2020.00037

[23] Jeffrey Ely, Andrea Galeotti, Ole Jann, and Jakub Steiner. 2021. Optimal test allocation. *Journal of Economic Theory* 193 (2021), 105236.

[24] Amos Fiat, Kira Goldner, Anna R Karlin, and Elias Koutsoupias. 2016. The fedex problem. In *Proceedings of the 2016 ACM Conference on Economics and Computation*. 21–22.

[25] Yiannis Giannakopoulos and Elias Koutsoupias. 2014. Duality and optimality of auctions for uniform distributions. In *ACM Conference on Economics and Computation, EC '14, Stanford , CA, USA, June 8-12, 2014*. 259–276. https://doi.org/10.1145/2600057.2602883

[26] Yiannis Giannakopoulos and Elias Koutsoupias. 2015. Selling Two Goods Optimally. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part II*. 650–662.

https://doi.org/10.1007/978-3-662-47666-6_52

[27] Nima Haghpanah and Jason Hartline. 2021. When is pure bundling optimal? *The Review of Economic Studies* 88, 3 (2021), 1127–1156.

[28] Sergiu Hart and Noam Nisan. 2012. Approximate Revenue Maximization with Multiple Items. In *the 13th ACM Conference on Electronic Commerce (EC).*

[29] Sergiu Hart and Noam Nisan. 2019. Selling multiple correlated goods: Revenue maximization and menu-size complexity. *Journal of Economic Theory* 183 (2019), 991–1029.

[30] Xinye Li and Andrew Chi-Chih Yao. 2013. On Revenue Maximization for Selling Multiple Independently Distributed Items. *Proceedings of the National Academy of Sciences* 110, 28 (2013), 11232–11237.

[31] Yingkai Li. 2021. Selling Data to an Agent with Endogenous Information. *arXiv preprint arXiv:2103.05788* (2021).

[32] Shuze Liu, Weiran Shen, and Haifeng Xu. 2021. *Optimal Pricing of Information.* Association for Computing Machinery, New York, NY, USA, 693. https://doi.org/10.1145/3465456.3467551

[33] David G Luenberger. 1997. *Optimization by vector space methods.* John Wiley & Sons.

[34] Roger B. Myerson. 1981. Optimal Auction Design. *Mathematics of Operations Research* 6, 1 (1981), 58–73.

[35] John Riley and Richard Zeckhauser. 1983. Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics* 98, 2 (1983), 267–289.

[36] Jean-Charles Rochet and Philippe Choné. 1998. Ironing, sweeping, and multidimensional screening. *Econometrica* (1998), 783–826.

[37] Moshe Shaked and J George Shanthikumar. 2007. *Stochastic orders.* Springer.

[38] Carl Shapiro, Shapiro Carl, Hal R Varian, et al. 1998. *Information rules: a strategic guide to the network economy.* Harvard Business Press.

[39] Juuso Toikka. 2011. Ironing without control. *Journal of Economic Theory* 146, 6 (2011), 2510–2526.

[40] Cédric Villani. 2009. *Optimal transport: old and new.* Vol. 338. Springer.

[41] Andrew Chi-Chih Yao. 2015. An n-to-1 Bidder Reduction for Multi-item Auctions and its Applications, In SODA. *CoRR* (2015). http://arxiv.org/abs/1406.3278

A ADDITIONAL PRELIMINARIES

DEFINITION 5 (RESPONSIVE EXPERIMENT [8]). *A buyer type θ is* responsive *to an experiment E if: There is a one-to-one mapping $g : S \rightarrow [m]$ from the signal set to the action space such that, for every $k \in [m]$, the buyer chooses the action k when receiving signal s iff $g(s) = k$.*

LEMMA 14 (ADAPTED FROM PROPOSITION 1 FROM [8]). *A mechanism $\mathcal{M} = \{(E(\theta), t(\theta))\}_{\theta \in \Theta}$ is* responsive *if every buyer type θ is responsive to $E(\theta)$. Then for any mechanism $\mathcal{M}$, there exists a responsive mechanism $\mathcal{M}' = \{E'(\theta), t(\theta)\}_{\theta \in \Theta}$ with the same payment rule, such that for every type θ, the buyer has the same value for both experiments $E(\theta)$ and $E'(\theta)$.*

LEMMA 15. *[8] Any optimal menu contains a fully informative experiment.*

B MISSING DETAILS FROM SECTION 3

Proof of Lemma 1: Since h is strictly decreasing in $[a, b]$, let h^{-1} be its inverse function. Define F as follows: $F(\theta) = 1 - \frac{c}{1-h(\theta)}, \forall \theta \in [a, b); F(b) = 1$. Here $c = 1 - h(a)$ to ensure that $F(a) = 0$. Consider the fully informative experiment with any price p. The utility for purchasing this experiment is $1 - p$ regardless of the buyer's prior. Thus the buyer purchases the fully informative experiment iff $h(\theta) \leq 1 - p$, which happens with probability $1 - F(h^{-1}(1 - p)) = c/p$. Thus FREV $= c$. Moreover, $f(\theta) = -c \cdot \frac{h'(\theta)}{(1-h(\theta))^2}, \forall \theta \in [a, b)$. [11] We have

$$\int_a^b f(\theta) \cdot (1 - h(\theta)) d\theta = -c \cdot \int_a^b \frac{h'(\theta)}{1 - h(\theta)} d\theta = c \cdot \log(1 - h(\theta)) \Big|_a^b = c \cdot \log\left(\frac{1 - h(b)}{1 - h(a)}\right)$$

$\square$

---

[11] As $h(\cdot)$ is a convex function, it is also continuously differentiable on except countably many points. For those points, we can choose $h'(\theta)$ to be any subdifferential at θ.

Proof of Lemma 2: We first bound the revenue of $\mathcal{M}$. Notice that for every $1 \leq i \leq m - 2$,

$$p_i = \sum_{j=0}^{i-1} h_j + \theta_i l_i = (2^i - 1) \cdot \varepsilon^{2^m} - \varepsilon^{2^i} \cdot (\varepsilon^{2^m} + \sum_{1 \leq j < i} 2^j \cdot \varepsilon^{2^m - 2^j}) \leq (2^i - 1) \cdot \varepsilon^{2^m} \tag{5}$$

On the other hand, when $\varepsilon < \frac{1}{2^m}$, we have

$$p_i = (2^i - 1) \cdot \varepsilon^{2^m} - \varepsilon^{2^m + 2^i} - \sum_{1 \leq j < i} 2^j \cdot \varepsilon^{2^m + 2^i - 2^j} \tag{6}$$

$$\geq (2^i - 1) \cdot \varepsilon^{2^m} - \varepsilon^{2^m + 2} \cdot (1 + \sum_{1 \leq j < i} 2^j) \geq (2^i - 1) \cdot \varepsilon^{2^m} - \varepsilon^{2^m + 1}$$

Moreover,

$$1 - u(\theta_{i+1}) = \sum_{j=0}^{i} h_j = (2^{i+1} - 1) \cdot \varepsilon^{2^m}$$

Thus for every θ such that $\theta_i < \theta \leq \theta_{i+1}$, the payment of $\mathcal{M}$ is at least

$$\min\{p_i, p_{i-1}, p_{i-2}\} = p_{i-2} \geq \frac{1}{9}(1 - u(\theta_{i+1})) \geq \frac{1}{9}(1 - u(\theta)).$$

Thus we have $\text{Rev}(\mathcal{M}) \geq \frac{1}{9} \cdot \int_{\theta_1}^{\theta_{m-1}} f(\theta) \cdot (1 - u(\theta)) d\theta$.

Next we prove that $\mathcal{M}$ is IR and δ-IC. Firstly, if the buyer follows the recommendation of experiment E_i, his expected utility is $\theta(1 + l_i) + 1 - \theta - p_i = 1 + \theta l_i - p_i = 1 - \sum_{j=0}^{i-1} h_j + (\theta - \theta_i) l_i$, which is $u(\theta_i)$ and $u(\theta_{i+1})$ if we choose θ to be θ_i and θ_{i+1} respectively. Since $u(\cdot)$ is linear on $(\theta_i, \theta_{i+1}]$, the buyer' expected utility for receiving (E_i, p_i) is exactly $u(\theta)$. Thus her utility by reporting truthfully is at least $u(\theta) \geq 0$, and $\mathcal{M}$ is IR.

Now fix any $\theta \in [\theta_1, \theta_{m-1}]$. Let i be the unique number such that $\theta_i < \theta \leq \theta_{i+1}$. Suppose that the buyer misreports and receives experiment E_j for $1 \leq j \leq m - 2$. We notice that when the experiment recommends action 1, the state ω_1 is fully revealed and the buyer will follow the recommendation, choosing action 1. If the buyer chooses action $m - r$ when being recommended action m, for $0 \leq r \leq m - 1$, then his utility for buying experiment j is

$$U_{j,r}(\theta) := \theta(1 + l_j) + \theta(-l_j) u_{1,m-r} + (1 - \theta) u_{2,m-r} - p_j = u_{2,m-r} + \theta(1 + l_j - l_j u_{1,m-r} - u_{2,m-r}) - p_j.$$

When $1 \leq r \leq m - 2$, using the fact that $u_{2,m-r} = 1 - p_r$ and $u_{1,m-r} = 1 - p_r + l_r$, we can rewrite $U_{j,r}(\theta) = 1 + \theta(p_r + l_j(p_r - l_r)) - p_r - p_j$.

We argue that for every $\theta \in (\theta_i, \theta_{i+1}]$, the following inequality holds.

$$\max_{j \in \{1,\ldots,m-2\}, r \in \{0,\ldots,m-1\}} U_{j,r}(\theta) \leq \max_{j' \in \{i,i-1,i-2\}, r' \in \{0,\ldots,m-1\}} U_{j',r'}(\theta) + 7 \cdot \varepsilon^{2^m + 1}. \text{ [12]} \tag{7}$$

Since a buyer with type θ receives utility $\max_{j' \in \{i,i-1,i-2\}, r'} U_{j',r'}(\theta)$ in $\mathcal{M}$ if he reports truthfully, Inequality (7) guarantees that $\mathcal{M}$ is δ-IC (recall that $\delta = 7 \cdot \varepsilon^{2^m + 1}$), as the buy can obtain utility at most $\max_{j \in \{1,\ldots,m-2\}, r} U_{j,r}(\theta)$ by misreporting his type.

Now we prove that for every $j \in \{1, \ldots, m - 2\}, r \in \{0, \ldots, m - 1\}$, $U_{j,r}(\theta)$ is no more than the RHS of Inequality (7) by a case analysis.

(1) $r = 0$. For every $j \in \{1, \ldots, m - 2\}$, $U_{j,0}(\theta) = V_\theta^*(E_j) - p_j$ is the agent's utility for purchasing E_j, if he follows the recommendation. By the definition of the IR curve $u(\cdot)$, the function $U_{\ell,0}(\cdot)$ coincides with $u(\cdot)$ on the interval $[\theta_\ell, \theta_{\ell+1}]$ for all $\ell \in \{1, \ldots, m - 2\}$. Since $u(\cdot)$ is a convex function and $\theta \in (\theta_i, \theta_{i+1}]$, $U_{i,0}(\theta) \geq U_{j,0}(\theta)$ for all $j \in \{1, \ldots, m - 2\}$.

---

[12]In the RHS of the inequality, j can only take value $i, i - 1$ when $i \leq 2$, and can only take value i when $i = 1$.

(2) $r = m-1$. For every $j \in \{1, \ldots, m-2\}$, since $\varepsilon < \frac{1}{2m}$, $U_{j,m-1}(\theta) = \theta - p_j \le \theta \le \theta_{m-1} < \varepsilon^{2^{m-1}-1}$. On the other hand, $U_{i,0}(\theta) = 1 + \theta\ell_i - p_i \ge 1 - 2^m \cdot \varepsilon^{2^m} > 1 - \varepsilon^{2^m-1} > U_{j,m-1}(\theta)$.

(3) $1 \le r \le m-2$. Recall $p_r - (2^r - 1) \cdot \varepsilon^{2^m} \in [-\varepsilon^{2^m+1}, 0]$ for every $r \in \{1, \ldots, m-2\}$ (Inequality (6)). For every $j, r \in \{1, \ldots, m-2\}$, $U_{j,r}(\theta)$ can be further bounded as follows:

$$U_{j,r}(\theta) \ge 1 - p_r - p_j + \theta \cdot [(2^r - 1)\varepsilon^{2^m} - \varepsilon^{2^m+1} - \varepsilon^{2^j}((2^r - 1)\varepsilon^{2^m} + \varepsilon^{2^r})]$$

$$\ge 1 - p_r - p_j - \theta \cdot \varepsilon^{2^j+2^r} - 2 \cdot \varepsilon^{2^m+1}$$

$$\ge 1 - (2^j + 2^r - 2) \cdot \varepsilon^{2^m} - \theta \cdot \varepsilon^{2^j+2^r} - 2 \cdot \varepsilon^{2^m+1}$$

The first inequality follows from the upper bound (Inequality (5)) and the lower bound (Inequality (6)) of p_i. The second inequality follows from $\theta < 1$ and $\varepsilon < \frac{1}{2m}$. The last inequality follows from Inequality (5). On the other hand,

$$U_{j,r}(\theta) \le 1 - p_r - p_j + \theta \cdot [(2^r - 1)\varepsilon^{2^m} - \varepsilon^{2^j}((2^r - 1)\varepsilon^{2^m} - \varepsilon^{2^m+1} + \varepsilon^{2^r})]$$

$$\le 1 - p_r - p_j - \theta \cdot \varepsilon^{2^j+2^r} + 2 \cdot \varepsilon^{2^m+1}$$

$$\le 1 - (2^j + 2^r - 2) \cdot \varepsilon^{2^m} - \theta \cdot \varepsilon^{2^j+2^r} + 4 \cdot \varepsilon^{2^m+1}.$$

The first and last inequalities follow from Inequality (5) and (6). The second inequality follows from $\theta \le \theta_{m-1} < \varepsilon^{2^{m-1}-1}$. There are two nice properties of the upper and lower bounds of $U_{j,r}(\theta)$: (i) they are within $O(\varepsilon^{2^m+1})$ and (ii) they are both symmetric with respect to j and r.

(a) If $2^j + 2^r > 2^i$, then either j or r must be at least i, which implies that $2^j + 2^r \ge 2^i + 2$. By Equation (2), $\theta \le \theta_{i+1} < \varepsilon^{2^m-2^i-1}$. Thus $\theta \cdot \varepsilon^{2^j+2^r} < \varepsilon^{2^m+1}$. We notice that if either j or r is at least i, $(2^j + 2^r - 2) \cdot \varepsilon^{2^m}$ is minimized when $(j, r) = (i, 1)$ or $(1, i)$. Thus $U_{j,r}(\theta) \le U_{i,1}(\theta) + 7 \cdot \varepsilon^{2^m+1}$.

(b) If $2^j + 2^r < 2^{i-1}$, then $i > 1$ and thus $2^j + 2^r \le 2^{i-1} - 2$ (as both are even numbers when $i > 1$). Since $\theta \ge \theta_i > 2^{i-1} \cdot \varepsilon^{2^m-2^{i-1}}$, we have $\theta \cdot \varepsilon^{2^j+2^r} \ge 2^{i-1} \cdot \varepsilon^{2^m-2}$. As $\varepsilon < \frac{1}{2m}$, the term $\theta \cdot \varepsilon^{2^j+2^r}$ dominates $(2^j + 2^r - 2) \cdot \varepsilon^{2^m}$ and $1 - (2^j + 2^r - 2) \cdot \varepsilon^{2^m} - \theta \cdot \varepsilon^{2^j+2^r}$ is maximized when $2^j + 2^r$ is maximized conditioned on $2^j + 2^r < 2^{i-1}$. In other words, it is maximized when $(j, r) = (i - 2, i - 3)$ or $(i - 3, i - 2)$. Thus $U_{j,r}(\theta) \le U_{i-2,i-3}(\theta) + 6 \cdot \varepsilon^{2^m+1}$.

(c) If $2^j + 2^r \in [2^{i-1}, 2^i]$, then either j or r is in $\{i-2, i-1, i\}$. Since $|U_{j,r}(\theta) - U_{r,j}(\theta)| \le 6 \cdot \varepsilon^{2^m+1}$, for all j and r in $\{1, \ldots, m-2\}$, we have $U_{j,r}(\theta) \le \max_{j' \in \{i,i-1,i-2\}, r' \in \{1,\ldots,m-2\}} U_{j',r'}(\theta) + 6 \cdot \varepsilon^{2^m+1}$.

□

To complete the proof of Theorem 1, we need the following lemma adapted from [12].

LEMMA 16 (ADAPTED FROM LEMMA 8 IN [12]). *Suppose $\mathcal{M}$ is a mechanism of option size ℓ, where the IC and IR constraints are violated by at most δ, for some $\delta > 0$ and finite $\ell > 0$. Then there exists an IC, IR mechanism $\mathcal{M}'$ of option size ℓ such that $\text{Rev}(\mathcal{M}') \ge (1 - \eta) \cdot \text{Rev}(\mathcal{M}) - \delta/\eta - \delta$, for any $\eta > 0$. Moreover, if every experiment E in $\mathcal{M}$ is semi-informative, then $\mathcal{M}'$ is a semi-informative menu.*

Proof of Lemma 16: Let $\mathcal{M} = (\{E(\theta), t(\theta)\})_{\theta \in [0,1]}$ and $\{(E_j, t_j)\}_{j \in [\ell]}$ be the ℓ options in $\mathcal{M}$. For every $j \in [\ell]$, let $t'_j = (1 - \eta)t_j - \delta$. Let $\mathcal{M}'$ be the menu with options $\{(E_j, t'_j)\}_{j \in [\ell]}$, i.e., the buyer with type θ purchase the experiment that maximizes $V_\theta(E_j) - t'_j$. Clearly $\mathcal{M}'$ is IC and IR.

For any type $\theta \in [0, 1]$, let $j \in [m]$ be the unique number such that $E(\theta) = E_j$. Since the IC constraints in $\mathcal{M}$ are violated by at most δ, we know that

$$V_\theta(E_j) - t_j \ge V_\theta(E_k) - t_k - \delta, \ \forall k \in [\ell]$$

Now suppose that the buyer with type θ purchases E_{k^*} in $\mathcal{M}'$. Then

$$V_\theta(E_{k^*}) - (1 - \eta)t_{k^*} \geq V_\theta(E_j) - (1 - \eta)t_j$$

Choosing k to be k^* in the first inequality and combining the two inequalities, we have that

$$V_\theta(E_{k^*}) - (1 - \eta)t_{k^*} \geq V_\theta(E_{k^*}) - t_{k^*} - \varepsilon + \eta t_j \implies t_j - t_{k^*} \leq \frac{\delta}{\eta}$$

Hence, for the revenue we have $\text{REV}(\mathcal{M}') \geq (1 - \eta) \cdot \text{REV}(\mathcal{M}) - \delta - \frac{\delta}{\eta}$. The second part of the statement directly follows from the fact that $\mathcal{M}'$ shares the same set of experiments with $\mathcal{M}$. □

Proof of Theorem 1: We consider the construction present in Section 3. By Lemma 1 and Lemma 2,

$$\text{REV}(\mathcal{M}) \geq \frac{1}{9} \cdot \log(2) \cdot (m - 2) \cdot \text{FREV}$$

We notice that $\text{FREV} = \varepsilon^{2^m}$. According to the construction of the mechanism, $\mathcal{M}$ is of option size at most $m - 2$. Thus by applying Lemma 16 with $\eta = \frac{1}{2}$ and $\delta = 7 \cdot \varepsilon^{2^m+1}$, we have

$$\text{SIREV} \geq \frac{1}{2} \cdot \text{REV}(\mathcal{M}) - 3 \cdot 7 \cdot \varepsilon^{2^m+1} \geq \varepsilon^{2^m} \cdot \left[\frac{\log(2)}{18}(m - 2) - 21\varepsilon \right] = \Omega(m) \cdot \text{FREV}$$

Since $\text{OPT} \geq \text{SIREV}$, we further have $\text{OPT} = \Omega(m) \cdot \text{FREV}$. □

Proof of Lemma 3: Let $E_1, \ldots, E_\ell$ be the experiments contained in mechanism $\mathcal{M}$, and $t(E_1), \ldots, t(E_\ell)$ be the prices. Let p_i be the revenue that is generated by experiment E_i, i.e., $p_i = \Pr_\theta[\mathcal{M}(\theta) = E_i] \cdot t(E_i)$. Then, $\text{REV}(\mathcal{M}) = \sum_{i=1}^\ell p_i \implies p_j \geq \frac{\text{REV}(\mathcal{M})}{\ell}$, for some $j \in [\ell]$. Now, consider menu $\mathcal{M}'$ that contains only the fully informative experiment E^* with price $t(E_j)$. We claim that for all types θ such that $\mathcal{M}(\theta) = E_j$, the agent with type θ is willing to purchase the experiment E^* at price $t(E_j)$, as $V_\theta(E^*) \geq V_\theta(E_j), \forall \theta \in [0, 1]$. Thus, for all θ such that $\mathcal{M}(\theta) = E_j$, $V_\theta(E^*) - t(E^*) \geq V_\theta(E_j) - t(E_j) \geq u(\theta)$, where the second inequality follows from $\mathcal{M}$ is IR. Hence, the revenue of $\mathcal{M}'$ is at least $p_j \geq \frac{\text{REV}(\mathcal{M})}{\ell}$. □

Proof of Theorem 2: By Theorem 1, there exists some environment such that $c := \frac{\text{OPT}}{\text{FREV}} = \Omega(m)$. Suppose there is an optimal menu $\mathcal{M}^*$ that consists of $c' < c$ different experiments. Then by applying Lemma 3 to $\mathcal{M}^*$, we have $\text{FREV} \geq \frac{\text{OPT}}{c'} > \frac{\text{OPT}}{c}$. Contradiction. Thus any optimal menu consists of at least c experiments. □

C MISSING DETAILS FROM SECTION 4

OBSERVATION 3. *[8] For any semi-informative experiment E, $q(E) = \pi_{11} \cdot u_{11} - \pi_{2m} \cdot u_{2m}$ satisfies the following: When $q(E) \leq u_{11} - u_{2m}$, E has Pattern 1 in Table 2, where $\pi_{11} = (q(E) + u_{2m})/u_{11}$. When $q(E) > u_{11} - u_{2m}$, E has Pattern 2 in Table 2, where $\pi_{2m} = (u_{11} - q(E))/u_{2m}$.*

OBSERVATION 4. *[8] For any semi-informative experiment E, and any $\theta \in [0, 1]$, $V_\theta^*(E) = \theta q(E) + u_{2m} + \min\{u_{11} - u_{2m} - q(E), 0\}$ (recall that $u_{11} \geq u_{2m} = 1$).*

C.1 Proof of Lemma 4

LEMMA 17. *There exists a responsive-IC and IR mechanism $\mathcal{M}$ that satisfies both of the following properties:*

(1) *Every experiment of $\mathcal{M}$ only recommends the fully informative actions. Formally, $\pi_{ij}(\mathcal{M}(\theta)) = 0$, for all $\theta \in \Theta, i \in \{1, 2\}, 2 \leq j \leq m - 1$.*
(2) *$\text{REV}(\mathcal{M}) = OPT^*$.*

PROOF. Fix any responsive-IC and IR mechanism $\mathcal{M}$, any experiment E that $\mathcal{M}$ offers, any action ℓ such that $2 \le \ell \le m - 1$, and any state $i = 1, 2$. Suppose $\pi_{i\ell}(E)$ is not zero. We modify E to get another experiment E' by moving all the probability mass from $\pi_{i\ell}(E)$ to $\pi_{i1}(E)$ and $\pi_{im}(E)$ in a proper way, such that the modified mechanism generates the same revenue. Let $\mathcal{M}^*$ be an optimal responsive-IC and IR mechanism. We show how to modify its experiments E to obtain a responsive-IC and IR menu that satisfies both properties in the statement.

For any $\theta \in [0, 1]$, recall that $V_\theta^*(E)$ is the agent's value for experiment E if she follows the recommendation:

$$V_\theta^*(E) = \theta \sum_{j \in [m]} \pi_{1j}(E) u_{1j} + (1 - \theta) \sum_{j \in [m]} \pi_{2j}(E) u_{2j}$$

Let $U_\theta(E) = V_\theta^*(E) - t(E)$ be her utility for E when following the recommendation. We only consider moving the probability mass from $\pi_{2\ell}(E)$ to $\pi_{21}(E)$ and $\pi_{2m}(E)$. The case where we move the probability mass of $\pi_{1\ell}(E)$ follows from a similar argument.

Without loss of generality, assume $\varepsilon = \pi_{2\ell}(E) > 0$. We move ε_1 mass from $\pi_{2\ell}(E)$ to $\pi_{21}(E)$ and ε_2 mass from $\pi_{2\ell}(E)$ to $\pi_{2m}(E)$. Both ε_1 and ε_2, which satisfy $\varepsilon_1 + \varepsilon_2 = \varepsilon$, will be determined later. Let E' be the modified experiment, and we keep its price as $t(E)$. For every θ, we show how to choose ε_1 and ε_2 appropriately so that $U_\theta(E') = U_\theta(E)$.

$$U_\theta(E') - U_\theta(E) = (1 - \theta) \sum_{j \in [m]} (\pi_{2j}(E') - \pi_{2j}(E)) u_{2j}$$
$$= (1 - \theta) \cdot (\varepsilon_1(u_{21} - u_{2\ell}) + \varepsilon_2(u_{2m} - u_{2\ell}))$$

Here the first equality follows from $\pi_{1j}(E') = \pi_{1j}(E), \forall j \in [m]$. Now choose $\varepsilon_1, \varepsilon_2$ such that $\varepsilon_1 + \varepsilon_2 = \varepsilon$ and $\varepsilon_1(u_{2\ell} - u_{21}) = \varepsilon_2(u_{2m} - u_{2\ell})$. In other words,

$$\varepsilon_1 = \frac{(u_{2m} - u_{2\ell})\varepsilon}{(u_{2\ell} - u_{21}) + (u_{2m} - u_{2\ell})}, \quad \varepsilon_2 = \frac{(u_{2\ell} - u_{21})\varepsilon}{(u_{2\ell} - u_{21}) + (u_{2m} - u_{2\ell})}$$

Notice that since $u_{21} < u_{2\ell} < u_{2m}$ we have that $\varepsilon_1, \varepsilon_2 > 0$. With the choices of $\varepsilon_1, \varepsilon_2$, we have $U_\theta(E) = U_\theta(E'), \forall \theta \in [0, 1]$. Thus, the modified mechanism is still responsive-IC and IR and has the same revenue. We modify $\mathcal{M}^*$ by applying the above procedure to every experiment E offered by $\mathcal{M}^*$, every $\ell \in \{2, ..., m - 1\}$, and every state $i = 1, 2$. Let $\mathcal{M}'$ be the mechanism after the modification. Then $\mathcal{M}'$ is responsive-IC and IR and satisfies both properties in the statement. □

Due to Lemma 17, we can focus on responsive-IC and IR mechanisms that only recommend fully informative actions. Let C_2 be the set of all such mechanisms. For ease of notation, we denote an experiment offered by the mechanism $E = \begin{pmatrix} \pi_{11} & 1 - \pi_{11} \\ 1 - \pi_{2m} & \pi_{2m} \end{pmatrix}$. Note that $V_\theta^*(E) = \theta u_{11} \cdot \pi_{11}(E) + (1 - \theta) u_{2m} \cdot \pi_{2m}(E)$.

Proof of Lemma 4: Let $\mathcal{M}$ be an arbitrary mechanism in C_2. Let E be any experiment offered by $\mathcal{M}$, such that $\pi_{11} < 1$ and $\pi_{2m} < 1$. Without loss of generality, assume that $(1 - \pi_{11}) u_{11} \le (1 - \pi_{2m}) u_{2m}$. The other case is similar. Let $\varepsilon_1 = 1 - \pi_{11}$ and $\varepsilon_2 = \frac{u_{11}}{u_{2m}} \cdot \varepsilon_1$. By our assumption, $\varepsilon_2 \le 1 - \pi_{2m}$. We modify experiment E and its price $t(E)$ in the following way: Move probability mass ε_1 from π_{1m} to π_{11}, and move probability mass ε_2 from π_{21} to π_{2m}. Denote E' the modified experiment. Since $\pi_{11} + \varepsilon_1 = 1$, we have

$$E' = \begin{pmatrix} 1 & 0 \\ 1 - \pi_{2m} - \varepsilon_2 & \pi_{2m} + \varepsilon_2 \end{pmatrix}.$$

Consider the mechanism $\widehat{\mathcal{M}}$ where we replace E by E' and let the price of experiment E' be $t(E') := t(E) + \varepsilon_2 \cdot u_{2m}$. Formally, for every θ such that $\mathcal{M}(\theta) = E$, define $\widehat{\mathcal{M}}(\theta) = E'$ and the payment $t'(\theta) = t(\theta) + \varepsilon_2 \cdot u_{2m}$. For any other types, $\widehat{\mathcal{M}}$ offers the same experiment with the same price as offered by $\mathcal{M}$.

For any θ, we have that

$$V_\theta^*(E') - V_\theta^*(E) = \theta u_{11} \cdot (\pi_{11}(E') - \pi_{11}(E)) + (1 - \theta)u_{2m} \cdot (\pi_{2m}(E') - \pi_{2m}(E))$$
$$= \theta \varepsilon_1 \cdot u_{11} + (1 - \theta)\varepsilon_2 \cdot u_{2m} = \varepsilon_2 \cdot u_{2m}.$$

Hence, we can immediately see that for every type θ, the agent's utility for experiment E' at price $t(E')$) is the same as the agent's utility for experiment E at price $t(E)$. Since $\mathcal{M}$ is responsive-IC and IR, $\widehat{\mathcal{M}}$ is also responsive-IC and IR. Moreover, we have that $\text{Rev}(\widehat{\mathcal{M}}) \geq \text{Rev}(\mathcal{M})$. By repeating this procedure for all the experiments in the optimal mechanism $\mathcal{M}' \in C_2$, we can construct a mechanism $\mathcal{M}^* \in C_2'$, such that $\text{Rev}(\mathcal{M}^*) \geq \text{Rev}(\mathcal{M}')$. By Lemma 17, $\text{Rev}(\mathcal{M}') = \text{OPT}^*$. Thus $\text{Rev}(\mathcal{M}^*) = \text{OPT}^*$. □

C.2 Proof of Lemma 5

We begin with necessity. For every θ, let $E(\theta)$ be the experiment that corresponds to $q(\theta)$ (Observation 3). For any two types $\theta_1, \theta_2 \in [0, 1]$, where $\theta_2 > \theta_1$, we have

$$V_{\theta_1}^*(E(\theta_1)) - t(\theta_1) \geq V_{\theta_1}^*(E(\theta_2)) - t(\theta_2)$$
$$V_{\theta_2}^*(E(\theta_2)) - t(\theta_2) \geq V_{\theta_2}^*(E(\theta_1)) - t(\theta_1)$$

Adding up both inequalities obtains

$$V_{\theta_1}^*(E(\theta_1)) + V_{\theta_2}^*(E(\theta_2)) \geq V_{\theta_1}^*(E(\theta_2)) + V_{\theta_2}^*(E(\theta_1)) \tag{8}$$

By Observation 4, for every $\theta, \theta' \in [0, 1]$,

$$V_\theta^*(E(\theta')) = \begin{cases} \theta \cdot q(\theta') + u_{2m}, & \text{if } q(\theta') \leq u_{11} - u_{2m} \\ (\theta - 1) \cdot q(\theta') + u_{11}, & \text{if } q(\theta') > u_{11} - u_{2m} \end{cases}$$

If both $q(\theta_1)$ and $q(\theta_2)$ are at most $u_{11} - u_{2m}$, Inequality (8) $\iff (\theta_1 - \theta_2)(q(\theta_1) - q(\theta_2)) \geq 0 \iff q(\theta_1) \leq q(\theta_2)$. Similarly, when both $q(\theta_1)$ and $q(\theta_2)$ are greater than $u_{11} - u_{2m}$, we also have $q(\theta_1) \leq q(\theta_2)$.

Without loss of generality, the only remaining case is when $q(\theta_2) \leq u_{11} - u_{2m} < q(\theta_1)$. Inequality (8) is equivalent to

$$(\theta_1 - 1)q(\theta_1) + u_{11} + \theta_2 \cdot q(\theta_2) + u_{2m} \geq \theta_1 \cdot q(\theta_2) + u_{2m} + (\theta_2 - 1) \cdot q(\theta_1) + u_{11}$$
$$\iff (\theta_1 - \theta_2)(q(\theta_1) - q(\theta_2)) \geq 0,$$

which contradicts with $\theta_1 < \theta_2$. Hence, $q(\theta_1) \geq q(\theta_2)$ and $q(\theta)$ is non-decreasing in θ.

For every $\theta \in [0, 1]$, let $U^*(\theta) = V_\theta^*(E(\theta)) - t(\theta)$ be the agent's utility by reporting truthfully and following the recommendation of the experiment. Let $\theta^* = \sup\{\theta : q(\theta) \leq u_{11} - u_{2m}\}$. Assume $q(\theta^*) \leq u_{11} - u_{2m}$.[13] When $\theta \in [0, \theta^*]$, $U^*(\theta) = \theta \cdot q(\theta) + u_{2m} - t(\theta)$ is a quasilinear function, and by Myerson's theory [34], $\mathcal{M}$ is responsive-IC implies that $U^*(\theta) = U^*(0) + \int_0^\theta q(x)dx$. Thus

$$t(\theta) = V_\theta^*(E(\theta)) - U^*(\theta) = \theta \cdot q(\theta) + u_{2m} - U^*(0) - \int_0^\theta q(x)dx \tag{9}$$

---

[13]The other case follows from a similar argument, where we replace both intervals $[0, \theta^*]$, $(\theta^*, 1]$ by $[0, \theta^*)$, $[\theta^*, 1]$.

Similarly, for all $\theta \in (\theta^*, 1]$, we have $U^*(\theta) = U^*(1) - \int_\theta^1 q(x)dx$. Hence

$$t(\theta) = (\theta - 1) \cdot q(\theta) + u_{11} - U^*(1) + \int_\theta^1 q(x)dx \tag{10}$$

According to Equation (9), when $\theta = 0$, $t(0) = u_{2m} - U^*(0) \geq 0$. Since $\mathcal{M}$ is IR, $U^*(0) \geq u_{2m}$. Thus $U^*(0) = u_{2m}$. Similarly, by Equation (10) we have $U^*(1) = u_{11}$.

When the agent has type θ^* and misreports to $\theta' > \theta^*$, responsive-IC implies that

$$V_{\theta^*}^*(E(\theta')) - t(\theta') = (\theta^* - \theta') \cdot q(\theta') + u_{11} - \int_{\theta'}^1 q(x)dx$$

$$\leq U^*(\theta^*) = u_{2m} + \int_0^{\theta^*} q(x)dx$$

Let $\theta' \to \theta^{*+}$,[14] we have $\int_0^1 q(x)dx \geq u_{11} - u_{2m}$. Similarly, when the agent has type $\theta' > \theta^*$ and misreports to θ^*, responsive-IC implies that

$$V_{\theta'}^*(E(\theta^*)) - t(\theta^*) = (\theta' - \theta^*) \cdot q(\theta^*) + u_{2m} + \int_0^{\theta^*} q(x)dx$$

$$\leq U^*(\theta') = u_{11} - \int_{\theta'}^1 q(x)dx$$

Let $\theta' \to \theta^{*+}$, we have $\int_0^1 q(x)dx \leq u_{11} - u_{2m}$. Thus $\int_0^1 q(x)dx = u_{11} - u_{2m}$. Now it's easy to verify that $t(\theta)$ satisfies Equation (4). Thus for every θ, $U^*(\theta) = V_\theta^*(E(\theta)) - t(\theta) = u_{2m} + \int_0^\theta q(x)dx$. The third property directly follows from the fact that $\mathcal{M}$ is IR.

For sufficiency, suppose $\mathbf{q}$ satisfies all of the properties in the statement, construct the payment $\mathbf{t}$ using Equation (4). Then by Observation 4, for every true type θ, misreporting to type θ' induces utility

$$V_\theta^*(E(\theta')) - t(\theta') = (\theta - \theta') \cdot q(\theta') + u_{2m} + \int_0^{\theta'} q(x)dx$$

Since $q(\theta')$ is non-decreasing, it is not hard to see that the utility is maximized when $\theta' = \theta$. Thus the mechanism is responsive-IC. Moreover, the utility for reporting truthfully her type θ is $u_{2m} + \int_0^\theta q(x)dx$. Thus by property 3, the mechanism is IR.

C.3 Proof of Lemma 6

By Lemma 5, for any $\mathbf{q}$ that can be implemented with some responsive-IC and IR mechanism $\mathcal{M}$, the revenue of $\mathcal{M}$ can be written as follows:

$$\text{REV}(\mathcal{M}) = \int_0^1 t(\theta)f(\theta)d\theta = \int_0^1 \left(\theta \cdot q(\theta) + \min\{u_{11} - u_{2m} - q(\theta), 0\} - \int_0^\theta q(x)dx \right) f(\theta)d\theta$$

$$= \int_0^1 [\theta f(\theta)q(\theta) + \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}]d\theta - \int_0^1 \int_0^\theta q(x)dx dF(\theta)$$

$$= \int_0^1 [\theta f(\theta)q(\theta) + \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}]d\theta - F(\theta) \int_0^\theta q(x)dx \Big|_0^1 + \int_0^1 q(\theta)F(\theta)d\theta$$

$$= \int_0^1 [(\theta f(\theta) + F(\theta))q(\theta) + \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}]d\theta - (u_{11} - u_{2m})$$

---

[14]$a \to b^+$ means that a approaches b from the right.

$$\textbf{sup} \int_0^1 [(\theta f(\theta) + F(\theta))q(\theta) + \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}]d\theta$$

$$\text{s.t.} \quad q(\theta) \text{ is non-decreasing in } \theta \in [0, 1]$$

$$q(0) \geq -u_{2m}, \quad q(1) \leq u_{11}$$

$$\int_0^1 q(\theta)d\theta = u_{11} - u_{2m}$$

$$u_{2m} + \int_0^\theta q(x)dx \geq u(\theta), \quad\quad\quad\quad \forall \theta \in [0, 1]$$

Fig. 6. The Optimization Problem with Explicit IR constraints

Thus the optimal mechanism $\mathcal{M}^* \in C_2'$ is captured by the optimization problem in Figure 6. Notice that there is an IR constraint for every type θ in Figure 6. To bound the size of the optimal responsive-IC and IR mechanism, we propose an equivalent optimization problem that only contains $O(m)$ constraints. Recall that for every $\theta \in [0, 1]$, $u(\theta) = \max_{k \in [m]}\{\theta \cdot u_{1k} + (1 - \theta) \cdot u_{2k}\}$. For every $k \in [m]$, let $h_k(\theta) = \theta \cdot u_{1,m+1-k} + (1 - \theta) \cdot u_{2,m+1-k} = \theta \cdot \ell_k + u_{2,m+1-k}, \forall \theta \in [0, 1]$ be the agent's value function when she has no additional information, and follows action $m + 1 - k$. Recall that $\ell_k = u_{1,m+1-k} - u_{2,m+1-k}$ is also the slope of the k-th piece (from the left) of the IR curve.

Given any feasible solution $\{q(\theta)\}_{\theta \in [0,1]}$ of the program in Figure 6, the IR constraint at type θ is equivalent to: $u_{2m} + \int_0^\theta q(x)dx \geq h_k(\theta), \forall k \in [m]$. For every $k \in [m]$, let $\theta_k := \sup\{\theta \in [0, 1] : q(\theta) \leq \ell_k\}$. Consider the following function

$$u_{2m} + \int_0^\theta q(x)dx - h_k(\theta) = u_{2m} + \int_0^\theta q(x)dx - \theta \cdot \ell_k - u_{2,m+1-k}$$

By taking the derivative on θ, we can see that the above function is minimized at θ_k, as $q(\theta) \leq \ell_k$ for all $\theta < \theta_k$ and $q(\theta) > \ell_k$ for all $\theta > \theta_k$. Thus the set of constraints $\left\{u_{2m} + \int_0^\theta q(x)dx \geq h_k(\theta), \forall \theta \in [0, 1]\right\}$ is captured by a single constraint $u_{2m} + \int_0^{\theta_k} q(x)dx \geq h_k(\theta_k)$. We notice that $\theta_k = \int_0^1 \mathbb{1}[q(x) \leq \ell_k]dx$. By the definition of h_k, the constraint can be rewritten as

$$\int_0^1 (q(x) - \ell_k) \cdot \mathbb{1}[q(x) \leq \ell_k]dx \geq u_{2,m+1-k} - u_{2m} \tag{11}$$

A feasible $\mathbf{q}$ must satisfy that $q(\theta) \in [-u_{2m}, u_{11}]$ and $\int_0^1 q(x)dx = u_{11} - u_{2m}$. Also $\ell_1 = -u_{2m}, \ell_m = u_{11}$. Thus, Inequality (11) is trivial when $k = 1$ or m. In our modified optimization problem, we only include the constraints for $2 \leq k \leq m - 1$.

Proof of Lemma 6:

It suffices to show that: A solution $\mathbf{q} = \{q(\theta)\}_{\theta \in [0,1]}$ is feasible in the optimization problem in Figure 6 if and only if it is feasible in the optimization problem in Figure 3.

We denote Q the optimization problem in Figure 6 and Q' the optimization problem in Figure 3. Suppose $\mathbf{q}$ is feasible in Q. For every $k \in [m]$, let $\theta_k = \sup\{\theta \in [0, 1] : q(\theta) \leq \ell_k\}$. Then since

$\theta_k = \int_0^1 \mathbb{1}[q(\theta) \leq \ell_k]dx$, we have

$$\int_0^1 (q(x) - \ell_k) \cdot \mathbb{1}[q(x) \leq \ell_k]dx = \int_0^{\theta_k} q(x)dx - \theta_k \ell_k$$
$$\geq u(\theta_k) - u_{2m} - \theta_k \ell_k \geq u_{2,m+1-k} - u_{2m}$$

Here the first inequality follows from the IR constraint of program Q at type θ_k, and the second inequality follows from the definition of the IR curve $u(\cdot)$. Thus $\mathbf{q}$ is feasible in Q'.

For the other direction, suppose $\mathbf{q}$ is feasible for Q'. For every $k \in [m]$, consider the following function $H_k : [0,1] \to \mathbb{R}$, where $H(\theta) := \int_0^\theta q(x)dx - \theta \cdot \ell_k$. Since $q(\theta)$ is non-decreasing in θ, $H_k(\theta)$ is minimized at $\theta = \theta_k = \sup\{\theta \in [0,1] : q(\theta) \leq \ell_k\}$. We notice that the last constraint of Q' implies that $H_k(\theta_k) \geq u_{2,m+1-k} - u_{2m}$ for all $k \in \{2,3,...,m-1\}$. It is not hard to verify that the inequality also holds for $k = 1$ and $k = m$. When $k = 1$, the inequality is trivial when $q(0) > -u_{2m}$. If $q(0) = -u_{2m}$, since $\ell_1 = -u_{2m}$ and $q(\theta) \in [-u_{2m}, u_{11}], \forall \theta \in [0,1]$, thus $q(\theta) = -u_{2m}, \forall \theta \in [0,\theta_1]$. Thus $H_1(\theta_1) = 0 = u_{2m} - u_{2m}$. When $k = m$, since $\theta_m = 1$ and $\ell_m = u_{11}$, $H_m(\theta_m) = u_{11} - u_{2m} - u_{11} = u_{21} - u_{2m}$.

Thus for every $k \in [m], \theta \in [0,1]$, we have

$$\int_0^\theta q(x)dx - \theta \cdot \ell_k \geq u_{2,m+1-k} - u_{2m}$$

Since $u(\theta) = \max_{k \in [m]}\{\theta \cdot \ell_k + u_{2,m+1-k}\}$, we have $\int_0^\theta q(x)dx \geq u(\theta) - u_{2m}$. Thus $\mathbf{q}$ is feasible for Q. □

Proof of Theorem 3: By Lemma 7, let $\mathcal{M}^*$ be the optimal semi-informative, responsive-IC and IR mechanism with option size at most $3m - 1$. By Lemma 4, $\text{REV}(\mathcal{M}^*) = \text{OPT}^* \geq \text{OPT}$. By Lemma 3, there exists a menu $\mathcal{M}$ that contains only the fully informative experiment, such that $\text{REV}(\mathcal{M}) \geq \frac{\text{REV}(\mathcal{M}^*)}{3m-1}$. Thus, $\text{FREV} \geq \frac{\text{OPT}}{3m-1}$. □

C.4 Proof of Lemma 7

We first consider the case where D is a discrete distribution. For simplicity we assume that $0 \in \text{SUPP}(D)$. Let $\text{SUPP}(D) = \{\theta_1, \ldots, \theta_N\}$, where $0 = \theta_1 < \theta_2 < \ldots < \theta_N \leq 1$ and N is the size of the support. For every $i \in N$, denote f_i the density of type θ_i. For ease of notations, we denote $\theta_{N+1} = 1$. The optimization problem in Figure 3 w.r.t. the discrete distribution D is shown in Figure 7. [15] The set of variables in the program is $\{q_i\}_{i \in [N]}$. Denote $\mathcal{P}(D)$ (or $\mathcal{P}$ when D is clear from context) the optimization problem in Figure 7.

Our goal is to turn the program $\mathcal{P}(D)$ into a collection of LPs, so that the highest optimum among the collection of LPs correspond to the optimum of $\mathcal{P}(D)$. Each LP is parametrized by a vector $(i^*, \mathbf{i} = \{i_2, \ldots, i_{m-1}\})$, and finds the optimal solution among all $(q_1, \ldots, q_N)$'s that satisfies the following conditions:

(1) $q_i \leq u_{11} - u_{2m}$ for all $i < i^*$ and $q_i \geq u_{11} - u_{2m}$ for all $i \geq i^*$.
(2) $q_i \leq \ell_k$ for all $i < i_k$ and $q_i \geq \ell_k$ for all $i \geq i_k$ for all $k \in \{2, \ldots, m-1\}$. Recall that $\ell_k = u_{1,m+1-k} - u_{2,m+1-k}$.

---

[15] For any solution $\{q_i\}_{i \in [N]}$, the problem in Figure 7 is clearly equivalent to the problem in Figure 3 where $q(\cdot) : [0,1] \to [-u_{2m}, u_{11}]$ is defined as $q(\theta) = q_i, \forall i \in [N], \theta \in [\theta_i, \theta_{i+1})$.

$$\max \sum_{i=1}^{N} \left[q_i \cdot \left(\theta_i \cdot f_i + (\theta_{i+1} - \theta_i) \cdot \sum_{j=1}^{i} f_j \right) + \min\{(u_{11} - u_{2m} - q_i) \cdot f_i, 0\} \right]$$

s.t.   (1) $q_i \le q_{i+1},$ $\forall i \in [N-1]$

         (2) $q_1 \ge -u_{2m}, \quad q_N \le u_{11}$

         (3) $\displaystyle\sum_{i=1}^{N} q_i \cdot (\theta_{i+1} - \theta_i) = u_{11} - u_{2m}$

         (4) $\displaystyle\sum_{i=1}^{N} \min\{q_i - \ell_k, 0\} \cdot (\theta_{j+1} - \theta_j) \ge u_{2,m+1-k} - u_{2m}, \quad \forall k \in \{2, 3, ..., m-1\}$

Fig. 7. The Optimal Responsive-IC and IR Mechanism for Discrete Distribution D

The reason that we consider solutions that satisfy these conditions is that now the objective and constraint (4) are linear. In particular, the objective becomes

$$\sum_{i=1}^{N} q_i \cdot \left(\theta_i \cdot f_i + (\theta_{i+1} - \theta_i) \cdot \sum_{j=1}^{i} f_j \right) + \sum_{i=i^*+1}^{N} (u_{11} - u_{2m} - q_i) \cdot f_i.$$

Constraint (4) becomes

$$\sum_{j=1}^{i_k} (q_j - \ell_k) \cdot (\theta_{j+1} - \theta_j) \ge u_{2,m+1-k} - u_{2m}.$$

To further simplify the program, we introduce variables $\hat{q}_i = q_{i+1} - q_i, \forall i \in [N-1]$, and replace q_i with $q_1 + \sum_{j=1}^{i-1} \hat{q}_j, \forall i \in [N]$. The monotonicity constraint of $\mathcal{P}$ is thus captured by all $\hat{q}_i$'s being non-negative. See Figure 8 for the LP we construct. In particular, constraint (1) corresponds to constraint (2) of $\mathcal{P}$; Constraint (2) corresponds to constraint (3) of $\mathcal{P}$. Constraints (3) and (4) follow from the definition of i^*. Constraints (5) and (6) follow from the definition of i_k, and (7) corresponds to constraint (4) of $\mathcal{P}$. In the LP in Figure 8, both i^* and $\mathbf{i} = \{i_k\}_{2 \le k \le m-1}$ are fixed parameters. We denote $\mathcal{P}'(D, i^*, \mathbf{i})$ the LP with those parameters.

We first show in Lemma 18 that the LP has an optimal solution $\hat{\mathbf{q}} = (q_1, \{\hat{q}_i\}_{i \in [N-1]})$ such that all but $(3m-1)$ $\hat{q}_i$'s are zero, for any discrete distribution D and any set of parameters $(i^*, \mathbf{i})$. We notice that it implies that the corresponding $\mathbf{q} = \{q_i\}_{i \in [N]}$ takes at most $3m-1$ different values, which then implies that the mechanism contains at most $3m-1$ different options.

LEMMA 18. *For any integer N and set of parameters $(i^*, \mathbf{i})$ where each parameter is in $[N]$, if $\mathcal{P}'(D, i^*, \mathbf{i})$ is feasible, then there exists an optimal solution of $\mathcal{P}'(D, i^*, \mathbf{i})$, denoted as $(q_1^*, \{\hat{q}_i^*\}_{i \in [N-1]})$, such that $|\{i \in [N-1] : \hat{q}_i^* > 0\}| \le 3m-1$.*

PROOF. The LP contains N variables (q_1 is unconstrained), 1 equality constraint, and $4+3(m-2) = 3m-2$ inequality constraints. We add slack variables to change the LP into the canonical form $\max\{\mathbf{c}^T \mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \ge \mathbf{0}\}$. In particular, we replace variable q_1 by two non-negative variables q_1^+, q_1^- such that $q_1 = q_1^+ - q_1^-$. For each inequality constraints, we add one slack variable and make the constraint an equality.

Now, we have a canonical form LP, with $N + 3m - 1$ variables, and $3m - 1$ equality constraints. Thus by the Fundamental Theorem of linear programming, if $\mathcal{P}'(D, i^*, \mathbf{i})$ is feasible, any basic

$$\max \sum_{i=1}^{N}\left(q_1 + \sum_{j=1}^{i-1}\hat{q}_j\right) \cdot \left(\theta_i f_i + (\theta_{i+1} - \theta_i) \cdot \sum_{r=1}^{i} f_r\right) + \sum_{i=i^*+1}^{N}\left(u_{11} - u_{2m} - q_1 - \sum_{j=1}^{i-1}\hat{q}_j\right) \cdot f_i$$

s.t. (1) $q_1 \geq -u_{2m}, \quad q_1 + \sum_{j=1}^{N-1}\hat{q}_j \leq u_{11}$

(2) $\sum_{i=1}^{N}\left(q_1 + \sum_{j=1}^{i-1}\hat{q}_j\right) \cdot (\theta_{i+1} - \theta_i) = u_{11} - u_{2m}$

(3) $q_1 + \sum_{j=1}^{i^*-1}\hat{q}_j \leq u_{11} - u_{2m}$

(4) $q_1 + \sum_{j=1}^{i^*}\hat{q}_j \geq u_{11} - u_{2m}$

(5) $q_1 + \sum_{j=1}^{i_k-1}\hat{q}_j \leq u_{1,m+1-k} - u_{2,m+1-k}, \qquad\qquad \forall k \in \{2, 3, \ldots, m-1\}$

(6) $q_1 + \sum_{j=1}^{i_k}\hat{q}_j \geq u_{1,m+1-k} - u_{2,m+1-k}, \qquad\qquad \forall k \in \{2, 3, \ldots, m-1\}$

(7) $\sum_{j=1}^{i_k}\left(q_1 + \sum_{r=1}^{j-1}\hat{q}_r - \ell_k\right) \cdot (\theta_{j+1} - \theta_j) \geq u_{2,m+1-k} - u_{2m}, \quad \forall k \in \{2, 3, \ldots, m-1\}$

$\hat{q}_i \geq 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall i \in [N-1]$

Fig. 8. The Linear Program with Parameters $(i^*, \mathbf{i})$

feasible solution in the canonical-form LP has at most $3m - 1$ non-zero entries. Since there must be an optimal solution of $\mathcal{P}'(D, i^*, \mathbf{i})$ that correspond to a basic feasible solution, so it must have at most $3m - 1$ non-zero entries. □

Proof of Lemma 7: We first prove the case when D is a discrete distribution. Let OPT′ be the maximum objective, over all parameters $(i^*, \mathbf{i})$, and all feasible solutions of $\mathcal{P}'(D, i^*, \mathbf{i})$.

CLAIM 1. *For any discrete distribution $D = (\{\theta_i\}_{i\in[N]}, \{f_i\}_{i\in[N]})$, a solution $\mathbf{q} = \{q_i\}_{i\in[N]}$ is feasible for $\mathcal{P}(D)$ if and only if there exists parameters $(i^*, \mathbf{i})$ such that the solution $\hat{\mathbf{q}} = (q_1, \{\hat{q}_i\}_{i\in[N-1]})$ is feasible for $\mathcal{P}'(D, i^*, \mathbf{i})$. Here $\hat{q}_i = q_{i+1} - q_i, \forall i \in [N-1]$. Moreover, OPT′ = OPT*.*

PROOF. Suppose $\mathbf{q}$ is feasible for $\mathcal{P}(D)$. Let $i^* = \max\{i \in [N] : q_i \leq u_{11} - u_{2m}\}$. For every $k \in \{2, \ldots, m-1\}$, let $i_k = \max\{i \in [N] : q_i \leq \ell_k\}$. We verify that $\hat{\mathbf{q}}$ satisfies all constraints of $\mathcal{P}'(D, i^*, \mathbf{i})$. Constraints (1), (2) follow from constraints (2), (3) of $\mathcal{P}(D)$ accordingly. Constraints (3)-(6) follow from the definition of i^* and all i_k's. Constraint (7) follows from the definition of i_k and (4) of $\mathcal{P}(D)$. Moreover, by the definition of i^* and the fact that $q_i = q_1 + \sum_{j=1}^{i-1}\hat{q}_j, \forall i \in [N]$, the objective of solution $\mathbf{q}$ in $\mathcal{P}(D)$ is equal to the objective of solution q_1 and $\hat{\mathbf{q}}$ in $\mathcal{P}'(D, i^*, \mathbf{i})$. Choosing $\mathbf{q}$ as the optimal solution for $\mathcal{P}(D)$ implies that OPT* $\leq$ OPT′.

On the other hand, suppose $\hat{\mathbf{q}}$ is feasible for $\mathcal{P}'(D, i^*, \mathbf{i})$ for some parameters $(i^*, \mathbf{i})$. For every $i \in \{2, \ldots, N\}$, define $q_i = q_1 + \sum_{j=1}^{i-1}\hat{q}_i$. We verify that $\mathbf{q}$ is feasible for $\mathcal{P}(D)$. Constraint (1) holds as all $\hat{q}_i$'s are non-negative; Constraints (2) and (3) follow from constraints (1) and (2) of $\mathcal{P}'(D, i^*, \mathbf{i})$

accordingly; Constraint (4) follows from constraints (5)-(7) of $\mathcal{P}'(D, i^*, \mathbf{i})$. Similarly, according to constraints (3) and (4) of $\mathcal{P}'(D, i^*, \mathbf{i})$ and the fact that $q_i = q_1 + \sum_{j=1}^{i-1} \hat{q}_i, \forall i \in [N]$, the objective of solution $\hat{\mathbf{q}}$ in $P'(D, i^*, \mathbf{i})$ is equivalent to the objective of solution $\mathbf{q}$ in $\mathcal{P}(D)$. Thus by choosing $\hat{\mathbf{q}}$ as the optimal solution among all parameters $(i^*, \mathbf{i})$ and all feasible solutions of $\mathcal{P}'(D, i^*, \mathbf{i})$, we have $\mathrm{OPT}' \leq \mathrm{OPT}^*$. Hence, $\mathrm{OPT}' = \mathrm{OPT}^*$. □

By Lemma 18 and the definition of OPT', there exist a set of parameters $(i^*, \mathbf{i})$ as well as a feasible solution $\hat{\mathbf{q}}$ of $\mathcal{P}'(D, i^*, \mathbf{i})$ such that: (i) the solution $\hat{\mathbf{q}}$ achieves objective OPT' in the program $\mathcal{P}'(D, i^*, \mathbf{i})$; (ii) $|\{i \in [N-1] : \hat{q}_i^* > 0\}| \leq 3m - 1$. By Claim 1 and property (i), the corresponding $\mathbf{q}$ is the optimal solution of $\mathcal{P}(D)$. By property (ii), there are at most $3m - 1$ different values among $\{q_i\}_{i \in [N]}$. Let $\mathcal{M} = \{(q_i, t_i)\}_{i \in [N]} \in C_2'$ be the responsive-IC and IR mechanism that implements $\mathbf{q}$. According to Equation (4) and the fact that $q_i \leq q_{i+1}, \forall i \in [N-1]$, $t_i = t_j$ for all i, j such that $q_i = q_j$. Thus $\mathcal{M}$ has option size at most $3m - 1$.

Next we prove the case for continuous distributions.

We consider a discretization of the continuous interval $[0, 1]$. Let $N \geq 2$ be any integer and $\varepsilon = \frac{1}{N}$. We consider the discretized space $\{0, \varepsilon, 2\varepsilon, ..., (N-1)\varepsilon\}$. We denote by $D^{(N)}$ the discretized distribution of D, and $\{f_i^{(N)}\}_{i \in [N]}$, $\{F_i^{(N)}\}_{i \in [N]}$ the pdf and cdf of $D^{(N)}$. Formally, $F_i^{(N)} = F(i\varepsilon)$, and $f_i^{(N)} = F(i\varepsilon) - F((i-1)\varepsilon), \forall i \in [N]$.

Denote $\mathcal{S} = \{\mathbf{q} : [0, 1] \to [-u_{2m}, u_{11}] : q(\cdot) \text{ is non-decreasing}\}$. For every N and every $\mathbf{q} \in \mathcal{S}$, denote $h_N(\mathbf{q})$ the objective of the program in Figure 7 under solution $\{q((i-1)\varepsilon)\}_{i \in [N]}$, with respect to the distribution $D^{(N)}$. Formally,

$$h_N(\mathbf{q}) = \sum_{i=1}^{N} \left[q((i-1)\varepsilon) \cdot ((i-1)\varepsilon \cdot f_i^{(N)} + \varepsilon \cdot F_i^{(N)}) + \min\{(u_{11} - u_{2m} - q((i-1)\varepsilon)) \cdot f_i^{(N)}, 0\} \right]$$

For every $\mathbf{q} \in \mathcal{S}$, denote $h(\mathbf{q})$ the objective of the program in Figure 3 under solution $\mathbf{q}$, with respect to the continuous distribution D, i.e.,

$$h(\mathbf{q}) = \int_0^1 [(\theta f(\theta) + F(\theta))q(\theta) + \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}]d\theta$$

CLAIM 2. $h_N(\cdot)$ uniformly converges to $h(\cdot)$ on $\mathcal{S}$, i.e., for every $\delta > 0$ there exists $N_0 \in \mathcal{N}$ such that $|h_N(\mathbf{q}) - h(\mathbf{q})| < \delta, \forall N \geq N_0, \mathbf{q} \in \mathcal{S}$.

PROOF. Fix any $\delta > 0$. Fix any $\mathbf{q} \in \mathcal{S}$. Consider the term

$$A_1^{(N)}(\mathbf{q}) = \sum_{i=1}^{N} \left[q((i-1)\varepsilon) \cdot ((i-1)\varepsilon \cdot f_i^{(N)} + \varepsilon \cdot F_i^{(N)}) \right] - \int_0^1 (\theta f(\theta) + F(\theta))q(\theta)d\theta$$

$$= \sum_{i=1}^{N} \left[q((i-1)\varepsilon) \cdot (i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon)) - \int_{(i-1)\varepsilon}^{i\varepsilon} (\theta f(\theta) + F(\theta))q(\theta)d\theta \right]$$

Here the second equality follows from the definition of $f_i^{(N)}, F_i^{(N)}$. For every $i \in [N]$, since $q(\cdot)$ is non-decreasing,

$$\int_{(i-1)\varepsilon}^{i\varepsilon} (\theta f(\theta) + F(\theta))q(\theta)d\theta \geq q((i-1)\varepsilon) \cdot (\theta F(\theta)) \Big|_{\theta=(i-1)\varepsilon}^{i\varepsilon}$$

$$= q((i-1)\varepsilon) \cdot (i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon))$$

Thus $A_1^{(N)}(\mathbf{q}) \le 0$. On the other hand, for every $i \in [N]$, similarly we have

$$\int_{(i-1)\varepsilon}^{i\varepsilon} (\theta f(\theta) + F(\theta))q(\theta)d\theta \le q(i\varepsilon) \cdot (i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon))$$

Thus

$$A_1^{(N)}(\mathbf{q}) \ge \sum_{i=1}^{N}(q((i-1)\varepsilon) - q(i\varepsilon)) \cdot (i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon))$$

$$\ge \max_{i\in[N]}\{i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon)\} \cdot (q(0) - q(1))$$

$$\ge -2u_{11} \cdot \max_{i\in[N]}\{i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon)\}$$

Here the last inequality follows from $q(0) \ge -u_{2m} \ge -u_{11}$ and $q(1) \le u_{11}$. Since F is continuous on the closed interval $[0, 1]$, the function $\theta \cdot F(\theta)$ is continuous on $[0, 1]$ and thus it is uniformly continuous. Hence there exists $N_1 \in \mathcal{N}$ such that $\max_{i\in[N]}\{i\varepsilon \cdot F(i\varepsilon) - (i-1)\varepsilon \cdot F((i-1)\varepsilon)\} < \frac{\delta}{4u_{11}}, \forall N \ge N_1$. Thus $A_1^{(N)}(\mathbf{q}) > -\frac{\delta}{2}$.

Now consider the other term

$$A_2^{(N)}(\mathbf{q}) = \sum_{i=1}^{N}\min\{(u_{11} - u_{2m} - q((i-1)\varepsilon)) \cdot f_i^{(N)}, 0\} - \int_0^1 \min\{(u_{11} - u_{2m} - q(\theta))f(\theta), 0\}d\theta$$

$$= \sum_{i=1}^{N}\left[f_i^{(N)} \cdot \min\{\bar{q}((i-1)\varepsilon), 0\} - \int_{(i-1)\varepsilon}^{i\varepsilon}\min\{\bar{q}(\theta), 0\}f(\theta)d\theta\right],$$

where $\bar{q}(\theta) = u_{11} - u_{2m} - q(\theta) \in [-u_{2m}, u_{11}]$ is non-increasing on θ. For every $i \in [N]$, we denote T_i the i-th term in the sum. Let $\theta^* = \sup\{\theta \in [0, 1] : \bar{q}(\theta) \le 0\}$ and $i^* = \max\{i \in [N] : (i-1)\varepsilon \le \theta^* < i\varepsilon\}$.

When $i < i^*$, since $\bar{q}(\cdot)$ is non-increasing and $f_i^{(N)} = F(i\varepsilon) - F((i-1)\varepsilon)$,

$$T_i \ge f_i^{(N)} \cdot \bar{q}((i-1)\varepsilon) - \bar{q}((i-1)\varepsilon) \cdot \int_{(i-1)\varepsilon}^{i\varepsilon} f(\theta)d\theta = 0$$

$$T_i \le f_i^{(N)} \cdot [\bar{q}((i-1)\varepsilon) - \bar{q}(i\varepsilon)]$$

When $i > i^*$. Clearly $T_i = 0$. When $i = i^*$,

$$|T_{i^*}| = \left|f_{i^*}^{(N)} \cdot \bar{q}((i^*-1)\varepsilon) - \int_{(i^*-1)\varepsilon}^{\theta^*}\bar{q}(\theta)f(\theta)d\theta\right| \le 2u_{11} \cdot f_{i^*}^{(N)},$$

where the inequality follows from the fact that $|\bar{q}(\theta)| \le \max\{u_{11}, u_{2m}\} = u_{11}, \forall \theta \in [0, 1]$ and that $F(\theta^*) - F((i^*-1)\varepsilon) \le f_{i^*}^{(N)}$. Thus

$$|A_2^{(N)}(\mathbf{q})| \le \left|\sum_{i=1}^{i^*-1} T_i\right| + |T_{i^*}| \le \left|\sum_{i=1}^{i^*-1} f_i^{(N)} \cdot [\bar{q}((i-1)\varepsilon) - \bar{q}(i\varepsilon)]\right| + 2u_{11} \cdot f_{i^*}^{(N)}$$

$$\le (|\bar{q}((i-1)\varepsilon) - \bar{q}(0)| + 2u_{11}) \cdot \max_{i\in[N]} f_i^{(N)}$$

$$\le 4u_{11} \cdot \max_{i\in[N]} f_i^{(N)}$$

Since $F(\cdot)$ is uniformly continuous on $[0, 1]$, there exists $N_2 \in \mathcal{N}$ such that $\max_i f_i^{(N)} = \max_i(F(i\varepsilon) - F((i-1)\varepsilon)) < \frac{\delta}{8u_{11}}, \forall N \ge N_2$. Thus $|A_2^{(N)}(\mathbf{q})| < \frac{\delta}{2}$.

Combining everything together, for every $N \geq N_0 = \max\{N_1, N_2\}$, for every $\mathbf{q} \in \mathcal{S}$, we have

$$|h_N(\mathbf{q}) - h(\mathbf{q})| = |A_1^{(N)}(\mathbf{q}) + A_2^{(N)}(\mathbf{q})| < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

Thus $h_N(\cdot)$ uniformly converges to $h(\cdot)$ on $\mathcal{S}$. □

Back to the proof of Lemma 7. Denote $\mathcal{P}_{cont}$ the program in Figure 3 and OPT* the supremum of the objective over all feasible solutions of $\mathcal{P}_{cont}$. We first argue that there exists a feasible solution $\mathbf{q}^*$ whose objective $h(\mathbf{q}^*)$ equals to OPT*.

CLAIM 3. *There exists a feasible solution $\mathbf{q}^*$ to $\mathcal{P}_{cont}$, whose objective $h(\mathbf{q}^*)$ equals to OPT*.*

PROOF. Let $\{\mathbf{q}_\ell^*\}_{\ell \in \mathcal{N}_+}$ be a sequence of feasible solutions such that $\lim_{\ell \to \infty} h(\mathbf{q}_\ell^*) = $ OPT*. We notice that for every ℓ, the feasible solution $\mathbf{q}_\ell^*(\cdot)$ is a non-decreasing function mapping $[0, 1]$ to a bounded interval $[-u_{2m}, u_{11}]$. Thus $\{\mathbf{q}_\ell^*\}_{\ell \in \mathcal{N}_+}$ is a sequence of non-decreasing functions and it's uniformly-bounded. By Helly's selection theorem (see for instance [2, 6]), there exists a subsequence $\{\mathbf{q}_{\ell_i}^*\}_{i \in \mathcal{N}_+}$ and a function $\mathbf{q}^* : [0, 1] \to [-u_{2m}, u_{11}]$ such that $\{\mathbf{q}_{\ell_i}^*\}_i$ pointwisely converges to $\mathbf{q}^*$.

For every $\theta \in [0, 1]$ and $q \in [-u_{2m}, u_{11}]$, define

$$G(\theta, q) = (\theta f(\theta) + F(\theta))q + \min\{(u_{11} - u_{2m} - q)f(\theta), 0\} \tag{12}$$

For every $\theta \in [0, 1]$, $i \in \mathcal{N}_+$, define $g_i(\theta) = G(\theta, q_{\ell_i}^*(\theta))$ and $g(\theta) = G(\theta, q^*(\theta))$. Then since $G(\theta, q)$ is continuous on q for every θ, $\{g_i\}_{i \in \mathcal{N}_+}$ pointwisely converges to g. Moreover, define $\bar{g}(\theta) = u_{11} \cdot (\theta f(\theta) + F(\theta) + f(\theta)) \geq 0, \forall \theta \in [0, 1]$. Then $\bar{g}(\cdot)$ is integrable in $[0, 1]$. And for every i and θ, since $q_{\ell_i}^*(\theta) \in [-u_{2m}, u_{11}]$ and $u_{11} \geq u_{2m}$, we have $|g_i(\theta)| \leq \bar{g}(\theta)$. Thus by the dominated convergence theorem,

$$\text{OPT}^* = \lim_{i \to \infty} h(\mathbf{q}_{\ell_i}^*) = \lim_{i \to \infty} \int_0^1 g_i(\theta)d\theta = \int_0^1 (\lim_{i \to \infty} g_i(\theta))d\theta = \int_0^1 g(\theta)d\theta = h(\mathbf{q}^*)$$

It remains to verify that $\mathbf{q}^*$ is a feasible solution to $\mathcal{P}_{cont}$. Since $\{\mathbf{q}_{\ell_i}^*\}_i$ is a sequence of feasible solutions, all constraints in $\mathcal{P}_{cont}$ hold for $\mathbf{q}_{\ell_i}^*$. For every inequality, we take the limit $i \to \infty$ on both sides. Constraints (1), (2) hold for $\mathbf{q}^*$ since $\{\mathbf{q}_{\ell_i}^*\}_i$ pointwisely converges to $\mathbf{q}^*$. In order for constraints (3) and (4) to hold for $\mathbf{q}^*$, it's sufficient to argue that we can swap the limit and integral for both inequalities. This is because for every i and θ, we have $|q_{\ell_i}^*(\theta)| \leq u_{11}$ and $|(q_{\ell_i}^*(\theta) - \ell_k) \cdot \mathbb{1}[q_{\ell_i}^*(\theta) \leq \ell_k]| \leq |q_{\ell_i}^*(\theta)| + |\ell_k| \leq 2u_{11}$ (recall that $|\ell_k| = |u_{1,m+1-k} - u_{2,m+1-k}| \leq \max\{u_{11}, u_{2m}\} = u_{11}$). Thus by the dominated convergence theorem, constraints (3) and (4) both hold for $\mathbf{q}^*$. Thus $\mathbf{q}^*$ is feasible for $\mathcal{P}_{cont}$. □

For every $N \geq 2$, denote OPT$_N$ the optimum of $\mathcal{P}(D^{(N)})$, the program in Figure 7 with respect to the distribution $D^{(N)}$. By Lemma 18 and Claim 1, there exists a set of parameters $(i^*, \mathbf{i})$ and a feasible solution $(q_1^{(N)}, \{\hat{q}_i^{(N)}\}_{i \in [N-1]})$ to $\mathcal{P}'(D^{(N)}, i^*, \mathbf{i})$ such that: (i) the solution has objective exactly OPT$_N$, and (ii) $|\{i \in [N-1] : \hat{q}_i^* > 0\}| \leq 3m - 1$. For every $i \in \{2, \ldots, N\}$, let $q_i^{(N)} = q_1^{(N)} + \sum_{j=1}^{i-1} \hat{q}_j^{(N)}$. Define $\mathbf{q}^{(N)} : [0, 1] \to [-u_{2m}, u_{11}]$ as follows: $q^{(N)}(\theta) = q_i^{(N)}, \forall i \in [N], \theta \in [(i-1)\varepsilon, i\varepsilon)$, and $q^{(N)}(1) = q_N^{(N)}$. By Claim 1, $\{q^{(N)}((i-1)\varepsilon)\}_{i \in [N]}$ is a feasible solution to program $\mathcal{P}(D^{(N)})$. Moreover, $h_N(\mathbf{q}^{(N)}) = $ OPT$_N$.

For every $i \in [N]$, by Claim 1, $\{q_i^{(N)}\}_{i \in [N]}$ is a feasible solution to $\mathcal{P}(D^{(N)})$. By the definition of $\mathbf{q}^{(N)}$, it's not hard to verify that $\mathbf{q}^{(N)}$ is a feasible solution to $\mathcal{P}_{cont}$. By property (ii), Im$(\mathbf{q}) = \{q(\theta) : \theta \in [0, 1]\}$, the image of $\mathbf{q}(\cdot)$, has size at most $3m - 1$. Let $n = 3m - 1$ and $\mathcal{S}' \subseteq \mathcal{S}$ be the set of all feasible solutions $\mathbf{q}$ to $\mathcal{P}_{cont}$ such that $|\text{Im}(\mathbf{q})| \leq n$. We notice that there is a mapping Φ from every $\mathbf{q} \in \mathcal{S}'$ to a set of $(2n - 1)$ real numbers $\xi = (\{\theta_i\}_{i \in [n-1]}, \{q_i\}_{i \in [n]})$ such that $q(\theta) = q_i, \forall i \in [n], \theta \in [\theta_{i-1}, \theta_i)$ and $q(1) = q_n$. Here $\theta_0 = 0, \theta_n = 1$.

Let $\mathcal{T}$ be the set of ξ's that satisfy all of the following properties:

(1) $0 \le \theta_1 \le \ldots \le \theta_{n-1} \le 1$.

(2) $-u_{2m} \le q_1 \le \ldots \le q_n \le u_{11}$.

(3) $\sum_{i=1}^n (\theta_i - \theta_{i-1}) \cdot q_i = u_{11} - u_{2m}$.

(4) $\sum_{i=1}^n (\theta_i - \theta_{i-1}) \cdot (q_i - \ell_k) \cdot \mathbb{1}[q_i \le \ell_k] \ge u_{2,m+1-k} - u_{2m}, \forall k \in \{2, 3, \ldots, m-1\}$.

Then it's not hard to verify that Φ is an 1-1 mapping from $\mathcal{S}'$ to $\mathcal{T}$. Denote Φ^{-1} the inverse of Φ. Moreover, since every constraint above is either equality or non-strict inequality among the sum and product of numbers ($\{\theta_i\}_{i \in [n-1]}, \{q_i\}_{i \in [n]}$), thus $\mathcal{T}$ is a closed subset of $\mathbb{R}^{2n-1}$. Thus $\mathcal{T}$ is a compact space under the L_∞-norm. Therefore, the sequence $\{\Phi(\mathbf{q}^{(N)})\}_N$ contains a subsequence $\{\Phi(\mathbf{q}^{(N_\ell)})\}_{\ell \in \mathbb{N}_+}$ that converges to some $\xi' \in \mathcal{T}$. Let $\mathbf{q}' = \Phi^{-1}(\xi') \in \mathcal{S}'$.

CLAIM 4. $\lim_{\ell \to \infty} h_{N_\ell}(\mathbf{q}^{(N_\ell)}) = h(\mathbf{q}')$.

PROOF. Fix any $\delta > 0$. Since both functions $\theta F(\theta)$ and $F(\theta)$ are continuous, there exists some $\eta(\delta) \in (0, \frac{\delta}{8n})$ such that $|\theta F(\theta) - \theta' F(\theta')| < \frac{\delta}{16n \cdot u_{11}}$ and $|F(\theta) - F(\theta')| < \frac{\delta}{16n \cdot u_{11}}$ for any θ and θ' with $|\theta - \theta'| < \eta(\delta)$.

Since $\{\Phi(\mathbf{q}^{(N_\ell)})\}_{\ell \in \mathbb{N}_+}$ converges to $\Phi(\mathbf{q}')$, then there exists ℓ_1 such that for every $\ell \ge \ell_1$, $\|\Phi(\mathbf{q}^{(N_\ell)}) - \Phi(\mathbf{q}')\|_\infty < \eta(\delta)$. Denote $\Phi(\mathbf{q}^{(N_\ell)}) = (\{\theta_i^\ell\}_{i \in [n-1]}, \{q_i^\ell\}_{i \in [n]})$ and $\Phi(\mathbf{q}') = (\{\theta_i'\}_{i \in [n-1]}, \{q_i'\}_{i \in [n]})$. Let $\theta_0^\ell = \theta_0' = 0$ and $\theta_n^\ell = \theta_n' = 1$. We are going to bound $|h(\mathbf{q}^{(N_\ell)}) - h(\mathbf{q}')|$. We notice that

$$h(\mathbf{q}^{(N_\ell)}) = \sum_{i=1}^n \left[q_i^\ell \cdot \int_{\theta_{i-1}^\ell}^{\theta_i^\ell} (\theta f(\theta) + F(\theta)) d\theta + (F(\theta_i^\ell) - F(\theta_{i-1}^\ell)) \cdot \min\{u_{11} - u_{2m} - q_i^\ell, 0\} \right]$$

$$= \sum_{i=1}^n \left[q_i^\ell \cdot (\theta_i^\ell F(\theta_i^\ell) - \theta_{i-1}^\ell F(\theta_{i-1}^\ell)) + (F(\theta_i^\ell) - F(\theta_{i-1}^\ell)) \cdot \min\{u_{11} - u_{2m} - q_i^\ell, 0\} \right]$$

Similarly,

$$h(\mathbf{q}') = \sum_{i=1}^n \left[q_i' \cdot (\theta_i' F(\theta_i') - \theta_{i-1}' F(\theta_{i-1}')) + (F(\theta_i') - F(\theta_{i-1}')) \cdot \min\{u_{11} - u_{2m} - q_i', 0\} \right]$$

Since $|\theta_i^\ell - \theta_i'| < \eta(\delta), \forall i \in [n-1]$, then for every $i \in [n]$, $|(\theta_i^\ell F(\theta_i^\ell) - \theta_{i-1}^\ell F(\theta_{i-1}^\ell)) - (\theta_i' F(\theta_i') - \theta_{i-1}' F(\theta_{i-1}'))| < 2 \cdot \frac{\delta}{16n \cdot u_{11}}$, and $|(F(\theta_i^\ell) - F(\theta_{i-1}^\ell)) - (F(\theta_i') - F(\theta_{i-1}'))| < 2 \cdot \frac{\delta}{16n \cdot u_{11}}$.

Since $\|\Phi(\mathbf{q}^{(N_\ell)}) - \Phi(\mathbf{q}')\|_\infty < \eta(\delta) < \frac{\delta}{8n}$, we have

$$|\min\{u_{11} - u_{2m} - q_i^\ell, 0\} - \min\{u_{11} - u_{2m} - q_i', 0\}| \le |q_i^\ell - q_i'| < \frac{\delta}{8n}$$

Moreover, $|q_i^\ell| \le u_{11}, |\min\{u_{11} - u_{2m} - q_i^\ell, 0\}| \le u_{11}$ and $\theta_i' F(\theta_i') - \theta_{i-1}' F(\theta_{i-1}') \in [0, 1], F(\theta_i') - F(\theta_{i-1}') \in [0, 1]$. Using the inequality $|ab - cd| \le |a| \cdot |b - d| + |d| \cdot |a - c|$ for every $a, b, c, d \in \mathbb{R}$, we have for every $\ell \ge \ell_1$,

$$|h(\mathbf{q}^{(N_\ell)}) - h(\mathbf{q}')| \le \sum_{i=1}^n 2 \cdot (u_{11} \cdot \frac{\delta}{8n \cdot u_{11}} + 1 \cdot \frac{\delta}{8n}) < \frac{\delta}{2}$$

Moreover, by Claim 2, there exists ℓ_2 such that for every $\ell \ge \ell_2$, $|h_{N_\ell}(\mathbf{q}^{(N_\ell)}) - h(\mathbf{q}^{(N_\ell)})| < \frac{\delta}{2}$. Thus when $\ell \ge \max\{\ell_1, \ell_2\}$, we have $|h_{N_\ell}(\mathbf{q}^{(N_\ell)}) - h(\mathbf{q}')| \le |h(\mathbf{q}^{(N_\ell)}) - h(\mathbf{q}')| + |h_{N_\ell}(\mathbf{q}^{(N_\ell)}) - h(\mathbf{q}^{(N_\ell)})| < \delta$. □

We prove $h(\mathbf{q}') \ge \text{OPT}^*$. For every N, we construct a feasible solution of $\mathcal{P}(D^{(N)})$. For every $i \in [N]$, let $p_i^{(N)} = \frac{1}{\varepsilon} \cdot \int_{(i-1)\varepsilon}^{i\varepsilon} q^*(\theta) d\theta$ (see Claim 3 for the definition of $\mathbf{q}^*$). We verify that $\{p_i^{(N)}\}_{i \in [N]}$ is a feasible solution to $\mathcal{P}(D^{(N)})$. Constraint (1) follows from $q^*(\cdot)$ being non-decreasing. For

constraint (2), since $q^*(\cdot)$ is non-decreasing, $p_1^{(N)} \geq \frac{1}{\varepsilon} \cdot \int_0^\varepsilon q^*(0)d\theta = q^*(0) \geq -u_{2m}$. Similarly $p_N^{(N)} \leq q^*(1) \leq u_{11}$. For constraint (3), we have

$$\varepsilon \cdot \sum_{i=1}^{N} p_i^{(N)} = \sum_{i=1}^{N} \int_{(i-1)\varepsilon}^{i\varepsilon} q^*(\theta)d\theta = \int_0^1 q^*(\theta)d\theta = u_{11} - u_{2m}$$

For constraint (4), for every $k \in \{2, \dots, m-1\}$. Let $\theta_k = \sup\{\theta \in [0,1] : q^*(\theta) \leq \ell_k\}$ and let i^* be the unique number such that $(i^*-1)\varepsilon \leq \theta_k < i^*\varepsilon$. Then $p_i^{(N)} \leq \ell_k, \forall i < i^*$ and $p_i^{(N)} > \ell_k, \forall i > i^*$. We have

$$
\begin{aligned}
\varepsilon \cdot \sum_{i=1}^{N} \min\{p_i^{(N)} - \ell_k, 0\} &= \sum_{i=1}^{i^*-1} \left(\int_{(i-1)\varepsilon}^{i\varepsilon} q^*(\theta)d\theta - \varepsilon\ell_k \right) + \varepsilon \cdot \min\{p_{i^*}^{(N)} - \ell_k, 0\} \\
&= \int_0^{(i^*-1)\varepsilon} (q^*(\theta) - \ell_k)d\theta + \varepsilon \cdot \min\{p_{i^*}^{(N)} - \ell_k, 0\} \\
&\geq \int_0^{(i^*-1)\varepsilon} (q^*(\theta) - \ell_k)d\theta + \int_{(i^*-1)\varepsilon}^{\theta_k} (q^*(\theta) - \ell_k)d\theta \\
&= \int_0^1 (q^*(\theta) - \ell_k)\mathbb{1}[q^*(\theta) \leq \ell_k]d\theta \geq u_{2,m+1-k} - u_{2m}
\end{aligned}
\tag{13}
$$

The first inequality of Equation (13) is because: By the definition of θ_k, $\int_{\theta_k}^{i^*\varepsilon} (q^*(\theta) - \ell_k)d\theta \geq 0$. Thus

$$\varepsilon \cdot \min\{p_{i^*}^{(N)} - \ell_k, 0\} \geq \varepsilon \cdot (p_{i^*}^{(N)} - \ell_k) = \int_{(i^*-1)\varepsilon}^{i^*\varepsilon} (q^*(\theta) - \ell_k)d\theta \geq \int_{(i^*-1)\varepsilon}^{\theta_k} (q^*(\theta) - \ell_k)d\theta$$

The third equality of Equation (13) follows from the definition of θ_k. The last inequality follows from that $\mathbf{q}^*$ is a feasible solution to $\mathcal{P}_{cont}$. Thus $\{p_i^{(N)}\}_{i\in[N]}$ is a feasible solution to $\mathcal{P}(D^{(N)})$. Define $\mathbf{p}^{(N)} : [0,1] \rightarrow [-u_{2m}, u_{11}]$ as follows: $p^{(N)}(\theta) = p_i^{(N)}, \forall i \in [N], \theta \in [(i-1)\varepsilon, i\varepsilon)$, and $p^{(N)}(1) = p_N^{(N)}$. Then $h_N(\mathbf{p}^{(N)})$ is exactly the objective of $\{p_i^{(N)}\}_{i\in[N]}$ with respect to the program $\mathcal{P}(D^{(N)})$. By the optimality of $\mathbf{q}^{(N)}$, $h_N(\mathbf{p}^{(N)}) \leq h_N(\mathbf{q}^{(N)})$. Thus $\forall \ell \in \mathcal{N}_+, h_{N_\ell}(\mathbf{p}^{(N_\ell)}) \leq h_{N_\ell}(\mathbf{q}^{(N_\ell)})$.

We will argue that $h_{N_\ell}(\mathbf{p}^{(N_\ell)})$ converges to OPT* as $\ell \rightarrow \infty$. If this is true, then by taking the limit on both sides of the above inequality, we have that OPT$^* \leq \lim_{\ell\to\infty} h_{N_\ell}(\mathbf{q}^{(N_\ell)}) = h(\mathbf{q}')$, where the equality follows from Claim 4.

Since $q^*(\cdot)$ is non-decreasing, for every $i \in [N]$ and $\theta \in [(i-1)\varepsilon, i\varepsilon)$, we have

$$|p^{(N)}(\theta) - q^*(\theta)| = |p_i^{(N)} - q^*(\theta)| \leq q^*(i\varepsilon) - q^*((i-1)\varepsilon).$$

Thus if $q^*(\cdot)$ is continuous at θ, then $\lim_{N\to\infty} |p^{(N)}(\theta) - q^*(\theta)| = 0$. Moreover, since $q^*(\cdot)$ is a monotone function in a closed interval $[0,1]$, the set of non-continuous points has zero measure. Thus $\lim_{N\to\infty} p^{(N)}(\theta) = q^*(\theta)$ almost everywhere.

Similar to the proof of Claim 3, we have

$$\lim_{\ell\to\infty} h(\mathbf{p}^{(N_\ell)}) = \lim_{\ell\to\infty} \int_0^1 G(\theta, p^{(N_\ell)}(\theta))d\theta = \int_0^1 G(\theta, q^*(\theta))d\theta = h(\mathbf{q}^*),$$

where the definition of G is shown in Equation (12). Thus, by Claim 2, we have $\lim_{\ell\to\infty} h_{N_\ell}(\mathbf{p}^{(N_\ell)}) = \lim_{\ell\to\infty} h(\mathbf{p}^{(N_\ell)}) = h(\mathbf{q}^*) = $ OPT*. We have proved that $h(\mathbf{q}') \geq$ OPT*. On the other hand, since $\mathbf{q}' \in \mathcal{S}'$, $\mathbf{q}'$ is feasible solution to $\mathcal{P}_{cont}$. Thus $h(\mathbf{q}') \leq$ OPT*. Thus we have $h(\mathbf{q}') =$ OPT*, which implies that $\mathbf{q}'$ is indeed an optimal solution to $\mathcal{P}_{cont}$. Let $\Phi(\mathbf{q}') = (\{\theta_i'\}_{i\in[n-1]}, \{q_i'\}_{i\in[n]})$. Then $q'(\theta) = q_i', \forall i \in [n], \theta \in [\theta_{i-1}', \theta_i')$ and $q'(1) = q_n'$. Since $\mathbf{q}'$ is a feasible solution to $\mathcal{P}_{cont}$,

let $\mathcal{M}' = \{q(\theta), t(\theta)\}_{\theta \in [0,1]} \in C_2'$ be the responsive-IC and IR mechanism that implements $\mathbf{q}'$ (Lemma 5). To show that $\mathcal{M}'$ contains at most $n = 3m - 1$ different options, it suffices to prove that for every $i \in [n]$, $t(\theta)$ is a constant in $[\theta'_{i-1}, \theta_i)$. Suppose there exist $\theta, \theta' \in [\theta'_{i-1}, \theta_i)$, $\theta \neq \theta'$ such that $t(\theta) \neq t(\theta')$. Without loss of generality, assume $t(\theta) < t(\theta')$. Then when the buyer has type θ', misreporting θ induces the same experiment, but a lower price. It contradicts with the fact that $\mathcal{M}'$ is responsive-IC. Thus $\mathcal{M}'$ is an optimal responsive-IC, IR mechanism that contains at most $3m - 1$ different options. We finish the proof.
□

D   MISSING DETAILS FROM SECTION 4.2

In this section we will provide a proof of Theorem 4. We first provide a sufficient and necessary condition under which selling complete information achieves the highest revenue among all responsive-IC and IR mechanisms (Theorem 9). Note that the same condition also guarantees that selling complete information is the optimal mechanism among all IC and IR mechanisms.

Since f is strictly positive on $[0, 1]$, $F(\cdot)$ is strictly increasing. Denote $F^{-1}(\cdot)$ the inverse of F. Recall that for every $\theta \in [0, 1]$, $\varphi^-(\theta) = \theta f(\theta) + F(\theta)$ and $\varphi^+(\theta) = (\theta - 1)f(\theta) + F(\theta)$. Intuitively, they can be viewed as the agent's virtual values when $q(\theta) \leq u_{11} - u_{2m}$ and when $q(\theta) > u_{11} - u_{2m}$ respectively. If either φ^- or φ^+ is not monotonically non-decreasing, we iron the functions using the generalized ironing procedure by Toikka [39], which is a generalization of the ironing procedure by Myerson [34]. Denote $\tilde{\varphi}^-$ and $\tilde{\varphi}^+$ the ironed virtual value functions of φ^- and φ^+ respectively. See Appendix D.1 for formal definitions of the ironing procedure and the ironed interval.

THEOREM 9. *Let $\mathcal{M}^*$ be the menu that contains only the fully informative experiment with price $p > 0$. Then $\mathcal{M}^*$ achieves the maximum revenue among all responsive-IC and IR mechanisms if and only if, there exist two multipliers η^* and $\lambda^* \geq 0$ such that:*

(1) *$p \leq \widehat{p}$, where $\widehat{p} = \min\left\{\frac{(u_{11}-u_{12})u_{2m}}{u_{11}-u_{12}+u_{22}}, \frac{(u_{2m}-u_{2,m-1})u_{11}}{u_{2m}+u_{1,m-1}-u_{2,m-1}}\right\}$. Moreover, $\lambda^* > 0$ only when $m = 3$, $u_{12} - u_{22} = u_{11} - 1$, and $p = \widehat{p} = 1 - u_{22}$.*

(2) *Let $\theta_p^L = \frac{p}{u_{11}}$ and $\theta_p^H = 1 - \frac{p}{u_{2m}}$ be the two points where the utility function of buying the fully informative experiment, i.e., a linear function, intersects with the IR curve. For every $\theta \in [0, \theta_p^L)$, $\tilde{\varphi}^-(\theta) - \eta^* + \lambda^* \leq 0$; For every $\theta \in [\theta_p^L, \theta_p^H)$, $\tilde{\varphi}^-(\theta) - \eta^* + \lambda^* \geq 0$ and $\tilde{\varphi}^+(\theta) - \eta^* \leq 0$; For every $\theta \in [\theta_p^H, 1]$, $\tilde{\varphi}^+(\theta) - \eta^* \geq 0$.*

(3) *θ_p^L is not in the interior of any ironed interval of $\tilde{\varphi}^-(\cdot)$. And θ_p^H is not in the interior of any ironed interval of $\tilde{\varphi}^+(\cdot)$.*

Here is a proof sketch of Theorem 9. We first prove an exact characterization of the optimal responsive-IC and IR mechanism in C_2', i.e., the optimal solution $\mathbf{q}^* = \{q^*(\theta)\}_{\theta \in [0,1]}$ of the program in Figure 3, by Lagrangian duality (Theorem 10). It is a generalization of the characterization by Bergemann et al. [8] to $m \geq 3$ actions. We Lagrangify constraints (3) and (4) in the program. $\mathbf{q} = \{q(\theta)\}_{\theta \in [0,1]}$ is an optimal solution iff there exist Lagrangian multipliers that satisfy the KKT conditions. As a second step, we apply this characterization to $\mathbf{q}$ that corresponds to selling the complete information. To simplify our characterization, we show that at most two of the Lagrangian multipliers can be non-zero (in contrast to $\Theta(m)$ non-zero multipliers as in Theorem 10), when the solution $\mathbf{q}$ corresponds to a mechanism that only sells complete information. This is enabled by showing that in order to be the optimal menu, the price of the complete information can not be too high (Lemma 20).

We notice that if a fully informative menu achieves the maximum revenue among all responsive-IC and IR mechanisms, then it must also achieve the maximum revenue among all IC and IR mechanisms. In the rest of this section, we focus on responsive-IC and IR mechanisms. By Lemma 4 and

Observation 3, we only need to consider mechanisms where every experiment is semi-informative (Table 2). Recall that C_2' is the set of responsive-IC, IR mechanisms that have this format. Throughout this section, we use $\mathbf{q} = \{q(\theta)\}_{\theta \in [0,1]}$ to represent the experiments of the mechanism. Recall that the optimal mechanism in C_2' is captured in the program in Figure 3.

In Appendix D.1, we come up with an exact characterization of the optimal responsive-IC and IR mechanism $\mathcal{M}^* \in C_2'$, in any environment with 2 states and m actions. In Appendix D.2, we apply this result to achieve a sufficient condition under which the fully informative menu is optimal.

D.1 Characterization of the Optimal Responsive-IC and IR Mechanism

In this section, we are going to present our exact characterization of the optimal mechanism in C_2', i.e., the optimal solution $\mathbf{q}^* = \{q^*(\theta)\}_{\theta \in [0,1]}$ of the program in Figure 3. We Lagrangify Constraints (3) and (4) in the program. Denote η and $\lambda = \{\lambda_k\}_{2 \le k \le m-1}$ the Lagrangian multipliers for both sets of constraints accordingly.

DEFINITION 6. *Given Lagrangian multipliers η and $\lambda = \{\lambda_k\}_{k \in \{2,\dots,m-1\}} \ge \mathbf{0}$, define the virtual surplus function $J^{(\eta,\lambda)} : [-u_{2m}, u_{11}] \times [0,1] \to \mathbb{R}$ as follows:*

$$J^{(\eta,\lambda)}(q, \theta) = (\theta f(\theta) + F(\theta)) \cdot q + f(\theta) \cdot \min\{u_{11} - u_{2m} - q, 0\}$$

$$+ \eta \cdot (u_{11} - u_{2m} - q) + \sum_{k=2}^{m-1} \lambda_k \cdot \left((q - \ell_k) \cdot \mathbb{1}[q \le \ell_k] - u_{2,m+1-k} + u_{2m}\right)$$

Then the Lagrangian $L(\mathbf{q}, \eta, \lambda)$ of the program in Figure 3 can be written as: $L(\mathbf{q}, \eta, \lambda) = \int_0^1 J^{(\eta,\lambda)}(q(\theta), \theta) d\theta$.

OBSERVATION 5. *For any fixed $\theta \in [0,1]$, all of the following functions are continuous, differentiable in $[-u_{2m}, u_{11}]$ except at finitely many points, and weakly concave in q (note that $g_2(q)$ is linear)*

(1) $g_1(q, \theta) = (\theta f(\theta) + F(\theta)) \cdot q + f(\theta) \cdot \min\{u_{11} - u_{2m} - q, 0\}$.
(2) $g_2(q) = u_{11} - u_{2m} - q$.
(3) $g_3^{(k)}(q) = (q - \ell_k) \cdot \mathbb{1}[q \le \ell_k] - u_{2,m+1-k} + u_{2m}, \forall k \in \{2, 3, \dots, m - 1\}$.

Thus for any Lagrangian multipliers $(\eta, \lambda \ge \mathbf{0})$, the function $J^{(\eta,\lambda)}(q, \theta)$ is continuous, differentiable in $q \in [-u_{2m}, u_{11}]$ except at finitely many points, and weakly concave in q, for any fixed $\theta \in [0,1]$. Besides,

$$\frac{\partial g_1}{\partial q}(q, \theta) = \begin{cases} \theta f(\theta) + F(\theta), & \forall q \in (-u_{2m}, u_{11} - u_{2m}) \\ (\theta - 1) f(\theta) + F(\theta), & \forall q \in (u_{11} - u_{2m}, u_{11}) \end{cases}$$

are both continuous in θ on $[0,1]$. Thus except at $q = -u_{2m}, u_{11} - u_{2m}, u_{11}$, $\frac{\partial J^{(\eta,\lambda)}}{\partial q}(q, \theta)$ exists and is continuous in θ on $[0,1]$.

Denote $\mathcal{S} = \{\mathbf{q} : [0,1] \to [-u_{2m}, u_{11}] : q(\cdot) \text{ is non-decreasing}\}$. We notice that $\mathcal{S}$ is convex, thus by Observation 5, the program in Figure 3 is a general convex programming problem. Strong duality holds according to Theorem 8.3.1 and Theorem 8.4.1 of [33].

LEMMA 19. *[33] Let $(\eta^*, \lambda^*) \in \arg\min_{\lambda \ge 0, \eta} \max_{\mathbf{q} \in \mathcal{S}} L(\mathbf{q}, \eta, \lambda)$ be the optimal Lagrangian multiplier. Then a solution $\mathbf{q}^*$ is optimal in the program in Figure 3 if and only if $\mathbf{q}^* \in \arg\max_{\mathbf{q} \in \mathcal{S}} L(\mathbf{q}, \eta^*, \lambda^*)$.*

Now we are ready to state the characterization of the optimal solution.

DEFINITION 7 (IRONING [34]). *Given any differentiable function $\varphi : [0,1] \to \mathbb{R}$. Let F be any continuous distribution on $[0,1]$, with density $f(\cdot)$. Define the ironed function $\tilde{\varphi}(\cdot)$ as follows: For every $r \in [0,1]$, let $H(r) := \int_0^r \varphi(F^{-1}(x)) dx$ and $G(\cdot)$ be the convex hull of $H(\cdot)$. [16] Let $g(\cdot)$ be the*

---

[16] $G(\cdot)$ is the convex hull of $H(\cdot)$ if G is the highest convex function such that $G(\theta) \le H(\theta), \forall \theta \in [0,1]$.

derivative of $G(\cdot)$, [17] and $\tilde{\varphi}(\theta) := g(F(\theta))$, $\forall \theta \in [0, 1]$. For any open interval $I \subseteq (0, 1)$, we call I an **ironed interval** of $\tilde{\varphi}(\cdot)$ if $G(F(\theta)) < H(F(\theta))$, $\forall \theta \in I$.

DEFINITION 8. *Let $F(\cdot)$ be the agent's type distribution on $[0, 1]$, with density $f(\cdot)$. For every $\theta \in [0, 1]$, let $\varphi^-(\theta) = \theta f(\theta) + F(\theta)$ and $\varphi^+(\theta) = (\theta - 1)f(\theta) + F(\theta)$ be the agent's two virtual values at type θ.[18] Denote $\tilde{\varphi}^-(\cdot)$ and $\tilde{\varphi}^+(\cdot)$ the ironed virtual value functions of $\varphi^-(\cdot)$ and $\varphi^+(\cdot)$ respectively. We say an open interval $I \subseteq (0, 1)$ an* ironed interval *of $\tilde{\varphi}^-$ (or $\tilde{\varphi}^+$) if $\int_0^{F(\theta)} \tilde{\varphi}^- \left(F^{-1}(x) \right) dx < \int_0^{F(\theta)} \varphi^- \left(F^{-1}(x) \right) dx$ (or $\int_0^{F(\theta)} \tilde{\varphi}^+ \left(F^{-1}(x) \right) dx < \int_0^{F(\theta)} \varphi^+ \left(F^{-1}(x) \right) dx$) for all $\theta \in I$.*

THEOREM 10. *Suppose $\mathbf{q}^* = \{q^*(\theta)\}_{\theta \in [0,1]}$ is a feasible solution of the program in Figure 3. Then $\mathbf{q}^*$ is optimal if and only if, there exist multipliers η^* and $\lambda^* = \{\lambda_k^*\}_{k \in \{2,...,m-1\}} \geq \mathbf{0}$, such that all of the following properties are satisfied:*

(1) $q^*(\cdot)$ is non-decreasing,

$$q^*(\theta) \in \underset{q \in [-u_{2m}, u_{11}]}{\arg\max} \left\{ \tilde{\varphi}^-(\theta) \cdot q^- + \tilde{\varphi}^+(\theta) \cdot q^+ - \eta^* \cdot q + \sum_{k=2}^{m-1} \lambda_k^*(q - \ell_k) \cdot \mathbb{1}\left[q \leq \ell_k\right] \right\},$$

where $q^- = \min\{q, u_{11} - u_{2m}\}, q^+ = \max\{q - u_{11} + u_{2m}, 0\}$, almost everywhere.

(2) $\mathbf{q}^*$ and (η^*, λ^*) satisfy the complementary slackness, i.e., for every $k \in \{2, 3, ..., m - 1\}$,

$$\lambda_k^* \cdot \left[\int_0^1 (q^*(x) - \ell_k) \cdot \mathbb{1}\left[q^*(x) \leq \ell_k\right] dx - (u_{2,m+1-k} - u_{2m}) \right] = 0$$

(3) $\mathbf{q}^*$ satisfies the generalized pooling property *(Definition 10): For any $x \in [-u_{2m}, u_{11}]$, define $q^{*-1}(x) = \inf\{\theta \in [0, 1] : q^*(\theta) \geq x\}$. Then $\mathbf{q}^*$ satisfies the* generalized pooling property *if*
(a) for every $x \in [-u_{2m}, u_{11} - u_{2m}]$ and open interval $I \subseteq (0, 1)$:

$$q^{*-1}(x) \in I \text{ and } I \text{ is an ironed interval for } \tilde{\varphi}^-(\cdot) \implies q^*(\cdot) \text{ is constant on } I.$$

(b) for every $x \in (u_{11} - u_{2m}, u_{11}]$ and open interval $I \subseteq (0, 1)$:

$$q^{*-1}(x) \in I \text{ and } I \text{ is an ironed interval for } \tilde{\varphi}^+(\cdot) \implies q^*(\cdot) \text{ is constant on } I.$$

Theorem 10 follows from the generalized ironing procedure by Toikka [39]. We first introduce the necessary background.

General Virtual Surplus Function: Given any function $J(\cdot, \cdot) : [a, b] \times [0, 1] \to \mathbb{R}$. We assume that J satisfies: (i) For every fixed $q \in [a, b]$, $J(q, \cdot)$ is continuously differentiable on $(0, 1)$, and (ii) For every fixed $\theta \in [0, 1]$, $J(\cdot, \theta)$ is continuous on $[a, b]$ and weakly concave. (iii) Except at finitely many points $q \in [a, b]$, $\frac{\partial J}{\partial q}(q, \theta)$ exists and is continuous in θ on $[0, 1]$. We define the ironed virtual surplus in Definition 9.

DEFINITION 9 (GENERALIZED IRONING PROCEDURE [39]). *Given any virtual surplus function J, for every $r \in [0, 1]$, let $h(q, r) := \frac{\partial J}{\partial q}\left(q, F^{-1}(r)\right)$. By assumption (iii), it is well-defined on $[a, b]$ except at finitely many points. We extend h to the whole interval $[a, b]$ by left-continuity. For every $q \in [a, b], r \in [0, 1]$, let $H(q, r) := \int_0^r h(q, x)dx$.[19] For every $q \in [a, b]$, define $G(q, \cdot) := \mathrm{conv}\, H(q, \cdot)$*

---

[17]Since $H(\cdot)$ is differentiable, $G(\cdot)$, as the convex hull of $H(\cdot)$, is continuously differentiable on $(0, 1)$. We extend g to $[0, 1]$ by continuity.

[18]$\varphi^-(\theta)$ is the virtual value when $q(\theta) \leq u_{11} - u_{2m}$, and $\varphi^+(\theta)$ is the virtual value when $q(\theta) > u_{11} - u_{2m}$.

[19]By assumption (iii) of $J(\cdot, \cdot)$ and the continuity of $F^{-1}(\cdot)$, $h(q, x)$ continuous in x and hence is integrable in x.

as the convex hull of function $H(q, \cdot)$. Let $g(q, r) := \frac{\partial G}{\partial r}(q, r)$, which is non-decreasing in r for every $q \in [a, b]$. [20] Define the ironed virtual surplus function $\tilde{J} : [a, b] \times [0, 1] \to \mathbb{R}$ as follows:

$$\tilde{J}(q, \theta) = J(0, \theta) + \int_0^q g(s, F(\theta)) ds$$

DEFINITION 10. [39] Given $\boldsymbol{q} = \{q(\theta)\}_{\theta \in [0,1]}$. For any $x \in [a, b]$, define $q^{-1}(x) = \inf\{\theta \in [0, 1] : q(\theta) \geq x\}$. Then $\boldsymbol{q}$ satisfies the generalized pooling property if for every $x \in [a, b]$ and open interval $I \subset [0, 1]$:

$$q^{-1}(x) \in I \text{ and } G(x, \theta) < H(x, \theta), \forall \theta \in I \implies q(\cdot) \text{ is constant on } I.$$

THEOREM 11. [39] Let $\mathcal{S} = \{\boldsymbol{q} : [0, 1] \to [a, b] : q(\cdot) \text{ is non-decreasing}\}$ and $\mathcal{S}' = \{\boldsymbol{q} : [0, 1] \to [a, b]\}$. Then

$$\sup_{q \in \mathcal{S}} \left\{ \int_0^1 J(q(\theta), \theta) d\theta \right\} = \sup_{q \in \mathcal{S}'} \left\{ \int_0^1 \tilde{J}(q(\theta), \theta) d\theta \right\}.$$

Moreover, $\boldsymbol{q} \in \mathcal{S}$ achieves the supremum of $\int_0^1 J(q(\theta), \theta) d\theta$ if and only if: $\tilde{J}(q(\theta), \theta) = \sup_{q \in [a,b]} \tilde{J}(q, \theta)$ almost everywhere, and $\boldsymbol{q}$ satisfies the generalized pooling property.

Proof of Theorem 10: By Observation 5, for any collection of Lagrangian multipliers $\eta, \lambda \geq \boldsymbol{0}$, the virtual surplus function $J^{(\eta, \lambda)}(\cdot, \cdot)$ satisfies our assumptions. We apply the ironing procedure in Definition 9 to $J^{(\eta, \lambda)}(\cdot, \cdot)$:

$$h(q, r) = \frac{\partial J^{(\eta, \lambda)}}{\partial q}\left(q, F^{-1}(r)\right) = \begin{cases} \varphi^-(F^{-1}(r)) - \eta + \sum_{j=k(q)+1}^{m-1} \lambda_j, & q \leq u_{11} - u_{2m} \\ \varphi^+(F^{-1}(r)) - \eta + \sum_{j=k(q)+1}^{m-1} \lambda_j, & q > u_{11} - u_{2m} \end{cases}$$

Recall that $\ell_k = u_{1,m+1-k} - u_{2,m+1-k}, \forall k \in [m]$. Here $k(q)$ is the unique number $j \in \{1, ..., m - 1\}$ such that $\ell_j < q \leq \ell_{j+1}$ (let $k(q) = 1$ when $q = \ell_1$). For every $q \in [-u_{2m}, u_{11}]$ and $r \in [0, 1]$,

$$H(q, r) = \int_0^r h(q, x) dx = \begin{cases} \int_0^r \varphi^-(F^{-1}(x)) dx + r \cdot \left(\sum_{j=k(q)+1}^{m-1} \lambda_j - \eta\right), & q \leq u_{11} - u_{2m} \\ \int_0^r \varphi^+(F^{-1}(x)) dx + r \cdot \left(\sum_{j=k(q)+1}^{m-1} \lambda_j - \eta\right), & q > u_{11} - u_{2m} \end{cases}$$

For every fixed $q \in [-u_{2m}, u_{11}]$, to obtain $G(q, \cdot)$, we take the convex hull of $H(q, \cdot)$. Note that $r \cdot \left(\sum_{j=k(q)+1}^{m-1} \lambda_j - \eta\right)$ is a linear function on r, thus we are effectively only taking the convex hull of the functions $\widehat{H}^-(r) := \int_0^r \varphi^-(F^{-1}(x)) dx$ and $\widehat{H}^+(r) := \int_0^r \varphi^+(F^{-1}(x)) dx$. Denote $\widehat{G}^-(\cdot)$ and $\widehat{G}^+(\cdot)$ the convex hull of $\widehat{H}^-(\cdot)$ and $\widehat{H}^+(\cdot)$ accordingly. Then

$$G(q, r) = \begin{cases} \widehat{G}^-(r) + r \cdot \left(\sum_{j=k(q)+1}^{m-1} \lambda_j - \eta\right), & q \leq u_{11} - u_{2m} \\ \widehat{G}^+(r) + r \cdot \left(\sum_{j=k(q)+1}^{m-1} \lambda_j - \eta\right), & q > u_{11} - u_{2m} \end{cases}$$

By definition of $\tilde{\varphi}^-, \tilde{\varphi}^+$ (Definition 8), we have

$$g(q, r) = \begin{cases} \tilde{\varphi}^-(F^{-1}(r)) + \sum_{j=k(q)+1}^{m-1} \lambda_j - \eta, & q \leq u_{11} - u_{2m} \\ \tilde{\varphi}^+(F^{-1}(r)) + \sum_{j=k(q)+1}^{m-1} \lambda_j - \eta, & q > u_{11} - u_{2m} \end{cases}$$

---

[20]Since $H(q, \cdot)$ is differentiable on $(0, 1)$ for every $q \in [a, b]$, $G(q, \cdot)$, as the convex hull of $H(q, \cdot)$, is continuously differentiable on $(0, 1)$ for every $q \in [a, b]$. We extend $g(q, \cdot)$ to $[0, 1]$ by continuity.

The ironed virtual surplus $\tilde{J}^{(\eta,\lambda)}(q,\theta) = J^{(\eta,\lambda)}(0,\theta) + \int_0^q g(s, F(\theta))ds$. We notice that $J^{(\eta,\lambda)}(0,\theta) = -\sum_{k=2}^{m-1} \lambda_k \ell_k$, and by definition of $k(q)$,

$$\int_0^q \sum_{j=k(s)+1}^{m-1} \lambda_j ds = \sum_{j=2}^{m-1} \lambda_j \cdot \int_0^q \mathbb{1}[s \le \ell_j]ds = \sum_{j=2}^{m-1} \lambda_j \cdot \min\{\ell_j, q\}$$

Thus we have

$$\tilde{J}^{(\eta,\lambda)}(q,\theta) = J^{(\eta,\lambda)}(0,\theta) + \int_0^q g(s, F(\theta))ds$$

$$= \begin{cases} \tilde{\varphi}^-(\theta) \cdot q - \eta \cdot q + \sum_{k=2}^{m-1} \lambda_k(q - \ell_k) \cdot \mathbb{1}[q \le \ell_k], & q \le u_{11} - u_{2m} \\ \tilde{\varphi}^-(\theta) \cdot (u_{11} - u_{2m}) + \tilde{\varphi}^+(\theta)(q - u_{11} + u_{2m}) - \eta \cdot q + \sum_{k=2}^{m-1} \lambda_k(q - \ell_k) \cdot \mathbb{1}[q \le \ell_k], & q > u_{11} - u_{2m} \end{cases}$$

$$= \tilde{\varphi}^-(\theta) \cdot q^- + \tilde{\varphi}^+(\theta) \cdot q^+ - \eta \cdot q + \sum_{k=2}^{m-1} \lambda_k(q - \ell_k) \cdot \mathbb{1}[q \le \ell_k],$$

where $q^- = \min\{q, u_{11} - u_{2m}\}, q^+ = \max\{q - u_{11} + u_{2m}, 0\}$.

We first prove that the properties are necessary for $\mathbf{q}^*$ to be optimal. Suppose $\mathbf{q}^*$ is an optimal solution. Let (η^*, λ^*) be the optimal Lagrangian multipliers. By Lemma 19, $\mathbf{q}^*$ maximizes $L(\mathbf{q}, \eta^*, \lambda^*) = \int_0^1 J^{(\lambda^*, \eta^*)}(q(\theta), \theta)d\theta$ over $\mathbf{q} \in \mathcal{S}$. By Theorem 11, $\tilde{J}^{(\eta^*, \lambda^*)}(q^*(\theta), \theta) = \sup_{q \in [-u_{2m}, u_{11}]} \tilde{J}^{(\eta^*, \lambda^*)}(q, \theta)$ almost everywhere, which is exactly the first property in the statement of Theorem 10, and $\mathbf{q}^*$ satisfies the generalized pooling property. Since $\mathbf{q}^*$ is the optimal solution and (η^*, λ^*) are the optimal Lagrangian multipliers, the second property in the statement of Theorem 10 directly follows from the KKT condition. Finally, we prove the last property. An important property of $H(\cdot, \cdot)$ and $G(\cdot, \cdot)$ is that for any $q \le u_{11} - u_{2m}, H(q, r) > G(q, r)$ iff $\widehat{H}^-(r) > \widehat{G}^-(r)$, and for any $q > u_{11} - u_{2m}$, iff $\widehat{H}^+(r) > \widehat{G}^+(r)$. Therefore, $\mathbf{q}^*$ satisfies the generalized pooling property is equivalent to the last property in the statement.

Now we prove sufficiency. For any feasible solution $\mathbf{q}^*$, suppose there exist multipliers $(\eta^*, \lambda^* \ge \mathbf{0})$ that satisfy all three properties in the statement. We argue that $\mathbf{q}^*$ is the optimal solution (and (η^*, λ^*) are the optimal Lagrangian multipliers) by verifying the KKT conditions. Firstly, both $\mathbf{q}^*$ and (η^*, λ^*) are feasible primal and dual solutions. Secondly, by the first property of the statement, $\tilde{J}^{(\eta^*, \lambda^*)}(q^*(\theta), \theta) = \sup_{q \in [-u_{2m}, u_{11}]} \tilde{J}^{(\eta^*, \lambda^*)}(q, \theta)$ almost everywhere. As argued above, the fact that $\mathbf{q}^*$ satisfies the third property implies that $\mathbf{q}^*$ satisfies the generalized pooling property. Thus by Theorem 11, $\mathbf{q}^*$ maximizes $L(\mathbf{q}, \eta^*, \lambda^*) = \int_0^1 J^{(\lambda^*, \eta^*)}(q(\theta), \theta)d\theta$ over $\mathbf{q} \in \mathcal{S}$. The stationary condition is satisfied. Thirdly, the second property of the statement implies that the complementary slackness is satisfied. Thus by the KKT conditions, $\mathbf{q}^*$ is the optimal solution. □

D.2 Optimality Conditions for Selling Complete Information

In this section, we characterize the conditions under which selling only the fully informative experiment is optimal using Theorem 10. Clearly, any responsive-IC and IR mechanism that only contains the fully informative experiment E^* is determined by the price $p > 0$ of E^*. And for every type $\theta \in [0, 1]$, the agent selects E^* if and only if $V_\theta(E^*) \ge u(\theta)$. Since $V_\theta(E^*) = \theta u_{11} + (1-\theta)u_{2m} - p$ is linear in θ and the IR curve is convex in θ, the agent selects E^* when her type θ is in some closed interval $[\theta_p^L, \theta_p^H]$ (see Figure 9a or Figure 9b for an illustration of θ_p^L, θ_p^H).[21] Clearly, the mechanism is also IC, since the agent will always follow the recommendation when receiving either no information or full information.

---

[21]We assume that θ_p^L and θ_p^H exist, otherwise the price is too high, and the mechanism has revenue 0.

We first prove the following corollary of Lemma 5. It states that any k-piecewise-linear "utility curve", which always stays above the IR curve, can be implemented by a responsive-IC and IR mechanism with k experiments.

COROLLARY 3. *Given any finite integer $k \geq 2$ and any continuous, piecewise linear function $h : [0, 1] \to \mathbb{R}_+$ with k pieces. Suppose $h(\cdot)$ satisfies all of the following: (i) $h(0) = u_{2m}, h(1) = u_{11}$; (ii) $-u_{2m} \leq \ell_1 < ... < \ell_k \leq u_{11}$, where for each $i \in [k]$, ℓ_i is the slope of the i-th piece of h; (iii) $h(\theta) \geq u(\theta), \forall \theta \in [0, 1]$. Then there exists a responsive-IC and IR mechanism $\mathcal{M} \in C'_2$ with k experiments, such that for every $\theta \in [0, 1]$, the agent's utility at type θ, when she follows the recommendation, is $h(\theta)$.*

PROOF. We define the mechanism $\mathcal{M} = (\mathbf{q}, \mathbf{t})$ as follows. Let $0 < \theta_1 < \theta_2 < ... < \theta_{k-1} < 1$ be the $k - 1$ kinks of h. Denote $\theta_0 = 0, \theta_k = 1$. For every $\theta \in [0, 1)$, define $q(\theta) = \ell_i$, where $i \in [k]$ is the unique number such that $\theta \in [\theta_{i-1}, \theta_i)$. Also define $q(1) = \ell_k$.

We will verify that $\mathbf{q}$ satisfies all the requirements of Lemma 5. By property (ii), we have that $q(\theta) \in [-u_{2m}, u_{11}]$ is non-decreasing in θ. Moreover, we notice that $q(\theta)$ is the right derivative of $h(\theta)$ at every $\theta \in [0, 1)$. Since $h(0) = u_{2m}$ and h is continuous at $\theta = 1$, we have $h(\theta) = u_{2m} + \int_0^\theta q(x)dx, \forall \theta \in [0, 1]$. Taking $\theta = 1$, we have $\int_0^1 q(x)dx = h(1) - u_{2m} = u_{11} - u_{2m}$. At last, by property (iii) we have $h(\theta) = u_{2m} + \int_0^\theta q(x)dx \geq u(\theta), \forall \theta \in [0, 1]$.

Thus by Lemma 5, there exists a payment rule $\mathbf{t}$ such that $\mathcal{M}$ is a responsive-IC and IR mechanism, and $\mathbf{t}$ satisfies Equation (4). By Observation 4, with such payment, the agent's utility at every type θ, when he follows the recommendation, is $u_{2m} + \int_0^\theta q(x)dx = h(\theta)$. Since $q(\theta)$ takes k different values, the mechanism contains k experiments. □

To obtain a simplified characterization, an important step is the following observation: If some fully informative menu with price p is the optimal menu, then both θ_p^L and θ_p^H have to stay in the first and last piece of the IR curve accordingly (see Figure 9a). To see the reason, consider a curve $h(\cdot)$ (a piecewise linear function) that is the maximum of the utility function of buying the fully informative experiment at price p and the IR curve. In other words, $h(\cdot)$ coincides with the utility function of buying the fully informative experiment on interval $[\theta_p^L, \theta_p^H]$, and coincides with the IR curve everywhere else. If θ_p^L and θ_p^H do not lie in the first and last piece of the IR curve respectively, there are pieces of the curve $h(\cdot)$ that belong to the i-th piece of the IR curve for some $2 \leq i \leq m - 1$ (the blue line in Figure 9b). We argue that we can change the mechanism to offer another experiment (based on this piece of the IR curve), so that (i) the mechanism is still responsive-IC and IR, and (ii) the revenue of the mechanism strictly increases. This contradicts with the optimality of selling only the fully informative experiment. See Figure 9b for an illustration and Lemma 20 for a formal statement.

LEMMA 20. *Consider the fully informative experiment E^* with any price $p > 0$. Assume there is some θ such that $V_\theta(E^*) > u(\theta)$. Let θ_p^L, θ_p^H ($0 < \theta_p^L < \theta_p^H < 1$) be the two θ's such that $V_\theta(E^*) = u(\theta)$. Let θ_1 and θ_2 be the first and last kink of the IR curve.[22] Suppose selling the full information at price p achieves the optimal revenue among all responsive-IC and IR mechanisms. Then $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$ (see Figure 9a).*

PROOF. Let $\mathcal{M}$ be the menu that only contains the fully informative experiment E^* with price p. We prove by contradiction. Without loss of generality, assume that $\theta_p^L > \theta_1$. Let $L : [0, 1] \to \mathbb{R}$ be the agent's utility function for the experiment E^*, i.e., $L(\theta) = V_\theta(E^*) - p = \theta \cdot u_{11} + (1 - \theta)u_{2m} - p$. Consider the function $h : [0, 1] \to \mathbb{R}$ such that $h(\theta) = \max\{u(\theta), L(\theta)\}, \forall \theta \in [0, 1]$. One can

---

[22]Formally, θ_1 satisfies $u_{2m}(1 - \theta_1) = u_{1,m-1}\theta_1 + (1 - \theta_1)u_{2,m-1}$, and θ_2 satisfies $u_{11}\theta_2 = u_{12}\theta_2 + (1 - \theta_2)u_{22}$.

easily verify that h is a piecewise linear function that satisfies the requirement of Corollary 3. By Corollary 3, there exists a responsive-IC and IR mechanism $\mathcal{M}' = (\mathbf{q}', \mathbf{t}') \in C_2'$, such that for every $\theta \in [0, 1]$, the agent's utility at type θ, when she follows the recommendation, is $h(\theta)$. When the agent's type $\theta \in [\theta_p^L, \theta_p^H]$, she purchases experiment E^* at price p in $\mathcal{M}'$, and thus the mechanism collects total revenue $\text{Rev}(\mathcal{M})$ at interval $[\theta_p^L, \theta_p^H]$. It suffices to show that $\mathcal{M}'$ collects strictly positive revenue from types in $[0, \theta_p^L) \cup (\theta_p^H, 1]$, which contradicts with the optimality of $\mathcal{M}$.

Since $\theta_1 < \theta_p^L$, there is at least one (linear) piece of the piecewise linear function $h(\cdot)$ is in $[\theta_1, \theta_p^L]$. Consider any such piece. The whole piece is also on the IR curve. Let $(\theta', u(\theta'))$ and $(\theta'', u(\theta''))$ be the two endpoints of this piece ($\theta' < \theta''$) and ℓ be the slope. Then $-u_{2m} < \ell < u_{11} - u_{2m}$. We design an experiment E offered by $\mathcal{M}'$, so that all types on interval (θ', θ'') purchase this experiment. Denote q the variable corresponds to the experiment E. Then by Observation 4, $q = \ell \leq u_{11} - u_{2m}$. Thus when $\theta \in (\theta', \theta'')$, $V_\theta(E) = q \cdot \theta + u_{2m}$. Moreover, notice that (i) $u(\theta') \geq u_{2m} - u_{2m}\theta'$, the RHS is agent's payoff for always choosing action m without receiving any experiment, and (ii) $h(\theta) = \ell \cdot (\theta - \theta') + u(\theta'), \forall \theta \in [\theta', \theta'']$. Hence

$$t'(\theta) = V_\theta(E) - h(\theta) = u_{2m} + \ell \cdot \theta' - u(\theta') = (u_{2m} + \ell)\theta' > 0, \forall \theta \in (\theta', \theta'')$$

Therefore, $\text{Rev}(\mathcal{M}') \geq \text{Rev}(\mathcal{M}) + (F(\theta'') - F(\theta')) \cdot (u_{2m} + \ell)\theta' > \text{Rev}(\mathcal{M})$, as $F(\cdot)$ is strictly increasing, contradicting with the optimality of $\mathcal{M}$. Hence, we must have $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$. □

(a) The fully informative experiment where $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$

(b) The scenario when the fully informative experiment is not optimal

(c) In the proof of Theorem 9: $\theta_p^H > \theta_2$ when $\ell_{m-1} = u_{11} - u_{2m}$
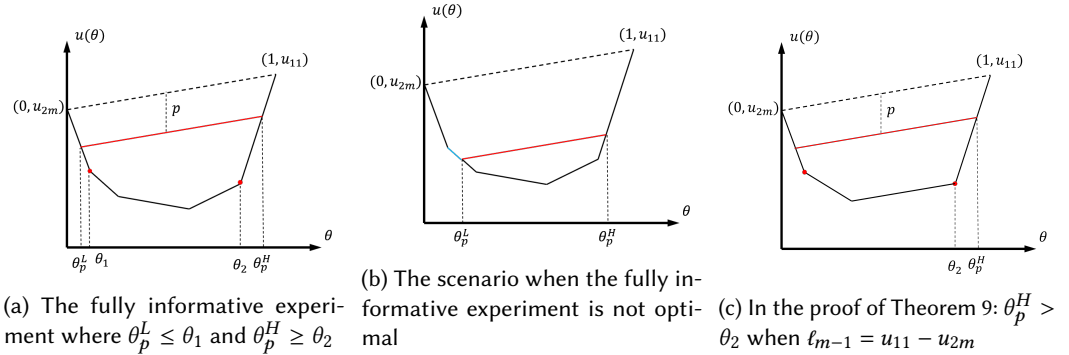
Fig. 9. Illustrations of notations and the utility curve of full information

To prove Theorem 9, we also need the following lemma for the ironed virtual value, which may be of independent interest.

LEMMA 21. *(Adapted from Lemma 4.11 in [39]) Let φ_1, φ_2 be any pair of differentiable real-valued functions on $[0, 1]$. Let F be any continuous distribution on $[0, 1]$. Denote $\tilde{\varphi}_1, \tilde{\varphi}_2$ the ironed functions of φ_1, φ_2 respectively, with respect to the distribution F (Definition 7). Suppose $\varphi_1(\theta) \geq \varphi_2(\theta), \forall \theta \in [0, 1]$. Then $\tilde{\varphi}_1(\theta) \geq \tilde{\varphi}_2(\theta), \forall \theta \in [0, 1]$.*

PROOF. Let f be the pdf of F. As shown in Definition 7, define $H_1(r) = \int_0^r \varphi_1(F^{-1}(x))dx, \forall r \in [0, 1]$, $G_1(\cdot)$ as the convex hull of $H_1(\cdot)$, and $g_1(\cdot)$ as the (extended) derivative of $G_1(\cdot)$ (see Footnote 17). Define H_2, G_2, g_2 similarly for φ_2.

For every $\theta \in [0, 1]$, $\tilde{\varphi}_1(\theta) = g_1(F(\theta)), \tilde{\varphi}_2(\theta) = g_2(F(\theta))$. Thus to show $\tilde{\varphi}_1(\theta) \geq \tilde{\varphi}_2(\theta)$, it suffices to prove that $g_1(r) \geq g_2(r), \forall r \in [0, 1]$. We prove the claim by contradiction. Suppose there exists $r_0 \in [0, 1]$ such that $g_1(r_0) < g_2(r_0)$. We notice that

$$\mathcal{H}(r) \triangleq H_1(r) - H_2(r) = \int_0^r \left[\varphi_1(F^{-1}(x)) - \varphi_2(F^{-1}(x))\right] dx,$$

which is non-decreasing in r, since $\varphi_1(F^{-1}(x)) \geq \varphi_2(F^{-1}(x)), \forall x \in [0, 1]$. Let $\rho = G_1(r_0) - G_2(r_0)$. To reach a contradiction, we show that there exists $r_1 \leq r_0$ such that $\mathcal{H}(r_1) \geq \rho$ and there exists $r_2 > r_0$ such that $\mathcal{H}(r_2) < \rho$, which contradicts with the fact that $\mathcal{H}$ is non-decreasing.

Let $L_1 : [0, 1] \to \mathbb{R}$ (or L_2) be the unique linear function tangent to G_1 (or G_2) at r_0. Then since G_1 (or G_2) is the convex hull of H_1 (or H_2), we have for every $r \in [0, 1]$, $L_1(r) \leq G_1(r) \leq H_1(r)$ and $L_2(r) \leq G_2(r) \leq H_2(r)$.

The existence of r_1: If $G_2(r_0) = H_2(r_0)$, then $H_1(r_0) - H_2(r_0) \geq G_1(r_0) - G_2(r_0) = \rho$. By choosing $r_1 = r_0$ we have $\mathcal{H}(r_1) \geq \rho$. Now assume $G_2(r_0) < H_2(r_0)$. Then r_0 is in the interior of some ironed interval I of φ_2. Let $r_1 < r_0$ be the left endpoint of the interval I. Then $H_2(r_1) = L_2(r_1)$. We have

$$\mathcal{H}(r_1) = H_1(r_1) - H_2(r_1) \geq L_1(r_1) - L_2(r_1) > L_1(r_0) - L_2(r_0) = G_1(r_0) - G_2(r_0) = \rho$$

Here the first inequality follows from $H_1(r_1) \geq L_1(r_1)$ and $H_2(r_1) = L_2(r_1)$. The second inequality follows from the fact that $d(L_1(r) - L_2(r))/dr = G_1'(r_0) - G_2'(r_0) = g_1(r_0) - g_2(r_0) < 0$. The second-last equality follows from the fact that L_1 (or L_2) meets G_1 (or G_2) at r_0.

The existence of r_2: The proof follows from a similar argument as r_1. If $G_1(r_0) = H_1(r_0)$, then $G_1'(r_0) = H_1'(r_0) = g_1(r_0)$. We notice that

$$\left. \frac{d(H_1(r) - L_2(r))}{dr} \right|_{r=r_0} = g_1(r_0) - g_2(r_0) < 0.$$

Thus there exists some $r_2 > r_0$ such that $H_1'(r) - g_2(r_0) < 0$ for all $r \in [r_0, r_2]$ due to the continuity of $H_1'(\cdot)$, which implies that

$$\rho = G_1(r_0) - G_2(r_0) = H_1(r_0) - L_2(r_0) > H_1(r_2) - L_2(r_2) \geq H_1(r_2) - H_2(r_2) = \mathcal{H}(r_2)$$

Now assume $G_1(r_0) < H_1(r_0)$. Then r_0 is in the interior of some ironed interval I of φ_1. Let $r_2 > r_0$ be the right endpoint of the interval I. Then $H_1(r_2) = L_1(r_2)$. We have

$$\mathcal{H}(r_2) = H_1(r_2) - H_2(r_2) \leq L_1(r_2) - L_2(r_2) < L_1(r_0) - L_2(r_0) = G_1(r_0) - G_2(r_0) = \rho$$

Here the first inequality follows from $H_1(r_2) = L_1(r_2)$ and $H_2(r_2) \geq L_2(r_2)$. The second inequality follows from the fact that $d(L_1(r) - L_2(r))/dr < 0$ and $r_2 > r_0$. The second-last equality follows from the fact that L_1 (or L_2) meets G_1 (or G_2) at r_0.

The existence of r_1 and r_2 contradicts with the fact that $\mathcal{H}$ is non-decreasing. Thus $\tilde{\varphi}_1(\theta) \geq \tilde{\varphi}_2(\theta), \forall \theta \in [0, 1]$. □

Proof of Theorem 9: Let L^* be the affine utility function for purchasing the fully informative experiment at price p, i.e. $L^*(\theta) = u_{11}\theta + u_{2m}(1 - \theta) - p = (u_{11} - u_{2m})\theta + u_{2m} - p$. Denote θ_p^L and θ_p^H the two points where L^* intersects with the IR curve. For every $k \in [m]$, let L_k be the affine function for the k-th piece of the IR curve, i.e., $L_k(\theta) = \theta \cdot u_{1,m+1-k} + (1 - \theta) \cdot u_{2,m+1-k} = \theta \cdot \ell_k + u_{2,m+1-k}, \forall \theta \in [0, 1]$.

Let $\mathbf{q}^*$ be the experiments purchased by the agent in $\mathcal{M}^*$, i.e.,

$$q^*(\theta) = \begin{cases} -u_{2m}, & \theta \in [0, \theta_p^L) \\ u_{11} - u_{2m}, & \theta \in [\theta_p^L, \theta_p^H) \\ u_{11}, & \theta \in [\theta_p^H, 1] \end{cases} \tag{14}$$

Necessary condition. We first prove that the properties in the statement are necessary. Suppose $\mathcal{M}^*$ is the optimal responsive-IC and IR mechanism, then $\mathbf{q}^*$ is the optimal solution to the program in Figure 3. We are going to verify each property of the statement by applying Theorem 10 to the solution $\mathbf{q}^*$.

Property 1. Recall that θ_1, θ_2 are the first and last kink of the IR curve. By Footnote 22,

$$\theta_1 = \frac{u_{2m} - u_{2,m-1}}{u_{1,m-1} + u_{2m} - u_{2,m-1}}, \quad \theta_2 = \frac{u_{22}}{u_{11} + u_{22} - u_{12}}$$

By Lemma 20, $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$. Thus θ_p^L, θ_p^H are the points where L^* intersects with the first and last piece of the IR curve, respectively. We have $L^*(\theta_p^L) = u_{1m}\theta_p^L + u_{2m}(1 - \theta_p^L) \Leftrightarrow \theta_p^L = \frac{p}{u_{11}}$. Similarly, $L^*(\theta_p^H) = u_{11}\theta_p^H + u_{21}(1 - \theta_p^H) \Leftrightarrow \theta_p^H = 1 - \frac{p}{u_{2m}}$. Thus $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$ are equivalent to

$$p \leq \widehat{p} = \min \left\{ \frac{(u_{11} - u_{12})u_{2m}}{u_{11} - u_{12} + u_{22}}, \frac{(u_{2m} - u_{2,m-1})u_{11}}{u_{2m} + u_{1,m-1} - u_{2,m-1}} \right\}.$$

To prove the second half of property 1, by Theorem 10, there exist multipliers $\lambda = \{\lambda_k\}_{k \in \{2,\ldots,m-1\}} \geq 0$ and η that satisfy all properties of Theorem 10. We choose $\eta^* = \eta$ and $\lambda^* = \lambda_2$. For every $k \in \{2, 3, \ldots, m - 1\}$, let $\theta_k^* = \sup\{\theta \in [0,1] : q^*(\theta) \leq \ell_k\}$. Then since $-u_{2m} < \ell_k = u_{1,m+1-k} - u_{2,m+1-k} < u_{11}$, we have $\theta_k^* = \theta_p^L$ iff $\ell_k < u_{11} - u_{2m}$, and $\theta_k^* = \theta_p^H$ otherwise. Notice that

$$\Delta_k := \int_0^1 (q^*(x) - \ell_k) \cdot \mathbb{1}[q^*(x) \leq \ell_k]dx - (u_{2,m+1-k} - u_{2m})$$

$$= \int_0^{\theta_k^*} (q^*(x) - \ell_k)dx - u_{2,m+1-k} + u_{2m} = L^*(\theta_k^*) - L_k(\theta_k^*)$$

(15)

Here the last inequality is because: $u_{2m} + \int_0^{\theta_p^L} q^*(x)dx = u_{2m}(1 - \theta_p^L) = u_{2m}(1 - \frac{p}{u_{11}}) = L^*(\theta_p^L)$ and $u_{2m} + \int_0^{\theta_p^H} q^*(x)dx = L^*(\theta_p^L) + (\theta_p^H - \theta_p^L) \cdot (u_{11} - u_{2m}) = L^*(\theta_p^H)$.

First consider the case when $m = 3$. We argue that if $\ell_2 \neq u_{11} - u_{2m}$, then $\lambda_2 = 0$. We only prove the case when $\ell_2 < u_{11} - u_{2m}$. The other case is similar. Suppose $\lambda_2 > 0$, the ironed virtual surplus function $\tilde{J}^{(\eta,\lambda)}$ can be written as follows (denote $\alpha = u_{11} - u_{2m}$):

$$\tilde{J}^{(\eta,\lambda)}(q, \theta) = \tilde{\varphi}^-(\theta) \cdot \min\{q, \alpha\} + \tilde{\varphi}^+(\theta) \cdot \max\{q - \alpha, 0\} - \eta q + \lambda_2 \cdot \min\{q - \ell_2, 0\}$$

For every fixed θ, the function $\tilde{J}^{(\eta,\lambda)}(\cdot, \theta)$ is continuous and 3-piecewise-linear. Thus at least one of the four values $-u_{2m}$, ℓ_2, $u_{11} - u_{2m}$, and u_{11} must be in $\arg\max_{q \in [-u_{2m}, u_{11}]} \tilde{J}^{(\eta,\lambda)}(q, \theta)$. We also have

$$\frac{\partial \tilde{J}^{(\eta,\lambda)}}{\partial q}(q, \theta) = \begin{cases} \tilde{\varphi}^-(\theta) - \eta + \lambda_2, & q < \ell_2 \\ \tilde{\varphi}^-(\theta) - \eta, & q \in (\ell_2, \alpha) \\ \tilde{\varphi}^+(\theta) - \eta, & q > \alpha \end{cases}$$

Define $\phi_1(\theta) = \tilde{\varphi}^-(\theta) - \eta + \lambda_2$, $\phi_2(\theta) = \tilde{\varphi}^-(\theta) - \eta$ and $\phi_3(\theta) = \tilde{\varphi}^+(\theta) - \eta$. By Definition 8, $\varphi^-(\theta) - \varphi^+(\theta) = f(\theta) > 0, \forall \theta \in [0, 1]$. Thus by Lemma 21, $\tilde{\varphi}^-(\theta) \geq \tilde{\varphi}^+(\theta), \forall \theta \in [0, 1]$. Thus $\phi_1(\theta) > \phi_2(\theta) \geq \phi_3(\theta), \forall \theta \in [0, 1]$. We notice that $-u_{2m} \in \arg\max_q \tilde{J}^{(\eta,\lambda)}(q, \theta)$ iff $\phi_1(\theta) \leq 0$. Since $q^*(\theta) = -u_{2m}, \forall \theta \in [0, \theta_p^L]$ and $\mathbf{q}^*$ is a pointwise maximizer of $\tilde{J}^{(\eta,\lambda)}(q, \theta)$ almost everywhere (the first property of Theorem 10), we have $\phi_1(0) \leq 0$. Since $\lambda_2 > 0$, $\phi_2(0) < 0$. Similarly, since $u_{11} \in \arg\max_q \tilde{J}^{(\eta,\lambda)}(q, \theta)$ iff $\phi_3(\theta) \geq 0$, we have $\phi_3(1) \geq 0$. Thus $\phi_2(1) \geq 0$. By Footnote 17, ϕ_2 is a continuous function on $[0, 1]$. Let $\varepsilon = \lambda_2/2 > 0$, then there exists an opened interval I such that $\phi_2(\theta) \in (-\varepsilon, 0), \forall \theta \in I$. At interval I, $\phi_1(\theta) = \phi_2(\theta) + \lambda_2 > 0$, and $\phi_3(\theta) \leq \phi_2(\theta) < 0$. This implies that for every $\theta \in I$, $\arg\max_q \tilde{J}^{(\eta,\lambda)}(q, \theta)$ contains a single value $q = \ell_2$, which contradicts with the fact that $\mathbf{q}^*$ is a pointwise maximizer of $\tilde{J}^{(\eta,\lambda)}(q, \theta)$ almost everywhere.

Thus, if $\lambda^* > 0$, then $\ell_2 = u_{11} - u_{2m} \implies u_{12} - u_{22} = u_{11} - 1$. Moreover, by the second property of Theorem 10, we have

$$0 = \Delta_2 = L^*(\theta_p^H) - L_2(\theta_p^H) = u_{2m} - p - u_{2,m-1} \implies p = 1 - u_{22}$$

Now consider the case when $m \geq 4$. We prove that $\lambda = \mathbf{0}$ and thus $\lambda^* = \lambda_2 = 0$. For every $k \in \{3, \ldots, m-2\}$, since $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$, we have $L^*(\theta_p^L) \geq L_2(\theta_p^L) > L_k(\theta_p^L)$ and $L^*(\theta_p^H) \geq L_{m-1}(\theta_p^H) > L_k(\theta_p^H)$. For every $k \in \{3, \ldots, m-2\}$, $\Delta_k > 0$. By the second property of Theorem 10, we have $\lambda_k = 0$. For $k = 2$, if $\ell_2 \geq u_{11} - u_{2m}$, then $\theta_2^* = \theta_p^H$. We thus have $L^*(\theta_p^H) \geq L_{m-1}(\theta_p^H) > L_2(\theta_p^H)$, since $m \geq 4$. By the second property of Theorem 10, $\lambda_2 = 0$. Similarly, if $\ell_{m-1} < u_{11} - u_{2m}$, $\lambda_{m-1} = 0$. If $\ell_{m-1} = u_{11} - u_{2m}$, we have $\theta_p^H > \theta_2$ (see Figure 9c for a proof by graph). We have $L^*(\theta_p^H) = L_m(\theta_p^H) > L_{m-1}(\theta_p^H)$, which implies that $\lambda_{m-1} = 0$. Since $\lambda_k = 0, \forall k \in \{3, \ldots, m-2\}$, then $\tilde{J}^{(\eta,\lambda)}(q, \theta)$ can be written as

$$\tilde{J}^{(\eta,\lambda)}(q,\theta) = \tilde{\varphi}^-(\theta) \cdot \min\{q, \alpha\} + \tilde{\varphi}^+(\theta) \cdot \max\{q-\alpha, 0\} - \eta q + \lambda_2 \cdot \min\{q-\ell_2, 0\} + \lambda_{m-1} \cdot \min\{q-\ell_{m-1}, 0\}$$

Now suppose $\lambda_2 > 0$ and $\lambda_{k-1} > 0$. The case when $\lambda_2 > 0 = \lambda_{k-1}$ and $\lambda_{k-1} > 0 = \lambda_2$ follows from a similar argument where either the term $\lambda_2 \cdot \min\{q-\ell_2, 0\}$ or $\lambda_{m-1} \cdot \min\{q-\ell_{m-1}, 0\}$ is redundant. Since $\lambda_2 > 0$ and $\lambda_{k-1} > 0$, we have $\ell_2 < u_{11} - u_{2m} < \ell_{m-1}$. Thus

$$\frac{\partial \tilde{J}^{(\eta,\lambda)}}{\partial q}(q,\theta) = \begin{cases} \tilde{\varphi}^-(\theta) - \eta + \lambda_2 + \lambda_{m-1}, & q < \ell_2 \\ \tilde{\varphi}^-(\theta) - \eta + \lambda_{m-1}, & q \in (\ell_2, \alpha) \\ \tilde{\varphi}^+(\theta) - \eta + \lambda_{m-1}, & q \in (\alpha, \ell_{m-1}) \\ \tilde{\varphi}^+(\theta) - \eta, & q > \ell_{m-1} \end{cases}$$

Similar to the proof for the $m = 3$ case, since $\lambda_2 > 0$, we can find an opened interval I such that for every $\theta \in I$, $\arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$ contains a single value $\theta = \ell_2$, which contradicts with the fact that $\mathbf{q}^*$ is a pointwise maximizer of $\tilde{J}^{(\eta,\lambda)}(q,\theta)$ almost everywhere.[23] Thus we have $\lambda_k = 0, \forall k \in \{2, \ldots, m-1\}$.

Property 2. In property 1, we show that when $m = 3$, either $\lambda_2 = 0$ or $\ell_2 = u_{11} - u_{2m}$. When $m \geq 4$, $\lambda = \mathbf{0}$. Recall that $\eta^* = \eta$ and $\lambda^* = \lambda_2$. In any of the scenarios, the ironed virtual surplus function $\tilde{J}^{(\eta,\lambda)}$ can be expressed as follows (recall that $\alpha = u_{11} - u_{2m}$):

$$\tilde{J}^{(\eta,\lambda)}(q,\theta) = \tilde{\varphi}^-(\theta) \cdot \min\{q, \alpha\} + \tilde{\varphi}^+(\theta) \cdot \max\{q-\alpha, 0\} - \eta^* \cdot q + \lambda^*(q-\alpha) \cdot \mathbb{1}[q \leq \alpha]$$

$$= \begin{cases} (\tilde{\varphi}^-(\theta) - \eta^* + \lambda^*) \cdot q - \lambda^* \alpha, & q \leq \alpha \\ (\tilde{\varphi}^+(\theta) - \eta^*) \cdot q + (\tilde{\varphi}^-(\theta) - \tilde{\varphi}^+(\theta)) \cdot \alpha, & q > \alpha \end{cases} \tag{16}$$

For every fixed θ, the function $\tilde{J}^{(\eta,\lambda)}(\cdot, \theta)$ is continuous and 2-piecewise-linear. Thus at least one of the three values $-u_{2m}$, $u_{11} - u_{2m}$, and u_{11} must be in $\arg\max_{q \in [-u_{2m}, u_{11}]} \tilde{J}^{(\eta,\lambda)}(q,\theta)$. We study the set $\arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$ by a case analysis.

- $\tilde{\varphi}^-(\theta) - \eta^* + \lambda^* \leq 0$: Since $\lambda^* \leq 0$, $\tilde{\varphi}^+(\theta) - \eta^* \leq \tilde{\varphi}^-(\theta) - \eta^* \leq 0$. Thus $-u_{2m} \in \arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$. On the other hand, $-u_{2m} \in \arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$ clearly implies that $\tilde{\varphi}^-(\theta) - \eta^* + \lambda^* \leq 0$.
- $\tilde{\varphi}^-(\theta) - \eta^* + \lambda^* \geq 0$ and $\tilde{\varphi}^+(\theta) - \eta^* \leq 0$: Now $q = u_{11} - u_{2m}$ maximizes $\tilde{J}^{(\eta,\lambda)}(q,\theta)$, and vice versa.
- $\tilde{\varphi}^+(\theta) - \eta^* \geq 0$: Then $\tilde{\varphi}^-(\theta) - \eta^* + \lambda^* \geq \tilde{\varphi}^+(\theta) - \eta^* \geq 0$. Thus $u_{11} \in \arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$. On the other hand, $u_{11} \in \arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$ clearly implies that $\tilde{\varphi}^+(\theta) - \eta^* \geq 0$.

Thus by the first property of Theorem 10, $\mathbf{q}^*$ is a pointwise maximizer of the ironed virtual surplus function almost everywhere. Thus property 2 holds almost everywhere. Moreover, if there exists some $\theta_0 \in [0, 1]$ such that $q^*(\theta_0) \notin \arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$. Suppose $q^*(\theta_0) = -u_{2m}$. Then according

---

[23]If $\lambda_2 = 0$ and $\lambda_{m-1} > 0$, we reach a contradiction by finding an opened interval I such that for every $\theta \in I$, $\arg\max_q \tilde{J}^{(\eta,\lambda)}(q,\theta)$ contains a single value $\theta = \ell_{m-1}$.

to the above case analysis, $\tilde{\varphi}^-(\theta_0) - \eta^* + \lambda^* > 0$. Since $\tilde{\varphi}^-(\cdot)$ is continuous at θ_0, there exists a neighborhood I of θ_0 such that $\tilde{\varphi}^-(\theta_0) - \eta^* + \lambda^* > 0, \forall \theta \in I$.[24] Thus $-u_{2m} \notin \arg\max_q \tilde{J}^{(\eta,\lambda)}(q, \theta), \forall \theta \in I$, contradicting with the fact that $\mathbf{q}^*$ is a pointwise maximizer of the ironed virtual surplus function almost everywhere. Similarly, we reach a contradition when $q^*(\theta_0) = u_{11} - u_{2m}$ or u_{11}. Hence, $\mathbf{q}^*$ is a pointwise maximizer of $\tilde{J}^{(\eta,\lambda)}$ everywhere on $[0, 1]$, which implies that property 2 holds for every $\theta \in [0, 1]$.

Property 3. $\mathbf{q}^*$ only takes three different values $-u_{2m}, u_{11} - u_{2m}, u_{11}$. Moreover, $q^{*^{-1}}(-u_{2m}) = 0$, $q^{*^{-1}}(u_{11} - u_{2m}) = \theta_p^L$, $q^{*^{-1}}(u_{11}) = \theta_p^H$. Thus by the third property of Theorem 10, $\mathbf{q}^*$ satisfies the generalized pooling property iff θ_p^L is not in the interior of any ironed interval of $\tilde{\varphi}^-(\cdot)$, and $\theta_p^H(\cdot)$ is not in the interior of any ironed interval of $\tilde{\varphi}^+$.

Sufficient Condition. Next, we prove that the properties in the statement are sufficient. Given a menu $\mathcal{M}$ that contains only the fully informative experiment with with price $p > 0$. Suppose there exist multiplier η^* and $\lambda^* \geq 0$ that satisfies all properties in the statement. We prove that $\mathbf{q}^*$ (defined in Equation (14)) is an optimal solution to the program in Figure 3.

Firstly, since $p \leq \hat{p}$, due to the analysis for Property 1, we know that $\theta_p^L \leq \theta_1$ and $\theta_p^H \geq \theta_2$. Thus $\theta_p^L < \theta_p^H$ and $\mathbf{q}^*$ is a feasible solution of the program, which induces strictly positive revenue.

To apply Theorem 10, we choose multipliers η' and $\lambda' = \{\lambda_k'\}_{k\in\{2,\dots,m-1\}} \geq \mathbf{0}$, and prove that all three properties in the statement of Theorem 10 are satisfied for $\mathbf{q}^*$ and (η', λ'), which implies that $\mathbf{q}^*$ is an optimal solution. Let $\eta' = \eta^*$. When $m = 3$, let $\lambda_2' = \lambda^*$. When $m \geq 4$, let $\lambda_k' = 0$ for all $k \in \{2, \dots, m - 1\}$.

For the first property of Theorem 10, we notice that the ironed virtual surplus function $\tilde{J}^{(\eta',\lambda')}(q, \theta)$ satisfies Equation (16). As argued earlier in the case analysis in property 2, when property 2 of Theorem 9 is satisfied, $\mathbf{q}^*$ is a pointwise maximizer of $\tilde{J}^{(\eta',\lambda')}(q, \theta)$. The first property of Theorem 10 is satisfied.

For the second property of Theorem 10, according to the choice of λ', we only have to argue that the equality holds when $m = 3$ and $\lambda^* > 0$. For other scenarios, the property trivially holds since the corresponding multiplier λ_k' is 0. Thus by property 1 (in the statement of Theorem 9), $u_{12} - u_{22} = u_{11} - 1$ and $p = 1 - u_{22}$. We follow the notation θ_k^* and Δ_k as used for proving property 1 is necessary. By Equation (15),

$$\Delta_2 = L^*(\theta_2^*) - L_2(\theta_2^*) = L^*(\theta_p^H) - L_2(\theta_p^H) = u_{23} - p - u_{22} = 0, \ (u_{23} = 1 \text{ as } m = 3).$$

Thus the second property of Theorem 10 is satisfied. As argued above, $\mathbf{q}^*$ satisfies the generalized pooling property iff θ_p^L is not in the interior of any ironed interval of $\tilde{\varphi}^-$, and θ_p^H is not in the interior of any ironed interval of $\tilde{\varphi}^+$. Thus the third property of Theorem 10 is also satisfied. Hence, $\mathbf{q}^*$ and (η', λ') satisfy all properties in Theorem 10, which implies that $\mathbf{q}^*$ is an optimal solution. □

Proof of Theorem 4: We are going to show that there exist $p > 0$ and multipliers $\eta^*, \lambda^* \geq 0$ that satisfy all properties of Theorem 9. Then by Theorem 9, selling the complete information at price p is the optimal menu.

We notice that in Definition 7, if $\varphi(\cdot)$ is non-decreasing, $H(\cdot)$ is a convex function. Thus $G(r) = H(r), \forall r \in [0, 1]$ and $\tilde{\varphi}(\theta) = \varphi(\theta), \forall \theta \in [0, 1]$. Since $\varphi^-(\cdot)$ and $\varphi^+(\cdot)$ are both monotonically non-decreasing, we have $\varphi^-(\theta) = \tilde{\varphi}^-(\theta)$ and $\varphi^+(\theta) = \tilde{\varphi}^+(\theta), \forall \theta \in [0, 1]$.

Define function $W : [0, 1] \to \mathbb{R}$ as: $W(p) = \varphi^-(\frac{p}{u_{11}}) - \varphi^+(1 - \frac{p}{u_{2m}})$. Then by Definition 8, we have $W(0) = \varphi^-(0) - \varphi^+(1) = -1$. If $\varphi^-(\frac{\hat{p}}{u_{11}}) \geq \varphi^+(1 - \frac{\hat{p}}{u_{2m}})$, $W(\hat{p}) \geq 0$. Since both φ^- and φ^+ are

---

[24]When $\theta_0 = 0$ (or 1), $\tilde{\varphi}^-(\cdot)$ is right-continuous (or left-continuous) at θ_0. We choose I to be $[0, \delta)$ or $(1 - \delta, 1]$ for some sufficiently small $\delta > 0$.

continuous, W is also continuous. Thus there exists $p_0 \in (0, \widehat{p}]$ such that $W(p_0) = 0$. Consider $p = p_0$ and multipliers $\eta^* = \varphi^-(\frac{p_0}{u_{11}}) = \varphi^+(1 - \frac{p_0}{u_{2m}}), \lambda^* = 0$. Property 1 of Theorem 9 holds since $p_0 \leq \widehat{p}$ and $\lambda^* = 0$. Property 3 of Theorem 9 holds since neither $\varphi^-(\cdot)$ nor $\varphi^+(\cdot)$ requires ironing. For property 2 of Theorem 9, since $\theta^L_{p_0} = \frac{p_0}{u_{11}}$ and $\theta^H_{p_0} = 1 - \frac{p_0}{u_{2m}}$, by our choice of η^*, $\varphi^-(\theta^L_{p_0}) - \eta^* = 0$ and $\varphi^+(\theta^H_{p_0}) - \eta^* = 0$. Property 2 then follows from the fact that $\varphi^-(\cdot)$ and $\varphi^+(\cdot)$ are both monotonically non-decreasing.

□

For several standard distributions, $\varphi^-(\cdot)$ and $\varphi^+(\cdot)$ are both monotonically non-decreasing. The following examples are some applications of Theorem 4.

EXAMPLE 1 (UNIFORM DISTRIBUTION). *Consider the uniform distribution $U[0, 1]$. Then $\varphi^-(\theta) = 2\theta$ and $\varphi^+(\theta) = 2\theta - 1$ are both increasing functions. Thus when $m = 3$ and $u_{12} - u_{22} = u_{11} - 1$, selling the complete information is optimal. When $u_{11} = u_{2m} = 1$, $\varphi^-(\frac{\widehat{p}}{u_{11}}) \geq \varphi^+(1 - \frac{\widehat{p}}{u_{2m}})$ if and only if $\widehat{p} \geq \frac{1}{4}$. In other words, if $u_{22} \leq 3(1 - u_{12})$ and $u_{1,m-1} \leq 3(1 - u_{2,m-1})$, selling the complete information is optimal.*

EXAMPLE 2 (EXPONENTIAL DISTRIBUTIONS RESTRICTED ON $[0, 1]$). *$f(\theta) = c \cdot \lambda e^{-\lambda\theta}$, where $c = 1 - e^{-\lambda}$. $f'(\theta) = -c\lambda^2 \cdot e^{-\lambda\theta}$. When $\lambda \leq 2$, $(\varphi^-)'(\theta) = 2f(\theta) + \theta f'(\theta) = c\lambda(2 - \theta\lambda) \cdot e^{-\lambda\theta} \geq 0$. And $(\varphi^+)'(\theta) = 2f(\theta) + (\theta - 1)f'(\theta) > 0$, since $f'(\theta) < 0$. Selling the complete information is optimal if either conditions in Theorem 4 is satisfied.*

EXAMPLE 3 (NORMAL DISTRIBUTIONS RESTRICTED ON $[0, 1]$). *Consider the normal distribution $\mathcal{N}(0, \sigma^2)$ restricted on $[0, 1]$. $f(\theta) = c \cdot \exp(-\frac{\theta^2}{2\sigma^2})$, where $c = 1/\int_0^1 \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp(-\frac{\theta^2}{2\sigma^2})d\theta$. $f'(\theta) = -c \cdot \frac{\theta}{\sigma^2}\exp(-\frac{\theta^2}{2\sigma^2}) < 0$. When $\sigma^2 \geq \frac{1}{2}$, $(\varphi^-)'(\theta) = 2f(\theta) + \theta f'(\theta) = c(2 - \frac{\theta}{\sigma^2}) \cdot \exp(-\frac{\theta^2}{2\sigma^2}) \geq 0$. And $(\varphi^+)'(\theta) = 2f(\theta) + (\theta - 1)f'(\theta) > 0$, since $f'(\theta) < 0$. Selling the complete information is optimal if either conditions in Theorem 4 is satisfied.*

E MISSING DETAILS FROM SECTION 5

E.1 Missing Details from Section 5.1

Proof of Theorem 6: Let $\beta = \int_0^\infty \frac{1}{r(x)}dx$. We first show that for any distribution D, $\frac{\text{REV}(\mathcal{M},D)}{\text{FREV}(D)} \leq \beta$. In fact, we have

$$\text{REV}(\mathcal{M}, D) = \mathbb{E}_{\theta \sim D}[t(\theta)] = \int_0^\infty \Pr_{\theta \sim D}[t(\theta) \geq x]dx \leq \int_0^\infty \Pr_{\theta \sim D}[U(\theta) \geq r(x)]dx$$

$$\leq \int_0^\infty \frac{\text{FREV}(D)}{r(x)}dx = \beta \cdot \text{FREV}(D)$$

Here the first inequality follows from the definition of $r(x)$: $t(\theta) \geq x$ implies that $U(\theta) \geq r(x)$. For the second inequality, we notice that by the definition of $U(\theta)$, the buyer will purchase the fully informative experiment at price p if and only if $U(\theta) \geq p$. Thus $\text{FREV}(D) \geq r(x) \cdot \Pr[U(\theta) \geq r(x)]$.

Now we prove that $\text{RATIO}(\mathcal{M}) \geq \beta$. In particular, we show that for any $\beta' < \beta$ there exists a distribution D such that $\text{REV}(\mathcal{M}, D) > \beta' \cdot \text{FREV}(D)$.

Since $1/r(x)$ is weakly decreasing, non-negative, and $\int_0^\infty \frac{1}{r(x)}dx = \beta$, there exist $0 = t_0 < t_1 < \cdots < t_N < \infty$ with $0 = r(t_0) < r(t_1) < \cdots < r(t_N) < \infty$ such that [25]

$$\beta'' = \sum_{k=1}^N \frac{t_k - t_{k-1}}{r(t_k)} > \beta'.$$

---

[25]We notice that $U(0) = 0$ as the payment $t(\cdot)$ is non-negative. Thus $r(t_0) = r(0) = 0$.

Let $\varepsilon > 0$ be small enough so that $\beta'' > (1+\varepsilon)\beta'$ and $r(t_{k+1}) > (1+\varepsilon)r(t_k)$ for all $1 \leq k < N$. By the definition of $r(\cdot)$, we can choose a sequence of types $\{\theta^{(k)}\}_{k\in[N]}$ in Θ such that $t(\theta^{(k)}) \geq t_k$ and $r(t_k) \leq U(\theta^{(k)}) < (1+\varepsilon)r(t_k)$. Then

$$\sum_{k=1}^{N} \frac{t_k - t_{k-1}}{U(\theta^{(k)})} > \sum_{k=1}^{N} \frac{t_k - t_{k-1}}{(1+\varepsilon)r(t_k)} = \frac{\beta''}{1+\varepsilon} > \beta'$$

For every $k \in [N]$, let $\xi_k = U(\theta^{(k)})$. We notice that $U(\theta^{(k)}) < (1+\varepsilon)r(t_k) < r(t_{k+1}) \leq U(\theta^{(k+1)})$ for $0 \leq k < N$. Consider the distribution D with support $\{\theta^{(1)}, \ldots, \theta^{(N)}\}$ and $\Pr[\theta = \theta^{(k)}] = \xi_1 \cdot (\frac{1}{\xi_k} - \frac{1}{\xi_{k+1}})$ for all $k \in [N]$, where $\xi_{N+1} = \infty$.

Recall that the buyer will purchase the fully informative experiment at price p if and only if $U(\theta) \geq p$. Thus under distribution D, it's sufficient to consider the price $p = U(\theta^{(k)}) = \xi_k$ for some $k \in [N]$. Since $\{U(\theta^{(k)})\}_{k\in[N]}$ is a strictly increasing sequence, then the buyer will purchase at price $p = \xi_k$ whenever $\theta \in \{\theta^{(k)}, \ldots, \theta^{(N)}\}$, which happens with probability $\sum_{j=k}^{N} \xi_1 \cdot (\frac{1}{\xi_k} - \frac{1}{\xi_{k+1}}) = \frac{\xi_1}{\xi_k}$. Hence we have $\text{FRev}(D) = \xi_1$.

$$\text{Rev}(\mathcal{M}, D) = \sum_{k=1}^{N} t(\theta^{(k)}) \cdot \Pr[\theta = \theta^{(k)}] = \sum_{k=1}^{N} t_k \cdot \xi_1 \cdot (\frac{1}{\xi_k} - \frac{1}{\xi_{k+1}})$$

$$= \xi_1 \cdot \sum_{k=1}^{N} \frac{t_k - t_{k-1}}{\xi_k} > \beta' \cdot \text{FRev}(D)$$

$\square$

Proof of Lemma 8: Recall that under this environment $U(\theta) = \sum_{i=1}^{3} \theta_i \cdot 1 - \max\{\theta_1, \theta_2, \theta_3\} = 1 - \max\{\theta_1, \theta_2, \theta_3\}$ (here $\theta_3 = 1 - \theta_1 - \theta_2$). We first construct the distribution D based on the given sequence $\{y_k\}_{k=1}^{N}$. Let $\{t_k\}_{k=1}^{N}$ be a sequence of positive numbers that increases fast enough so that: (i) $(t_k/\text{gap}_k) \cdot U(y_k)$ is increasing, and (ii) $t_{k+1}/t_k \geq 1/\varepsilon$ for all $1 \leq k < N$. Such a sequence must exist since after $t_1, \ldots, t_{k-1}$ are decided, we can choose a large enough t_k to satisfy both properties. Let $\delta > 0$ be any number such that $1/\delta > \max_{k\in[N]}\{t_k/\text{gap}_k\}$. For every $k \in [N]$, define $x_k \in [0,1]^2$ as $x_k = (\delta t_k/\text{gap}_k) \cdot y_k$. For every $k \in [N]$, since $\delta t_k/\text{gap}_k < 1$ and $y_k \in \Theta$, we have $x_k \in \Theta$.

By Property 2 and 3 of the statement, $0.4 \geq \sqrt{y_{k,1}^2 + y_{k,2}^2} \geq \sqrt{y_{k,1}^2 + (0.9y_{k,1})^2}$, which implies that $y_{k,1} \leq \sqrt{0.4^2/1.81} < \frac{1}{3}$. Similarly, $y_{k,2} < \frac{1}{3}$. Let $y_{k,3} = 1 - y_{k,1} - y_{k,2}$ and $x_{k,3} = 1 - x_{k,1} - x_{k,2}$. Then $U(y_k) = 1 - \max\{y_{k,1}, y_{k,2}, y_{k,3}\} = 1 - y_{k,3} = y_{k,1} + y_{k,2}$. Moreover, since $\delta t_k/\text{gap}_k < 1$, we have $x_{k,i} < y_{k,i} < \frac{1}{3}$ for $i = 1, 2$. Thus $U(x_k) = 1 - x_{k,3} = x_{k,1} + x_{k,2} = (\delta t_k/\text{gap}_k) \cdot U(y_k)$. For every $k \in [N]$, let $\xi_k = U(x_k)$. Then $\{\xi_k\}_{k=1}^{N}$ is an increasing sequence due to (i).

Consider the distribution D with support $\{x_1, \ldots, x_N\}$. $\Pr[\theta = x_k] = \xi_1 \cdot (\frac{1}{\xi_k} - \frac{1}{\xi_{k+1}})$ for all $k \in [N]$, where $\xi_{N+1} = \infty$. In the next step, we construct a sequence of experiments $\{\Pi^{(k)}\}_{k\in[N]}$ and a mechanism $\mathcal{M}$. For every $k \in [N]$, define experiment $\Pi^{(k)}$ as follows:

$\Pi^{(k)}$	1	2	3
ω_1	$y_{k,1}$	0	$1 - y_{k,1}$
ω_2	0	$y_{k,2}$	$1 - y_{k,2}$
ω_3	0	0	1

Now consider the following mechanism $\mathcal{M}$. For every buyer's type θ in the support, the buyer chooses the experiment $\Pi^{(k^*)}$ where $k^* = \arg\max_k \{V_\theta^*(\Pi^{(k)}) - \delta \cdot t_k\}$, and pays $\delta \cdot t_{k^*}$. In other

words, he chooses the option (with experiment $\Pi^{(k^*)}$ and price $\delta \cdot t_{k^*}$) that obtains the highest utility, when he follows the recommendation.

CLAIM 5. $\mathcal{M}$ is IC and IR.

PROOF. We first prove that $\mathcal{M}$ is IR. For every buyer's type x_k, the buyer's utility after purchasing $\Pi^{(k^*)}$ is

$$
\begin{aligned}
V_{x_k}(\Pi^{(k^*)}) - \delta \cdot t_{k^*} &\geq V_{x_k}^*(\Pi^{(k^*)}) - \delta \cdot t_{k^*} \\
&\geq V_{x_k}^*(\Pi^{(k)}) - \delta \cdot t_k \\
&= x_{k,1} y_{k,1} + x_{k,2} y_{k,2} + x_{k,3} - \delta \cdot t_k \\
&= \delta t_k \cdot \left(\frac{y_{k,1}^2 + y_{k,2}^2}{\mathrm{gap}_k} - 1 \right) + x_{k,3} \\
&\geq x_{k,3}
\end{aligned}
$$

Here the second inequality follows from the definition of k^*; The first equality follows from the fact that: under the matching utility environment, $V_\theta^*(\Pi^{(k)}) = \theta_1 y_{k,1} + \theta_2 y_{k,2} + \theta_3$ for every type $\theta \in \Theta$; The second equality follows from the definition of x_k; The last inequality follows from Property 1 and 3 of the statement. We notice that at type x_k, the buyer has value $\max_j \{ \sum_i x_{k,i} \cdot u_{ij} \} = \max\{x_{k,1}, x_{k,2}, x_{k,3}\} = x_{k,3}$ before purchasing any experiment. Thus $\mathcal{M}$ is IR.

By the definition of $\mathcal{M}$, it's clear that $\mathcal{M}$ is responsive-IC. To prove that $\mathcal{M}$ is also IC, it suffices to show that for any type θ from the distribution D and any experiment $\Pi^{(k)}$, following the recommendation always maximizes the buyer's utility, i.e. $V_\theta(\Pi^{(k)}) = V_\theta^*(\Pi^{(k)})$. In fact, according to the construction of $\Pi^{(k)}$, when action 1 or 2 is recommended, the corresponding state (1 or 2, respectively) is fully revealed and thus following the recommendation clearly maximizes the buyer's expected payoff. When action 3 is recommended, the buyer with type x_q (for some $q \in [N]$) has expected payoff $x_{q,1}(1 - y_{q,1})$, $x_{q,2}(1 - y_{q,2})$ and $x_{q,3}$, by choosing action 1, 2 and 3 respectively. Since $x_{q,1} < \frac{1}{3}$ and $x_{q,2} < \frac{1}{3}$, we have $x_{q,3} > \frac{1}{3} \geq \max\{x_{q,1}(1 - y_{q,1}), x_{q,2}(1 - y_{q,2})\}$. Thus the buyer will always follow the recommendation for every experiment $\Pi^{(k)}$. Since $\mathcal{M}$ is responsive-IC, $\mathcal{M}$ is also IC. $\square$

Back to the proof of Lemma 8. Now we compute $\mathrm{REV}(\mathcal{M}, D)$. We notice that for every $k \in [N]$ and every type $\theta \in \Theta$,

$$
V_\theta^*(\Pi^{(k)}) = \theta_1 \cdot y_{k,1} + \theta_2 \cdot y_{k,2} + \theta_3.
$$

Consider any buyer's type x_k. For any $1 \leq q < k$, by the definition of x_k and gap_k, we have

$$
\begin{aligned}
V_{x_k}^*(\Pi^{(k)}) - V_{x_k}^*(\Pi^{(q)}) &= x_{k,1} \cdot (y_{k,1} - y_{q,1}) + x_{k,2} \cdot (y_{k,2} - y_{q,2}) \\
&= \frac{\delta t_k}{\mathrm{gap}_k} \cdot \{ y_{k,1} \cdot (y_{k,1} - y_{q,1}) + y_{k,2} \cdot (y_{k,2} - y_{q,2}) \} \\
&\geq \delta t_k > \delta t_k - \delta t_q
\end{aligned}
$$

Recall that in mechanism $\mathcal{M}$, the buyer chooses the experiment $\Pi^{(q)}$ that maximizes $\{ V_\theta^*(\Pi^{(q)}) - \delta \cdot t_q \}$. Thus the above inequality implies that in $\mathcal{M}$, the buyer must purchase an experiment $\Pi^{(q)}$ where $q \geq k$. Since $\{t_k\}_{k=1}^N$ is an increasing sequence, the buyer's payment is at least $\delta \cdot t_k$. By Theorem 6,

$$\text{REV}(\mathcal{M}, D) \geq \sum_{k=1}^{N} \delta \cdot t_k \cdot \xi_1 \cdot \left(\frac{1}{\xi_k} - \frac{1}{\xi_{k+1}}\right) = \delta \xi_1 \cdot \sum_{k=1}^{N} \frac{t_k - t_{k-1}}{\xi_k}$$

$$\geq (1 - \varepsilon) \xi_1 \cdot \sum_{k=1}^{N} \frac{\delta t_k}{\xi_k} = (1 - \varepsilon) \cdot \sum_{k=1}^{N} \frac{\text{gap}_k}{y_{k,1} + y_{k,2}} \cdot \text{FREV}(D)$$

$$\geq \frac{3}{2}(1 - \varepsilon) \cdot \sum_{k=1}^{N} \text{gap}_k \cdot \text{FREV}(D)$$

Here the second inequality follows from $t_{k+1}/t_k \geq 1/\varepsilon$. The third inequality follows from the fact that both $y_{k,1}$ and $y_{k,2}$ are at most $\frac{1}{3}$, as shown in the beginning of the proof. The second equality follows from $\xi_k = U(x_k) = (\delta t_k/\text{gap}_k) \cdot U(y_k) = (\delta t_k/\text{gap}_k) \cdot (y_{k,1} + y_{k,2})$ and the fact that $\text{FREV}(D) = \xi_1$, which is proved in Theorem 6. We include the proof here for completeness. Recall that a buyer with type x will purchase the fully informative experiment at price p if and only if $U(x) \geq p$. Thus under distribution D, it's sufficient to consider the price $p = U(x_k) = \xi_k$ for some $k \in [N]$. Since $\{U(x_k)\}_{k \in [N]}$ is a strictly increasing sequence, then the buyer will purchase at price $p = \xi_k$ whenever $x \in \{x_k, \ldots, x_N\}$, which happens with probability $\sum_{j=k}^{N} \xi_1 \cdot \left(\frac{1}{\xi_k} - \frac{1}{\xi_{k+1}}\right) = \frac{\xi_1}{\xi_k}$. Hence, $\text{FREV}(D) = \xi_1$.

The proof is done by noticing that $\text{REV}(\mathcal{M}, D) \leq \text{OPT}(D)$, since $\mathcal{M}$ is IC and IR by Claim 5. □

Proof of Theorem 5: For any integer N, we construct a sequence $\{y_k\}_{k=1}^{N}$ that satisfies all three properties in the statement of Lemma 8, and $\text{gap}_k = \Theta(k^{-6/7})$. Then by Lemma 8, there exists a distribution D with support size N such that (by choosing $\varepsilon = \frac{1}{3}$)

$$\frac{\text{OPT}(D)}{\text{FREV}(D)} \geq \sum_{k=1}^{N} \text{gap}_k = \Omega\left(\sum_{k=1}^{N} k^{-6/7}\right) = \Omega(N^{1/7})$$

All the points $\{y_k\}_{k=1}^{N}$ are placed in a sequence of "shells". As k goes larger, y_k stays in a shell with a larger and larger radius. For every $i = 1, \ldots, M$, the i-th shell has radius $0.3 + \sum_{\ell=1}^{i} \ell^{-3/2}/\alpha$, where $\alpha = 0.1 \cdot \sum_{\ell=1}^{M} \ell^{-3/2}$. For every shell i, let r_i be the arc from angle $\arctan(\frac{9}{10})$ to angle $\arctan(\frac{10}{9})$. The i-th shell contains $i^{3/4}$ points. All points are evenly spread in the arc r_i, so that angle between any two of them is at least $c \cdot i^{-3/4}$ for some absolute constant c.

We notice that the radius of every shell is in the range $[0.3, 0.4]$. Thus any point y_k on each shell satisfies $y_{k,1} + y_{k,2} \leq 1$ and $||y_k||_2 = \sqrt{y_{k,1}^2 + y_{k,2}^2} \in [0.3, 0.4]$. Moreover, according to the definition of r_i, all points y_k on this arc must satisfy $\frac{y_{k,1}}{y_{k,2}} \in [\frac{9}{10}, \frac{10}{9}]$. It remains to analyze $\text{gap}_k = \min_{0 \leq j < k} \{(y_{k,1} - y_{j,1}) \cdot y_{k,1} + (y_{k,2} - y_{j,2}) \cdot y_{k,2}\} = \min_{0 \leq j < k} \{y_k \cdot (y_k - y_j)\}$. We have that $y_j \cdot y_k = ||y_j||_2 \cdot ||y_k||_2 \cdot \cos(\alpha)$ where α is the angle between y_k and y_j. Let i and i' be the shell that y_k and y_j are placed in respectively. Then $i' \leq i$. Suppose $i' = i$. Since $\alpha = \Omega(i^{-3/4})$, we have $\cos(\alpha) = 1 - \Omega(i^{-3/2})$ (because $\cos(\alpha) = 1 - \alpha^2/2 + \alpha^4/24 - \ldots$). $(y_k - y_j) \cdot y_k = ||y_k||_2^2 \cdot \Omega(i^{-3/2}) = \Omega(i^{-3/2})$. Now suppose $i' < i$. $||y_k||_2 - ||y_j||_2 = \sum_{\ell=i'+1}^{i} \ell^{-3/2}/\alpha \geq i^{-3/2}/\alpha$. Since $\sum_{\ell=1}^{\infty} \ell^{-3/2}$ converges, $\alpha = 0.1 \cdot \sum_{\ell=1}^{M} \ell^{-3/2}$ is bounded by some absolute constant. Thus

$$(y_k - y_j) \cdot y_k \geq ||y_k||_2^2 - ||y_j||_2 ||y_k||_2 \geq 0.3(||y_k||_2 - ||y_j||_2) = \Omega(i^{-3/2}),$$

where the second inequality follows from $||y_j||_2 \geq 0.3$. Combining both cases, we have proved that $\text{gap}_k = \Theta(i^{-3/2})$. Since the first i shells together contain $\sum_{\ell=1}^{i} i^{3/4} = \Theta(i^{7/4})$ points, we have $k = \Theta(i^{7/4})$ and $\text{gap}_k = \Theta(k^{-6/7})$. □

E.2 Missing Details from Section 5.2

The following observation directly follows from the definition of responsive mechanisms.

OBSERVATION 6. *In the matching utilities environment, a mechanism $\mathcal{M}$ is responsive if and only if $\theta_i \pi_{ii}(\theta) \geq \theta_j \pi_{ji}(\theta), \forall \theta \in T, i, j \in [n]$. Here $\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$.*

Proof of Lemma 9: Let $\mathcal{M}$ be any optimal responsive mechanism. Then the buyer's value for experiment $E(\theta)$ is ($\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$):

$$V_\theta^*(E(\theta)) = \sum_{i=1}^{n} \theta_i \pi_i(\theta) = \sum_{i=1}^{n-1} \theta_i \cdot (\pi_i(\theta) - \pi_n(\theta)) + \pi_n(\theta) \tag{17}$$

Now for every type θ, let $c(\theta) = 1 - \max_i \{\pi_i(\theta)\} \geq 0$. For every $i \in [n]$, we arbitrarily move a total mass of $c(\theta)$ from $\{\pi_{ij}(\theta)\}_{j \neq i}$ to $\pi_{ii}(\theta)$. We also increase the payment of type θ by $c(\theta)$. Let $\mathcal{M}' = \{E'(\theta), t'(\theta)\}_{\theta \in \Theta}$ be the induced mechanism. By Observation 6, $\mathcal{M}'$ is responsive since $\mathcal{M}$ is responsive and $\pi_{ii}(\theta)$ is (weakly) increased while $\pi_{ij}(\theta)$ is (weakly) decreased for $i \neq j$.

Moreover, for every $\theta \in \Theta$, the buyer's utility after purchasing $E'(\theta)$, $V_\theta^*(E'(\theta)) - t'(\theta)$, is the same as $V_\theta^*(E(\theta)) - t(\theta)$. Since $\mathcal{M}$ is IC, IR, $\mathcal{M}'$ is also IC, IR. And $\text{REV}(\mathcal{M}') \geq \text{REV}(\mathcal{M})$. Thus $\mathcal{M}'$ is also the optimal responsive mechanism. The proof is finished by noticing that $\mathcal{M}'$ satisfies the property in the statement due to the choice of $c(\theta)$. □

Proof of Observation 1: For any $\theta, \theta' \in \Theta$, since the mechanism is responsive, IC and IR, we have

$$G(\theta') - G(\theta) \geq [V_{\theta'}^*(E(\theta)) - t(\theta)] - [V_\theta^*(E(\theta)) - t(\theta)]$$

$$= V_{\theta'}^*(E(\theta)) - V_\theta^*(E(\theta)) = \sum_{i=1}^{n-1} (\pi_i(\theta) - \pi_n(\theta))(\theta_i' - \theta_i)$$

Taking derivative on both sides at $\theta' = \theta$ finishes the proof. □

Proof of Lemma 10: For any θ, the buyer's utility function $G(\theta)$ is the maximum of a collection of linear functions over θ. Thus $G(\cdot)$ is convex. Moreover, for any $\theta, \theta' \in \Theta$,

$$G(\theta) - G(\theta') \leq [V_\theta(E(\theta)) - t(\theta)] - [V_{\theta'}(E(\theta)) - t(\theta)] \leq c(\theta, \theta')$$

Here the first inequality follows from the fact that $G(\theta') = V_{\theta'}(E(\theta')) - t(\theta') \geq V_{\theta'}(E(\theta)) - t(\theta)$ since $\mathcal{M}$ is IC. The second inequality follows from the definition of $c(\theta, \theta')$.

It remains to show that $G(\mathbf{0}) = 1$. Since $\mathcal{M}$ is IR, $G(\mathbf{0}) \geq u(\mathbf{0}) = 1$. On the other hand, the buyer's value for any experiment is at most 1. Thus $G(\mathbf{0}) \leq 1$. □

Proof of Lemma 11: We use the notations from Definition 4. For every $\theta \in \Theta$, by the product rule, we have [26]

$$\text{div}(G(\theta)f(\theta) \cdot \theta) = \theta \cdot \nabla(G(\theta) \cdot f(\theta)) - G(\theta)f(\theta) \cdot \nabla\theta$$

$$= \theta \cdot [G(\theta) \cdot \nabla f(\theta) + f(\theta) \cdot \nabla G(\theta)] + (n-1) \cdot G(\theta)f(\theta)$$

Integrating over Θ for both sides, by the divergence theorem, we have

$$\int_\Theta (\nabla G(\theta) \cdot \theta)f(\theta)d\theta = \int_{\partial\Theta} G(\theta)f(\theta)(\theta \cdot \boldsymbol{n})d\theta - \int_\Theta G(\theta) \cdot [\nabla f(\theta) \cdot \theta + (n-1)f(\theta)] \, d\theta \tag{18}$$

Similarly, for every $i \in [n-1]$,

$$\text{div}(G(\theta)f(\theta) \cdot \boldsymbol{e}_i) = \boldsymbol{e}_i \cdot [G(\theta) \cdot \nabla f(\theta) + f(\theta) \cdot \nabla G(\theta)]$$

---

[26]For any real-valued function $F(\cdot)$ over Θ, $\text{div}(F)$ is the divergence of F, defined as $\text{div}(F) = \sum_{i=1}^{n-1} \frac{\partial F}{\partial \theta_i}$.

Thus

$$
\int_{\Theta_i} \frac{\partial G(\theta)}{\partial \theta_i} f(\theta) d\theta = \int_{\Theta_i} (\nabla G(\theta) \cdot \boldsymbol{e}_i) f(\theta) d\theta
$$
$$
= \int_{\partial \Theta_i} G(\theta) f(\theta) (\boldsymbol{e}_i \cdot \boldsymbol{n}_i) d\theta - \int_{\Theta_i} G(\theta) \cdot (\boldsymbol{e}_i \cdot \nabla f(\theta)) d\theta \tag{19}
$$

Combining Equation (18), Equation (19) and the fact that $\int_\Theta f(\theta) d\theta = 1$ completes the proof. □

Proof of Lemma 12: By Lemma 11,

$$
\int_\Theta G(\theta) d\mu^P = \int_\Theta [-G(\theta) + \nabla G(\theta) \cdot \theta + G(\boldsymbol{0})] f(\theta) d\theta - \sum_{i=1}^{n-1} \int_{\Theta_i} \frac{\partial G(\theta)}{\partial \theta_i} f(\theta) d\theta
$$
$$
\geq \int_\Theta \left[-G(\theta) + \nabla G(\theta) \cdot \theta - \max \left\{ \frac{\partial G(\theta)}{\partial \theta_1}, \ldots, \frac{\partial G(\theta)}{\partial \theta_{n-1}}, 0 \right\} + G(\boldsymbol{0}) \right] f(\theta) d\theta
$$

The last inequality achieves equality if and only if Condition (1) is satisfied. It suffices to prove that $\int_\Theta G(\theta) d\mu^P \leq \int_{\Theta \times \Theta} c(\theta, \theta') d\gamma(\theta, \theta')$. In fact, since $G(\theta) - G(\theta') \leq c(\theta, \theta'), \forall \theta, \theta' \in \Theta$, we have

$$
\int_{\Theta \times \Theta} c(\theta, \theta') d\gamma(\theta, \theta') \geq \int_{\Theta \times \Theta} (G(\theta) - G(\theta')) d\gamma(\theta, \theta')
$$
$$
= \int_\Theta G(\theta) d\gamma_1 - \int_\Theta G(\theta') d\gamma_2
$$
$$
\geq \int_\Theta G(\theta) d\mu^P
$$

The first inequality achieves equality if and only if Condition (2) is satisfied. The last inequality follows from the fact that $\gamma_1 - \gamma_2 \succeq_{cux} \mu^P$ and G is convex. The last inequality achieves equality if and only if Condition (3) is satisfied. □
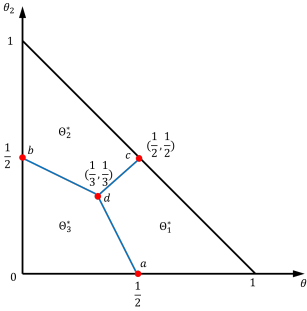
(a) Illustration of P^*. Line segments ad, bd, cd correspond to S_{13}, S_{23}, S_{12} respectively.
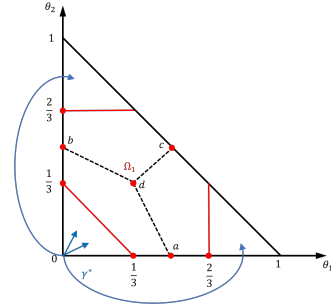
(b) Ω_1 is the hexagon separated by the three red lines. γ^* has a positive density between $\boldsymbol{0}$ and each point in Ω_2.

Fig. 10. Illustrations of the optimal dual when $n = 3$

Proof of Observation 2: Firstly, $\nabla f(\theta) = \boldsymbol{0}$ for every θ since $f(\theta)$ is a constant. Thus the last two terms in Definition 4 is always 0. Moreover, the term $\mathbb{1}_A(\boldsymbol{0})$ contributes a point mass of $+1$ at $\boldsymbol{0}$. The term $-n \cdot \int_\Theta \mathbb{1}_A(\theta) f(\theta) d\theta$ contributes a total mass of -3 uniformly distributed through out Θ. By Lemma 11, the other two terms contribute a total mass of 2 on the boundary of triangle Θ and line segments ad, bd, cd. We notice that when θ is on the x-axis (or y-axis), both $\theta \cdot \boldsymbol{n}$ and $\boldsymbol{e}_1 \cdot \boldsymbol{n}_1$ (or

$e_2 \cdot \boldsymbol{n}_2$) are 0. Moreover, for any point θ on the line segment $\theta_1 + \theta_2 = 1$, the outward pointing unit vector is $(1, 1)$. Thus $\theta \cdot \boldsymbol{n} = 1 = e_i \cdot \boldsymbol{n}_i$ for any $i \in \{1, 2\}$. Thus there is no mass on the boundary of the big triangle Θ and a total mass of 2 is distributed among line segments ad, bd, cd. The mass on each segment is then obtained by calculating the line integral. $\square$

Proof of Lemma 13: We verify feasibility and all three conditions in the statement of Lemma 12:

Primal Solution Feasibility. Since G^* is induced by a responsive, IC and IR mechanism $\mathcal{M}^*$, by Lemma 10, G^* is a feasible solution to the primal.

Condition 1. By Observation 1, for every $i \in \{1, 2\}$, $\frac{\partial G^*(\theta)}{\partial \theta_i} = \pi_i^*(\theta) - \pi_3^*(\theta), \forall \theta \in \Theta$. Now for any $i \in \{1, 2, 3\}$ and $\theta \in \Theta_i$, if $\theta \in \Omega_1$, $\pi_j^*(\theta) = 1, \forall j \in \{1, 2, 3\}$. Thus $\max\{\frac{\partial G^*(\theta)}{\partial \theta_1}, \frac{\partial G^*(\theta)}{\partial \theta_2}\} = 0$; If $\theta \in \Omega_2$, the buyer does not purchase any experiment. Since $\mathcal{M}^*$ is responsive, by Observation 6 and the fact that $\theta_i \geq \theta_j, \forall j \neq i$, we have that $\pi_{ji}^*(\theta) = 1, \forall j \in \{1, 2, 3\}$. Thus $\max\{\frac{\partial G^*(\theta)}{\partial \theta_1}, \frac{\partial G^*(\theta)}{\partial \theta_2}\}$ equals to 1 if $i \in \{1, 2\}$, and -1 otherwise. Thus Condition (1) is satisfied.

Condition 2. $\gamma^*(\theta, \theta') > 0$ if and only if $\theta = \boldsymbol{0}$ and $\theta' \in \Omega_2$. Thus $G^*(\theta) - G^*(\theta') = 1 - \max\{\theta_1', \theta_2', 1 - \theta_1' - \theta_2'\}$, since the buyer with any type θ'' has utility $u(\theta'') = \max\{\theta_1'', \theta_2'', 1 - \theta_1'' - \theta_2''\}$ if she does not purchase any experiment. Consider the experiment E such that $\pi_{j1}(E) = 1, \forall j \in \{1, 2, 3\}$ and 0 elsewhere, which is one of the no information experiments. Then $V_{\boldsymbol{0}}(E) = u(\boldsymbol{0}) = 1$ and $V_{\theta'}(E) = u(\theta') = \max\{\theta_1', \theta_2', 1 - \theta_1' - \theta_2'\}$. $G^*(\theta) - G^*(\theta') = V_{\boldsymbol{0}}(E) - V_{\theta'}(E)$. Moreover, $G^*(\theta) - G^*(\theta') \leq c(\theta, \theta')$ since G^* is feasible. Thus the inequality achieves equality for any (θ, θ') such that $\gamma^*(\theta, \theta') > 0$.

Condition 3 and Dual Solution Feasibility. Denote γ_1^*, γ_2^* the marginals of γ. We prove that we can transform μ^P to $\gamma_1^* - \gamma_2^*$ through a sequence of mean-preserving spreads in the region where G^* is linear. By the definition of γ^*, γ_1^* has a point mass of +1 at $\boldsymbol{0}$. $-\gamma_2^*$ has a mass of -1 uniformly distributed on Ω_2. Recall that $\Omega_2 = \{\theta \in \Theta : \max\{\theta_1, \theta_2\} \geq \frac{2}{3}\} \cup \{\theta \in \Theta : \theta_1 + \theta_2 \leq \frac{1}{3}\}$, which consists of 3 right triangles with side length $\frac{1}{3}$. Thus $\text{VoL}(\Omega_2) = \frac{1}{6} = \frac{1}{3} \cdot \text{VoL}(\Theta)$. Thus by Observation 2, $\mu^{P^*}(\Omega_2) = (\gamma_1^* - \gamma_2^*)(\Omega_2)$. Hence $\eta = \mu^{P^*} - (\gamma_1^* - \gamma_2^*)$ can be written as $\eta_1 - \eta_2$, where

(1) η_1: A mass of $\frac{2}{3}$ uniformly distributed on each line segment ad, bd, cd, having a total mass of 2.

(2) η_2: A mass of 2 uniformly distributed through out Ω_1.

We notice that for every $\theta \in \Omega_1$, the buyer purchases the fully informative experiment and thus $G^*(\theta) = \frac{2}{3}$ is constant throughout the region Ω_1. Hence, to prove $\gamma_1^* - \gamma_2^* \succeq_{cux} \mu^{P^*}$ and that Condition 3 is satisfied, it suffices to show that we can spread the positive mass on ad, bd, cd to the whole region Ω_1 via mean-preserving spreads, to "zero out" the negative mass (in $-\eta_2$).

To visualize the proof, we map each point $(\theta_1, \theta_2) \in \Theta$ to the point $(\theta_1, \theta_2, \theta_3 = 1 - \theta_1 - \theta_2) \in [0, 1]^3$, so that the type space becomes a regular triangle (Figure 11). Now Ω_1 is separated into three pentagons by ad, bd and cd.

For every θ on the line segment cd, denote $w(\theta)$ the width of the pentagon efhdg at θ (Figure 12). Similarly, define $w(\theta)$ for θ on the line segment ad, bd to be the width of the corresponding pentagon at θ. Let η_2' be a measure that has a total mass of 2 on the line segments ad, bd, cd, such that the density of each point θ is proportional to $w(\theta)$. We claim that $\eta_2 \succeq_{cux} \eta_2'$ by transforming η_2' to η_2 by mean-preserving spreads. For each point θ on the line segment cd, we spread the mass of θ uniformly to all points θ' in the pentagon efhdg such that $\theta_3 = \theta_3'$. This is a mean-preserving spread since by symmetry, the mean of those points is θ. Moreover, since the density of θ in η_2' is proportional to the width of the pentagon at θ, all the mass are uniformly distributed in the
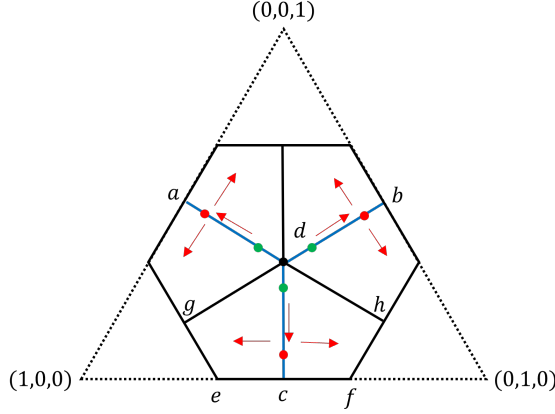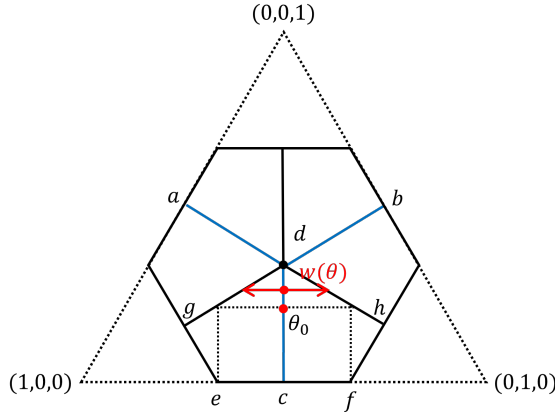
Fig. 11. Proof of Condition 3

pentagon. We perform similar operations for the other two pentagons. Thus we can transform η_2' to η_2 by mean-preserving spreads.

Fig. 12. Illustration of $w(\theta)$ and θ_0 in the proof

Now it suffices to prove that we can transform η_1 to η_2' by mean-preserving spreads. We prove the following claim.

CLAIM 6. *For any $x, y \in \mathbb{R}_+$ such that $x < y$, let $\theta_1, \theta_2, \theta_3$ be the three points on the line segments ad, bd, cd respectively, such that the distance between each θ_i and $d = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is x (green points in Figure 11). Similarly, let $\theta_1', \theta_2', \theta_3'$ be the points that have distance y from point d (red points in Figure 11). Then any positive mass uniformly distributed in $\theta_1, \theta_2, \theta_3$ can be transformed to the same amount of mass uniformly distributed in $\theta_1', \theta_2', \theta_3'$, via a mean-preserving spread.*

PROOF. For every $i \in \{1, 2, 3\}$, we notice that θ_i is a convex combination of d and θ_i'. Thus we can spread the mass at θ_i to d and θ_i' for every $i \in \{1, 2, 3\}$. Next, since d is a convex combination of $\theta_1', \theta_2', \theta_3'$, we can then spread of mass at d to $\theta_1', \theta_2', \theta_3'$. By symmetry, the mass is uniformly distributed in those three points. □

Back to the proof of Lemma 13. The density of each point in η_1 is $\frac{2/3}{cd} = \frac{2\sqrt{6}}{3}$. For η_2', by symmetry, the density of each point θ is

$$\frac{2w(\theta)}{3\int w(\theta')d\theta'} = \frac{2/3 \cdot w(\theta)}{\text{Area of pentagon efhdg}} = 2\sqrt{3} \cdot w(\theta)$$

Let θ_0 be the (unique) point in the interior of cd, such that $w(\theta_0) = \frac{\sqrt{2}}{3}$ (Figure 12). [27] Then for any point θ in the line segment $d\theta_0$, the density of θ in η_1 is at least the density of θ in η_2'; For any point θ in the line segment $\theta_0 c$, the density of θ in η_1 is at most the density of θ in η_2'. By Claim 6, we can transform η_1 to η_2' via mean-preserving spreads, by keeping spreading the mass of points in $d\theta_0$ to points in $\theta_0 c$. Thus we have $\eta_2' \succeq_{cux} \eta_1$. Combining with the fact that $\eta_2 \succeq_{cux} \eta_2'$, we have $\eta_2 \succeq_{cux} \eta_1$, which implies that $\gamma_1^* - \gamma_2^* \succeq_{cux} \mu^{P^*}$.

□

---

[27] The point is unique since $\frac{\sqrt{2}}{3} = ef < gh$.