Constrained Information Design[*]

Laura Doval[†]   Vasiliki Skreta[‡]

June 10, 2021

Abstract

We provide tools to analyze information design problems subject to constraints. We do so by showing that the techniques in Le Treust and Tomala (2019) extend to the case of multiple inequality and equality constraints. This showcases the power of the results in that paper to analyze problems of information design subject to constraints. We illustrate our results with applications to mechanism design with limited commitment (Doval and Skreta, 2020) and persuasion of a privately informed receiver (Kolotilin et al., 2017).

[†]Columbia University and CEPR. E-mail: laura.doval@columbia.edu
[‡]University of Texas at Austin, University College London, and CEPR. E-mail: vskreta@gmail.com.

1 Introduction

We provide tools to solve constrained information design problems. These problems are becoming common: Since Kamenica and Gentzkow (2011) seminal paper on Bayesian persuasion, the literature on information design has grown steadily. A bulk of new work analyzes *constrained* information design problems, which can be classified in three groups:

1. The information designer faces constraints additional to the Bayes' plausibility constraint in Kamenica and Gentzkow (2011), like in Boleslavsky and Kim (2018) on persuasion and moral hazard, and Le Treust and Tomala (2019) on information transmission with capacity constraints.

2. The information designer is designing a mechanism that satisfies incentive and participation constraints, like in mechanism design with aftermarkets (e.g., Calzolari and Pavan, 2006; Dworczak, 2020), in optimal monitoring in moral hazard (e.g., Georgiadis and Szentes, 2020), and in mechanism design with limited commitment (e.g., Doval and Skreta, 2020).

3. Mechanism design problems that do not involve information design and still can be solved using information design tools, like in Dworczak et al. (2019).

A natural approach to tackle these constrained information design problems is to set up a Lagrangian to incorporate the constraints into the objective function, except for the Bayes' plausibility constraint. If each constraint can be written as the expectation over posteriors of some function, then the Lagrangian itself can be written as an expectation over posteriors of some function given the Lagrange multiplier. If there are N possible states of the world, one may be tempted to apply Carathéodory's theorem (Rockafellar, 1970) and conclude from this that the optimal information policy uses at most N posteriors. After all, the solution to the problem would correspond to the concavification of the function whose expectation over posteriors determines the Lagrangian.

In an inspiring contribution, Le Treust and Tomala (2019) show that the above reasoning is flawed when the information designer faces one inequality constraint. At the heart of their result is the observation that the Lagrange multiplier is also part of the solution to the optimization problem. Indeed, they show that the solution corresponds to concavifying a function of $N + 1$ variables: the first N correspond to a belief and the last corresponds to the inequality constraint. It follows then that the optimal policy may involve $N + 1$ posteriors. The authors also show that the Lagrangian approach is valid for their problem.

Many information design problems involve multiple inequality and equality constraints. For instance, in Doval and Skreta (2020), the designer designs both an allocation rule and an information structure; both have to satisfy the agent's participation and incentive compatibility constraints. As another example, consider the problem of persuading a privately informed receiver in Kolotilin et al. (2017): the designer designs a *menu* of information structures, which has to satisfy the agent's incentive compatibility constraints. Finally, consider the problem of a designer who designs a menu of offers for a privately informed agent, but is limited in how much information the allocation can reveal because of privacy concerns, as in Eilat et al. (2021).

We extend the results in Le Treust and Tomala (2019) to the case of multiple equality and inequality constraints. Theorem 3.1 shows that the information design problem subject to constraints is equivalent to the solution of a *standard*, but higher dimensional, Bayesian persuasion problem, where the dimensions represent the number of states together with the number of constraints. We use this to derive an upper bound on the number of posteriors used in an optimal experiment (Corollary 3.1): An optimal experiment induces at most $N + K$ posteriors, where N is the number of states and K the number of constraints. Corollary 3.2 then shows that this upper bound can be refined whenever a constraint does not bind. We also show that the Lagrangian approach is valid. Indeed, Proposition 3.1 shows that there exists a Lagrange multiplier such that the solution to the constrained information design problem follows from the concavification of the Lagrangian at that multiplier. Example 3.1 illustrates how Proposition 3.1 can be used to solve constrained information design problems, even without solving for the optimal multiplier.

Section 4 shows how Theorem 3.1 can be leveraged to obtain useful results in two important settings:

Section 4.1 considers the problem of mechanism design with limited commitment. In this application, the set of states of the world corresponds to the agent's private information. Doval and Skreta (2020) show that it is without loss of generality to consider mechanisms in which the designer designs both an information structure and an allocation. Furthermore, the mechanism must satisfy the agent's participation and incentive compatibility constraints. Theorem 3.1 implies that it is without loss of generality to focus on mechanisms that induce information structures with finite support and provides an upper bound on the number of posteriors induced by the mechanism. Proposition 4.1 shows that when the agent's payoff satisfies a version of single-crossing for lotteries (Bester and Strausz (2007);

3

Celik (2015); Kartik et al. (2017)) the upper bound implied by Theorem 3.1 can be further reduced (Corollary 4.1). The assumption of transferable utility provides another way in which this bound can be reduced: When the optimal mechanism can be obtained by maximizing the virtual surplus, the information structure associated to the optimal mechanism uses at most as many posteriors as the number of states of the world (Proposition 4.2).

Section 4.2 considers the problem of persuading a privately informed receiver (Kolotilin et al. (2017); Guo and Shmaya (2019); Candogan and Strack (2021)). Here the information designer designs a menu of information structures subject to the incentive compatibility constraints of the agent. We show how the designer's problem can be separated into different problems, one for each type of the receiver.[1] We use this decomposition and Theorem 3.1 to derive an upper bound on the number of posteriors employed in an optimal experiment. Since we make no assumption on the cardinality of the set of receiver actions, the bounds in Proposition 4.4 are the most useful when the set of actions is larger than the set of types.

Related Literature: The paper builds and expands on the results in Le Treust and Tomala (2019). Given the prevalence of *constrained* information design this simple extension is bound to be useful to other researchers. Furthermore, we provide novel applications where these results greatly simplify the analysis.

In the context of mechanism design with limited commitment, Bester and Strausz (2007) provide analogues of Propositions 4.2 and 4.3, using tools of infinite dimensional linear programming. While this allows them to conclude that mechanisms in their paper use finitely many output messages, Bester and Strausz (2007) do not provide a characterization of the set of output messages. Therefore, in order to characterize an optimal mechanism, the analyst still has to identify the optimal message space. Instead, we leverage the characterization in Doval and Skreta (2020), which allows us to equate the message space of the mechanism to the set of beliefs the designer holds about the agent's type. We then use the results in Section 3, which are based on the tools of convex analysis employed in the information design literature, to conclude that the principal's optimal mechanism employs finitely many posteriors.[2]

---

[1]Candogan and Strack (2021) make a similar observation in their problem.

[2]Salamanca (2021) studies communication equilibria in sender-receiver games. Salamanca (2021) develops Lagrangian techniques to study the sender optimal communication equilibrium. This allows him to derive an upper bound like the one in Corollary 3.1 in the context of his model. As it will be clear from the analysis in Section 3, Lagrangian techniques limit the use of the stan-

Since the first circulation of our draft (see, Doval and Skreta (2018)), there has been renewed interest in providing tools to solve constrained (information) design problems. Dworczak and Kolotilin (2019) apply our results in their study on duality in Bayesian persuasion. Kang (2020) provides a set of tools complementary to the one in this paper, by combining results in Bauer (1958) and Szapiel (1975). Babichenko et al. (2020) derive the results in Section 3 using infinite dimensional linear programming tools as Bester and Strausz (2007). They complement the results in our paper by providing computational complexity results. Azrieli (2021) illustrates the difference between unconstrained rational inattention problems and those subject to a capacity constraint.

2   Setting

Consider the following problem.[3] Let Ω be a finite set of states, $\Omega = \{\omega_1, \ldots, \omega_N\}$. Let $f, g_1, \ldots, g_r, g_{r+1}, \ldots, g_K : \Delta(\Omega) \mapsto \mathbb{R} \cup \{-\infty\}$ be a tuple of functions defined on $\Delta(\Omega)$. For $\mu \in \Delta(\Omega)$ and $\gamma_1, \ldots, \gamma_K \in \mathbb{R}$, consider

$$\text{cav}_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K) := \sup \left\{ \sum_m \lambda_m f(\mu_m) : \begin{array}{l} \sum_m \lambda_m \mu_m = \mu, \\ \sum_m \lambda_m g_l(\mu_m) \geq \gamma_l, l \in \{1, \ldots, r\}, \\ \sum_m \lambda_m g_l(\mu_m) = \gamma_l, l \in \{r+1, \ldots, K\} \end{array} \right\}.$$
(OPT)

Le Treust and Tomala (2019) consider the above problem for $r = 1$ and no equality constraints.

3   Main results

The main result of this section, Theorem 3.1 relates the solution to OPT to the concavification of the function $f^{g_1,\ldots,g_K} : \Delta(\Omega) \times \mathbb{R}^K \mapsto \mathbb{R} \cup \{-\infty\}$ defined as follows:

$$f^{g_1,\ldots,g_K}(\mu, \gamma_1, \ldots, \gamma_K) = \begin{cases} f(\mu) & \text{if } \gamma_i \leq g_i(\mu), i \in \{1, \ldots, r\} \wedge \gamma_i = g_i(\mu), i \in \{r+1, \ldots, K\} \\ -\infty & \text{otherwise} \end{cases}.$$

---

dard toolkit in information design since the Lagrange multiplier is endogenous.

[3]To make the comparison with Le Treust and Tomala (2019) simple, we follow their notation as much as possible. However, while they present their results for any convex set X, to make the presentation closer to information design, we let X be the space of beliefs over the set of states Ω.

Theorem 3.1. *For each $(\mu, \gamma_1, \ldots, \gamma_K) \in \Delta(\Omega) \times \mathbb{R}^K$,*

$$cav_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K) = cav f^{g_1,\ldots,g_K}(\mu, \gamma_1, \ldots, \gamma_K).$$

Proof. The function cav $f^{g_1,\ldots,g_K}(\mu, \gamma_1, \ldots, \gamma_K)$ is given by the following program:

$$\sup \sum_m \lambda_m f^{g_1,\ldots,g_K}(\mu_m, \gamma_{1,m}, \ldots, \gamma_{K,m})$$

$$\text{s.t.} \begin{cases} \sum_m \lambda_m \mu_m = \mu \\ \sum_m \lambda_m \gamma_{l,m} = \gamma_l, l \in \{1, \ldots, K\} \end{cases}.$$

Take a family $(\lambda_m, \mu_m, \gamma_{1,m}, \ldots, \gamma_{K,m})$ that is feasible for this program. Then, for $k \leq r$, we have $\sum_m \lambda_m g_k(\mu_m) \geq \sum_m \lambda_m \gamma_{k,m} = \gamma_k$, and for $k \in \{r+1, \ldots, K\}$, we have $\sum_m \lambda_m g_k(\mu_m) = \sum_m \lambda_m \gamma_{k,m} = \gamma_k$. Thus, $(\lambda_m, \mu_m, \gamma_{1,m}, \ldots, \gamma_{K,m})$ is feasible for cav $_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K)$. Thus, $cav_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K) \geq cav f^{g_1,\ldots,g_K}(\mu, \gamma_1, \ldots, \gamma_K)$.

On the other hand, let (λ_m, μ_m) such that $\sum_m \lambda_m \mu_m = \mu$ and $\sum \lambda_m g_k(\mu_m) \geq \gamma_k, k \leq r$ and $\sum_m \lambda_m g_k(\mu_m) = \gamma_k, k \in \{r+1, \ldots, K\}$. For each k, let $\overline{\gamma}_k = \sum_m \lambda_m g_k(x_m)$ and for each m, let $\gamma_{k,m} = g_k(\mu_m) + \gamma_k - \overline{\gamma}_k$. Then, $\sum_m \lambda_m \gamma_{k,m} = \gamma_k$. Because $\overline{\gamma}_k \geq \gamma_k$ (with equality for $k \in \{r+1, \ldots, K\}$), $g_k(\mu_m) \geq \gamma_{k,m}$ for $k \in \{1, \ldots, r\}$ and $g_k(\mu_m) = \gamma_{k,m}$ for $k \in \{r+1, \ldots, K\}$. Thus, $(\lambda_m, \mu_m, \gamma_{1,m}, \ldots, \gamma_{K,m})$ is feasible for cav $f^{g_1,\ldots,g_K}$. Hence, $cav_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K) \leq cav f^{g_1,\ldots,g_K}(\mu, \gamma_1, \ldots, \gamma_K)$. $\square$

While Theorem 3.1 is written in terms of the solution to OPT, the proof actually shows that the convex hull of the graph of $f^{g_1,\ldots,g_K}$ coincides with the convex hull of the graph of the function $h = (f, g_1, \ldots, g_K)$ over the set

$$C = \{\mu \in \Delta(\Omega) : (\forall i \leq r)\gamma_i \leq g_i(\mu) \wedge (\forall i \in \{r+1, \ldots, K\})g_i(\mu) = \gamma_i\}.$$

Indeed, for the case in which $r = 1$ and no equality constraints, Boleslavsky and Kim (2018) use the convex hull of the graph of h to derive results in their model.

Theorem 3.1 together with Carathéodory's theorem (see, e.g., Rockafellar (1970)) implies the following:

Corollary 3.1. *The solution to problem* (OPT) *uses at most $N + K$ posteriors.*

Furthermore, we can relate the upper bound on the number of posteriors at an optimal solution to the number of binding constraints:

Corollary 3.2. *Suppose that in problem* (OPT), *only $M < r$ inequality constraints bind. Then the solution to problem* (OPT) *uses at most $N + M + K - r$ posteriors.*

Proof. Suppose that in the solution to program $\text{cav}_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K)$, $M \leq r$ constraints bind and $r - M$ are slack. Then

$$\text{cav}_{g_1,\ldots,g_K} f(\mu, \gamma_1, \ldots, \gamma_K) = \text{cav}_{g_M, g_{r+1}, \ldots, g_K} f(\mu, \gamma_M, \gamma_{r+1}, \ldots, \gamma_K),$$

where γ_M is the projection of vector $(\gamma_1, \ldots, \gamma_r)$ on the binding constraints and g_M is the projection of vector $(g_1, \ldots, g_r)$ on the set M. It follows from Theorem 3.1 that $\text{cav}_{g_M, g_{r+1}, \ldots, g_K} f(\mu, \gamma_M, \gamma_{r+1}, \ldots, \gamma_K) = \text{cav } f^{g_M, g_{r+1}, \ldots, g_K}(\mu, \gamma_M, \ldots, \gamma_{r+1}, \ldots, \gamma_K)$. Thus the solution to (OPT) uses at most $N + M + K - r$ beliefs. $\qquad\square$

3.1 Validating the Lagrangian approach

Proposition 3.1.

$$cav_{g_1,\ldots,g_K} f(\mu, \gamma) = \inf \left\{ cav \left(f + \sum_{k=1}^{K} t_k g_k \right)(\mu) - \sum_{k=1}^{K} t_k \gamma_k : t \in \mathbb{R}_+^r \times \mathbb{R}^{K-r} \right\}$$

(3.1)

The proof is in Appendix A.

Proposition 3.1 states that a multiplier t^* exists such that the solution to OPT corresponds to the concavification of the Lagrangian at t^*. An alternative application of Proposition 3.1 is the following. As mentioned in Section 1, for each t, the concavification of the Lagrangian can be achieved by considering convex combinations of at most N posteriors. If one can show that, for each t, convex combinations involving more points do not achieve the value of the concavified Lagrangian, then one can conclude that the solutions for (OPT) involve at most N posterior beliefs. Example 3.1 illustrates this point:

Example 3.1. *Consider the following version of the prosecutor example in Kamenica and Gentzkow (2011). Everything is as in Kamenica and Gentzkow (2011) except that the judge has access to outside information. However, the judge has limited time, so that the judge can either listen to the prosecution or their own source of information. We model this as the prosecutor facing a constraint: the judge has to receive the payoff they can achieve by using their own source of information.*

Formally, let the set of states $\Omega = \{0, 1\}$ and let the set of actions for the judge be $A = \{0, 1\}$. The payoffs are $u(a, \omega) = \mathbb{1}[a = \omega]$ and $v(a, \omega) = \mathbb{1}[a = 1]$ for the judge and the prosecutor, respectively. Let $\mu_0 \in [0, 1]$ denote the prior probability that $\omega = \omega_1$. Assume that $\mu_0 < 1/2$.

The judge has access to another experiment, $\tau^J \in \Delta\Delta\Omega$. Let $a^*(\mu)$ denote the judge's optimal action choice when the posterior belief is μ and $\hat{v}(\mu) = \sum_{\omega \in \Omega} \mu(\omega)v(a^*(\mu), \omega)$. Then, the prosecutor's optimal payoff follows from solving the following problem:

$$\max_{\tau \in \Delta\Delta\Omega: \mathbb{E}_\tau \mu = \mu_0} \mathbb{E}_\tau \hat{v}(\mu)$$

$$s.t. \mathbb{E}_\tau \left[\sum_{\omega \in \Omega} u(a^*(\mu), \omega) \right] \geq \mathbb{E}_{\tau^J} \left[\sum_{\omega \in \Omega} u(a^*(\mu), \omega) \right].$$

Since the prosecutor can design any experiment, the prosecutor can always replicate the judge's source of information. It is therefore without loss of generality to assume that the judge uses the prosecutor's experiment, while the prosecutor offers an experiment to the judge that is at least as valuable to the judge as their own source of information.

Proposition 3.1 shows that there exists $t^* \geq 0$ such that the optimal experiment solves:

$$\max_{\tau \in \Delta\Delta(\Omega)} \mathbb{E}_\tau [\hat{v}(\mu) + t^* \sum_\omega \mu(\omega)u(a^*(\mu), \omega)]$$

As we illustrate graphically, for all $t \geq 0$, the concavification of the function $\hat{v}(\mu) + t \sum_\omega \mu(\omega)u(a^*(\mu), \omega)$ is attained by experiments that use exactly 2 posteriors. It then follows that in this constrained Bayesian persuasion problem, there is no need for a third posterior.
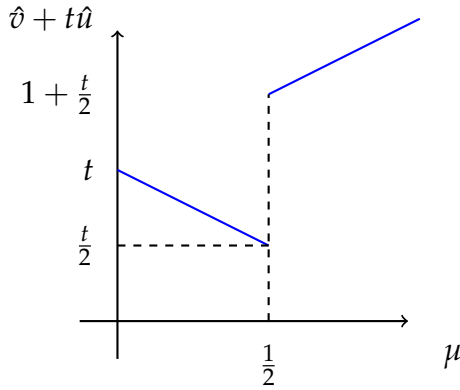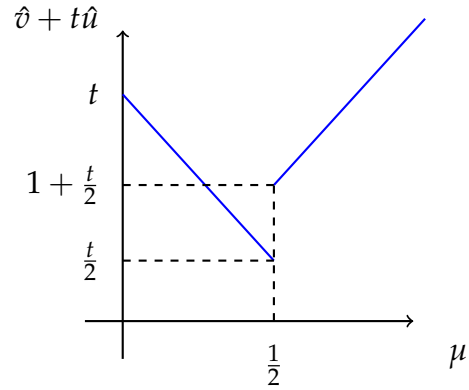

Figure (a) $t \leq 2$ Figure (b) $t > 2$

Figure 1: The Lagrangian for $t \leq 2$ (left) and $t > 2$ (right)

8

4 Applications

4.1 Mechanism design with limited commitment

Section 4.1 showcases how Theorem 3.1 can be leveraged to inform the characterization of optimal mechanisms under limited commitment. To keep the presentation simple and to facilitate the comparison with other work in the literature, we present the results in the context of a model based on Bester and Strausz (2007).

Consider the problem of a principal who interacts with a privately informed agent, who knows the state of the world. Let $\mu_0 \in \Delta(\Omega)$ denote the principal's prior belief about the state of the world. The interaction lasts for two periods, $t \in \{1, 2\}$. In each period t, as a result of the interaction, an allocation $y_t \in Y_t$ is determined, where Y_t is the set of allocations in period t. There is a correspondence $\mathcal{Y} : Y_1 \rightrightarrows Y_2$ that describes the set of feasible period 2 allocations as a function of the allocation in period 1. Let $v(y_1, y_2, \omega)$ and $u(y_1, y_2, \omega)$ denote the principal and the agent's payoff, respectively, when the allocation is (y_1, y_2) and the state of the world is ω. We assume there exists an allocation (y_1^*, y_2^*) such that $u(y_1^*, y_2^*, \omega) = 0$ for all $\omega \in \Omega$. This allocation plays the role of the outside option in what follows.[4]

The interaction between the principal and the agent proceeds as follows:

In period 1, after the agent observes the state of the world, the principal offers the agent a mechanism, **M**. A mechanism **M** consists of a set of input messages M, a set of output messages S, and a device $\varphi : M \mapsto \Delta(S \times Y_1)$, which associates to each input message $m \in M$ a distribution over output messages and allocations.

After observing the mechanism, the agent decides whether to accept or reject. If she rejects the mechanism, an allocation $(y_1^*, y_2^*) \in Y_1 \times Y_2$ is implemented. If instead she accepts the mechanism, she *privately* submits an input message to the mechanism. This message determines the distribution $\varphi(\cdot|m)$ from which the out-

---

[4]Throughout, we make the following technical assumptions. First, the set of allocations Y_1, Y_2 are compact Polish spaces, we endow them with their Borel σ-algebra. Second, endowing product sets with their product σ-algebra, we assume that the principal and the agent's utility functions are bounded measurable functions. We assume that the principal's utility is continuous in y_2 for each (y_1, ω). Third, the correspondence $\mathcal{Y}$ is measurable and for each $y_1 \in Y_1$, $\mathcal{Y}(y_1)$ is compact. Fourth, for a Polish space X, we let $\Delta(X)$ denote the set of Borel probability measures over X, endowed with the weak* topology. Thus, $\Delta(X)$ is also Polish (Aliprantis and Border (2006)). Finally, for any two measurable spaces X and Y, a mapping $\zeta : X \mapsto \Delta(Y)$ is a *transition probability* from X to Y if, for any measurable $C \subseteq Y$, $\zeta(C|x) \equiv \zeta(x)(C)$ is a measurable real valued function of $x \in X$.

put message and the allocation is drawn. In period 2, after observing the output message and the allocation, the principal chooses an allocation $y_2 \in Y_2$.

Our objective is to characterize the optimal mechanism for the principal under the solution concept of Perfect Bayesian equilibrium. In particular, the principal's choice of the allocation in period 2 must be sequentially rational. The allocation $y_2 \in Y_2$ then captures in reduced form the principal's limited commitment.

Theorem 1 in Doval and Skreta (2020) implies that the principal's optimal Perfect Bayesian equilibrium can be characterized as the solution to a constrained optimization program (see ($\mathcal{P}$) below). Indeed, Theorem 1 in Doval and Skreta (2020) implies that it is without loss of generality to consider mechanisms, $\mathbf{M}$, such that the following hold. First, the set of input and output messages are the set of states and posterior beliefs, respectively, i.e., $M = \Omega$ and $S = \Delta(\Omega)$. Second, the device can be decomposed into two transition probabilities $\beta : \Theta \mapsto \Delta\Delta(\Omega)$ and $\alpha : \Delta(\Omega) \mapsto \Delta(Y_1)$. Third, it is optimal for the agent to participate and truthfully report the state of the world. Finally, when the mechanism outputs a belief μ, this is the belief that would result from Bayes' rule when the principal observes output message μ, and the agent participates and truthfully reports her type. When β has finite support, this is equivalent to requiring that

$$\mu(\omega) = \frac{\mu_0(\omega)\beta(\mu|\omega)}{\sum_{\omega' \in \Omega} \mu_0(\omega')\beta(\mu|\omega')}.$$

For each $y_1 \in Y_1$, let

$$y_2(y_1, \mu) \in Y_2(y_1, \mu) \equiv \arg \max_{y_2 \in \mathcal{Y}(y_1)} \sum_{\omega \in \Omega} \mu(\omega) v(y_1, y_2, \omega),$$

denote a solution to the principal's problem in period 2 when his belief about the state of the world is μ. Given the technical assumptions listed in footnote 4, the above problem is well-defined. In a slight abuse of notation, let Y_2 denote the set of all selections from the principal's best response correspondence $Y_2(\cdot, \cdot)$.

We can write the principal's problem as follows:[5]

$$\max_{\beta:\Theta\mapsto\Delta\Delta(\Omega),\alpha:\Delta(\Omega)\mapsto\Delta(Y_1),y_2\in Y_2} \sum_{\omega\in\Omega} \mu_0(\omega)\mathbb{E}_{\beta(\cdot|\omega)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[v(y_1,y_2(y_1,\mu),\omega)\right]\right] \qquad (\mathcal{P})$$

$$\text{s.t.}\begin{cases} (\forall\omega\in\Omega) & \mathbb{E}_{\beta(\cdot|\omega)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[u(y_1,y_2(y_1,\mu),\omega)\right]\right]\geq 0 \\ (\forall\omega\in\Omega)(\forall\omega'\neq\omega) & \mathbb{E}_{\beta(\cdot|\omega)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[u(y_1,y_2(y_1,\mu),\omega)\right]\right]\geq \mathbb{E}_{\beta(\cdot|\omega')}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[u(y_1,y_2(y_1,\mu),\omega)\right]\right] \end{cases}$$

Furthermore, the transition probability β must satisfy that for all measurable subsets $\tilde{U}$ of $\Delta(\Omega)$ and all subsets $\tilde{\Omega}$ of Ω,

$$\sum_{\omega'\in\tilde{\Omega}} \beta(\tilde{U}|\omega')\mu_0(\omega') = \sum_{\omega\in\Omega}\int \mu(\tilde{\Omega})\beta(d\mu|\omega)\mu_0(\omega)$$

To show how Theorem 3.1 can inform the solution to $\mathcal{P}$ we first show how to write the principal's optimization problem as one in which he chooses a Bayes' plausible distribution over posteriors and an allocation rule $\alpha : \Delta(\Omega) \mapsto \Delta(Y_1)$. For any measurable subset $\tilde{U}$ of $\Delta(\Omega)$, and for any subset $\tilde{\Omega}$ of Ω, let $\mathbb{P} \in \Delta(\Omega \times \Delta(\Omega))$ denote the following measure:

$$\mathbb{P}(\tilde{\Omega}\times\tilde{U}) = \sum_{\omega\in\tilde{\Omega}} \beta(\tilde{U}|\omega)\mu_0(\omega).$$

The disintegration theorem (see Crauel (2002)) implies that there exists $\tau \in \Delta\Delta(\Omega)$ such that

$$\mathbb{P}(\tilde{\Omega}\times\tilde{U}) = \int_{\tilde{U}}\left(\sum_{\omega\in\tilde{\Omega}}\mu(\omega)\right)\tau(d\mu).$$

It follows that for all $\omega \in \Omega$ and all measurable subsets $\tilde{U}$ of $\Delta(\Omega)$, we have

$$\beta(\tilde{U}|\omega)\mu_0(\omega) = \int_{\tilde{U}}\mu(\omega)\tau(d\mu).$$

---

[5]Bester and Strausz (2007) analyze a version of the problem ($\mathcal{P}$) with the following differences. First, instead of letting the set of output messages be the set of beliefs, they leave the set S unspecified. Therefore, they only analyze the principal's optimal mechanism within those that use signals in S. Second, because the set S is unspecified, their program has an additional constraint: the choice of y_2 has to be optimal given the realization of s and the period-1 allocation. Finally, they do not allow for randomized allocations. Doval and Skreta (2020) show that this may be with loss of generality (Strausz (2003) also shows the importance of allowing for randomization for the standard version of the revelation principle to hold).

Therefore, we can write the agent's payoff when the state of the world is ω and she reports ω' as follows:

$$\mathbb{E}_{\beta(\cdot|\omega')}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[u(y_1,y_2(y_1,\mu),\omega)\right]\right] = \mathbb{E}_\tau\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\frac{\mu(\omega')}{\mu_0(\omega')}u(y_1,y_2(y_1,\mu),\omega)\right]\right].$$

It follows that the principal's problem can be written similar to the problem in OPT:

$$\max_{\tau\in\Delta\Delta(\Omega),\alpha:\Delta(\Omega)\mapsto\Delta(Y_1),y_2\in Y_2} \mathbb{E}_\tau\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\sum_{\omega\in\Omega}\mu(\omega)v(y_1,y_2(y_1,\mu),\omega)\right]\right] \tag{4.1}$$

$$s.t. \begin{cases} & \mathbb{E}_\tau[\mu] = \mu_0 \\ (\forall\omega\in\Omega) & \mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\frac{\mu(\omega)}{\mu_0(\omega)}u(y_1,y_2(y_1,\mu),\omega)\right]\right] \geq 0 \\ (\forall\omega\in\Omega)(\forall\omega'\neq\omega) & \mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\left(\frac{\mu(\omega)}{\mu_0(\omega)}-\frac{\mu(\omega')}{\mu_0(\omega')}\right)u(y_1,y_2(y_1,\mu),\omega)\right]\right] \geq 0 \end{cases}$$

Observation 1. *Corollary 3.1 implies that it is without loss of generality to focus on mechanisms where for each $\omega\in\Omega$, the support of the communication device is finite.*[6]

When the principal has full commitment, the revelation principle implies that it is without loss of generality to focus on *finite* mechanisms, where one associates with each type a (possibly randomized) allocation. Observation 1 affords a similar simplification for the case of limited commitment, by ensuring that we associate to each agent type an experiment with finite support.

The resulting program in Equation 4.1 allows us to highlight the connection between mechanism design with limited commitment and the literature on information design. After all, the designer can be thought of as a sender who designs an information structure for a receiver, who happens to be his period-2 self. However, there are differences. In our setting, the period-1 principal (the sender in Kamenica and Gentzkow (2011)) also takes an action for each posterior he induces. In addition, the first-period principal cannot implement any Bayes' plausible distribution over posteriors, but only those that satisfy the incentive compatibility and participation constraints of the agent.

In what follows, we focus on the case in which the agent's preferences satisfy a version of *increasing differences*, which takes into account that the agent faces lotteries over allocations. To introduce the condition, assume that the states are ordered from low to high, $\omega_1 < \cdots < \omega_N$.

---

[6] The formal argument follows the same lines as the proof of Corollary 4.1, so we omit it.

Definition 4.1. *[Bester and Strausz (2007); Celik (2015); Kartik et al. (2017)] The family $\{u(\cdot, \omega) : \omega \in \Omega\}$ satisfies* monotonic expectational differences *if for any two distributions $P, Q \in \Delta(Y_1 \times Y_2)$, $\int u(\cdot, \omega_i) d(P - Q)$ is monotone in i.*

Kartik et al. (2017) show that u satisfies monotonic expectational differences if, and only if, it takes the form, $u(y_1, y_2, \omega_i) = b(\omega_i) h_1(y_1, y_2) + h_2(y_1, y_2) + c(\omega_i)$, where h_1, h_2 are finitely integrable and b is monotonic. Without loss of generality, assume that b is weakly increasing, so that ω_1 is the agent's "lowest type."[7]

Like increasing differences in mechanism design with commitment, monotonic expectational differences imply that the solutions to $\mathcal{P}$ coincide with the solutions to a much simpler program, which imposes only a subset of the incentive compatibility constraints:

Proposition 4.1. *If $\{u(\cdot, \omega) : \omega \in \Omega\}$ satisfies monotonic expectational differences, then to characterize the solution to $\mathcal{P}$, it suffices to guarantee that the following hold:*

1. *The agent's participation constraint when the state is ω_1,*

2. *Adjacent incentive compatibility constraints are satisfied.*

See Appendix B for a proof. We then obtain the following corollary:

Corollary 4.1. *Any solution to $(\mathcal{P})$ utilizes at most $3N - 1$ posteriors.*

Transferable utility: Transferable utility is a common assumption in mechanism design. In what follows, we show how this assumption further simplifies the characterization of an optimal mechanism by reducing in some instances the number of posteriors that the mechanism employs. Therefore, we make the following assumptions in the remainder of this section. First, the set of period 1 allocations is given by $Y_1 = Y_1' \times \mathbb{R}_+$, where the second coordinate denotes a payment from the agent to the principal. We denote an element of Y_1 by $y_1 = (y_1', p)$. Second, we assume that $\mathcal{Y}((y_1', p)) = \mathcal{Y}(y_1')$. Finally, we assume that the agent and the principal's payoffs can be written as follows:

$$v(y_1, y_2, \omega) = \tilde{v}(y_1', y_2, \omega) + p$$
$$u(y_1, y_2, \omega) = \tilde{u}(y_1', y_2, \omega) - p.$$

---

[7]Moreover, we assume that $h_1(y_1^*, y_2^*) = \min_{(y_1, y_2): y_2 \in \mathcal{Y}(y_1)} h_1(y_1, y_2)$. This allows us to conclude that whenever the lowest type, ω_1 participates of the mechanism, then all types participate of the mechanism.

Transferable utility implies that focusing on mechanisms that do not randomize on transfers is without loss of generality. Hereafter, we replace p with its expectation under the mechanism when the posterior is μ, $p(\mu)$.

The following result follows from Proposition 4.1:

Corollary 4.2. *Suppose that the family $\{u(\cdot, \omega) : \omega \in \Omega\}$ satisfies monotonic expectational differences and utility is transferable. Then, the participation constraint is binding for ω_1.*

Under the assumptions of monotonic expectational differences and transferable utility, we could further simplify $\mathcal{P}$ by showing that *downward-looking* incentive constraints always bind at the optimum. This then justifies the study of the so-called relaxed program:

$$\max_{\tau \in \Delta\Delta(\Omega), \alpha: \Delta(\Omega) \mapsto \Delta(Y_1' \times \mathbb{R}), y_2 \in Y_2} \mathbb{E}_\tau \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[\sum_{\omega \in \Omega} \mu(\omega) \tilde{v}(y_1', y_2(y_1', \mu), \omega) + p(\mu) \right] \right] \tag{$\mathcal{R}$}$$

$$\text{s.t.} \begin{cases} \mathbb{E}_\tau[\mu] = \mu_0 \\ \mathbb{E}_\tau \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[\frac{\mu(\omega_1)}{\mu_0(\omega_1)} \left(\tilde{u}(y_1', y_2(y_1', \mu), \omega) - p(\mu) \right) \right] \right] = 0 \\ (\forall i \in \{2, \ldots, N\}) \quad \mathbb{E}_\tau \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[\left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i-1})}{\mu_0(\omega_{i-1})} \right) \left(\tilde{u}(y_1', y_2(y_1', \mu), \omega_i) - p(\mu) \right) \right] \right] = 0 \end{cases}$$

which is obtained by dropping the *monotonicity constraints*:[8]

$$\mathbb{E}_\tau \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[\left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i-1})}{\mu_0(\omega_{i-1})} \right) \left(\tilde{u}(y_1', y_2(y_1', \mu), \omega_i) - \tilde{u}(y_1', y_2(y_1', \mu), \omega_{i-1}) \right) \right] \right] \geq 0, \tag{M}$$

for each $i \in \{2, \ldots, N\}$. In mechanism design with commitment, it suffices to check that the solution to the relaxed program satisfies the monotonicity constraints, (M), to show it is the solution to ($\mathcal{P}$) (see the discussion in footnote 8).

However, in mechanism design with limited commitment, the solution to the relaxed program is not necessarily a solution to ($\mathcal{P}$) even if it satisfies the monotonicity constraints, when the type space is finite and there are three or more types. Whereas in the relaxed program the binding downward-looking incentive constraints together with ω_1's participation constraint impose N restrictions on the

---

[8]The constraints in equation (M) are obtained from combining the restriction that ω_i does not want to report ω_{i-1} and ω_{i-1} does not want to report ω_i. Under Definition 4.1, the binding downward-looking incentive constraints together with the monotonicity constraints imply the local constraints in Proposition 4.1.

transfers $\{p(\mu) : \mu \in \Delta(\Omega)\}$, the solution to the relaxed program might use less than N posteriors. Therefore, finding transfers $p(\mu)$ that satisfy all constraints may not possible.[9] Alternatively, not all downward-looking constraints may bind in the optimal mechanism.

Fortunately, the above is not an issue when there are two types or a continuum of types. In both cases, it is possible to show that downward looking constraints bind (see Doval and Skreta (2020)). Because most of the literature focuses on one of these cases, and because the relaxed program provides a useful benchmark, the rest of this section studies its properties.

We can use the binding constraints to substitute the transfers out of the principal's program and obtain the following:

Proposition 4.2. *The solution to the relaxed program uses at most N posteriors.*

Then, if the solution to the relaxed program satisfies the monotonicity constraints *and* it is possible to find transfers $(p(\mu))$ that satisfy the downward looking binding incentive constraints, we have found a solution to the principal's problem, ($\mathcal{P}$). The proof of Proposition 4.2 is in Appendix B. It follows from two observations. First, once we substitute the transfers out of the principal's payoff, we are left with an expression that only depends on the distribution over posteriors induced by the mechanism and the portion of the allocation rule that corresponds to Y_1'. Second, since we are ignoring the monotonicity constraints, one can solve for the optimal α by pointwise maximization. We are then left with a function that depends only on the distribution over posterior beliefs, that is, a standard Bayesian persuasion problem. The proof of Proposition 4.2 also suggests how the principal chooses y_2 when he is indifferent: ties are broken in favor of maximizing the virtual surplus.

In many instances, however, the solution to ($\mathcal{R}$) will fail to satisfy the monotonicity constraints, (M). As we show next, adding as many posteriors as binding monotonicity constraints at the optimum may be necessary:

Proposition 4.3. *Consider the program obtained by adding the monotonicity constraints (M) to the relaxed program ($\mathcal{R}$). The solution to the new program uses at most $N + K$ posteriors, where K is the number of binding constraints at the optimum.*

The proof of Proposition 4.3 follows immediately from Corollary 3.1 and Corol-

---

[9]This is never an issue in mechanism design with commitment: Without loss of generality, we can always have one transfer for each type.

4.2 Persuasion of a privately informed receiver

Consider an information designer who controls the release of information about a state of the world $\omega \in \Omega$ and faces a privately informed agent. Let Θ denote a finite set of agent types and let q_θ denote the probability that the agent is of type θ. Let $M = |\Theta|$. As in Section 4.1, let μ_0 denote the prior belief over Ω. We assume that the state of the world ω and the agent's type θ are independently distributed.

While the designer controls the release of information about the state of the world, the agent is the one who ultimately takes actions. That is, after observing the information released by the designer, the agent selects an action a from a compact set A. Let $u(a, \theta, \omega)$ and $v(a, \theta, \omega)$ denote the agent and the designer's payoffs, respectively, when the agent takes action a, the agent's type is θ, and the state of the world is ω. We assume that both functions are continuous in a for each $(\theta, \omega) \in \Theta \times \Omega$.

For each $\theta \in \Theta$, let

$$a^*(\mu, \theta) \in \arg\max_{a \in A} \sum_{\omega \in \Omega} \mu(\omega) u(a, \theta, \omega)$$

denote the agent's optimal action choice when her type is θ and her belief about ω is given by μ. Let

$$U(\mu, \theta) = \sum_{\omega \in \Omega} \mu(\omega) u(a^*(\mu, \theta), \theta, \omega),$$

denote the agent's optimal payoff when her type is θ and her belief about ω is given by μ. Whenever it is necessary, we assume that the agent breaks ties in favor of the designer.

The information designer designs a menu of experiments $\tau : \Theta \mapsto \Delta(\Delta(\Omega))$ to solve:[10]

$$\max_{\tau : \Theta \mapsto \Delta\Delta(\Omega)} \sum_{\theta \in \Theta} q_\theta \sum_{\mu \in \Delta(\Omega)} \tau(\mu, \theta) \sum_{\omega \in \Omega} \mu(\omega) v(a^*(\mu, \theta), \theta, \omega) \tag{4.2}$$

$$\text{s.t.} \begin{cases} (\forall \theta \in \Theta) \mathbb{E}_{\tau(\theta, \cdot)}[\mu] = \mu_0 \\ \mathbb{E}_{\tau(\theta, \cdot)}[U(\mu, \theta)] \geq \mathbb{E}_{\tau(\theta', \cdot)}[U(\mu, \theta)] \end{cases}.$$

---

[10] As we show in Lemma C.1 in Appendix C, standard arguments imply that it is without loss of generality to focus on experiments where the set of signals is the space of beliefs over Ω.

That is, the designer chooses an experiment to maximize his payoff subject to two constraints. First, for each type $\theta \in \Theta$, the experiment must induce a Bayes' plausible distribution over posteriors. Second, each type $\theta \in \Theta$ must prefer their experiment over the one offered to types θ' other than θ.

Proposition 4.4 illustrates how Theorem 3.1 can be used to simplify the solution to the problem in Equation 4.2:

Proposition 4.4. *The designer's optimal payoff can be found from the solution to*

$$\max_{\{u_\theta\}_{\theta \in \Theta}} \max_{\{\tau:\Theta \mapsto \Delta(\Delta(\Omega)):\mathbb{E}_{\tau(\cdot,\theta)}[\mu]=\mu_0\}} \sum_{\theta \in \Theta} q_\theta \sum_{\mu \in \Delta(\Omega)} \tau(\mu,\theta) \sum_{\omega \in \Omega} \mu(\omega)v(a^*(\mu,\theta),\theta,\omega)$$

$$\tag{4.3}$$

$$s.t. \begin{cases} (\forall \theta \in \Theta) & \mathbb{E}_{\tau(\theta,\cdot)}[U(\cdot,\theta)] \geq u_\theta \\ (\forall \theta \in \Theta)(\forall \theta' \neq \theta) & u_{\theta'} \geq \mathbb{E}_{\tau(\theta,\cdot)}[U(\mu,\theta')] \end{cases}$$

It follows that for each θ, the experiment induces at most $N + M$ posteriors.

The result in Proposition 4.4 affords two simplifications for the designer's problem. First, while the incentive compatibility constraints in Equation 4.2 impose conditions across the experiments for different types, the optimization problem in Equation 4.3 decouples the problem of designing the experiment for θ from the problem of designing the experiment for θ'. Second, Proposition 4.4 states that each experiment uses at most $N + M$ posteriors. This second simplification is useful when the set of actions available to the agent is rich. Consider, for instance, the case in which $A = \mathbb{R}$ and the agent's payoff is $u(a,\theta,\omega) = -(a - (\omega + \theta))^2$. In this case, even if the space of actions is a continuum, Proposition 4.4 implies that the designer can focus, without loss of generality, on experiments that induce finitely many beliefs.

References

ALIPRANTIS, C. D. AND K. BORDER (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Berlin: Springer-Verlag,.

AZRIELI, Y. (2021): "Constrained Versus Unconstrained Rational Inattention," *Games*, 12, 3.

BABICHENKO, Y., I. TALGAM-COHEN, AND K. ZABARNYI (2020): "Bayesian Persuasion under Ex Ante and Ex Post Constraints," *arXiv preprint arXiv:2012.03272*.

BAUER, H. (1958): "Minimalstellen von Funktionen und Extremalpunkte," *Archiv der Mathematik*, 9, 389–393.

BESTER, H. AND R. STRAUSZ (2007): "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory*, 136, 236–259.

BOLESLAVSKY, R. AND K. KIM (2018): "Bayesian persuasion and moral hazard," Tech. rep.

CALZOLARI, G. AND A. PAVAN (2006): "Monopoly with resale," *The RAND Journal of Economics*, 37, 362–375.

CANDOGAN, O. AND P. STRACK (2021): "Optimal Disclosure of Information to a Privately Informed Receiver," *arXiv preprint arXiv:2101.10431*.

CELIK, G. (2015): "Implementation by gradual revelation," *The Rand Journal of Economics*, 46, 271–296.

CRAUEL, H. (2002): *Random probability measures on Polish spaces*, vol. 11, CRC press.

DOVAL, L. AND V. SKRETA (2018): "Constrained Information Design: Toolkit," .

——— (2020): "Mechanism Design with Limited Commitment," .

DWORCZAK, P. (2020): "Mechanism design with aftermarkets: Cutoff mechanisms," *Econometrica*, 88, 2629–2661.

DWORCZAK, P. AND A. KOLOTILIN (2019): "The persuasion duality," *arXiv preprint arXiv:1910.11392*.

DWORCZAK, P., S. D. KOMINERS, AND M. AKBARPOUR (2019): "Redistribution through markets," *Becker Friedman Institute for Research in Economics Working Paper*.

EILAT, R., K. ELIAZ, AND X. MU (2021): "Bayesian Privacy," .

GEORGIADIS, G. AND B. SZENTES (2020): "Optimal monitoring design," *Econometrica*, 88, 2075–2107.

GUO, Y. AND E. SHMAYA (2019): "The interval structure of optimal disclosure," *Econometrica*, 87, 653–675.

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615.

KANG, Z. Y. (2020): "Markets for Goods with Externalities," *Available at SSRN*.

KARTIK, N., S. LEE, AND D. RAPPOPORT (2017): "Single-crossing differences on distributions," Tech. rep., Working paper.

KOLOTILIN, A., T. MYLOVANOV, A. ZAPECHELNYUK, AND M. LI (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85, 1949–1964.

LE TREUST, M. AND T. TOMALA (2019): "Persuasion with limited communication capacity," *Journal of Economic Theory*, 184, 104940.

ROCKAFELLAR, R. T. (1970): *Convex analysis*, Princeton University Press.

SALAMANCA, A. (2021): "The value of mediated communication," *Journal of Economic Theory*, 105191.

STRAUSZ, R. (2003): "Deterministic mechanisms and the revelation principle," *Economics Letters*, 79, 333–337.

SZAPIEL, W. (1975): "Points extrémaux dans les ensembles convexes (I), Théorie générale," *Bull. Acad. Polon. Sci. Math. Phys*, 22, 939–945.

A Proofs of Section 3

Proof of Proposition 3.1. The proof follows similar steps as in Le Treust and Tomala (2019). Given a function $h : \mathbb{R}^d \mapsto \mathbb{R}$, let h^* denote its Fenchel conjugate. That is $h^*(t) = \sup_x \{xt - h(x)\}$. Thus, $-\text{cav}(-h)(x) = (h^*)^*(x)$. Thus,

$$\text{cav}\, h(x) = \inf_p \left\{ xp + \sup_y (h(y) - py) \right\}. \tag{A.1}$$

We now apply this to the function $f^{g_1,\dots,g_K}(\mu, \gamma)$. Thus, letting $p_\mu \in \mathbb{R}^N$ and $p_\gamma \in \mathbb{R}^K$,

$$\text{cav}\, f^{g_1,\dots,g_K}(\mu, \gamma) = \inf_{p_\mu, p_\gamma} \left\{ p_\mu \mu + p_\gamma \gamma + \sup_{\nu, \eta} \left(f^{g_1,\dots,g_K}(\nu, \eta) - p_\mu \nu - p_\gamma \eta \right) \right\}$$

$$= \inf_{p_\mu, p_\gamma} \left\{ p_\mu \mu + p_\gamma \gamma + \sup_{\nu, \eta : (\forall k \leq r) g_k(\nu) \geq \eta_k \wedge (\forall k \geq r+1) g_k(\nu) = \eta_k} \left(f(\nu, \eta) - p_\mu \nu - p_\gamma \eta \right) \right\}$$

19

If there exists $k \in \{1, \ldots, r\}$ such that $p_{\gamma,k} > 0$, then letting $\eta_k \to -\infty$, the sup is $+\infty$. Therefore, we can restrict attention to $p_{\gamma,k} \leq 0$ for $k \in \{1, \ldots, r\}$. Setting $t_\gamma = -p_\gamma$ we get,

$$
\operatorname{cav} f^{g_1, \ldots, g_K}(\mu, \gamma) = \inf_{p_\mu, t_\gamma} \left\{ p_\mu \mu - t_\gamma \gamma + \sup_{\nu, \eta : (\forall k \leq r) g_k(\nu) \geq \eta_k \vee (\forall k \geq r+1) g_k(\nu) = \eta_k} \left(f(\nu, \eta) - p_\mu \nu + t_\gamma \eta \right) \right\}
$$

$$
= \inf_{p_\mu, p_\gamma} \left\{ p_\mu \mu - t_\gamma \gamma + \sup_\nu \left(f(\nu, \eta) - p_\mu \nu + \sum_{k=1}^K t_{\gamma,k} g_k(\nu) \right) \right\}
$$

$$
= \inf_{t_\gamma \in \mathbb{R}_+^r \times \mathbb{R}^{K-r}} \left\{ \inf_{p_\mu} \left[p_\mu \mu + \sup_\nu \left(f(\nu) + \sum_{k=1}^K t_{\gamma,k} g_k(\nu) - p_\mu \nu \right) \right] - t_\gamma \gamma \right\}
$$

$$
= \inf_{t_\gamma \in \mathbb{R}_+^r \times \mathbb{R}^{K-r}} \left\{ \operatorname{cav}\left(f + \sum_{k=1}^K t_{\gamma,k} g \right)(\mu) - t_\gamma \gamma \right\},
$$

where (i) the second line follows from noticing that $\eta_k = g_k(\nu)$, whenever $k > r$, while it is optimal to set $\eta_k = g_k(\nu)$ whenever $k \leq r$ since $t_{\gamma,k} \geq 0$, (ii) the third line is just a rewriting, and (iii) the fourth line follows from noting that the infimum in the square brackets is the definition of the concavification of $f + \sum t_{\gamma,k} g_k$ (see Equation A.1). The statement of Proposition 3.1 then follows. □

B Proofs of Section 4.1

Proof of Proposition 4.1. Consider the following program:

$$
\max_{\beta : \Theta \mapsto \Delta\Delta(\Omega), \alpha : \Delta(\Omega) \mapsto \Delta(Y_1)} \sum_{\omega \in \Omega} \mu_0(\omega) \mathbb{E}_{\beta(\cdot|\theta)} \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[v(y_1, y_2, \omega) \right] \right] \tag{$\mathcal{A}$}
$$

$$
\text{s.t.} \begin{cases} \mathbb{E}_{\beta(\cdot|\omega_1)} \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[u(y_1, y_2(y_1, \mu), \omega_1) \right] \right] \geq 0 \\ (\forall i \in \{2, \ldots, N\}) \mathbb{E}_{\beta(\cdot|\omega_i)} \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[u(y_1, y_2(y_1, \mu), \omega_i) \right] \right] \geq \mathbb{E}_{\beta(\cdot|\omega_{i-1})} \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[u(y_1, y_2(y_1, \mu), \omega_i) \right] \right] \\ (\forall i \in \{1, \ldots, N-1\}) \mathbb{E}_{\beta(\cdot|\omega_i)} \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[u(y_1, y_2(y_1, \mu), \omega_i) \right] \right] \geq \mathbb{E}_{\beta(\cdot|\omega_{i+1})} \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[u(y_1, y_2(y_1, \mu), \omega_i) \right] \right] \end{cases}
$$

We show that the solution to $(\mathcal{A})$ satisfies all the constraints of $(\mathcal{P})$. To simplify notation, in what follows, let

$$
u(\mu, \omega_i) = \mathbb{E}_{\alpha(\cdot|\mu)} \left[u(y_1, y_2(y_1, \mu), \omega_i) \right].
$$

Note first that the solution to $(\mathcal{A})$ satisfies that for all $i \geq 2$,

$$\mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_i)] \geq \mathbb{E}_{\beta(\cdot|\omega_{i-1})}[u(\mu, \omega_i)]$$
$$\mathbb{E}_{\beta(\cdot|\omega_{i-1})}[u(\mu, \omega_{i-1})] \geq \mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_{i-1})],$$

so that for all $i \geq 2$, we have

$$\mathbb{E}_{\beta(\cdot|\omega_i)-\beta(\cdot|\omega_{i-1})}[u(\mu, \omega_i) - u(\mu, \omega_{i-1})] \geq 0. \tag{B.1}$$

We note two implications of Equation B.1. First, monotonic expectational differences together with equation Equation B.1 implies that if $k < i$, then it cannot be the case that

$$\mathbb{E}_{\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1})}[u(\mu, \omega_i) - u(\mu, \omega_k)] < 0.$$

Hence, we must have $\mathbb{E}_{\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1})}u(\mu, \omega_i) \geq \mathbb{E}_{\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1})}u(\mu, \omega_k)$, when $k < i$. Second, if $k > i$, Equation B.1 evaluated at k together with monotonic expectational differences implies that $\mathbb{E}_{\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1})}u(\mu, \omega_k) \geq \mathbb{E}_{\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1})}u(\mu, \omega_i)$. We use these two implications in what follows.

To show that the statement of the proposition holds, consider i and $j < i - 1$. The solution to $(\mathcal{A})$ satisfies

$$\mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_i)] \geq \mathbb{E}_{\beta(\cdot|\omega_{i-1})}[u(\mu, \omega_i)]$$
$$\mathbb{E}_{\beta(\cdot|\omega_{i-1})}[u(\mu, \omega_{i-1})] \geq \mathbb{E}_{\beta(\cdot|\omega_{i-2})}[u(\mu, \omega_{i-1})]$$
$$\cdots$$
$$\mathbb{E}_{\beta(\cdot|\omega_{j+1})}[u(\mu, \omega_{j+1})] \geq \mathbb{E}_{\beta(\cdot|\omega_j)}[u(\mu, \omega_{j+1})].$$

Adding up, we obtain

$$\sum_{k=j+1}^{i} \mathbb{E}_{(\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1}))}u(\mu, \omega_k) \geq 0. \tag{B.2}$$

As discussed above, monotonic expectational differences together with equation (B.1) implies the left-hand side is bounded above by

$$\sum_{k=j+1}^{i} \mathbb{E}_{(\beta(\cdot|\omega_k)-\beta(\cdot|\omega_{k-1}))}[u(\mu, \omega_i)] = \mathbb{E}_{\beta(\cdot|\omega_i)-\beta(\cdot|\omega_j)}[u(\mu, \omega_i)]. \tag{B.3}$$

Equations ([B.2](#)) and ([B.3](#)) imply

$$\mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_i) - t_h] \geq \mathbb{E}_{\beta(\cdot|\omega_j)}[u(\mu, \omega_i)].$$

Therefore, the constraint that i does not report $j < i - 1$ holds.

Similarly, consider i and $j > i + 1$. The solution to ($\mathcal{A}$) satisfies

$$\mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_i)] \geq \mathbb{E}_{\beta(\cdot|\omega_{i+1})}[u(\mu, \omega_i)]$$
$$\mathbb{E}_{\beta(\cdot|\omega_{i+1})}[u(\mu, \omega_{i+1})] \geq \mathbb{E}_{\beta(\cdot|\omega_{i+2})}[u(\mu, \omega_{i+1})]$$
$$\cdots$$
$$\mathbb{E}_{\beta(\cdot|\omega_{j-1})}[u(\mu, \omega_{j-1})] \geq \mathbb{E}_{\beta(\cdot|\omega_j)}[u(\mu, \omega_{j-1})].$$

Adding up, we obtain

$$\sum_{k=i}^{j-1} \mathbb{E}_{(\beta(\cdot|\omega_k) - \beta(\cdot|\omega_{k+1}))} u(\mu, \omega_k) \geq 0. \tag{B.4}$$

As discussed above, monotonic expectational differences together with equation ([B.1](#)) imply that the left-hand side is bounded above by

$$\sum_{k=i}^{j-1} \mathbb{E}_{(\beta(\cdot|\omega_k) - \beta(\cdot|\omega_{k+1}))} u(\mu, \omega_i) = \mathbb{E}_{(\beta(\cdot|\omega_i) - \beta(\cdot|\omega_j))} u(\mu, \omega_i). \tag{B.5}$$

Equation ([B.5](#)) follows because equation ([B.1](#)) implies $\mathbb{E}_{(\beta(\cdot|\omega_k) - \beta(\cdot|\omega_{k+1}))} u(\mu, \omega_k)$ is decreasing in k.

Equations ([B.4](#)) and ([B.5](#)) imply

$$\mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_i)] \geq \mathbb{E}_{\beta(\cdot|\omega_j)}[u(\mu, \omega_i)].$$

Therefore, the incentive constraint that i does not report j, $j > i + 1$ holds.

Finally, because we have all incentive compatibility constraints, it follows that, when u_i satisfies [Definition 4.1](#), the participation constraints for $i \geq 2$ are implied by the participation constraint for $i = 1$. To see this, note the following. First, because all incentive compatibility constraints are satisfied, we have that for all $i \geq 2$,

$$\mathbb{E}_{\beta(\cdot|\omega_i)}[u(\mu, \omega_i)] \geq \mathbb{E}_{\beta(\cdot|\omega_1)}[u(\mu, \omega_i)]. \tag{B.6}$$

We can write the right-hand side of equation (B.6) as

$$\mathbb{E}_{\beta(\cdot|\omega_1)}[u(\mu,\omega_1) + u(\mu,\omega_i) - u(\mu,\omega_1)]. \tag{B.7}$$

Now, since the family $\{u(\cdot,\omega) : \omega \in \Omega\}$ satisfies monotonic expectational differences, we have that

$$u(\mu,\omega_i) - u(\mu,\omega_1) = (b(\omega_i) - b(\omega_1))\mathbb{E}_{\alpha(\cdot|\mu)}[h_1(y_1,y_2(y_1,\mu))] + c(\omega_i) - c(\omega_1).$$

Moreover, recall from footnote 7, that we assume that $u(y_1^*,y_2^*,\omega_i) = 0$ for all $i \in \{1,\ldots,N\}$. Hence, we can rewrite the above as:

$$u(\mu,\omega_i) - u(\mu,\omega_1) = (b(\omega_i) - b(\omega_1))\mathbb{E}_{\alpha(\cdot|\mu)}[h_1(y_1,y_2(y_1,\mu))] + c(\omega_i) - c(\omega_1)$$
$$= (b(\omega_i) - b(\omega_1))(\mathbb{E}_{\alpha(\cdot|\mu)}[h_1(y_1,y_2(y_1,\mu))] - h_1(y_1^*,y_2^*)) \geq 0$$

since $b(\cdot)$ is increasing in ω_i and h_1 is minimized at (y_1^*,y_2^*) by the assumption in footnote 7. $\qquad\square$

Proof of Corollary 4.1. Proposition 4.1 implies that under monotonic expectational differences, it is enough to consider the solution to $\mathcal{A}$. Writing it in terms of the distribution of posteriors it induces, we obtain:

$$\max_{\alpha:\Delta(\Omega)\mapsto\Delta(Y_1),y_2\in Y_2} \max_{\tau\in\Delta\Delta(\Omega):\mathbb{E}_\tau\mu=\mu_0} \mathbb{E}_\tau\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\sum_{\omega\in\Omega}\mu(\omega)v(y_1,y_2(y_1,\mu),\omega)\right]\right] \quad (\mathcal{A}')$$

$$s.t. \begin{cases} \mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\frac{\mu(\omega)}{\mu_0(\omega)}u(y_1,y_2(y_1,\mu),\omega_1)\right]\right] \geq 0 \\ (\forall i \in \{2,\ldots,N\})\mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i-1})}{\mu_0(\omega_{i-1})}\right)u(y_1,y_2(y_1,\mu),\omega_i)\right]\right] \geq 0 \\ (\forall i \in \{1,\ldots,N-1\})\mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i+1})}{\mu_0(\omega_{i+1})}\right)u(y_1,y_2(y_1,\mu),\omega_i)\right]\right] \geq 0 \end{cases}$$

Fix $\alpha : \Delta(\Omega) \mapsto \Delta(Y_1)$ and a selection $y_2 \in Y_2$, and consider the program:

$$\max_{\tau\in\Delta\Delta(\Omega):\mathbb{E}_\tau\mu=\mu_0} \mathbb{E}_\tau\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\sum_{\omega\in\Omega}\mu(\omega)v(y_1,y_2(y_1,\mu),\omega)\right]\right] \quad (\mathcal{A}'_\alpha)$$

$$s.t. \begin{cases} \mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\frac{\mu(\omega)}{\mu_0(\omega)}u(y_1,y_2(y_1,\mu),\omega_1)\right]\right] \geq 0 \\ (\forall i \in \{2,\ldots,N\})\mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i-1})}{\mu_0(\omega_{i-1})}\right)u(y_1,y_2(y_1,\mu),\omega_i)\right]\right] \geq 0 \\ (\forall i \in \{1,\ldots,N-1\})\mathbb{E}_{\tau(\cdot)}\left[\mathbb{E}_{\alpha(\cdot|\mu)}\left[\left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i+1})}{\mu_0(\omega_{i+1})}\right)u(y_1,y_2(y_1,\mu),\omega_i)\right]\right] \geq 0 \end{cases}$$

23

Note that there might be allocations α for which there is no τ that satisfies the incentive compatibility and/or participation constraints. To address this issue, let C_α denote the policies τ that satisfy the constraints in $(\mathcal{A}'_\alpha)$. Let $f^0_\alpha(\tau)$ denote

$$\mathbb{E}_\tau \left[\mathbb{E}_{\alpha(\cdot|\mu)} \left[\sum_{\omega \in \Omega} \mu(\omega) v(y_1, y_2(y_1, \mu), \omega) \right] \right],$$

and let

$$f_\alpha(\tau) = \begin{cases} f^0_\alpha(\tau) & \text{if } \tau \in C_\alpha \\ -\infty & \text{otherwise} \end{cases}.$$

In what follows, $f_\alpha(\tau)$ is the objective function under consideration. Note that letting,

$$g_i(\mu) = \left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i+1})}{\mu_0(\omega_{i+1})} \right) u(\mu, \omega_i), \quad i \in \{1, \ldots, N-1\}$$

$$g_{N-2+i}(\mu) = \left(\frac{\mu(\omega_i)}{\mu_0(\omega_i)} - \frac{\mu(\omega_{i-1})}{\mu_0(\omega_{i-1})} \right) u(\mu, \omega_i), \quad i \in \{2, \ldots, N\}$$

$$g_{2N-1}(\mu) = \frac{\mu(\omega_1)}{\mu_0(\omega_1)} u(\mu, \omega_1),$$

we can write $\mathcal{A}'_\alpha$ as a special case of OPT, with $r = 2N - 2$. Corollary 3.1 implies that any finite solution to $\mathcal{A}'_\alpha$ uses at most $3N - 1$ beliefs. $\qquad \square$

Proof of Corollary 4.2. Towards a contradiction, suppose the participation constraint of ω_1 is not binding. Then, let $\epsilon = \mathbb{E}_{\beta(\cdot|\omega_1)} \left[\mathbb{E}_{\alpha(\cdot|\mu)} [\tilde{u}(y'_1, y_2(\cdot), \omega_1)] \right]$. Consider a mechanism that increases all transfers, $p(\mu) + \epsilon$. All incentive constraints continue to be satisfied, the participation constraint for ω_1 binds, and revenue increases, contradicting that the solution was optimal. $\qquad \square$

Proof of Proposition 4.2. Given a selection $y_2(y'_1, \mu)$ from the principal's best response correspondence in period 2 when his belief is μ, let

$$\bar{u}(y'_1, y_2(y'_1, \mu), \omega_i) = \tilde{u}(y'_1, y_2(y'_1, \mu), \omega_i)$$
$$- \frac{1 - \sum_{n \le i} \mu_0(\omega_n)}{\mu_0(\omega_i)} (\tilde{u}(y'_1, y_2(y'_1, \mu), \omega_i) - \tilde{u}(y'_1, y_2(y'_1, \mu), \omega_{i-1}))$$

24

Then, replacing the constraints in ($\mathcal{R}$) in the principal's objective function, we obtain the following expression:

$$\mathbb{E}_\tau[\underbrace{\mathbb{E}_{\alpha(\cdot|\mu)} \sum_{\omega \in \Omega} \mu(\omega) \left(\tilde{v}(y_1', y_2(y_1', \mu), \omega_i) + \overline{u}(y_1', y_2(y_1', \mu), \omega_i) \right)]}_{\check{v}(\alpha, y_2, \mu)}$$

Therefore, we can write ($\mathcal{R}$) as

$$\max_{\tau, \alpha, y_2} \mathbb{E}_\tau \left[\check{v}(\alpha, y_2, \mu) \right], \tag{B.8}$$

where the distribution over posteriors must satisfy the Bayes' plausibility constraint and $y_2 \in Y_2$. That is, the solution to the relaxed problem is obtained by maximizing a version of the virtual surplus, represented by $\check{v}$, and then choosing a distribution over posteriors that averages out to the prior. The following remark is in order:

Remark 1 (Tie-breaking in favor of the principal). *So far we have remained silent about how $y_2(y_1', \mu)$ is chosen, beyond the restriction that $y_2(\cdot) \in Y_2(\cdot)$. We can use the function $\tilde{v}(y_1', y_2(y_1', \mu), \omega_i) + \overline{u}(y_1', y_2(y_1', \mu), \omega_i)$ to determine how to break the possible ties in $Y_2(\cdot)$ and make the principal's objective function upper-semicontinuous. In fact, if $y_2, y_2' \in Y_2(y_1', \mu)$, then in the relaxed program, y_2 is selected as long as*

$$\sum_{\omega \in \Omega} \mu(\omega) \left[\tilde{v}(y_1', y_2, \omega_i) + \overline{u}(y_1', y_2, \omega_i) \right] \geq \sum_{\omega \in \Omega} \mu(\omega) \left[\tilde{v}(y_1', y_2', \omega_i) + \overline{u}(y_1', y_2', \omega_i) \right].$$

In other words, ties are broken in favor of the virtual surplus.

We now illustrate how to solve the program in Equation B.8. Towards this, fix the selection y_2^* as in Remark 1. Because the program is separable in the allocation α across posteriors μ, the solution can be obtained in two steps. First, for each posterior μ, we maximize $\check{v}(\cdot, y_2^*, \mu)$ with respect to α. Denote the value of this problem $\hat{v}(\mu)$. Second, we choose τ to maximize the expectation of $\hat{v}(\cdot)$ subject to the constraint that τ is Bayes' plausible. A straightforward application of Carathéodory's theorem implies that the solution to ($\mathcal{R}$) involves at most N posteriors. $\qquad\square$

C Proofs of Section 4.2

A menu of experiments consists of a finite[11] set of signals S and a collection of distributions $\{\pi_\theta : \Omega \mapsto \Delta(S) : \theta \in \Theta\}$. Under experiment π_θ, when the agent observes signal $s \in S$, the agent updates her belief about the state of the world as follows:

$$\mu_s(\omega) = \frac{\mu_0(\omega)\pi_\theta(s|\omega)}{\sum_{\omega' \in \Omega} \pi_\theta(s|\omega)\mu_0(\omega')} \equiv \frac{\mu_0(\omega)\pi_\theta(s|\omega)}{Pr_{\pi,\theta}(s)}.$$

A menu of experiments is incentive compatible if the following holds for all $\theta \in \Theta$ and $\theta' \neq \theta$:

$$\sum_{\mu \in \Delta(\Omega)} \sum_{\{s \in S : \mu_s = \mu\}} Pr_{\pi,\theta}(s)U(\mu,\theta) \geq \sum_{\mu \in \Delta(\Omega)} \sum_{\{s \in S : \mu_s = \mu\}} Pr_{\pi,\theta'}(s)U(\mu,\theta) \qquad (C.1)$$

Lemma C.1. *It is without loss of generality to focus on experiments such that $S = \Delta(\Omega)$.*

Proof. The statement follows from Equation C.1. To see this, let $\langle \{\pi_\theta\}_{\theta \in \Theta}, S \rangle$ denote an experiment. Consider the following experiment, $\langle \{\pi'_\theta\}_{\theta \in \Theta}, \Delta(\Omega) \rangle$

$$\pi'_\theta(\mu|\omega) = \sum_{\{s \in S : \mu_s = \mu\}} \pi_\theta(s|\omega). \qquad (C.2)$$

Note that

$$Pr_{\pi',\theta}(\mu) = \sum_{\omega \in \Omega} \mu_0(\omega)\pi'_\theta(\mu|\theta) = \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{\{s \in S : \mu_s = \mu\}} \pi_\theta(s|\omega) = \sum_{\{s \in S : \mu_s = \mu\}} Pr_{\pi,\theta}(s).$$

Thus, $\langle \{\pi'_\theta\}_{\theta \in \Theta}, \Delta(\Omega) \rangle$ yields the same payoff to the designer and the agent. Furthermore, it is incentive compatible. □

Proof of Proposition 4.4. The proof proceeds in two steps. We first argue that the solution to the problems in Equations 4.2 and 4.3 are the same. We then apply Theorem 3.1 to the problem in Equation 4.3 to argue for the upper bound in the number of posteriors induced in an optimal experiment.

To see that the solutions to both problems are the same, consider the following argument. Let τ^* denote a solution to Equation 4.2. For each $\theta \in \Theta$, let

$$u_\theta^* = \mathbb{E}_{\tau^*(\theta,\cdot)}[U(\mu,\theta)].$$

---

[11]Proposition 4.4 implies this is without loss of generality.

Then, it is immediate to check that $(\tau^*, (u_\theta^*)_{\theta \in \Theta})$ solves the problem in Equation 4.3.

Let $(\tau^*, (u_\theta^*)_{\theta \in \Theta})$ denote a solution to the problem in Equation 4.3. Note that without loss of generality we can take

$$u_\theta^* = \mathbb{E}_{\tau^*(\theta, \cdot)}[U(\mu, \theta)].$$

Note that for each θ this relaxes the incentive compatibility constraint for $\theta' \neq \theta$ and it does not affect the first constraint for θ's experiment. It then follows that τ^* solves the problem in Equation 4.2.

Consider now the problem in Equation 4.3. Fix $\{u_\theta\}_{\theta \in \Theta}$. Note that the problem of finding an optimal $\tau : \Theta \mapsto \Delta\Delta(\Omega)$ given $\{u_\theta\}_{\theta \in \Theta}$ is separable across $\theta \in \Theta$. That is, given $\{u_\theta\}_{\theta \in \Theta}$, it is enough to solve M optimization problems:

$$\max_{\{\tau(\theta, \cdot) \in \Delta\Delta(\Omega) : \mathbb{E}_{\tau(\theta, \cdot)}[\mu] = \mu_0\}} \mathbb{E}_{\tau(\theta, \cdot)}[V(\mu, \theta)] \tag{C.3}$$

$$\text{s.t.} \begin{cases} (\forall \theta \in \Theta) & \mathbb{E}_{\tau(\theta, \cdot)}[U(\cdot, \theta)] \geq u_\theta \\ (\forall \theta \in \Theta)(\forall \theta' \neq \theta) & u_{\theta'} \geq \mathbb{E}_{\tau(\theta, \cdot)}[U(\mu, \theta')] \end{cases},$$

where $V(\mu, \theta) = \sum_{\omega \in \Omega} \mu(\omega) v(a^*(\mu, \theta), \theta, \omega)$. For each θ, the problem in Equation C.3 is a special case of the problem in OPT. Corollary 3.1 implies that there exists a solution that uses at most $N + M$ posteriors. $\qquad \square$