

Delegated Information Provision^{*}

Francesco Bilotta[†] Christoph Carnehl[‡] Justus Preusser[§]

May 2, 2026

Abstract

A designer relies on an experimenter to provide information to a decision maker, but the experimenter has incentives to persuade rather than merely transmit information. Anticipating this motive, the designer can restrict the set of admissible experiments, but cannot prevent the experimenter from garbling any admissible experiment. We model this situation as delegation over experiments. The optimal delegation set is obtained by comparing maximally informative experiments among those the experimenter has no incentive to garble. When the experimenter's preferences are *S*-shaped, we characterize these experiments as double censorship. Relative to the full-delegation benchmark, double censorship features an intermediate pooling region, inducing a smaller pooling region for the highest states. We show that the designer strictly benefits from imposing a nontrivial delegation set that constrains persuasion while retaining information provision. Applying our results to recommender systems, we show that privacy constraints can arise endogenously to protect consumers against persuasion.

1 Introduction

In many environments, delegation separates the production of information from its use in decision making. In courts, prosecutors generate evidence that informs judges; in regulatory settings, firms conduct tests that guide approval decisions; and on platforms,

^{*}We thank Martino Banchio, Gregorio Curello, Alkis Georgiadis-Harris, Nicola Pavoni, Marco Ottaviani, Mark Whitmeyer, and seminar audiences at Bielefeld, Bocconi, and Essex for their comments. Preusser acknowledges financial support by the European Research Council (HEUROPE 2022 ADG, GA No. 101055295 – InfoEcoScience).

[†]Università Bocconi. Email: francesco.bilotta2@phd.unibocconi.it

[‡]Università Bocconi and IGER. Email: christoph.carnehl@unibocconi.it

[§]Università Bocconi and IGER. Email: justus.preusser@unibocconi.it

algorithms recommend products to consumers. In each case, the party who produces information has incentives to influence the decision maker rather than merely transmit facts. While a large literature on Bayesian persuasion characterizes how the experimenter optimally designs information to exploit this influence, much less attention has been paid to the complementary problem of a designer who anticipates persuasion and seeks to limit its force. To do so, a designer may restrict the experimenter's discretion by specifying which experiments are admissible. Yet even under such restrictions, delegation remains subject to moral hazard: the experimenter can garble any admissible experiment before it reaches the decision maker. We study the optimal design of delegated information provision, which trades off the informativeness of admissible experiments against the experimenter's ability to persuade.

In practice, such control over informational discretion often takes the form of restrictions on admissible experiments rather than direct control of actions. In the examples above, legislators restrict which evidence prosecutors may present, regulators specify testing procedures for firms, and privacy protection limits which data algorithms can use. These restrictions impose an upper bound on informativeness while leaving substantial discretion to the experimenter. As a result, the experimenter may still suppress available evidence, implement tests imperfectly, or communicate only coarse information.

We therefore cast the design of restrictions as a delegation problem over experiments. The designer specifies a delegation set: a set of admissible Blackwell-experiments. The experimenter can choose an arbitrary garbling of any admissible experiment. Therefore, the designer can only impose an upper bound on information. Given a delegation set, the experimenter faces a standard Bayesian persuasion problem with a one-dimensional state and posterior-mean preferences. The experimenter picks an experiment to persuade a decision maker (DM) to take a binary action. The DM's optimal choice depends on a privately known outside option. We assume that the designer has convex preferences in the posterior mean and thus values information. Due to this assumption, the designer need not necessarily be distinct from the DM but can also be interpreted as the DM themselves who commits *ex ante* to an admissibility rule as a form of self-discipline against persuasion.

Our central insight is that effective limits on the experimenter's discretion cannot deliver experiments that are uniformly more informative than the full-delegation benchmark—the outcome of the traditional Bayesian persuasion problem. Any uniformly more informative restriction will be garbled by the experimenter, while any uniformly less informative restriction is dominated by the persuasion outcome. The designer's only viable strategy is therefore to reshape the admissible information structures by selecting an experiment that is Blackwell-incomparable to persuasion, and prevents persuasion from being replicated

through garbling. Such restrictions redistribute informativeness across states: by deliberately inducing pooling of some regions of the state space, the designer dampens incentives for concealment and induces finer disclosure where persuasion would otherwise be most severe. We show that optimally designed restrictions strictly dominate full delegation, adding value not via more information per se, but with a different shape of information revelation.

As a first step, we show that the designer’s problem reduces to choosing among *maximal incentive-compatible* (MIC) experiments. An experiment F is incentive compatible if it is self-enforcing: when the designer sets F as the upper bound, the experimenter optimally implements F rather than garble it. An incentive-compatible experiment is maximal if no strictly more informative experiment is also incentive compatible. Since the designer values information, the designer never imposes a restriction that is incentive compatible but not maximal.

While incentive compatibility and maximality are intuitive properties of optimal delegation, they already yield significant structure on the designer’s problem. A key implication of MIC is that any profitable restriction must be Blackwell-incomparable to the persuasion outcome under full delegation F^* . Blackwell-more informative restrictions are garbled back to F^* , while Blackwell-less informative restrictions are dominated by F^* . Therefore, any strict gain from restricting discretion must arise from Blackwell-incomparable experiments that reallocate informativeness across regions of the state space.

This fact is particularly evident when the experimenter’s payoff is S -shaped, which is our baseline assumption. In this case, we fully characterize maximal incentive-compatible (MIC) experiments. They take the form of *double censorship*: there exist two thresholds such that all states below the first threshold are fully revealed, while states between the thresholds and above the second threshold are pooled into one intermediate and one high signal, respectively. Full delegation is itself MIC and corresponds to the unrestricted S -shaped persuasion outcome; it arises as the limiting case in which the two thresholds coincide, yielding *upper censorship*.

Double censorship makes the designer’s informational trade-off transparent. Incentive compatibility implies that shrinking the top pooling region—i.e., improving information precisely in the high states the experimenter would otherwise censor—requires expanding the intermediate pooling region. Thus, better information in high states comes at the cost of worse information in intermediate ones. MIC experiments are therefore ordered by the size of the top pooling interval. However, they are not Blackwell-comparable: each MIC experiment redistributes informativeness between high and intermediate states rather than uniformly increasing or decreasing it. This redistribution logic and the ensu-

ing incomparability are the key factors generating scope for improvements beyond full delegation.

In fact, leveraging the structure of MIC experiments, we show that the designer strictly prefers a nontrivial restriction: full delegation is never optimal. Intuitively, starting from full delegation, introducing an arbitrary small amount of pooling around the censorship cutoff entails only a second-order informational loss, yet it disciplines the experimenter's incentives and reduces the pooling region at the highest states. The resulting gain is of first order, as it yields a benefit proportional to the mass of states pooled under persuasion. Hence, improving information in high states strictly outweighs the local loss on intermediate states, yielding a higher payoff for the designer.

We illustrate our results in the context of platforms persuading consumers to purchase a product. Platforms have access to vast data about consumer demographics and their browsing behavior that they can use to estimate their valuation of products. As platforms profit from sales, they have an incentive to persuade consumers into purchasing. We show that this setting can be mapped into our environment and that a regulatory authority that cares about consumers has a strict incentive to restrict the data that the platform's recommendation algorithm can use.¹ We show that our restrictions can be implemented using prominent notions of privacy constraints, such as privacy sets and differential privacy. Hence, the tension between information provision and persuasion gives rise to endogenous privacy constraints—even in the absence of intrinsic preferences for privacy.

We then ask whether nontrivial restrictions remain optimal beyond the S-shaped case. In the baseline environment, profitable restrictions work by shifting the location of the atom in the top interval of the experimenter's best reply. This channel may be infeasible under alternative payoff shapes. If the DM's outside option is degenerate, for instance, the experimenter's best reply never assigns mass to states strictly above that outside option; in this case, restrictions cannot move any top atom and full delegation is optimal. Similarly, if the experimenter's payoff is M-shaped and the prior is sufficiently dispersed, the persuasion outcome is a binary experiment supported on the two peaks of the M. Any restriction can then only induce the experimenter to shift mass into the valley between the peaks, which reduces informativeness and lowers the designer's payoff. Building on these observations, we provide general conditions under which full delegation is suboptimal for general posterior-mean payoffs of the experimenter.

Moreover, we show that the designer's problem is generally solved by an incentive-

¹Indeed, such restrictions may also arise if the platform cares about consumer surplus in addition to commissions from sales. If the platform is organized hierarchically and the recommendation algorithm is designed to maximize sales, then management may want to limit the data that the algorithm can use.

compatible bi-pooling experiment. In fact, we show that all MIC experiments are extreme points of the set of all experiments. Essentially, these results demonstrate how the literature’s extreme-point and linear-duality approaches can be combined to solve information design problems with the novel incentive-compatibility and maximality constraints that we introduce.

Related Literature. Most closely related are [Ichihashi \(2019\)](#), [Arieli et al. \(2022\)](#), and [Mylovanov and Zapechelnyuk \(2024\)](#). [Ichihashi \(2019\)](#) studies endogenously constrained information of an experimenter in a binary-action persuasion environment with an uninformed DM.² In our setting, by contrast, the DM has private information, which generates S-shaped posterior-mean payoffs for the experimenter. This introduces fundamentally different incentives for the experimenter and delivers a key implication that is absent in the uninformed benchmark: the designer strictly benefits from imposing a nontrivial restriction. Moreover, while [Ichihashi \(2019\)](#) focuses on characterizing attainable utility profiles, we characterize optimal restrictions, show how they can be implemented and how they trade off information revelation across states.

[Arieli et al. \(2022\)](#) analyze persuasion with strategic mediators who sequentially garble a sender’s experiment. They provide geometric characterizations for abstract settings. We instead focus on the setting with a single experimenter (a single mediator in their framework) and posterior-mean preferences. We use this structure to obtain a sharp description of MIC restrictions and the informational trade-offs.

[Mylovanov and Zapechelnyuk \(2024\)](#) compare sequential obfuscation and sequential disclosure. Sequential obfuscation corresponds to designing Blackwell-upper bounds on experiments, while sequential disclosure corresponds to designing Blackwell-lower bounds. Their analysis is tailored to conflict in debates, with the designer and the intermediary having opposing preferences over posterior means; this differs from our environment, where the designer values information.

Our framework also connects to the literature on attention management. In attention-management models, a receiver may rationally ignore a sender’s information due to attention costs. Our model nests a class of such problems.³ Within that class, our characterization implies that full disclosure is not optimal. A close precedent is an example due to [Lipnowski et al. \(2020\)](#) that likewise emphasizes informational trade-offs across states.⁴

²[Bird and Neeman \(2022\)](#) study a related problem in which a designer restricts an experimenter who persuades a DM. In contrast to our setting, the experimenter can restrict the actions available to the DM.

³Namely, by identifying the designer with the sender, and the experimenter with the receiver. The sender has convex posterior-mean payoffs, and the receiver’s costs are posterior-mean separable.

⁴Other related papers in this strand of literature are [Wei \(2021\)](#); [Bloedel and Segal \(2021\)](#); [Lipnowski](#)

A related strand studies information design under exogenously given constraints on information structures, such as limits on the number of messages or monotonicity requirements; see [Le Treust and Tomala \(2019\)](#); [Babichenko et al. \(2021\)](#); [Ivanov \(2021\)](#); [Mensch \(2021\)](#); [Tsakas and Tsakas \(2021\)](#); [Ball and Espín-Sánchez \(2022\)](#); [Onuchic and Ray \(2023\)](#); [Aybas and Turkel \(2025\)](#); [Lyu et al. \(2025\)](#). We instead make the restriction itself the design object.

The persuasion literature has studied other reasons why information transmission may improve in environments with persuasion incentives. [Tsakas et al. \(2021\)](#) ask whether the DM can gain by committing to burning money when taking certain actions. [Curello and Sinander \(2025\)](#) ask which changes to the experimenter’s payoffs induce the experimenter to choose a more informative experiment.⁵ We instead focus on informational restrictions on admissible experiments and highlight that profitable restrictions necessarily induce experiments that are Blackwell-incomparable to unrestricted experimentation.

[Kleiner et al. \(2021\)](#) and [Kolotilin and Zapechelnyuk \(2025\)](#) establish equivalences between a class of delegation and persuasion problems. These equivalences do not apply to our setting since the designer delegates persuasion itself to the experimenter.

Finally, we contribute to the literature on information design subject to privacy constraints. We explain in [Section 6](#) how our restrictions can capture both differential privacy constraints, as proposed by [Dwork et al. \(2006\)](#) and widely used in practice, and privacy sets, as proposed by [Strack and Yang \(2024\)](#).⁶ These privacy notions are motivated by an intrinsic interest in protecting individual-specific characteristics. [Schmutte and Yoder \(2025\)](#) and [Pan et al. \(2025\)](#) study information design subject to differential privacy constraints. In contrast, we study the optimal design of such restrictions by a designer who has no intrinsic preference for privacy but seeks to steer the incentives of a biased experimenter. We show that privacy restrictions may arise nevertheless to protect individuals from persuasion.

2 Model

Our model features a designer, an experimenter, and a decision maker (DM).

State, outside option, and payoffs. There is a state $\omega \in [0, 1]$ which is the realization of a random variable ω with prior CDF H and strictly positive, continuous density h . We

et al. (2022); [Song \(2025\)](#); [Jain and Whitmeyer \(2026\)](#); [Dall’Ara \(2026\)](#).

⁵[Curello and Sinander \(2025\)](#) also provide a result concerning changes to the prior. We comment on how this result relates to our problem in [Footnote 9](#) further ahead.

⁶See also [He et al. \(2026\)](#) for the more distant notion of private information.

denote the prior mean of the state by $\mu = \mathbb{E}_H[\omega]$. There is also an outside option $r \in [0, 1]$ which is the realization of a random variable r with prior CDF G and strictly positive, continuously differentiable density g . The outside option is privately observed by the DM.

The DM decides whether or not to act, $a \in \{0, 1\}$, where $a = 0$ represents inaction and $a = 1$ represents action. The DM's payoff from inaction is normalized to zero. When the state is ω and the outside option is r , the DM's payoff from action is $\omega - r$. Accordingly, when the DM's belief about the state has mean m , the DM takes the action if and only if $m \geq r$.⁷

The experimenter's payoff from inaction is zero and that from action is one. Hence, the experimenter has a state-independent preference for action.

We describe the designer's payoffs in reduced form given the DM's posterior mean about the state: if the posterior mean is m , the designer's interim expected payoff equals $u_D(m)$, where the expectation is taken over the DM's outside option and action. We assume that $u_D: [0, 1] \rightarrow \mathbb{R}$ is strictly convex and differentiable with a bounded derivative. Note that one interpretation of our model is for the designer and the DM to represent the same agent. In this case, we obtain $u_D(m) = \mathbb{E}_G[\max\{0, m - r\}]$.

We make two assumptions on the distribution of the outside option G and the prior mean of the state μ .

Assumption 1. The density g is strictly quasiconcave with an interior maximum $r_0 \in (0, 1)$. Accordingly, the CDF G is *S-shaped*: strictly convex on $[0, r_0]$, and strictly concave on $[r_0, 1]$.

Assumption 2. It holds $g(\mu)\mu > G(\mu)$.

[Assumption 1](#) is a common assumption in the persuasion literature (see, for example, [Kolotilin et al., 2022](#)). For us, it yields tractability in the experimenter's problem as described momentarily. We study more general preferences in [Section 7](#).

[Assumption 2](#) states that the prior mean μ lies in the convex part of G or close to the inflection point r_0 in the concave part. The assumption rules out trivial cases: if [Assumption 2](#) fails, no information transmission would arise; see [Section 3](#).

Information. While the DM privately observes the realized outside option r , the DM does not observe the realized state ω . Instead, the experimenter provides information about the state to the DM via an *experiment*. Due to the assumed posterior-mean preferences, we may identify each experiment with its induced CDF F of posterior means. It is convenient to consider *integrated CDFs (ICDFs)* and *mean-preserving contractions (MPCs)*. Given a CDF F , we define its ICDF I_F by $I_F(m) = \int_0^m F(m') dm'$ for all $m \in [0, 1]$. A CDF F is an MPC of a

⁷The DM's tie-breaking will be irrelevant as the CDF G of the outside option is atomless.

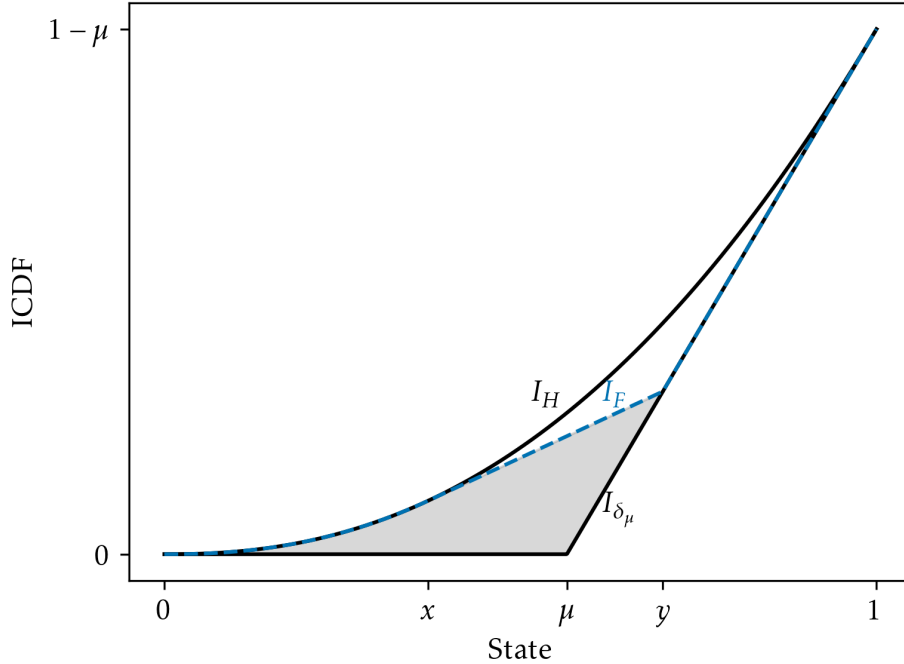


Figure 1: The ICDFs of the prior H , the degenerate experiment δ_μ , and an experiment F that pools H above a threshold x to a point y , and else coincides with H .

CDF \bar{F} if and only if $I_F \leq I_{\bar{F}}$ and $I_F(1) = I_{\bar{F}}(1)$; analogously, \bar{F} is a mean-preserving spread (MPS) of F . As is well-known, the set of experiments coincides with the set of MPCs of the prior H , denoted $\text{MPC}(H)$. We equip $\text{MPC}(H)$ with the L^1 -norm.

Let δ_μ be the degenerate (uninformative) experiment, i.e., the Dirac measure on the prior mean μ . Then, we can express $\text{MPC}(H)$ as the set of CDFs F satisfying $I_{\delta_\mu} \leq I_F \leq I_H$. Figure 1 depicts the ICDF representation and constraint.

Restrictions. Before the experimenter chooses an experiment, the designer imposes a restriction $\bar{F} \in \text{MPC}(H)$: given restriction \bar{F} , the experimenter must choose an experiment from the set $\text{MPC}(\bar{F})$ rather than from $\text{MPC}(H)$. Note, a restriction is an MPS-upper bound on admissible experiments, not an MPS-lower bound: the experimenter can always garble information. We discuss this model of informational restrictions along with real-world analogues at the end of this section. In Figure 1, if the restriction is F , the experimenter can choose among experiments whose ICDFs lie between I_F and I_{δ_μ} , as indicated by the shaded area. For example, F is admissible, but H is not.

Incentive-compatibility and optimality. Given a posterior mean m and outside option r , the DM chooses to act ($a = 1$) if and only if $m \geq r$. It follows from the experimenter's state-independent payoffs that the experimenter's interim expected payoff from inducing

a posterior mean m equals $G(m)$. Importantly, this payoff is S -shaped ([Assumption 1](#)). The experimenter’s ex-ante payoff from an experiment F equals $\int G(m)dF(m)$.

Given restriction \bar{F} , let $\text{BR}(\bar{F})$ be the set of the experimenter’s *best replies*:⁸

$$\text{BR}(\bar{F}) = \operatorname{argmax}_{F \in \text{MPC}(\bar{F})} \int G(m)dF(m).$$

An alternative interpretation of $\text{MPC}(\bar{F})$ is simply the set of all experiments when \bar{F} is the underlying prior distribution of the state. Accordingly, $F \in \text{BR}(\bar{F})$ means that the experimenter finds F optimal given full delegation and the fictitious prior \bar{F} .⁹

We define the set of incentive-compatible experiments as those that are best replies to some restriction.

Definition 1. An experiment $F \in \text{MPC}(H)$ is *incentive compatible* if there exists a restriction $\bar{F} \in \text{MPC}(H)$ such that $F \in \text{BR}(\bar{F})$. We denote the set of incentive-compatible experiments by \mathcal{F}^{IC} .

A useful observation is that an experiment F is incentive compatible if and only if $F \in \text{BR}(F)$; that is, when F is the imposed restriction, the experimenter finds it optimal not to garble the restriction. To see why, suppose $F \in \text{BR}(\bar{F})$ for some restriction \bar{F} , implying that F is optimal for the experimenter among the admissible experiments $\text{MPC}(\bar{F})$. Since $F \in \text{MPC}(\bar{F})$, all mean-preserving contraction of F are admissible too, that is, $\text{MPC}(F) \subseteq \text{MPC}(\bar{F})$. Thus, imposing F as a restriction instead of \bar{F} possibly shrinks the set of admissible experiments while F remains admissible. Naturally, F remains optimal among $\text{MPC}(F)$, implying $F \in \text{BR}(F)$.

The designer’s utility from an experiment F equals $\int u_D(m)dF(m)$. An incentive-compatible experiment is *optimal* if it maximizes the designer’s utility across the set of incentive-compatible experiments \mathcal{F}^{IC} .¹⁰ An optimal incentive-compatible experiment exists; see [Appendix A.1](#).

2.1 Examples of Restrictions

We illustrate through examples how our framework captures restrictions on real-world information provision. In these examples, we consider the experimenter’s available in-

⁸Note that $\text{BR}(\bar{F})$ is non-empty as $\text{MPC}(\bar{F})$ is compact and $F \mapsto \int G(m)dF(m)$ is continuous.

⁹Interpreting the restriction \bar{F} as a fictitious prior, we explain how our analysis relates to a result of [Curello and Sinander \(2025\)](#). Their Proposition 5 implies that there is no shift of the prior \bar{F} that makes $\text{BR}(\bar{F})$ more informative (in the weak set order) for all S -shaped G . This result has no direct bearing on our analysis since we fix G . Moreover, as we will highlight in [Section 4](#), optimizing over restrictions \bar{F} entails optimizing over a certain set of incomparable experiments.

¹⁰This optimality notion assumes that the experimenter breaks ties in favor of the designer.

formation and possible communication more explicitly using the language of signals. A *signal* consists of a measurable space \mathcal{S} and a random variable $\pi: [0, 1] \times [0, 1] \rightarrow \mathcal{S}$; here, $\pi(\omega, z)$ represents the signal realization given the state ω and a uniformly distributed random variable z that is independent of ω . It is apparent how choices of signals and restrictions on them map back to our reduced-form notation in terms of distributions of the posterior means and mean-preserving contractions. For any signal (\mathcal{S}, π) , let $F_{(\mathcal{S}, \pi)}$ be the induced distribution of posterior means. Any garbling of (\mathcal{S}, π) induces a mean-preserving contraction of $F_{(\mathcal{S}, \pi)}$.

Privacy constraints. An important class of restrictions arises through privacy notions studied in the literature and applied in practice (see, for example, [Strack and Yang, 2024](#); [Dwork et al., 2006](#), for the notion of privacy sets and differential privacy, respectively). Such constraints define sets of signals that are closed with respect to garblings and are therefore captured by our model. We make this connection more explicit in [Section 6](#).

Restrictions on communication. Consider a model in which the designer restricts what the experimenter can reveal about the state. Suppose that the designer first chooses a set Π of signals. Then, the experimenter can choose any signal from Π or a Blackwell-garbling thereof. With the initial choice of Π , the designer effectively places an upper bound on what the experimenter can reveal about the state, even if the experimenter can observe the state perfectly. Now, if the experimenter finds it optimal to induce an experiment F via some signal in Π , then we must have $F \in \text{BR}(F)$ since any mean-preserving contraction of F can also be induced via some signal in Π .¹¹ Thus, in our model the designer can implement F by imposing F as a restriction. This argument also shows that nothing is gained by imposing multiple restrictions, i.e. restricting the experimenter to $\bigcap_{\bar{F} \in \mathcal{F}} \text{MPC}(\bar{F})$ for a collection \mathcal{F} of restrictions. The experimenter will choose some F satisfying $F \in \text{BR}(F)$, meaning it suffices to impose F directly.

For example, suppose that the state is distributed uniformly on $[0, 1]$. A legislator requires a three-category claim structure with $\mathcal{S} = \{i, sg, vg\}$ (innocent, somewhat guilty, and very guilty) and imposes the deterministic map

$$\pi(\omega) = \begin{cases} i, & \text{if } \omega \in [0, 1/3) \\ sg, & \text{if } \omega \in [1/3, 2/3) \\ vg, & \text{if } \omega \in [2/3, 1]. \end{cases} \quad (1)$$

This corresponds to a restriction $\bar{F} = \frac{1}{3}\delta_{1/6} + \frac{1}{3}\delta_{1/2} + \frac{1}{3}\delta_{5/6}$. The prosecutor can however

¹¹See [Lemma 8](#) in [Appendix OA](#) for the details.

choose to communicate coarser by merging sg and vg into a joint guilty signal g . The resulting experiment F would correspond to $F = \frac{1}{3}\delta_{1/6} + \frac{2}{3}\delta_{2/3}$.¹²

Restrictions on measurement. Consider a model in which the designer restricts what the experimenter can measure about the state. Suppose that the designer can choose to reveal only partial information about the state to the experimenter, that is, the designer chooses a signal $([0, 1], \pi)$ in which the signal space coincides with the state space $[0, 1]$. The experimenter then chooses an arbitrary signal (S', π') that defines what the DM will eventually observes. However, the input into the experimenter’s signal is not the realized state ω but the signal realization that the experimenter receives through the designer’s signal. The experimenter can then induce any experiment that is a mean-preserving contraction of the experiment \bar{F} that is induced by $([0, 1], \pi)$. Thus, in our model, this situation is captured by imposing \bar{F} as a restriction.

For example, a regulator specifies the measurement technology that a producer must use to estimate a product’s safety. Suppose that the underlying state (the product’s safety) ω is uniformly distributed on $[0, 1]$. Then, one restriction is that the measurement technology can only produce binary outcomes $\pi \in \{0, 1\}$ (safe or not safe) with $\Pr(\pi = 1 | \omega) = \omega$ and $\Pr(\pi = 0 | \omega) = 1 - \omega$. Thus, the regulator’s restriction corresponds to $\bar{F} = \frac{1}{2}\delta_{1/3} + \frac{1}{2}\delta_{2/3}$. The producer can then process these estimated states further. For instance, the producer can hide the test result with probability $1 - \alpha$ and truthfully reveal it with probability α . This induces the experiment $F = (1 - \alpha)\delta_{1/2} + \frac{\alpha}{2}\delta_{1/3} + \frac{\alpha}{2}\delta_{2/3}$.

3 Benchmark: Full Delegation

In this section, we discuss the benchmark in which the designer fully delegates the experiment choice to the experimenter, that is, when $\bar{F} = H$. Hence, there is no restriction in place and the experimenter chooses among all mean-preserving contractions of the prior H , $MPC(H)$. The results in [Kolotilin et al. \(2022\)](#) imply that upper censorship of the prior is the unique best reply of the experimenter: the experimenter reveals all states below a threshold x , and pools all remaining states to an atom $y = \mathbb{E}_H[\omega | \omega \geq x]$. The dashed blue experiment F depicted in [Figure 1](#) is an instance of an upper censorship experiment.

¹²Note that the experimenter is not restricted to a weakly lower number of messages in our environment. The experimenter is permitted to use an arbitrary amount of messages, but these are still restricted to be a mean-preserving contraction of \bar{F} . For example, the prosecutor can create further subcategories of the guilt level by post-processing the maximal signals $\pi(\omega)$ given ω . However, the prosecutor is cannot reveal more information than by directly reported $\pi(\omega)$. Post-trial audits or preventing certain types of evidence can be used to prevent such claims.

Definition 2. An experiment F is *upper censorship* (of the prior H) with threshold x and atom y if F coincides with the prior H on the interval $[0, x]$ and assigns mass $1 - H(x)$ to the point $y = \mathbb{E}_H[\omega \mid \omega \geq x]$.

We restate the characterization from [Kolotilin et al. \(2022\)](#) of the experimenter’s best reply.

Proposition 1 ([Kolotilin et al. \(2022\)](#)). *There is a unique best reply F^* to H . In particular, F^* is an upper censorship experiment with a threshold x^* and an atom $y^* = \mathbb{E}_H[\omega \mid \omega \geq x^*]$ satisfying $0 < x^* < r_0 < y^* < 1$ and*

$$G(x^*) + g(y^*)(y^* - x^*) = G(y^*). \quad (2)$$

We recall the intuition from [Kolotilin et al. \(2022\)](#). The atom y^* lies in the concave region of the experimenter’s payoff G . Due to this concavity, the experimenter optimally pools some states around y^* to y^* . As pooling additional states into y^* raises the probability of the pooling outcome, the experimenter also adds some states from the convex part into the pooling region. [Equation \(2\)](#) is the first-order condition that characterizes the optimal pooling region, trading off the value of revelation in the convex region against the likelihood of the pooling outcome. [Figure 2](#) depicts this optimality condition and illustrates the optimal upper censorship experiment.

Importantly, the threshold x^* is interior: the experimenter reveals a non-trivial interval of states. The interiority of x^* follows from [Assumption 2](#). If this assumption fails, the results in [Kolotilin et al. \(2022\)](#) imply that the uninformative experiment δ_μ is the unique best reply to H .

When the designer can pick a restriction, [Assumption 2](#) remains necessary for information transmission. Trivially, δ_μ is admissible under all restrictions and imposing a restriction only shrinks the set of experiments from which the experimenter can choose. Thus, if the experimenter’s best reply to the prior H is unique and uninformative—i.e. $\delta_\mu = \text{BR}(H)$ —, then δ_μ is also the unique best reply to all restrictions. For this reason, we maintain [Assumption 2](#).

The optimal upper censorship experiment provides the relevant benchmark for the designer. The designer can always guarantee the payoff from the optimal censorship experiment by full delegation, not restricting the set of admissible experiments. However, under upper censorship a significant set of states is pooled into a point and, hence, substantial information is lost. We next study which alternative experiments the designer can induce by introducing non-trivial delegation sets.

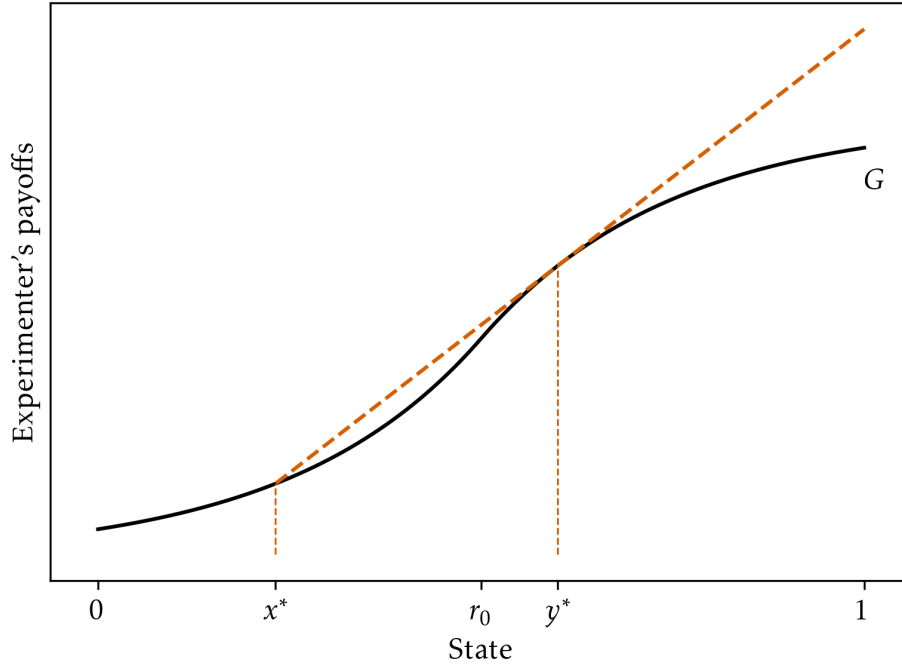


Figure 2: Equation (2) defines the optimal upper censorship experiment, illustrated by the red dotted line. The tangent to G at y^* intersects G at x^* . Since $x^* < r_0$, some states in the convex region of the experimenter's payoff are also pooled into the high atom y^* .

4 Maximal Incentive-Compatible Experiments

In this section, we first show that it suffices to consider a subset of incentive-compatible experiments that we call *maximal*. Our first main result characterizes the entire set of maximal incentive-compatible experiments.

4.1 Motivation and Definition

Incentive-compatible experiments represent all experiments that the designer can implement with some delegation set. However, some incentive-compatible experiments are clearly suboptimal. For example, the uninformative experiment δ_μ is straightforwardly incentive-compatible: when the designer imposes δ_μ as a restriction, the delegation set contains only the uninformative experiment δ_μ and the experimenter can only choose δ_μ . However, the designer never finds it optimal to implement δ_μ since δ_μ is strictly less informative than the full delegation outcome F^* , and since the designer has strictly convex payoffs.

This observation suggests to refine the candidate set of optimal experiments by considering only those incentive-compatible experiments that are not mean-preserving contrac-

tions of other incentive-compatible experiments. We next formalize this refinement and show that it is indeed without loss for optimality.

Definition 3. An incentive-compatible experiment F is *maximal* (MIC) if there does not exist an incentive-compatible experiment \tilde{F} of which F is a proper mean-preserving contraction; that is, there does not exist an incentive-compatible experiment \tilde{F} such that $F \in \text{MPC}(\tilde{F})$ and $\tilde{F} \notin \text{MPC}(F)$.

Economically, the set of MIC experiments is the relevant boundary of what the designer can implement. An MIC experiment represents a restriction that the experimenter has no incentive to garble, whereas the experimenter would garble every strictly more informative restriction.

The full delegation outcome F^* is an example of an MIC experiment. To see why, let F be an incentive-compatible experiment of which F^* is a mean-preserving contraction. On the one hand, the assumption that F is incentive compatible and a mean-preserving spread of F^* implies that the experimenter weakly prefers F to F^* . On the other hand, [Proposition 1](#) asserts that the experimenter strictly prefers F^* over all other experiments. Hence, the experiments F and F^* must be identical, implying that F^* is MIC.¹³

The following lemma justifies restricting attention to MIC experiments.

Lemma 1. *For all incentive-compatible experiments F there exists a maximal incentive-compatible experiment \hat{F} that is a mean-preserving spread of F .*

[Lemma 1](#) implies that it is without loss to focus on MIC experiments when searching for restrictions delivering optimal delegation sets. The MIC experiment \hat{F} is incentive compatible. Hence, the experimenter will optimally choose the experiment \hat{F} when the designer's restriction is \hat{F} . Further, since \hat{F} is a mean-preserving spread of F , the designer and DM, who have convex posterior-mean payoffs, prefer \hat{F} over F .

To gain intuition for the existence argument, suppose F is not MIC. By definition, this means there is an incentive-compatible experiment that is a mean-preserving spread of F . If this experiment is also not MIC, there is a further incentive compatible mean-preserving spread, and so on. A continuity-compactness argument shows that this reasoning delivers an MIC experiment \hat{F} that is a mean-preserving spread of F .

The notion and optimality of MIC experiments highlight how the designer may potentially benefit from imposing a restriction despite having a preference for information. All

¹³In environments in which the experimenter's best-reply set under full delegation, $\text{BR}(H)$, is multi-valued, all designer-preferred experiments in $\text{BR}(H)$ must be MIC if the designer's posterior-mean payoffs are strictly convex.

MIC experiments are incomparable to one another, by definition of maximality. Since MIC experiments are without loss, the designer thus trades-off informativeness about different regions of the state space when choosing among MIC experiments. In particular, since the full delegation outcome F^* is also an MIC experiment, the candidate set of experiments is incomparable to F^* .

Remark 1. The experiment \hat{F} in [Lemma 1](#) Pareto dominates F . The DM and the designer with convex payoffs prefer \hat{F} to F since $F \in \text{MPC}(\hat{F})$. Turning to the experimenter, incentive compatibility of \hat{F} implies that the experimenter finds \hat{F} optimal when \hat{F} is given as a restriction. In particular, since $F \in \text{MPC}(\hat{F})$, the experimenter prefers \hat{F} to F .

4.2 Double Censorship

In our main characterization result, we show that all MIC experiments take the form of *double censorship* experiments. We introduce them formally in the definition below. Intuitively, a double-censorship experiment divides the state space into three intervals: a fully revealing interval, an intermediate pooling interval, and a high pooling interval. [Figure 3](#) depicts the ICDF of a double-censorship experiment.

Definition 4. An experiment F is *double censorship* with thresholds (s, t) and atoms (x, y) if $s \leq t$ and $x = \mathbb{E}_H[\omega \mid \omega \in [s, t]]$ and $y = \mathbb{E}_H[\omega \mid \omega \in [t, 1]]$ and

$$\forall m \in [0, 1], \quad F(m) = \begin{cases} H(m) & \text{if } m \in [0, s), \\ H(s) & \text{if } m \in [s, x), \\ H(t) & \text{if } m \in [x, y), \\ 1 & \text{if } m \in [y, 1]. \end{cases}$$

Each upper censorship experiment, including the full-delegation outcome F^* , is a special double censorship experiment with $s = x = t$, and such that all states below x are fully revealed while all states above x are pooled to $y = \mathbb{E}_H[\omega \mid \omega \in [x, 1]]$.

4.3 Characterization

Our first main result shows that all MIC experiments take the form of double censorship. To do so, we introduce the set P , which relates the two atoms of a double censorship experiment to each other through a tangency condition similar to the one in [Proposition 1](#)

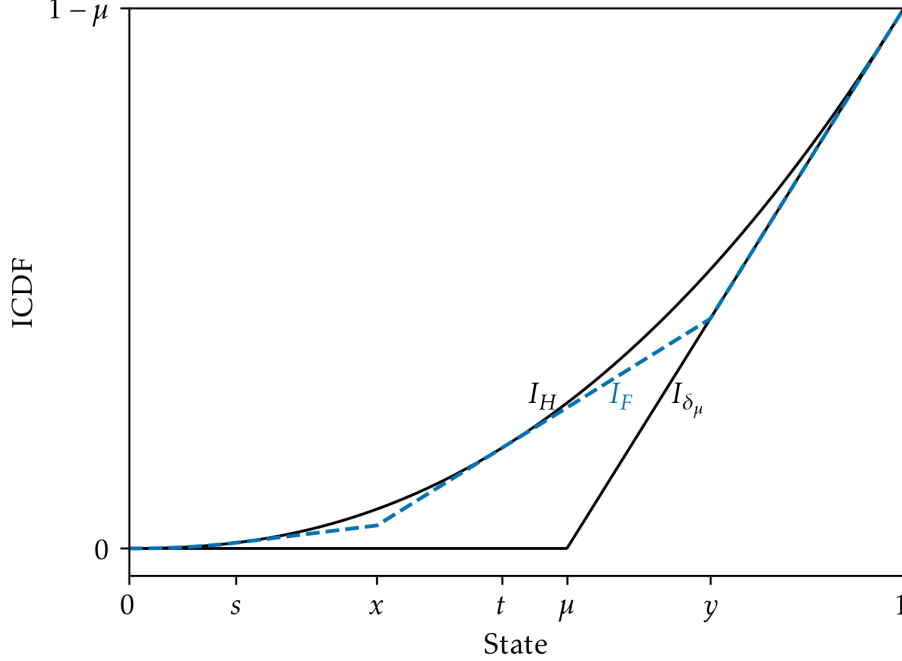


Figure 3: The ICDF of double censorship F with thresholds (s, t) and atoms (x, y) .

and depicted in Figure 2. We will see that this tangency condition characterizes incentive-compatibility of double censorship experiments. Formally, let

$$P = \{(x, y): 0 \leq x \leq r_0 \leq y \leq 1 \text{ and } G(x) + g(y)(y - x) = G(y)\}. \quad (3)$$

Theorem 1. *An experiment F is MIC if and only if F is a double censorship experiment with thresholds (s, t) and atoms (x, y) satisfying*

$$(x, y) \in P \text{ and } s \leq x^* \leq t \text{ and } 0 < x < y < 1,$$

where x^* denotes the threshold of the upper censorship experiment under full delegation.

Theorem 1 shows that double censorship experiments are the only candidates for MIC experiments. MIC experiments are characterized by interpretable properties of the thresholds s and t , and the atoms x and y . The first property, $(x, y) \in P$, characterizes incentive-compatibility for double censorship experiments, as we explain in more detail in the discussion of the proof. The second property, $s \leq x^* \leq t$, reflects maximality and highlights how a given MIC experiment F compares to the full-delegation outcome F^* . On the one hand, the pooling region $[t, 1]$ is contained in the full-delegation pooling region $[x^*, 1]$. On the other hand, there is a new pooling region $[s, t]$ vis-à-vis full revelation of states below x^* under full delegation. Finally, the inequalities $0 < x < y < 1$ assert that x and y are indeed distinct atoms and imply that F is non-degenerate.

Theorem 1 shows how each given MIC experiment compares to the full delegation outcome F^* (itself an MIC experiment). More generally, choosing among MIC experiments entails trading off the experiment's informational content about different regions of the state space. We can express this trade-off via a total order:

Corollary 1. *For all two MIC double censorship experiments with respective parameters (s, t, x, y) and (s', t', x', y') , if $y' < y$, then $[s', t'] \subsetneq [s, t]$ and $[t, 1] \subsetneq [t', 1]$ and $x < x'$.*

Thus, the designer can shrink the upper pooling region $[t, 1]$ (an informational gain) only by expanding the lower pooling region $[s, t]$ (an informational loss). This trade-off is driven by the need to maintain the constraint $(x, y) \in P$ that reflects incentive compatibility. **Corollary 1** implies that MIC experiments are a one-parameter family indexed and ordered by the top atom y . Given y , we compute (s, t, x) using the condition $(x, y) \in P$ from **Theorem 1** and the definition of x and y as conditional expectations for double censorship experiments.

To gain intuition for the designer's trade-off, consider the *experimenter's* trade-off when choosing an upper censorship threshold under the prior H . The experimenter's choice of the upper censorship threshold x^* trades-off the location of the atom $y^* = \mathbb{E}_H[\omega \mid \omega \in [x^*, 1]]$ versus the probability assigned to the atom, $1 - H(x^*)$. This trade-off is captured by the first-order condition (2) that defines the set P . If the designer imposes some restriction \bar{F} , the designer effectively imposes \bar{F} as the prior, thereby affecting the trade-off between the location of the atom and the probability assigned to it. To be more concrete, suppose \bar{F} obtains from H by pooling a small interval $(s, t]$ around x^* to x , where x lies *below* x^* , i.e. $s < x < x^* < t < y^*$. Suppose the experimenter would continue to reveal all states weakly below x^* while pooling all states strictly above x^* , and when states are drawn from the fictitious prior \bar{F} . Since \bar{F} pools states in $[x^*, t]$ to a point x below x^* , the new atom lies above y^* —i.e. $y = \mathbb{E}_{\bar{F}}[\omega \mid \omega \in [x^*, 1]] > y^*$ —, and the probability assigned to y is below the probability assigned to y^* —i.e. $1 - \bar{F}(x^*) = 1 - H(t) < 1 - H(x^*)$. Therefore, from the perspective of the experimenter's trade-off, the atom y is now (weakly) too high and the probability assigned to y is (weakly) too low. Therefore, the experimenter's best reply is to (weakly) reduce the upper censorship threshold from x^* . In fact, we can show that, if the points x and y are precisely such that $(x, y) \in P$, as in the experiments described by **Theorem 1**, then a new optimal upper censorship threshold is x^* itself (and that the resulting experiment is optimal among all experiments that are feasible given \bar{F}). Given the construction of \bar{F} , revealing all states weakly below x^* while pooling all strictly states above x^* effectively means that, when states are drawn according to H , the experimenter reveals all states weakly below s , while pooling all states in the respective intervals $(s, t]$

and $(t, 1]$. We have obtained a double censorship experiment.

4.4 Sketch of the Proof

In the following, we outline the proof of [Theorem 1](#), focusing on the characterization of an arbitrary MIC experiment as a double censorship experiment. In the appendix, we also prove the other direction that all double censorship experiments with the properties in [Theorem 1](#) are MIC experiments.

First, we characterize incentive-compatible experiments in our environment using the price-theoretic approach of [Dworczak and Martini \(2019\)](#).

Lemma 2. *An experiment F is IC if and only if there exists a convex, continuous function $p: [0, 1] \rightarrow \mathbb{R}$ such that $p \geq G$ and $\text{supp } F \subseteq \{m \in [0, 1]: p(m) = G(m)\}$.*

Proof of Lemma 2. [Dworczak and Martini \(2019, Corollary 1\)](#) provide the following necessary and sufficient condition for an experiment F to be optimal in an unconstrained persuasion problem when the prior is some distribution \bar{F} : *Given a prior \bar{F} and $F \in \text{MPC}(\bar{F})$, it holds $F \in \text{BR}(\bar{F})$ if and only if there exists a convex, continuous function $p: [0, 1] \rightarrow \mathbb{R}$ such that $p \geq G$ and $\text{supp } F \subseteq \{m \in [0, 1]: p(m) = G(m)\}$ and $\int p(m) dF(m) = \int p(m) d\bar{F}(m)$.*

Our lemma follows from this result and recalling that an experiment F is incentive-compatible if and only if F is a best reply to itself, that is, if $F \in \text{BR}(F)$. Thus, F would be optimal in the full-delegation problem if F was the prior. Indeed, if $F = \bar{F}$, then the integral condition $\int p(m) dF(m) = \int p(m) d\bar{F}(m)$ trivially holds.¹⁴ \square

To further characterize incentive-compatible experiments, we combine [Lemma 2](#) with S -shapedness of the experimenter's payoff.

Lemma 3. *A non-degenerate experiment F is incentive-compatible if and only if there exist $(x, y) \in P$ such that F assigns no mass to the set $(x, y) \cup (y, 1]$.*

To gain intuition, suppose F is an incentive-compatible experiment. Then, F cannot assign mass to multiple states in the concave part $[r_0, 1]$ of the experimenter's payoff G since the experimenter would strictly prefer to pool such states. Thus, there is at most one point y in support of F above r_0 . In fact, as discussed in the context of full delegation in [Section 3](#), the experimenter would also pool some states just below the inflection point

¹⁴[Lemma 2](#) does not rely on S -shapedness of G . To apply Corollary 1 of [Dworczak and Martini \(2019\)](#), the following suffices: G is upper semicontinuous with at most finitely many discontinuities at interior points $y_1, \dots, y_k \in (0, 1)$, and is Lipschitz continuous in each interval (y_i, y_{i+1}) , with $y_0 = 0$ and $y_{k+1} = 1$; this is part (i) of the regularity condition of [Dworczak and Martini \(2019\)](#).

r_0 to shift more probability mass to the pooling outcome. The point $x < r_0$ represents the cutoff value where the experimenter finds it optimal to pool no further, as reflected in the condition $G(x) + g(y)(y - x) = G(y)$ that defines the set P (recall (3)). We conclude that F cannot assign mass to $(x, y) \cup (y, 1]$. Note, however, that we cannot yet conclude that, as in double censorship, y is obtained by pooling the prior above some threshold, or that x or y are even in the support of F .

Lemma 3 shows that incentive-compatibility is fully characterized by regions of the state space on which an experiment does not assign any mass. Thus, any MIC experiment should reveal as much information as possible while respecting these forbidden regions. We next show how double censorship emerges from combining incentive-compatibility with maximality considerations. To that end, we first prove that the set P , which characterizes the forbidden regions, is ordered by set inclusion.

Lemma 4. *If $(x, y) \in P$ and $(x', y') \in P$, then $[x, y] \subseteq [x', y']$ or $[x', y'] \subseteq [x, y]$.*

This lemma directly follows from the S-shapedness of G , as illustrated in [Figure 4](#). Whenever there is a feasible (x, y) -combination, that is, one that satisfies (3), the tangent of the experimenter's value at y has to intersect the experimenter's value at x . As $y \geq r_0$, the point y lies in the concave part of the experimenter's value. Now, consider a marginal increase in y . By concavity, the tangent of the experimenter's value becomes flatter. Hence, the intersection with the experimenter's value must occur at a lower value. It follows that the intervals $[x, y]$ defined by the elements of the set P are nested.

Next, we derive the remaining properties of the characterization of MIC experiments that result from maximality. Consider an arbitrary MIC experiment F . Incentive-compatibility implies, by [Lemma 3](#), that F does not allocate any mass to $(x, y) \cup (y, 1]$, which implies that the ICDF I_F of the experiment F is affine on $[x, y]$ and $[y, 1]$. Maximality implies then that F is as informative as possible subject to not assigning any mass to the forbidden regions. Moreover, due to [Lemma 4](#), we only have to distinguish two cases for how any MIC experiment F relates to the full-delegation outcome F^* : the experiment F is such that $[x, y] \subseteq [x^*, y^*]$ or it is such that $[x^*, y^*] \subseteq [x, y]$.

First, suppose $[x, y] \subseteq [x^*, y^*]$. In this case, we show that F must be an MPC of the full-delegation outcome F^* and therefore not maximal unless $F = F^*$. We illustrate graphically in [Figure 5a](#). As the experiments F and F^* assign no mass to $[y, 1]$ and $[y^*, 1]$, respectively, their ICDFs both hit the ICDF of the uninformative experiment at y and y^* , respectively. In particular, since $y^* \geq y$, we find $I_F(y^*) = I_{F^*}(y^*)$. In this sense, the experiment F pools weakly more states at the top than F^* . Moreover, I_F lies weakly below I_{F^*} on $[0, x^*]$ since the full-delegation F^* outcome fully reveals all states in this region. Therefore, F^* is weakly

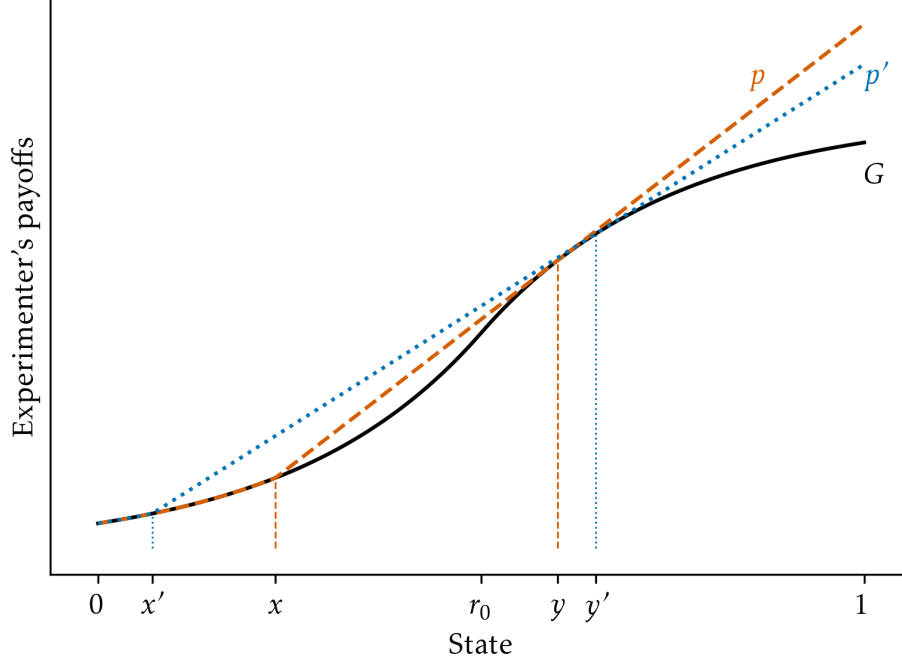
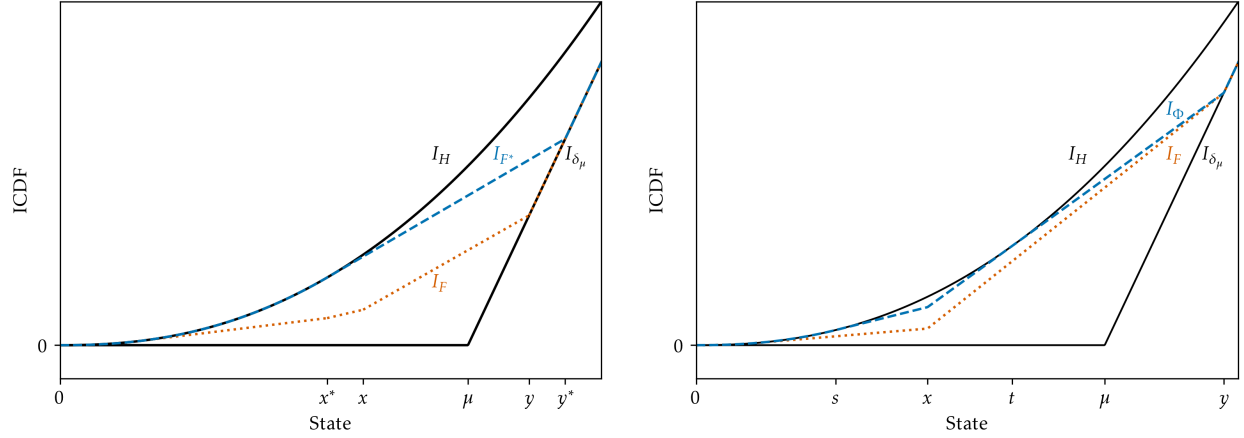


Figure 4: *The order of the set P . For a higher value of y , the slope of G at y is lower, meaning the tangent to G at y intersects G below r_0 at a lower point x .*

more informative at the bottom as well. Finally, note that I_{F^*} is affine on $[x^*, y^*]$, while I_F is convex. Since $I_F(x^*) \leq I_{F^*}(x^*)$ and $I_F(y^*) = I_{F^*}(y^*)$, we obtain $I_F(m) \leq I_{F^*}(m)$ for all $m \in [x^*, y^*]$. Summarizing, we find that I_F lies below I_{F^*} on all of $[0, 1]$, implying that F is an MPC of F^* . The experiment F^* is incentive-compatible. Thus, for F to be maximal, we must have $F = F^*$, implying that F is upper censorship, which is a special case of double censorship.

Second, suppose $[x^*, y^*] \subseteq [x, y]$. We illustrate this case graphically in [Figure 5b](#). In this case, I_F hits the ICDF of the uninformative experiment at $y \geq y^*$ and, hence, after I_{F^*} . It follows from incentive-compatibility that I_F is affine on $[x, y]$ and coincides with I_H at most once on this interval. For the purpose of this sketch, suppose I_F lies strictly below I_H on $[x, y]$. In this case, we can obtain a new experiment Φ by rotating the affine segment of I_F on $[x, y]$ clockwise, while anchoring it $I_F(y)$, until it is tangent to I_H at some point t . Then, we can find a point s such that the tangent to I_H at s intersects the new affine segment at x . Finally, we let Φ reveal all states below s . In the proof, we show that $[x^*, y^*] \subseteq [x, y]$ guarantees the existence of such points s and t . The constructed Φ is incentive-compatible, a mean-preserving spread of F , and a double-censorship experiment. Incentive-compatibility follows from [Lemma 3](#) as Φ does not allocate mass to the set $(x, y) \cup (y, 1]$, while $\Phi \in \text{MPS}(F)$ follows since $I_\Phi \geq I_F$. Hence, for F to be maximal, we must have $F = \Phi$, implying that F is double censorship.



(a) An incentive-compatible experiment F that is an MPC of F^* .

(b) An incentive-compatible experiment F and a double censorship experiment Φ that is an MPS of F .

Figure 5: Characterization of MIC experiments. For readability, the horizontal axes are truncated at the top.

5 Optimal Delegation

5.1 Full delegation is not optimal

We now revisit the designer's problem that is at the core of our environment. Anticipating the persuasion motive of the experimenter, can the designer gain from restricting the set of experiments available to the experimenter even though the designer has a preference for informative experiments? To answer this, [Lemma 1](#) implies that it suffices to consider restrictions that are MIC experiments, i.e. the experimenter has no incentive to garble the restriction but would garble any more informative restriction. We now apply our characterization of MIC experiments as double censorship experiments. The following theorem is our second main result.

Theorem 2. *The full delegation outcome F^* is not optimal.*

We prove [Theorem 2](#) using a local perturbation argument at the full-delegation solution. Recall that the family of MIC experiments is a one-dimensional family parametrized by the atom y in the top pooling interval ([Corollary 1](#)). Moreover, full delegation leads to the upper censorship experiment with $y = y^*$ and $s = t = x^*$. Denote by $U_D(y)$ the designer's value from choosing the double censorship experiment with y as restriction. Due to [Corollary 1](#), $y \geq y^*$. Hence, we consider the right-derivative of $U_D(y)$ and evaluate it at

$y = y^*$. After simplifications, we obtain

$$\left(u'_D(y^*) - \frac{u_D(y^*) - u_D(x^*)}{y^* - x^*} \right) (1 - H(x^*)) > 0, \quad (4)$$

which is strictly positive since u_D is strictly convex and $x^* < y^*$.

To interpret this derivative, consider first the following hypothetical scenario: the designer is presented with a binary experiment with support $\{x^*, y^*\}$ and mean μ , and can costlessly perturb the support subject to the mean constraint. The derivative (4) equals (up to a multiplicative factor) the change in the designer's payoff when increasing y^* marginally while holding x^* fixed and adjusting the probability of the realization y^* to satisfy the mean constraint. The term $u'_D(y^*)$ reflects the marginal increase in the value of the atom y^* , and the term $\frac{u_D(y^*) - u_D(x^*)}{y^* - x^*}$ reflects the mass that is shifted from y^* to x^* to maintain the mean. The designer gains from this perturbation because an increase in y^* while holding x^* fixed makes the binary experiment more informative.

In our actual model, the designer cannot freely pick the experiment but must additionally satisfy incentive compatibility. To achieve this, the designer introduces the double censorship thresholds s and t . Since the states in the interval $[s, t]$ are now pooled to x , there is a cost to the perturbation. Importantly though, we show that this cost is of second order. Intuitively, the thresholds s and t remain close to x^* for any small perturbation, and hence, are an order of magnitude smaller than $y^* - x^*$. Formally, this cost is

$$\int_{[s,t]} (u_D(x) - u_D(m)) dH(m), \quad (5)$$

reflecting the informational loss from pooling states to x instead of revealing them. For s and t close to x^* , this integral is an order of magnitude smaller than the gain reflected in (4).

Remark 2. There exists a solution to the designer's problem, and it is attained by an incentive-compatible double censorship experiment. Viewed as a best reply to itself, this experiment is a fictitious prior with two atoms. [Theorem 2](#) does not apply to this fictitious prior since the theorem assumes an absolutely continuous prior.

5.2 A Parametric Example

To illustrate our results, we consider a parametrized example. Suppose that the state is uniformly distributed on the unit interval, that is, $H(\omega) = \omega$, and that the outside option distribution is a *Beta*(2,2) distribution, that is, $G(r) = 3r^2 - 2r^3$ and $g(r) = 6r(1 - r)$ with

$r_0 = 1/2$.¹⁵ The designer's and the DM's payoffs coincide.

Full-delegation benchmark. First, we consider the full-delegation benchmark, that is, the persuasion solution. By [Proposition 1](#), the optimal experiment will be upper censorship (with cutoff x^* and top atom $y^* = \mathbb{E}[\omega \mid \omega \geq x^*]$) and determined through the tangency condition. Due to the uniform assumption, we obtain for the top atom $y^* = (x^* + 1)/2$ (and equivalently $x^* = 2y^* - 1$). Applying the tangency condition (2) yields the optimal upper censorship experiment:

$$x^* = \frac{1}{4}, \quad y^* = \frac{5}{8}.$$

MIC experiments. By [Theorem 1](#), we can restrict attention to double censorship experiments. With our parametric assumptions, we obtain that for a given top atom y , the double-censorship experiment is characterized by

$$x(y) = \frac{3}{2} - 2y, \quad t(y) = 2y - 1, \quad s(y) = 4 - 6y$$

as long as the solution is feasible (i.e., $x(y) \leq x^*$, $0 < x < y < 1$, and $s \geq 0$). Thus, we can characterize the set of implementable top atoms as the constraint

$$y \in \left[\frac{5}{8}, \frac{2}{3} \right].$$

Optimal delegation. For any feasible double-censorship experiment with top atom y , the designer's (and DM's) ex-ante payoff is

$$\begin{aligned} U_D(y) &= \int_0^{s(y)} \int_0^\omega (\omega - r) dG(r) dH(\omega) \\ &\quad + (H(t(y)) - H(s(y))) \int_0^{x(y)} (x(y) - r) dG(r) \\ &\quad + (1 - H(t(y))) \int_0^y (y - r) dG(r). \end{aligned}$$

Applying the expressions for $s(y)$, $x(y)$ and $t(y)$, we obtain that $U_D(y)$ is strictly increasing on the feasible range of y . Hence, the optimal restriction is $y^{\text{opt}} = 2/3$ and the corresponding optimal double-censorship experiment is characterized by

$$s^{\text{opt}} = 0, \quad x^{\text{opt}} = \frac{1}{6}, \quad t^{\text{opt}} = \frac{1}{3}, \quad y^{\text{opt}} = \frac{2}{3}.$$

¹⁵To ensure that $g(0) > 0$ and $g(1) > 0$ to satisfy [Assumption 1](#), we could mix the $Beta(2, 2)$ with a uniform distribution and make the uniform component vanishingly small. This does not affect the results and for this reason, we suppress this technical detail.

The optimal delegation of information provision induces simple information provision endogenously: a binary signal whether the state is above or below a threshold of $t^{\text{opt}} = 1/3$ is the optimal experiment.

6 Application: Endogenous Privacy Constraints

In this section, we show how our framework gives rise to endogenous privacy constraints in the context of recommender systems. Algorithms are becoming increasingly important in making recommendations to decision makers; for example, in suggesting products to consumers. Recommendation algorithms often take as input user-specific data, such as their location, age, gender, or browsing history. Policy discussions frequently raise privacy concerns regarding the use of such data, suggesting a tension between information transmission and privacy protection. Surprisingly, our results imply the opposite: a user who has no intrinsic preference for privacy may be strictly better off under optimally designed privacy constraints than without privacy constraints. Even without intrinsic preferences for privacy, such privacy constraints may be optimal to protect consumers against persuasion without giving up on valuable information provision.¹⁶

In the following, we first introduce two prominent notions of privacy constraints. Second, we link these notions to restrictions in our model. Third, we provide a stylized model of a recommender system persuading a consumer to purchase a product; we show how the consumer can benefit from well-designed privacy constraints, even without intrinsic preferences for privacy. Fourth, we illustrate how privacy constraints can endogenously arise as a form of self-regulation of a hierarchical organization with misaligned preferences over information provision and persuasion.

6.1 Privacy Constraints

We consider two notions of privacy constraints: differential privacy (Dwork et al. (2006)) and privacy-preserving signals (Strack and Yang (2024)). To that end, we slightly extend our model. There is a random variable θ taking values in a measurable space Θ , which is correlated with the payoff-relevant state ω . In particular, we assume that $\omega = v(\theta)$ with v a strictly increasing function. We give an interpretation of this condition in Section 6.3, in which we apply our discussion of privacy constraints to recommender systems. The variable θ represents privacy-protected characteristics but is itself payoff-irrelevant. A

¹⁶Liang et al. (2026) consider the distinct but related issue of regulating algorithmic inputs to manage a trade-off between fairness and accuracy.

signal is a pair (\mathcal{S}, π) , where \mathcal{S} is a measurable space and $\pi: [0, 1] \times \Theta \times [0, 1] \rightarrow \mathcal{S}$ maps the realizations of the state ω , the privacy-protected characteristic θ , and a randomization device z to signals. As before, an experiment refers to a distribution of posterior means about the state. We next state the formalizations of the two privacy notions that we study.

Differential privacy. Given a function $\delta: \Theta^2 \rightarrow \mathbb{R}$, a signal (\mathcal{S}, π) satisfies *differential privacy* if for all $\theta, \theta' \in \Theta$ and measurable $S \subseteq \mathcal{S}$ it holds $\Pr(\pi \in S \mid \theta) \leq \Pr(\pi \in S \mid \theta')\delta(\theta, \theta')$.¹⁷ Intuitively, the likelihood ratio of θ and θ' is bounded by $\delta(\theta, \theta')$, bounding the inference one can draw from observing the event S . In practice, δ often reflects physical distances between characteristics.

Privacy-preserving signals. Let \mathcal{P} be a family of events P in Θ that is closed with respect to finite intersections, called privacy sets. A signal (\mathcal{S}, π) is *privacy preserving* if for all $P \in \mathcal{P}$ it holds $\Pr(\theta \in P \mid \pi) = \Pr(\theta \in P)$ almost surely; that is, the signal cannot reveal information about any privacy set. Moreover, given some event P_0 , a signal (\mathcal{S}, π) is *conditionally privacy-preserving* signal, if for all $P \in \mathcal{P}$ it holds $\Pr(\theta \in P \mid \pi, P_0) = \Pr(\theta \in P \mid P_0)$ almost surely; i.e. the privacy-constraint is required only conditional on the event P_0 .

6.2 Delegation Sets as Privacy Constraints

In the following, we show that the delegation sets we study are closely linked to privacy constraints as introduced above. To do so, we relate the objects defining the privacy constraints, δ and \mathcal{P} , to restrictions on the set of admissible experiments.

First, suppose that the experimenter can pick any signal satisfying differential privacy, for some choice of δ . Importantly, note that if a signal (\mathcal{S}, π) satisfies δ -differential privacy, then all Blackwell-garblings of this signal satisfy δ -differential privacy as well. Therefore, it follows that if the experimenter optimally picks a signal (\mathcal{S}, π) that induces an experiment F , then this experiment F is a best response to the restriction F , $F \in \text{BR}(F)$. To see why, note that the experimenter could have chosen any other signal inducing experiment $\hat{F} \in \text{MPC}(F)$, but preferred F . Hence, in our model, F is incentive compatible and the designer can implement the F by imposing F itself as a restriction. The same argument applies for privacy-preserving signals given some choice of privacy set \mathcal{P} , since the constraint set that \mathcal{P} induces is also closed with respect to garblings.¹⁸

¹⁷To be precise, this is a notion of *metric* differential privacy (Xie et al. (2025)), which generalizes differential privacy (Dwork et al. (2006)) by allowing δ to vary across characteristics.

¹⁸In fact, Strack and Yang (2024) show that the experiments, that can arise under some privacy set, correspond exactly to the MPCs of a certain experiment.

Second, we argue that optimal restrictions can be implemented via privacy constraints under appropriate conditions on the relation between the payoff-relevant state ω and the random variable θ on which the privacy constraints are imposed. Since the restrictions in our model concern the payoff-relevant state ω , the implementation is easiest when ω is strongly correlated with θ . Given the strictly monotonic mapping between the two random variables θ and ω that is known to the experimenter, imposing a privacy constraint on the individual-specific data θ is formally equivalent to imposing a privacy constraint on the estimated match payoff-relevant state ω .

The next corollary shows that any MIC experiment F can be implemented as a best reply to a simple restriction \bar{F} , which we shall use to relate restrictions to privacy constraints.

Corollary 2. *If F is MIC double censorship with parameters (s, t, x, y) , then F is a best reply to the restriction \bar{F} that is obtained by pooling H on $[s, t]$ to x and otherwise revealing the state.*

Using [Corollary 2](#), we can implement F via differential privacy as follows: Choose $\delta(\omega, \omega') = 1$ if ω and ω' are both in $[s, t]$, and otherwise put $\delta(\omega, \omega') = \infty$. The differential-privacy constraint reduces to requiring that the signal realization be constant in the payoff-relevant state ω whenever the state is in $[s, t]$. Thus, every differentially private signal induces an experiment (a distribution of posterior means for the state) that is a mean-preserving contraction of the experiment that reveals the state whenever the state is outside $[s, t]$, and otherwise only reveals that the state is in $[s, t]$. [Corollary 2](#) then implies that imposing this experiment as a restriction yields the given MIC experiment F as a best reply. Hence, the chosen differential privacy constraint implements F .

Next, we describe how to implement F via privacy sets and conditionally privacy-preserving signals. First, we choose as conditioning event P_0 the event that the payoff-relevant state ω falls into the interval $[s, t]$ and let \mathcal{P} be the family of all Borel subsets of $[s, t]$. Conditional on $\omega \notin [s, t]$, the privacy constraint holds trivially, since for all subsets P of $[s, t]$ the probability of the event $\omega \in P$ equals 0. Conversely, conditional on $\omega \in [s, t]$, the privacy constraint requires any signal to reveal no additional information beyond the state's falling into $[s, t]$. Therefore, every conditionally privacy-preserving signal induces an experiment that is a mean-preserving contraction of the experiment that reveals the state whenever the state is outside $[s, t]$, and otherwise only reveals that the state is in $[s, t]$. As before, it follows that the experimenter chooses the desired MIC experiment F when constrained to use signals that are conditionally privacy-preserving for \mathcal{P} .

6.3 Recommender Systems and Privacy Constraints

We now illustrate how consumers can strictly benefit from privacy constraints in the context of recommender systems, even in the absence of intrinsic preferences for privacy. A platform receives a commission for each sale and wishes to persuade consumers to purchase a product. As products are priced by sellers, the platform regards the price $p \in [0, 1]$ as exogenous. The platform has access to consumer-specific data that is generated, for example, from the consumers demographic characteristics and their browsing and purchase history. We assume that the consumer-specific information is the realization of a random variable θ which is distributed according to a full-support distribution T on $[0, 1]$. For a consumer described by θ , the platform can estimate her match value $\omega(\theta)$ from purchasing the product $\omega = v(\theta)$, where $v(\cdot)$ is a strictly increasing function.¹⁹ The consumer has an idiosyncratic preference component ε distributed according to a full-support distribution E , which she observes privately prior to her purchase decision. We assume that her value from not purchasing the object is zero. Hence, the consumer purchases the product given a belief m about her match value and price p if

$$m + \varepsilon - p \geq 0 \iff m \geq p - \varepsilon. \quad (6)$$

We can map this environment back into our framework by regarding $p - \varepsilon$ as the outside option. Then, the random outside option $r \equiv p - \varepsilon$ is distributed according G , where G is supported on $[0, 1]$ and satisfies [Assumption 1](#) whenever E has full support on $[p - 1, p]$ and is S -shaped on that interval. The distribution T of the random variable θ describing the consumer data gives rise to the distribution H of estimated match values when setting $T(t) = H(v(t))$ for $t \in [v^{-1}(0), v^{-1}(1)]$.

Optimal recommendations. An unregulated platform that maximizes its commissions will use the consumer data to persuade consumers to buy. Thus, it will implement an information policy as in [Proposition 1](#): send a simple *buy* recommendation for the highest consumer types, and give a more nuanced recommendation to the lower types revealing the estimated match value for the consumer.

Optimal privacy constraints. A regulator who may be concerned with consumers being persuaded too much by the platform’s recommendation system, may want to impose a privacy constraint, limiting how the platform can use consumer-specific data. [Theorem 2](#) shows that a designer aligned with consumers always imposes a strict restriction to

¹⁹Note that this assumes that the platform has sufficient information to predict the consumer’s match value perfectly. It is straightforward to relax this assumption by assuming that $\omega = v(\theta) + \eta$ where η is unobservable mean-zero noise orthogonal to the consumer type θ .

increase the consumer's payoff. One way to implement this restriction is to introduce a privacy constraint corresponding to the optimal restriction, as derived above. Hence, our framework reveals that such privacy constraints may arise endogenously not due to intrinsic privacy constraints but as a means to balance the tension between information provision and persuasion.

6.4 Privacy Constraints as Self-Regulation

In the preceding discussion, we assumed that the platform chooses its recommendations to maximize sales. This may be a good approximation of the objective under which engineers design the algorithm, yet, the platform may also be concerned with the consumer's payoffs. To ensure a sufficient customer base while retaining high profits, we assume that the platform maximizes a weighted sum of commissions and the consumer's payoffs. This corresponds to a designer maximizing a weighted sum of the experimenter's and DM's payoffs, and may lead to non-convex preferences of the designer. In the following, we show that even under such payoffs, the platform (designer) will optimally choose to restrict the algorithm (experimenter). Economically, we assume that the platform's leadership can constrain the data use of its recommendation algorithm.

Formally, given a posterior mean m about the match value, the consumer's interim expected payoff equals $\int_0^1 \max\{m - r, 0\} dG(r)$ for all $m \in [0, 1]$. Integrating by parts, we obtain for the consumer's payoff $I_G(m) = \int_0^m G(r) dr$, and we observe for later reference that this payoff is strictly convex in m . The platform's payoff from commissions remains $G(m)$. Thus, the platform's total payoff is $u_D(m) = \lambda I_G(m) + (1 - \lambda)G(m)$, for some weight $\lambda \in [0, 1]$ on the consumer's payoffs.

Under this objective, focusing on MIC experiments remains without loss for the platform. Indeed, in [Remark 1 \(Section 4.1\)](#) we noted that for every incentive-compatible experiment there exists a MIC and preferred to the given experiment by both the DM and experimenter. The characterization of MIC experiments (as double censorship) remains unchanged, as it is entirely independent of the platform's preferences.

Finally, the platform finds full delegation to the algorithm suboptimal provided that it assigns a non-zero weight λ to the consumer's payoff. To see this, recall the derivative (4) that we derived via a perturbation argument:

$$\left(u'_D(y^*) - \frac{u_D(y^*) - u_D(x^*)}{y^* - x^*} \right) (1 - H(x^*)).$$

From the characterization of the full delegation outcome in [Proposition 1](#), we know $g(y^*) = \frac{G(y^*) - G(x^*)}{y^* - x^*}$, reflecting the optimal size of the full delegation pooling region from

the algorithm’s (experimenter’s) perspective. Thus, the derivative from the perturbation simplifies to:

$$\lambda \left(G(y^*) - \frac{I_G(y^*) - I_G(x^*)}{y^* - x^*} \right) (1 - H(x^*)).$$

This derivative is strictly positive since the consumer’s payoff I_G is strictly convex. Thus, the platform wants to constrain the algorithm’s use of information.

Hence, in a hierarchical organization, privacy constraints may arise endogenously as a form of self-regulation when the preferences regarding information provision and persuasion across the organization are misaligned.

7 Beyond S-shaped Persuasion

This section first discusses how our analysis of MIC experiments and optimal restrictions extends beyond our S-shaped benchmark. Second, we also provide a connection between our problem and the extreme-point approach in the information-design literature.

7.1 Profitable Restrictions Beyond S-shaped Persuasion

In this subsection, we investigate whether the designer can benefit from imposing a restriction when the experimenter’s payoff G (i.e. the outside option distribution) is not necessarily S-shaped. Equivalently, under which conditions is full delegation suboptimal? We argue that two conditions are important for the profitability of restrictions and we formalize the discussion in [Appendix OB](#). First, the experimenter has *locally strict curvature preferences*: G is locally strictly convex or strictly concave, and satisfies a technical smoothness condition. Second, the full delegation outcome admits *incomplete full revelation*: under full delegation, the experimenter fully reveals a non-trivial interval of states, but does not reveal all states.

We begin with two examples to demonstrate the importance of both conditions. We then describe how to generalize the insights from these examples. In this discussion, the definitions of incentive compatibility and optimality are as in the baseline model.

Uninformed DM. Suppose the DM has no private information: the outside option distribution G is a step function with step at r_0 . In particular, the experimenter does not have locally strict preferences, being indifferent between revealing and obfuscating the states within the subintervals below and above r_0 , respectively.

In this example, the experimenter maximizes the probability of drawing a posterior mean above r_0 . If $\mu > r_0$, the experimenter can induce the DM to take action with probability one regardless of the restriction by choosing the experiment δ_μ ; it is then easy to see that the designer cannot gain by imposing a restriction. If $\mu \leq r_0$, then, for an arbitrary restriction \bar{F} and best reply F , the largest point in the support of F is at most r_0 . This implies that F is an MPC of the experiment that reveals all states up to a threshold s and pools the rest to r_0 (i.e. $r_0 = \mathbb{E}_H[\omega \mid \omega \in [s, 1]]$, which is feasible since $\mu \leq r_0$). This experiment, in turn, is a best reply to the prior H . Thus, full delegation is optimal. Ichihashi (2019, Corollary 1) makes the same observation.

***M*-shaped experimenter preferences.** Suppose the experimenter’s payoff G is *M*-shaped as in Figure 6:²⁰ concave on an interval of the lowest states, then convex, and finally concave on an interval of the highest states. In this example, G has locally strict preferences. We now discuss incomplete full revelation, which also involves the prior H of the state.

Figure 6a depicts a situation in which the binary experiment F^* that is supported on the peaks of the M is a valid MPC of H . Then, F^* is also the experimenter’s unique best reply under full delegation (apply Corollary 2 of Dworczak and Martini (2019) with the price function in Figure 6a). The full delegation outcome F^* does not admit incomplete full revelation since it does not fully reveal the state on any non-trivial interval. Let us argue why no profitable restriction exists. Lemma 2 implies that an arbitrary incentive-compatible experiment F must be supported between the two peaks, i.e. $\text{supp } F \subseteq [y_1, y_2]$.²¹ Among all such experiments, the binary experiment F^* with support $\{y_1, y_2\}$ (and mean μ) is the most informative. Thus, full delegation is optimal for the designer.

Conversely, suppose the prior H is such that the full delegation outcome F^* reveals states in an interval $[x_1, x_2]$ in the valley between the peaks, and pools the remaining states to two points, y_1 and y_2 , on the left and right slopes, as in Figure 6b. In particular, F^* admits incomplete full revelation, and F^* roughly looks like a lower censorship experiment and an upper censorship experiment patched together. Following our perturbation argument for upper censorship experiments from the *S*-shaped case, the designer can strictly improve on full delegation by pooling a small interval of states around each of x_1 and x_2 .

General experimenter preferences. The preceding examples suggest that profitable restrictions exist when the experimenter has locally strict preferences, and the full delega-

²⁰The depicted G is not a CDF. However, transformations of the form $m \mapsto aG(m) + bm$ with $a > 0$ do not alter the experimenter’s preferences over experiments since all experiments have mean μ .

²¹Indeed, since F^* is feasible, the prior mean μ lies between the two peaks. Thus, the support of F cannot be entirely to the left of the left peak or to the right of the right peak. Since all price functions are convex, and since F is supported on a set where a price function touches G , the support of F lies between the peaks.

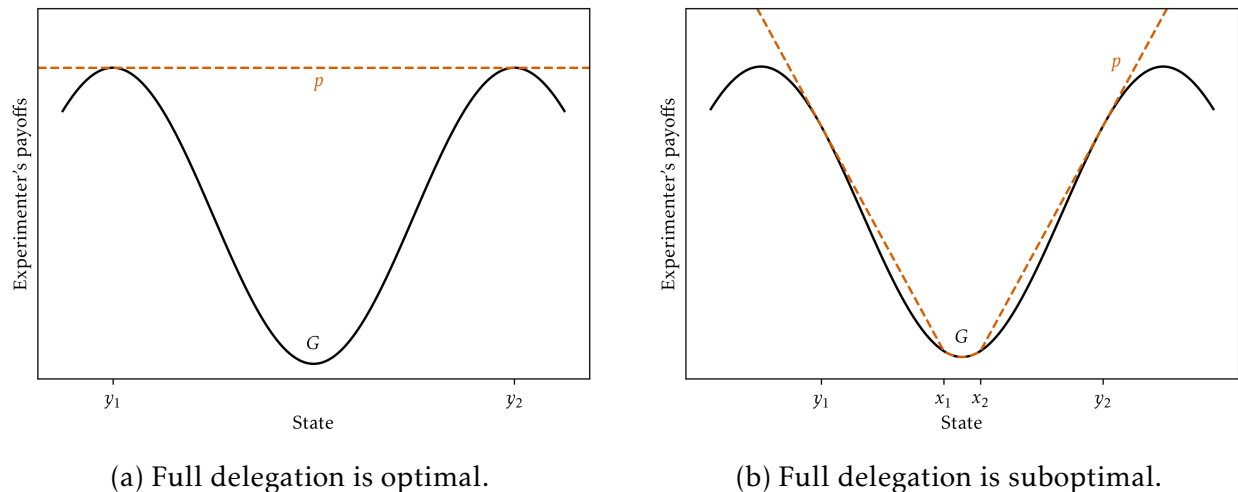


Figure 6: Examples of *M-shaped* delegation of information provision.

tion outcome reveals some but not all states. We formalize these notions and the claim in [Appendix OB](#). Here, we give an informal description.

First, we formalize *locally strict preferences* as follows. There is a partition of the state space $[0, 1]$ into finitely many intervals on each of which G is either strictly convex or strictly concave, and if G is tangential to a price function at specific points, then G is also tangential to a (possibly distinct) price function at nearby *distinct* points. With price functions characterizing incentive-compatibility, the key implication is that relevant incentive-compatible experiments can be approximated by distinct incentive-compatible experiments. This condition is violated, for example, if the DM is uninformed— G is a step function—since then the largest point in the support of any incentive-compatible experiment is at the step of G .

Second, we conceptualize *incomplete full revelation* under full delegation by considering the extreme points of $\text{MPC}(H)$. To that end, we first note that there exists an extreme point of $\text{MPC}(H)$ that is designer-preferred among the experimenter’s best replies under full delegation. We refer to such an extreme point as a *DEF* (**d**esigner-**o**ptimal **e**xtr**e**me point under full delegation).

Let us recall the extreme-point characterization as *bi-pooling* experiments; see [Kleiner et al. \(2021\)](#); [Arieli et al. \(2023\)](#). For each extreme point F , there are countably many, disjoint *bi-pooling intervals* $[x_1, x_2]$, such that F pools $[x_1, x_2]$ to at most two atoms and I_F lies strictly below I_H on the interior (x_1, x_2) . Outside the union of the bi-pooling intervals, I_F coincides with I_H . We say a non-degenerate interval $[a, b]$ is *fully revealing* if I_F coincides with I_H on $[a, b]$. Note, an extreme point may not admit any fully revealing intervals, such as the binary experiment depicted in [Figure 6a](#).

We say that an extreme point F of $\text{MPC}(H)$ admits *incomplete full revelation* if F fully reveals the state on non-degenerate intervals to the left and right of some bi-pooling interval, or only to one side of it if the bi-pooling interval admits a single atom and lies at the boundary of the state space.²² The experiment depicted in [Figure 6a](#) does not admit incomplete full revelation since it admits no fully revealing intervals. The experiment depicted in [Figure 6b](#) and all non-degenerate upper censorship experiments from the S-shaped case admit incomplete full revelation.

Our claim is thus that full delegation is suboptimal if the experimenter has locally strict preferences and there exists a DEF that admits incomplete full revelation. The formal details of the definitions and the proof are in [Appendix OB](#).

7.2 MIC Experiments as Extreme Points

We establish a link between extreme points of $\text{MPC}(H)$ and MIC experiments for general experimenter payoffs G . The literature has demonstrated how unrestricted persuasion problems are solved via extreme points of $\text{MPC}(H)$. The contribution of this section is to show that extreme points of $\text{MPC}(H)$ likewise solve the problem of designing restrictions subject to incentive constraints.

Let G be upper semicontinuous with at most finitely many discontinuities at interior points $y_1, \dots, y_k \in (0, 1)$, and Lipschitz continuous in each interval (y_i, y_{i+1}) , with $y_0 = 0$ and $y_{k+1} = 1$; this is part (i) of the regularity condition of [Dworczak and Martini \(2019\)](#).

The definitions of incentive-compatibility and maximality are as in the baseline model, and MIC experiments are without loss (see [Appendix A.2](#)). Our price-theoretic characterization of incentive-compatibility ([Lemma 2](#)) remains applicable given the regularity assumption on G . In contrast, our characterization of MIC experiments as double censorship makes explicit use of S-shapedness of G and, hence, does not apply here.

At first sight, the extreme points of $\text{MPC}(H)$ may seem unrelated to MIC experiments. The extreme point characterization reflects the feasibility constraints of being a mean-preserving contraction of H ; in particular, the characterization is independent of preferences. In contrast, MIC experiments are precisely characterized by the experimenter's incentives to garble information. Nevertheless:

Theorem 3. *Every MIC experiment is an extreme point of $\text{MPC}(H)$.*

²²These cases do not describe all constellations for how a bi-pooling interval may neighbor a full revelation interval. We do not handle bi-pooling intervals with two atoms where one neighboring interval is the boundary of the state space. We also do not handle two consecutive bi-pooling intervals, each with a single atom, that are preceded and succeeded by full revelation intervals.

Since [Theorem 3](#) shows that MIC experiments are extreme points of the *entire* set $\text{MPC}(H)$, the literature’s characterization of the extreme points as bi-pooling experiments applies under the novel MIC constraint.

Our proof leverages the characterization of extreme points as bi-pooling experiments. For a high-level intuition, note that any experiment $F \in \text{MPC}(H)$ is weakly increasing, and that the ICDF of F lies weakly below the ICDF of the prior H . Roughly, the extreme point characterization states that at least one of these two constraints must bind locally. When the monotonicity constraint binds on an interval, F assigns no mass to this interval; when the ICDF constraint binds, F is maximally informative on this interval. Likewise, incentive-compatibility and maximality are characterized by two constraints on the respective experiment. Incentive-compatibility requires that F assigns no mass to certain intervals ([Lemma 2](#)); this is akin to the binding monotonicity constraint. Maximality requires that F is maximally informative otherwise; this is akin to the binding informativeness constraint.

[Theorem 3](#) showcases how the literature’s extreme-point and linear-duality approaches can be combined to solve information design problems under incentive-compatibility and maximality constraints. In [Appendix OC](#), we show that incentive-compatible bi-pooling experiments are without loss more generally: each incentive-compatible experiment can be represented as a mixture of incentive-compatible bi-pooling experiments. This representation is useful when the designer’s preferences do not justify MIC experiments.

8 Conclusions

We analyze delegated information provision when the party producing information, the experimenter, has persuasion incentives and can always garble any admissible experiment before it reaches a decision maker. The designer cannot impose experiments upon the experimenter, but exerts control by imposing a Blackwell-upper bound on informativeness.

The main conceptual contribution is to reduce optimal restriction design to the study of maximal incentive-compatible (MIC) experiments. An experiment is incentive compatible if it is self-enforcing: when the designer sets it as the upper bound, the experimenter optimally implements it rather than garbling. It is maximal if no strictly more informative experiment is also incentive compatible. Because the designer values information, we show that it is without loss to compare only MIC experiments.

In the baseline case of S -shaped experimenter preferences, MIC experiments are fully characterized as double censorship experiments. These reveal low states, pool an intermediate region to an interior atom, and pool high states to a top atom. Crucially,

full delegation produces upper censorship (pooling all sufficiently high states), which is itself an extreme case of double censorship. Yet, we show that the designer always strictly prefers a nontrivial double-censorship restriction to full delegation.

The key takeaway is that the designer can protect the decision maker from some unwanted persuasion, while retaining the benefits of information provision. Optimal restrictions reallocate informativeness across states to discipline persuasion: sacrificing some precision in intermediate states relaxes the experimenter’s incentives to censor high states, yielding a strict gain for the designer.

We show how our results can be applied in the context of recommender systems and rationalize privacy constraints to steer the tension between information provision and persuasion by a biased experimenter.

A Proofs

A.1 Existence of an Optimal Restriction

Equip $\text{MPC}(H)$ with the L^1 -norm, making it compact.

Lemma 5. *The set of IC experiments is a compact subset of $\text{MPC}(H)$. There exists an optimal IC experiment.*

Existence can be shown by adapting Lemma 1 of [Lipnowski et al. \(2020\)](#), who consider settings with an abstract state. For the sake of completeness, we give a simple argument for our setting with posterior-mean payoffs. The proof does not use S -shapedness, but only upper semicontinuity of the designer’s and the experimenter’s payoffs.

Proof of Lemma 5. Existence of an optimal IC experiment follows from Berge’s Maximum Theorem and upper semicontinuity of the designer’s payoffs if we can show that the set of IC experiments is compact. Compactness follows if we can show that BR is an upper hemicontinuous, nonempty- and compact-valued correspondence from $\text{MPC}(H)$ to itself. These properties of BR follow from Lemma 17.30 of [Aliprantis and Border \(2006\)](#) and upper semicontinuity of the experimenter’s payoffs if we can show that MPC is a continuous, compact-valued correspondence from $\text{MPC}(H)$ to itself. Compact-valuedness and upper hemicontinuity are straightforward.

As for lower hemicontinuity, let $(\bar{F}_n)_{n \in \mathbb{N}}$ be a sequence in $\text{MPC}(H)$ converging to $\bar{F} \in \text{MPC}(H)$, and let $\Phi \in \text{MPC}(\bar{F})$. We have to find a subsequence of $(\Phi_n)_{n \in \mathbb{N}}$ converging to Φ such that $\Phi_n \in \text{MPC}(\bar{F}_n)$ for all n in the subsequence. Note, L^1 -convergence

of $(\tilde{F}_n)_{n \in \mathbb{N}}$ to \tilde{F} implies pointwise convergence of $(I_{\tilde{F}_n})_{n \in \mathbb{N}}$ to $I_{\tilde{F}}$. For all n and m , let $\tilde{J}_n(m) = \min\{I_{\Phi}(m), I_{\tilde{F}_n}(m)\}$.

For all n , since ICDFs are 1-Lipschitz and \tilde{J}_n is the pointwise minimum of ICDFs, also \tilde{J}_n is 1-Lipschitz. Thus, the Arzelà–Ascoli Theorem delivers a uniformly convergent subsequence of $(\tilde{J}_n)_n$. By possibly relabeling, let this be the entire sequence. The limit of the sequence $(\tilde{J}_n)_n$ equals I_{Φ} since $I_{\tilde{F}_n} \rightarrow I_{\tilde{F}}$ pointwise and since $I_{\Phi} \leq I_{\tilde{F}}$. Now, for all n , let J_n be the lower convex hull of \tilde{J}_n . Let $\varepsilon_n = \|\tilde{J}_n - I_{\Phi}\|_{\infty}$, so that $I_{\Phi} - \varepsilon_n \leq \tilde{J}_n \leq I_{\Phi}$. Since $I_{\Phi} - \varepsilon_n$ is convex, and since J_n is the lower convex hull of \tilde{J}_n , we have $I_{\Phi} - \varepsilon_n \leq J_n \leq I_{\Phi}$. In particular, $(J_n)_n$ converges uniformly to I_{Φ} . Since, for all n , the function J_n is convex and weakly increasing, it admits a weakly positive, weakly increasing derivative Φ_n ; i.e. $J_n = I_{\Phi_n}$. Since \tilde{J}_n is 1-Lipschitz and J_n is its lower convex hull, also J_n is 1-Lipschitz. In particular, $\Phi_n(1) \leq 1$ so that Φ_n is a CDF and, hence, in $\text{MPC}(H)$. In fact, since $I_{\Phi_n} \leq J_n \leq \tilde{J}_n \leq I_{\tilde{F}_n}$, we have $\Phi_n \in \text{MPC}(\tilde{F}_n)$. Finally, by possibly passing to a further subsequence and invoking Helly’s Selection Theorem, let $(\Phi_n)_n$ converge pointwise to, say, $\tilde{\Phi}$. By Dominated Convergence, the sequence $(I_{\Phi_n})_n$ (i.e. the sequence $(J_n)_n$) converges pointwise to $I_{\tilde{\Phi}}$. We have already argued that the sequence $(I_{\Phi_n})_n$ (i.e. the sequence $(J_n)_n$) converges pointwise to I_{Φ} . Thus, $I_{\Phi} = I_{\tilde{\Phi}}$, implying $\Phi = \tilde{\Phi}$ almost everywhere. Thus, $(\Phi_n)_n$ converges pointwise a.e. to Φ and, hence, also with respect to the L^1 -norm. \square

A.2 Proof of Lemma 1

Consider the auxiliary problem of maximizing $\int \exp(m) d\tilde{F}(m)$ across incentive-compatible experiments \tilde{F} in $\text{MPC}(H) \cap \text{MPS}(F)$. Incentive-compatible experiments form a compact subset of $\text{MPC}(H)$ when G is upper semicontinuous (Lemma 5). In particular, the constraint set of the auxiliary problem is also compact. Hence, a solution exists. By construction, the solution is incentive-compatible and a mean-preserving spread of F . The solution is maximal since \exp is strictly convex, and since any experiment that is a mean-preserving spread of \tilde{F} is also a mean-preserving spread of F . \square

Remark 3. Our proof of Lemma 1 only uses upper semicontinuity of G .

A.3 Proof of Lemma 3

To prove one part of the equivalence, let $F \in \text{MPC}(H)$ be IC and non-degenerate. Let p be as in Lemma 2. If $\max \text{supp} F \leq r_0$, then, definitionally, F assigns no mass to $(r_0, 1]$. The proof is complete for $x = y = r_0$, since $(r_0, r_0) \in P$. In what follows, let $\max \text{supp} F > r_0$. Let

$y = \max \text{supp} F$, so that $p(y) = G(y)$. We find a point $x < r_0$ such that $(x, y) \in P$ and such that F assigns no mass to $(x, y) \cup (y, 1]$.

Recall that G is strictly concave on $[r_0, 1]$. Since p is convex and $p \geq G$ holds, there exists at most one point in $[r_0, 1]$ at which p and G coincide. Since $y > r_0$, this point is y . Thus, $\text{supp} F \cap (r_0, 1]$ contains only y . Since F is non-degenerate, the intersection $\text{supp} F \cap [0, r_0]$ is non-empty. Let $\tilde{x} = \max(\text{supp} F \cap [0, r_0])$. In particular, $p(\tilde{x}) = G(\tilde{x})$ and $\tilde{x} \leq r_0$.

Now, recall $p \geq G$ and $p(y) = G(y)$, so that $p - G$ is minimized at y . Since p is convex while G is concave on a neighborhood of y , the subdifferential of p at y contains $g(y)$. Since also $p(\tilde{x}) = G(\tilde{x})$, we infer $G(\tilde{x}) + g(y)(y - \tilde{x}) = p(\tilde{x}) + g(y)(y - \tilde{x}) \geq p(y) = G(y)$, where the last inequality follows by definition of the subdifferential. We also know $G(r_0) + g(y)(y - r_0) < G(y)$ since G is strictly concave on $[r_0, 1]$. Thus, $\tilde{x} < r_0$ and, by continuity, there exists $x \in [\tilde{x}, r_0)$ such that $G(x) + g(y)(y - x) = G(y)$. That is, $(x, y) \in P$. Since $\tilde{x} = \max(\text{supp} F \cap [0, r_0]) \leq x$ and $\text{supp} F \cap (r_0, 1] = \{y\}$, we conclude that F assigns no mass to $(x, y) \cup (y, 1]$.

To prove the other part of the equivalence, let $(x, y) \in P$ and $F \in \text{MPC}(H)$ be such that F is non-degenerate and assigns no mass to $(x, y) \cup (y, 1]$. In view of [Lemma 2](#), we can show that F is IC by finding a convex function p such that $p \geq G$ and such that $p = G$ holds on $\text{supp} F$. Let $p(m) = G(m)$ for all $m \in [0, x]$, and $p(m) = G(x) + g(y)(m - x)$ for all $m \in [x, 1]$. Note $p(y) = G(y)$ since $(x, y) \in P$. Clearly, $p = G$ holds on $\text{supp} F$ since F assigns no mass to $(x, y) \cup (y, 1]$. We next argue $p \geq G$. On $[r_0, 1]$, the function G is concave whereas p is affine. Since $p(y) = G(y)$, it holds $p \geq G$ on $[r_0, 1]$. On $[x, r_0]$ (recall $x \leq r_0$ from the definition of P), the function G is convex whereas p is affine. Since $p(x) = G(x)$ but $p(r_0) \geq G(r_0)$ (as already argued), we find $p \geq G$ on $[x, r_0]$. It remains to argue that p is convex. Clearly, p is convex on $[0, x]$. Since p is affine on $[x, 1]$ with slope $g(y)$, global convexity follows if we can show $g(y) \geq g(x)$. This inequality, in turn, holds since $p \geq G$ and $p(x) = G(x)$ both hold. \square

A.4 Proof of [Lemma 4](#)

For all (x, y) such that $0 \leq x \leq r_0 \leq y \leq 1$, let $\rho(x, y) = G(x) + g(y)(y - x) - G(y)$. By definition of P , if $(x, y) \in P$, then $\rho(x, y) = 0$. We show that ρ is strictly single-crossing from above in each argument (holding the other fixed). Therefore, if (x, y) and (x', y') are both in P , then $[x, y]$ and $[x', y']$ are ordered by set inclusion.

For all (x, y) such that $0 \leq x \leq r_0 \leq y \leq 1$, it holds $\partial \rho(x, y) / \partial y = g'(y)(y - x) \leq 0$, where the inequality holds since $x \leq r_0 \leq y$ and since G is concave on $[r_0, 1]$; further, $g'(y) < 0$ if $r_0 < y$. Thus, $\tilde{y} \mapsto \rho(x, \tilde{y})$ (defined on $[r_0, 1]$) is strictly single-crossing from above.

For all (x, y) such that $0 \leq x \leq r_0 \leq y \leq 1$, it holds $\partial \rho(x, y) / \partial x = g(x) - g(y)$. We claim $g(x) < g(y)$ if $\rho(x, y) = 0$ and $x < y$, which suffices to establish that $\tilde{x} \mapsto \rho(\tilde{x}, y)$ (defined on

$[0, r_0]$) is strictly single-crossing from above. Towards a contradiction, let $g(x) \geq g(y)$. Since g is strictly quasiconcave, all $r \in (x, y)$ satisfy $g(r) > g(y)$, implying $G(y) - G(x) > g(y)(y - x)$, which contradicts $\rho(x, y) = 0$. \square

A.5 Proof of Theorem 1

For reference, we record the ICDF of a double censorship experiment F with thresholds (s, t) and atoms (x, y) :

$$\forall m \in [0, 1], \quad I_F(m) = \begin{cases} I_H(m) & \text{if } m \in [0, s], \\ I_H(s) + H(s)(m - s) & \text{if } m \in (s, x], \\ I_H(t) + H(t)(m - t) & \text{if } m \in (x, y], \\ m - \mu, & \text{if } m \in (y, 1]. \end{cases}$$

Lemmata 6 and 7 below together prove Theorem 1.

Lemma 6. *If F is an MIC experiment, then F is non-degenerate double censorship for a pair of thresholds (s, t) and atoms (x, y) such that $(x, y) \in P$ and $s \leq x^* \leq t$ and $0 < x < y < 1$.*

Proof of Lemma 6. The experiment must be non-degenerate since, otherwise, it is a strict MPC of F^* , contradicting maximality. Hence, we can appeal to Lemma 3 to find $(x, y) \in P$ such that F assigns no mass to $(x, y) \cup (y, 1]$.

We next claim $[x^*, y^*] \subseteq [x, y]$. By Lemma 4, it holds $[x^*, y^*] \subseteq [x, y]$ or $[x, y] \subseteq [x^*, y^*]$. In the first case, we are done, so let $[x, y] \subseteq [x^*, y^*]$. We show that F is an MPC of F^* ; by maximality, it follows $F = F^*$ and, hence, $[x, y] = [x^*, y^*]$. To show F is an MPC of F^* , we show $I_F \leq I_{F^*}$ on $[0, 1]$. Clearly, $I_F \leq I_H = I_{F^*}$ on $[0, x^*]$. Next, we know I_F coincides with I_{δ_μ} on $[y, 1]$, whereas I_{F^*} coincides with I_{δ_μ} on $[y^*, 1]$. Since $y \leq y^*$, we infer $I_F = I_{F^*}$ on $[y^*, 1]$. In summary, $I_F \leq I_{F^*}$ on $[0, x^*] \cup [y^*, 1]$. Finally, since F^* assigns no mass to (x^*, y^*) , the integrated CDF I_{F^*} is affine on $[x^*, y^*]$. Since I_F is convex and lies below I_{F^*} at the endpoints of this interval (as already argued), it follows $I_F \leq I_{F^*}$ also on $[x^*, y^*]$. Thus, $I_F \leq I_{F^*}$ on $[0, 1]$.

We next show that F is double censorship. The idea is to construct an MPS \tilde{F} of F that is double censorship and IC. Since F is MIC, it will follow $\tilde{F} = F$.

Our candidate double censorship will use (x, y) as atoms. We thus show next that (x, y) can arise as atoms for some choice of thresholds (s, t) . First, since $y^* = \mathbb{E}_H[\omega \mid \omega \in [x^*, 1]]$ and $y^* \leq y \leq 1$, there exists $t \in [x^*, y]$ such that $y = \mathbb{E}_H[\omega \mid \omega \in [t, 1]]$. Note, $t \in [x, y]$ since $x \leq x^*$. Second, we argue there exists $s \in [0, x^*]$ such that $x = \mathbb{E}_H[\omega \mid \omega \in [s, t]]$. By construction, it holds $y = \mathbb{E}_H[\omega \mid \omega \in [t, 1]]$, which rearranges to $I_F(y) = I_H(t) + H(t)(y - t)$. The affine map $m \mapsto I_H(t) + H(t)(m - t)$ lies above I_F at $m = t$ (since $I_F(t) \leq I_H(t)$) and coincides with I_F at

$m = y$ (by construction of t). Since also I_F is affine on $[x, y]$ (since F assigns no mass to (x, y)), and since two affine maps on \mathbb{R} either coincide or intersect at most once, we conclude $I_F(x) \leq I_H(t) + H(t)(x - t)$. Let $\xi = I_H(t) + H(t)(x - t)$. In particular, $\xi \geq I_F(x) \geq 0$. We also have $\xi \leq I_H(x)$ since the affine map $\tilde{t} \mapsto I_H(t) + H(t)(\tilde{t} - t)$ is tangent to the convex map I_H at t . Finally, find $s \in [0, x]$ such that $I_H(s) + H(s)(x - s) = \xi$; this is possible since $\xi \in [0, I_H(x)]$. Thus, $I_H(s) + H(s)(x - s) = I_H(t) + H(t)(x - t)$, which rearranges to $x = \mathbb{E}_H[\omega \mid \omega \in [s, t]]$.

Now let Φ be double censorship with thresholds (s, t) and atoms (x, y) . Notice that Φ is IC; indeed, Φ assigns no mass to the set $(x, y) \cup (y, 1]$ (by definition of double censorship), and it holds $(x, y) \in P$ (as argued at the very beginning of the proof); now invoke [Lemma 3](#).

We now argue Φ is an MPS of F . Clearly, $I_\Phi = I_H \geq I_F$ on $[0, s]$, and $I_\Phi = I_{\delta_\mu} = I_F$ on $[y, 1]$. For $\tilde{t} \in [x, y]$ it holds $I_\Phi(\tilde{t}) = I_H(t) + H(t)(\tilde{t} - t) \geq I_F(\tilde{t})$, where the inequality follows from the argument that constructed s . Finally, it holds $I_\Phi \geq I_F$ on $[s, x]$ since I_Φ is affine on this interval, I_F is convex, and since $I_\Phi \geq I_F$ holds at the endpoints of this interval (as already argued). Thus, Φ is an MPS of F .

It remains to show $s \leq x^* \leq t$ and $0 < x < y < 1$. The inequalities $s \leq x^* \leq t$ follow from $[x^*, y^*] \subseteq [x, y]$ and the expressions for x , y , and y^* as conditional expectations. From $[x^*, y^*] \subseteq [x, y]$ and $x^* < r_0 < y^*$, we also obtain $x < y$.

Next, towards a contradiction, let $y = 1$. Then also $t = 1$ (since $y = \mathbb{E}_H[\omega \mid \omega \in [t, 1]]$), meaning x is the largest point in the support of F . From $y = 1 > y^*$ we also get $x < x^*$ ([Lemma 4](#)). It now follows easily that F is a proper MPC of F^* ; indeed, $I_F = I_{\delta_\mu} \leq I_{F^*}$ on $[x, 1]$; this inequality is strict on (x, x^*) since I_F is affine on this interval whereas I_{F^*} is strictly convex (it coincides with I_H); finally, $I_F \leq I_{F^*} \leq I_H$ on $[0, x]$. Thus, F is a proper MPC of F^* , contradicting maximality of F . Thus, $y < 1$.

Finally, towards a contradiction, let $x = 0$. This requires $s = t = 0$ since $x = \mathbb{E}_H[\omega \mid \omega \in [s, t]]$. It follows that F is degenerate on y ; contradiction. \square

Lemma 7. *If F is double censorship for a pair of thresholds (s, t) and atoms (x, y) such that $(x, y) \in P$ and $s \leq x^* \leq t$ and $0 < x < y < 1$, then F is MIC.*

Proof of Lemma 7. [Lemma 3](#) implies that F is IC. To prove that F is maximal, let Φ be IC and such that $I_F \leq I_\Phi$. We show $I_F = I_\Phi$. Let $\hat{y} = \max \text{supp } \Phi$.

First, we show $\hat{y} < y$ cannot hold. Indeed, the function I_Φ coincides with $m \mapsto m - \mu$ on $[\hat{y}, 1]$. Meanwhile, I_F coincides with $m \mapsto m - \mu$ on $[y, 1]$ but has slope of at most $H(t)$ on $[0, y)$. Since $t \leq y < 1$, we have $H(t) < 1$. Thus, if $\hat{y} < y$, then I_F lies strictly above I_Φ on (\hat{y}, y) , contradicting $I_F \leq I_\Phi$.

Thus, let $\hat{y} \geq y$. Since $\hat{y} = \max \text{supp } \Phi$ and $y > r_0$, the IC characterization ([Lemma 3](#)) implies there exists $\hat{x} \leq r_0$ such that $(\hat{x}, \hat{y}) \in P$ and such that Φ assigns no mass to $(\hat{x}, \hat{y}) \cup (\hat{y}, 1]$.

Since $\hat{y} \geq y$, [Lemma 4](#) implies $\hat{x} \leq x$.

Since Φ assigns no mass to (\hat{x}, \hat{y}) , and since $[x, y] \subseteq [\hat{x}, \hat{y}]$ and $t \in [x, y]$, we conclude that I_Φ is affine on $[\hat{x}, \hat{y}]$ and this interval contains t . Now, since F is double censorship, I_F is tangential to I_H at t . Thus, also $I_\Phi(t) = I_H(t)$ since $I_F \leq I_\Phi \leq I_H$. Thus, I_Φ is affine with slope $H(t)$ on $[\hat{x}, \hat{y}]$. Since I_F is affine $[x, y]$ (a subset of $[\hat{x}, \hat{y}]$) and tangential to I_H at t , it follows that I_Φ and I_F coincide on $[x, y]$. From here and the inclusion $[x, y] \subseteq [\hat{x}, \hat{y}]$, it readily follows $I_F = I_\Phi = I_{\delta_\mu}$ on $[y, 1]$.

We also know that I_F and I_Φ coincide on $[0, s]$ since on this interval $I_F = I_H \geq I_\Phi \geq I_F$. It now follows $I_F = I_\Phi$ on $[s, x]$ since I_F is affine on this interval, I_Φ is convex, and since the two coincide at the endpoints (as already argued). Thus, $I_F = I_\Phi$. \square

A.6 Proof of [Corollary 1](#)

Let (s, t, x, y) and (s', t', x', y') be the parameters of two MIC experiments such that $y' < y$. Since $(x, y) \in P$ and $(x', y') \in P$ ([Theorem 1](#)), [Lemma 4](#) implies $x \leq x'$; the proof of this lemma delivers $x < x'$. Using the expressions for x, x', y , and y' as conditional expectations, it also follows $[s', t'] \subsetneq [s, t]$ and $[t, 1] \subsetneq [t', 1]$ and $x < x'$. \square

A.7 Proof of [Corollary 2](#)

We apply [Corollary 1](#) of [Dworczak and Martini \(2019\)](#). Let p be the function that coincides with G on $[0, x]$, and coincides with the affine map $m \mapsto G(x) + g(y)(m - x)$ for $m \in (x, 1]$. Then, p is convex, weakly above G , and coincides with G on $[0, x]$ and y ; see the proof of [Lemma 3](#) for this argument. Moreover, $\text{supp } F \subseteq \{m \in [0, 1]: p(m) = G(m)\}$ since F is double censorship with parameters (s, t, x, y) . Finally, it holds $\int p dF = \int p d\bar{F}$. Indeed, we have $F = \bar{F} = H$ on $[0, s]$. Conditioned on the interval $[s, t]$, both F and \bar{F} are degenerate on x . Finally, on the interval $[t, 1]$, the function p is affine and F is obtained by pooling \bar{F} . Thus, $\int p dF = \int p d\bar{F}$. By [Corollary 1](#) of [Dworczak and Martini \(2019\)](#), the experiment F is a best reply to \bar{F} . \square

A.8 Proof of [Theorem 2](#)

The idea is to consider MIC double censorship, parametrized by (s, t, x, y) , such that y is close to the full delegation atom y^* ; in this case, also s, t , and x are all close to x^* . We evaluate the marginal gain from this perturbation.

[Lemma 3](#) implies that a tuple (s, t, x, y) with $0 \leq s \leq x \leq t \leq y < 1$ and $x \leq r_0 \leq y$

represents an IC double censorship if

$$G(x) + g(y)(y - x) - G(y) = 0, \quad (7a)$$

$$\int_{[t,1]} (\omega - y) dH(\omega) = 0, \quad (7b)$$

$$\int_{[s,t]} (\omega - x) dH(\omega) = 0, \quad (7c)$$

all hold. (In fact, in this case the experiment is MIC.) We know that one such tuple is $(s, t, x, y) = (x^*, x^*, x^*, y^*)$, where x^* is the threshold and y^* is the atom under full delegation. We next argue that for all y sufficiently close to y^* there exist (s, t, x) satisfying the above.

The proof of [Lemma 4](#) shows $g(x^*) < g(y^*)$, which implies that the derivative of $x \mapsto G(x) + g(y^*)(y^* - x) - G(y^*)$ is strictly negative at x^* . By continuous differentiability of g , moreover, the function $(x, y) \mapsto G(x) + g(y^*)(y^* - x) - G(y^*)$ is continuously differentiable. Thus, by the Implicit Function Theorem, for all $y > y^*$ sufficiently close to y^* there exists $\hat{x}(y)$ such that $G(\hat{x}(y)) + g(y)(y - \hat{x}(y)) = G(y)$, i.e. (7a) holds. Moreover, $\hat{x}(y) \rightarrow x^*$ as $y \rightarrow y^*$ and $y \mapsto \hat{x}(y)$ is differentiable at y^* . Finally, $\hat{x}(y) \leq x^*$ for all y ([Lemma 4](#)).

Next, using $y^* < 1$, it is easy to see (using, e.g., the Intermediate Value Theorem) that for all y sufficiently close to y^* there exists $\hat{t}(y) \in [0, y)$ solving (7b). Using the Implicit Function Theorem, on a neighborhood of y^* the function \hat{t} has the derivative

$$\frac{\partial \hat{t}(y)}{\partial y} = \frac{1 - H(\hat{t}(y))}{(y - \hat{t}(y))h(\hat{t}(y))} > 0.$$

For later reference, for $y > y^*$, equation (7b) requires $\hat{t}(y) > x^*$.

Finally, using also $0 < x^* < y^* < 1$, it is easy to see that for all y sufficiently close to y^* there exists $\hat{s}(y) \in (0, \hat{x}(y))$ such that $(\hat{s}(y), \hat{t}(y), \hat{x}(y), y)$ satisfies (7c). (We cannot use the Implication Function Theorem to conclude that \hat{s} is differentiable, which complicates bounding the integral (9) further ahead.)

Thus, for some $\varepsilon > 0$ and all $y \in (y^*, y^* + \varepsilon]$, the tuple $(\hat{s}(y), \hat{t}(y), \hat{x}(y), y)$ defines an MIC double censorship experiment. We show that for y sufficiently close to y^* the designer is strictly better off than under unrestricted persuasion. Note, for $y \searrow y^*$, we have $(\hat{s}(y), \hat{t}(y), \hat{x}(y)) \rightarrow (x^*, x^*, x^*)$ since, following the above arguments, there is a unique solution for eqs. (7a) to (7c) that involves y^* , and (x^*, x^*, x^*, y^*) is a solution.

The designer's utility from $(\hat{s}(y), \hat{t}(y), \hat{x}(y), y)$, denoted $U_D(y)$, is given by

$$U_D(y) = \int_{[0, \hat{s}(y)]} u_D(\omega) dH(\omega) + u_D(\hat{x}(y))(H(\hat{t}(y)) - H(\hat{s}(y))) + u_D(y)(1 - H(\hat{t}(y))).$$

The expected utility from (x^*, x^*, x^*, y^*) is given by

$$U_D(y^*) = \int_{[0, x^*]} u_D(\omega) dH(\omega) + u_D(y^*)(1 - H(x^*)),$$

which also obtains as the limit as $y \rightarrow y^*$ from above. Thus,

$$U_D(y) - U_D(y^*) = (u_D(y) - u_D(y^*))(1 - H(x^*)) + (u_D(y) - u_D(\hat{x}(y)))(H(x^*) - H(\hat{t}(y))) \quad (8)$$

$$+ \int_{[\hat{s}(y), x^*]} (u_D(\hat{x}(y)) - u_D(\omega)) dH(\omega). \quad (9)$$

Dividing by $y - y^*$ and taking $y \rightarrow y^*$, the sum of the two terms in (8) converges to

$$\begin{aligned} & u'_D(y^*)(1 - H(x^*)) - (u_D(y^*) - u_D(x^*))h(x^*) \frac{\partial \hat{t}(y)}{\partial y} \Big|_{y=y^*} \\ & = \left(u'_D(y^*) - \frac{u_D(y^*) - u_D(x^*)}{y^* - x^*} \right) (1 - H(x^*)) > 0, \end{aligned}$$

where the strict inequality follows from strict convexity of u_D and since $y^* > x^*$ holds. Thus, to show $U_D(y) > U_D(y^*)$ for y sufficiently close to y^* , it suffices to show that the integral in (9) vanishes when normalized by $y - y^*$. We have

$$\begin{aligned} & \frac{1}{y - y^*} \left| \int_{[\hat{s}(y), x^*]} (u_D(\hat{x}(y)) - u_D(\omega)) dH(\omega) \right| \\ & \leq \frac{1}{y - y^*} \left| (u_D(\hat{x}(y)) - u_D(x^*))(H(x^*) - H(\hat{s}(y))) \right| + \frac{1}{y - y^*} \left| \int_{[\hat{s}(y), x^*]} (u_D(x^*) - u_D(\omega)) dH(\omega) \right| \end{aligned}$$

Here, $\frac{1}{y - y^*} \left| (u_D(\hat{x}(y)) - u_D(x^*))(H(x^*) - H(\hat{s}(y))) \right|$ vanishes since \hat{x} is differentiable at y^* , and since $\hat{s}(y) \rightarrow x^*$ as $y \rightarrow y^*$. Consider the integral. Find $L > 0$ such that $|u'_D| \leq L$. Recall that $\hat{s}(y)$ is chosen such that $\int_{[\hat{s}(y), x^*]} (\hat{x}(y) - \omega) dH(\omega) = \int_{[x^*, \hat{t}(y)]} (\omega - \hat{x}(y)) dH(\omega)$. We then have

$$\begin{aligned} & \frac{1}{y - y^*} \left| \int_{[\hat{s}(y), x^*]} (u_D(x^*) - u_D(\omega)) dH(\omega) \right| \\ & \leq \frac{1}{y - y^*} \int_{[\hat{s}(y), x^*]} \left| \int_{[\omega, x^*]} u'_D(r) dr \right| dH(\omega) \\ & \leq \frac{L}{y - y^*} \int_{[\hat{s}(y), x^*]} (x^* - \omega) dH(\omega) \\ & = \frac{L}{y - y^*} (x^* - \hat{x}(y))(H(x^*) - H(\hat{s}(y))) + \frac{L}{y - y^*} \int_{[\hat{s}(y), x^*]} (\hat{x}(y) - \omega) dH(\omega) \\ & = \frac{L}{y - y^*} (x^* - \hat{x}(y))(H(x^*) - H(\hat{s}(y))) + \frac{L}{y - y^*} \int_{[x^*, \hat{t}(y)]} (\omega - \hat{x}(y)) dH(\omega) \end{aligned}$$

$$\leq \frac{L}{y-y^*}(x^* - \hat{x}(y))(H(x^*) - H(\hat{s}(y))) + \frac{L}{y-y^*}(\hat{t}(y) - \hat{x}(y))(H(\hat{t}(y)) - H(x^*))$$

Now, $\frac{1}{y-y^*}(x^* - \hat{x}(y))(H(x^*) - H(\hat{s}(y)))$ vanishes as $y \rightarrow y^*$ since \hat{x} is differentiable at y^* , and since $\hat{s}(y) \rightarrow x^*$ as $y \rightarrow y^*$. Moreover, $\frac{1}{y-y^*}(\hat{t}(y) - \hat{x}(y))(H(\hat{t}(y)) - H(x^*))$ vanishes since $\hat{t}(y)$ is differentiable at y^* , and since $\hat{t}(y)$ and $\hat{x}(y)$ both converge to x^* as $y \rightarrow y^*$. Thus, the integral in (9) vanishes when normalized by $y - y^*$. \square

A.9 Proof of Theorem 3

Let F be MIC. For later reference, by Lemma 2, there exists a continuous, convex function $p: [0, 1] \rightarrow \mathbb{R}$ such that $p \geq G$ and $\text{supp } F \subseteq \{m \in [0, 1]: p(m) = G(m)\}$.

By Kleiner et al. (2021, Theorem 2), to show that F is an extreme point of $\text{MPC}(H)$, we have to show that, for every open subinterval (a, b) on which the ICDF of F lies strictly below the ICDF of H , the support of F has at most two points in (a, b) .

Let $[a, b]$ be a subinterval of $[0, 1]$ such that $I_F(m) < I_H(m)$ for all $m \in (a, b)$.

Let $m \in (\text{supp } F) \cap (a, b)$. We claim that there is a neighborhood O of m such that $(\text{supp } F) \cap O = \{m\}$. Towards a contradiction, suppose not. Then, there is a sequence $(m_n)_{n \in \mathbb{N}}$ in $(\text{supp } F) \setminus \{m\} \cap (a, b)$ converging to m . By possibly passing to a subsequence, let the sequence be strictly monotone. Assume the sequence is strictly increasing, the case where it is strictly decreasing being similar. Consider the line segment between $(m_n, I_F(m_n))$ and $(m, I_F(m))$; it is given by $t \mapsto I_F(m_n) + (I_F(m) - I_F(m_n))(t - m_n)/(m - m_n)$ for $t \in [m_n, m]$. Find n sufficiently large such that this line segment does not intersect I_H on $[m_n, m]$; such n exists since I_F lies strictly below I_H on (a, b) and since ICDFs are 1-Lipschitz. Consider the function that coincides with I_F below m_n and above m , and that coincides with the chosen line segment on $[m_n, m]$. This function is weakly increasing and convex, coincides with I_H at 0 and 1, and lies below I_H . Thus, its derivative \tilde{F} is an experiment. Since m_n and m are in the support of F , and since $I_{\tilde{F}}$ and I_F coincide outside $[m_n, m]$, it holds $\text{supp } \tilde{F} \subseteq \text{supp } F \subseteq \{m' \in [0, 1]: p(m') = G(m')\}$. In particular, Lemma 2 implies that \tilde{F} is IC. Finally, \tilde{F} is an MPS of F since I_F is convex on $[m_n, m]$ while $I_{\tilde{F}}$ is affine on this interval and coincides with I_F at m_n and m . Since F is maximal, we infer $F = \tilde{F}$. In particular, I_F is affine on $[m_n, m]$, meaning F assigns no mass to the open interval (m_n, m) . However, by strict monotonicity, $m_{n+1} \in (m_n, m)$, contradicting $m_{n+1} \in \text{supp } F$.

We now claim that for all closed subintervals $[c, d]$ of (a, b) the intersection $(\text{supp } F) \cap [c, d]$ is finite. Indeed, for all $m \in (\text{supp } F) \cap [c, d]$, let O_m be as in the previous paragraph, i.e. $(\text{supp } F) \cap O_m = \{m\}$. This yields an open cover of the compact set $(\text{supp } F) \cap [c, d]$. Now find a finite subcover to conclude that $(\text{supp } F) \cap [c, d]$ is finite.

Let $[c, d] \subseteq (a, b)$ be arbitrary. We claim $(\text{supp} F) \cap [c, d]$ contains at most two points. Towards a contradiction, suppose not. Since $(\text{supp} F) \cap [c, d]$ is finite, there exist three consecutive points x, y, z in $(\text{supp} F) \cap [c, d]$; by “consecutive” we mean that there is no other point in $\text{supp} F$ between x and y , nor between y and z . In particular, I_F is affine between x and y ; call the slope α . Also, I_F is affine between y and z ; call the slope β . Since y is in the support of F , it holds $\alpha < \beta$. Consider the function \tilde{f} obtained from I_F by raising α by $\varepsilon_\alpha > 0$ while decreasing β by $\varepsilon_\beta > 0$, holding fixed I_F at x and at z . (Thus, I_F is shifted up at y .) Finally, let \tilde{f} coincide with I_F outside $[x, z]$. For ε_α and ε_β sufficiently small, the function \tilde{f} is convex since $\alpha < \beta$. Moreover, since I_F lies strictly below I_H on (a, b) , and since I_F is affine on each of $[x, y]$ and $[y, z]$, for ε_α and ε_β sufficiently small also \tilde{f} lies below I_H on $[x, z]$ (and thus on all of $[0, 1]$).²³ Thus, the derivative of \tilde{f} , call it \tilde{F} , is an experiment. By construction \tilde{F} is a proper MPS of F . Finally, since $\text{supp} \tilde{F} \subseteq \text{supp} F \subseteq \{m \in [0, 1] : p(m) = G(m)\}$, Lemma 2 implies that \tilde{F} is IC. Contradiction to maximality of F .

Since in the previous paragraph $[c, d] \subseteq (a, b)$ was arbitrary, the intersection $(\text{supp} F) \cap (a, b)$ contains at most two points. Since $[a, b]$ was an arbitrary interval such that $I_F < I_H$ on (a, b) , Theorem 2 of Kleiner et al. (2021) implies that F is an extreme point of $\text{MPC}(H)$. \square

B References

- ALIPRANTIS, CHARALAMBOS D. AND KIM C. BORDER (2006): *Infinite Dimensional Analysis : A Hitchhiker’s Guide*, Springer Berlin Heidelberg.
- ARIELI, ITAI, YAKOV BABICHENKO, AND FEDOR SANDOMIRSKIY (2022): “Bayesian Persuasion with Mediators,” Working paper.
- ARIELI, ITAI, YAKOV BABICHENKO, RANN SMORODINSKY, AND TAKURO YAMASHITA (2023): “Optimal persuasion via bi-pooling,” *Theoretical Economics*, 18 (1), 15–36.
- AYBAS, YUNUS C. AND ERAY TURKEL (2025): “Persuasion with Coarse Communication,” Working paper.
- BABICHENKO, YAKOV, INBAL TALGAM-COHEN, AND KONSTANTIN ZABARNYI (2021): “Bayesian Persuasion under Ex Ante and Ex Post Constraints,” Working paper.
- BALL, IAN AND JOSÉ-ANTONIO ESPÍN-SÁNCHEZ (2022): “Experimental persuasion,” Working paper.
- BIRD, DANIEL AND ZVIKA NEEMAN (2022): “What should a firm know? Protecting consumers’ privacy rents,” *American Economic Journal: Microeconomics*, 14 (4), 257–295.

²³To see this, note $I_H - I_F$ is positive and bounded away from 0 across $[x, z]$, by continuity. As $\varepsilon_\alpha \rightarrow 0$ and $\varepsilon_\beta \rightarrow 0$, the function \tilde{f} converges uniformly to I_F on $[x, z]$ since I_F is piecewise affine.

- BLOEDEL, ALEXANDER W. AND ILYA R. SEGAL (2021): “Persuading a Rationally Inattentive Agent,” Working paper.
- CURELLO, GREGORIO AND LUDVIG SINANDER (2025): “The comparative statics of persuasion,” Working paper.
- DALL’ARA, PIETRO (2026): “Persuading an inattentive and privately informed receiver,” Working paper.
- DWORCZAK, PIOTR AND GIORGIO MARTINI (2019): “The simple economics of optimal persuasion,” *Journal of Political Economy*, 127 (5), 1993–2048.
- DWORK, CYNTHIA, FRANK MCSHERRY, KOBBI NISSIM, AND ADAM SMITH (2006): “Calibrating noise to sensitivity in private data analysis,” in *Theory of cryptography conference*, Springer, 265–284.
- HE, KEVIN, FEDOR SANDOMIRSKIY, AND OMER TAMUZ (2026): “Private private information,” *Journal of Political Economy*, 134 (5), 000–000.
- ICHIHASHI, SHOTA (2019): “Limiting Sender’s information in Bayesian persuasion,” *Games and Economic Behavior*, 117, 276–288.
- IVANOV, MAXIM (2021): “Optimal monotone signals in Bayesian persuasion mechanisms,” *Economic Theory*, 72 (3), 955–1000.
- JAIN, VASUDHA AND MARK WHITMEYER (2026): “Competitive disclosure of information to a rationally inattentive agent,” *Journal of Economic Behavior & Organization*, 242, 107410.
- KLEINER, ANDREAS, BENNY MOLDOVANU, AND PHILIPP STRACK (2021): “Extreme points and majorization: Economic applications,” *Econometrica*, 89 (4), 1557–1593.
- KOLOTILIN, ANTON, TIMOFIY MYLOVANOV, AND ANDRIY ZAPECHELNYUK (2022): “Censorship as optimal persuasion,” *Theoretical Economics*, 17 (2), 561–585.
- KOLOTILIN, ANTON AND ANDRIY ZAPECHELNYUK (2025): “Persuasion meets delegation,” *Econometrica*, 93 (1), 195–228.
- LE TREUST, MAËL AND TRISTAN TOMALA (2019): “Persuasion with limited communication capacity,” *Journal of Economic Theory*, 184, 104940.
- LIANG, ANNIE, JAY LU, XIAOSHENG MU, AND KYOHEI OKUMURA (2026): “Algorithm Design: A Fairness-Accuracy Frontier,” *Journal of Political Economy*, forthcoming.
- LIPNOWSKI, ELLIOT, LAURENT MATHEVET, AND DONG WEI (2020): “Attention management,” *American Economic Review: Insights*, 2 (1), 17–32.
- (2022): “Optimal attention management: A tractable framework,” *Games and Economic Behavior*, 133, 170–180.
- LYU, QIANJUN, WING SUEN, AND YIMENG ZHANG (2025): “Coarse Information Design,” Working paper.
- MENSCH, JEFFREY (2021): “Monotone persuasion,” *Games and Economic Behavior*, 130,

521–542.

- MYLOVANOV, TYMOFIY AND ANDRIY ZAPECHELNYUK (2024): “Constructive versus Toxic Argumentation in Debates,” *American Economic Journal: Microeconomics*, 16 (1), 262–292.
- ONUCHIC, PAULA AND DEBRAJ RAY (2023): “Conveying value via categories,” *Theoretical Economics*, 18 (4), 1407–1439.
- PAN, YUQI, ZHIWEI STEVEN WU, HAIFENG XU, AND SHURAN ZHENG (2025): “Differentially private bayesian persuasion,” in *Proceedings of the ACM on Web Conference 2025*, 1425–1440.
- SCHMUTTE, IAN M AND NATHAN YODER (2025): “Information Design for Differential Privacy,” Working paper.
- SONG, DOYOUNG (2025): “Robust Persuasion under Costly Attention,” Working paper.
- STRACK, PHILIPP AND KAI HAO YANG (2024): “Privacy-Preserving Signals,” *Econometrica*, 92 (6), 1907–1938.
- TSAKAS, ELIAS AND NIKOLAS TSAKAS (2021): “Noisy persuasion,” *Games and Economic Behavior*, 130, 44–61.
- TSAKAS, ELIAS, NIKOLAS TSAKAS, AND DIMITRIOS XEFTERIS (2021): “Resisting persuasion,” *Economic Theory*, 72 (3), 723–742.
- WEI, DONG (2021): “Persuasion under costly learning,” *Journal of Mathematical Economics*, 94, 102451.
- XIE, XINPENG, CHENYANG YU, YAN HUANG, YANG CAO, AND CHENXI QIU (2025): “A Decade of Metric Differential Privacy: Advancements and Applications,” Working paper.

Online Appendix: Delegated Information Provision

Francesco Bilotta

Christoph Carnehl

Justus Preusser

OA Ordered Experiments and Garblings

Lemma 8. *If $F \in \text{MPC}(H)$ and $F' \in \text{MPC}(F)$, then there exist a signal (S, π) inducing F , and a signal (S', π') inducing F' such that (S', π') is a garbling of (S, π) .*

Proof of Lemma 8. Since $F \in \text{MPC}(H)$, there exists $K \in \Delta([0, 1]^2)$ whose marginals are F and H , and such that $m = \int \omega dK(\omega|m)$ for F -almost all m . Similarly, since $F' \in \text{MPC}(F)$, there exists $K' \in \Delta([0, 1]^2)$ whose marginals are F' and F , and such that $m' = \int m dK'(m|m')$ for F' -almost all m' . Now define a joint CDF K^* on $[0, 1]^3$ for all m', m, ω by $K^*(m', m, \omega) = \int_{[0, 1]^2} \mathbf{1}_{\tilde{m} \leq m} \mathbf{1}_{\tilde{\omega} \leq \omega} K'(m'|\tilde{m}) dK(\tilde{m}, \tilde{\omega})$. The joint K^* yields random variables ω, m, m' such that (i) $\omega \sim H$, (ii) conditional on $\omega = \omega$, the CDF of m is $K(\cdot|\omega)$, and (iii), conditional on $\omega = \omega$ and $m = m$ the CDF of m' is $K'(\cdot|m)$ (note the conditional independence). The random variable m defines a signal whose distributions of posterior means is F since $m = \int \omega dK(\omega|m)$ holds for F -almost all m . Similarly, m' defines a signal whose distribution of posterior means is F' . By the conditional independence in (iii), the signal induced by m' is a garbling of the signal induced by m . \square

OB Profitable Restrictions Beyond S-shaped Persuasion

In this appendix, we provide the definition of locally strict curvature preferences and incomplete full revelation that we alluded to in [Section 7.1](#). Then, we prove that these conditions are sufficient for the existence of profitable restrictions when the designer has strict preferences for information, and under smoothness assumptions on H , G , and u_D .

OB.1 Definitions

OB.1.1 Locally strict curvature preferences

We first introduce an auxiliary definition. Let $T = (x_1, y_1, y_2, x_2) \in (0, 1)^4$ be such that $x_1 < y_1 \leq y_2 < x_2$. Let $p: [x_1, x_2] \rightarrow \mathbb{R}$ be a function. Say T is *supported by* p if p is continuous and convex, lies weakly above G on $[x_1, x_2]$, and coincides with G at each point in $\{x_1, y_1, y_2, x_2\}$. The tuple T can be supported by a price function if T is supported by at least one function p . In this definition, y_1 and y_2 should be thought of the atoms in a bi-pooling interval $[x_1, x_2]$ of an extreme point (possibly, $y_1 = y_2$).

Definition 5. The experimenter has *locally strict curvature preferences* if the following hold.

(I) The state space $[0, 1]$ can be partitioned into a finite number of intervals such that, on each interval, G is either strictly convex or strictly concave. Let \mathcal{I}_{cvx} and \mathcal{I}_{ccv} , respectively, be the union of intervals on which G is strictly convex and strictly concave, respectively.

(II) For all $T = (x_1, y_1, y_2, x_2)$ such that $x_1, x_2 \in \mathcal{I}_{\text{cvx}}$ and $y_1, y_2 \in \mathcal{I}_{\text{ccv}}$ and $0 < x_1 < y_1 \leq y_2 < x_2 < 1$, if T can be supported by an affine function, then there exists a continuous function $\tilde{T} = (\tilde{x}_1, \tilde{y}_1, \tilde{y}_2, \tilde{x}_2): [0, 1] \rightarrow (0, 1)^4$ such that:

- it holds $\tilde{T}(0) = T$ and $\tilde{x}_1(r) \leq x_1$ and $\tilde{x}_2(r) \geq x_2$ for all $r \in (0, 1]$;
- \tilde{y}_1 and \tilde{y}_2 are continuously differentiable, and at $r = 0$ their derivatives are non-zero; further, \tilde{x}_1 and \tilde{x}_2 are differentiable at $r = 0$;
- for all $r \in (0, 1]$, the tuple $\tilde{T}(r)$ can be supported by a price function.

In plain words, [Definition 5](#) says that if some tuple can be supported by an affine price function, then there are nearby tuples that “smoothly approximate” the given one and that can also be supported by price functions. [Figures 7](#) and [8](#) depict such approximations schematically. In both figures, there is an affine function p (dashed blue) that coincides with G at x_1, y_1, y_2 , and x_2 , and otherwise lies above G . Further, p and G are tangential at y_1 and y_2 , where G is concave. The points y_1 and y_2 are distinct in [Figure 7](#); they coincide in [Figure 8](#). As alluded to earlier, y_1 and y_2 should be thought of the atoms in a bi-pooling interval $[x_1, x_2]$ of an extreme point (possibly, $y_1 = y_2$). The affine function p is a price function (restricted to the interval $[x_1, x_2]$) verifying that this extreme point is a best reply of the experimenter to the prior.

The dotted orange line depicts a price function \tilde{p} that touches G at nearby but distinct points $\tilde{x}_1, \tilde{y}_1, \tilde{y}_2$, and \tilde{x}_2 . The points \tilde{y}_1 and \tilde{y}_2 are distinct in [Figure 7](#); they coincide in [Figure 8](#). A substantive assumption in [Figure 7](#) is that the kinked part of \tilde{p} between \tilde{y}_1 and \tilde{y}_2 lies above G ; likewise, in [Figure 8](#) the kinked part of \tilde{p} between \tilde{y}_1 and \tilde{x}_2 lies above G . As a further substantive assumption, one can choose $(\tilde{x}_1, \tilde{y}_1, \tilde{y}_2, \tilde{x}_2)$ arbitrarily close to

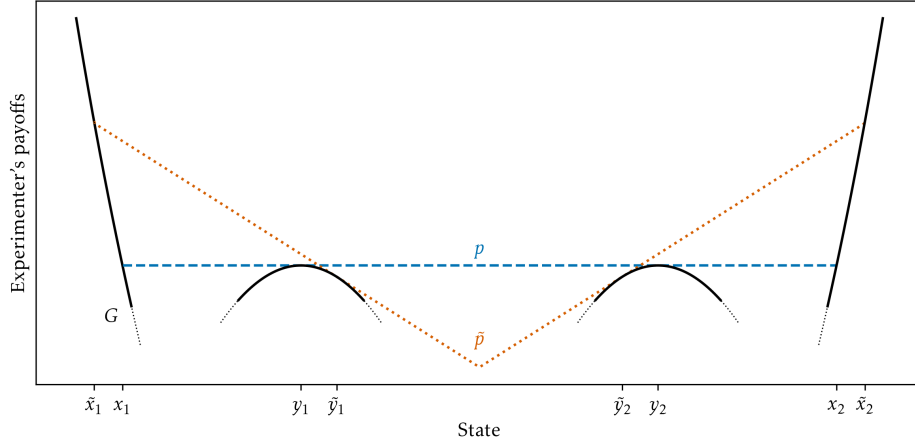


Figure 7: Two tuples and supporting price functions (colored dashed and dotted). The experimenter's payoff G is shown in black. The points y_1 and y_2 lie on two distinct intervals on which G is concave.

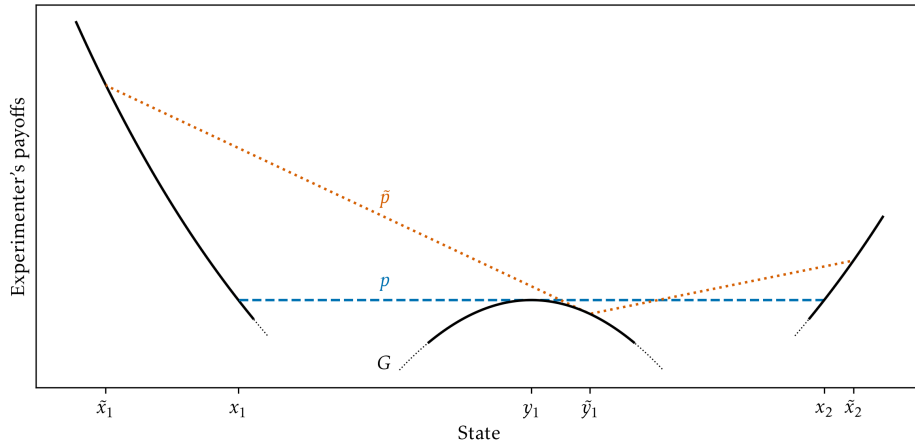


Figure 8: Two tuples and supporting price functions (colored dashed and dotted). The experimenter's payoff G is shown in black. Both tuples admit a single point that lies in an interval where G is concave.

(x_1, y_1, y_2, x_2) to meet the differentiability assumptions of [Definition 5](#). We shall use the price function \tilde{p} to construct other IC experiments whose support contains $\tilde{x}_1, \tilde{y}_1, \tilde{y}_2,$ and \tilde{x}_2 ; the fact that \tilde{p} coincides with G at these points allows to verify IC.

OB.1.2 Incomplete full revelation

Incomplete full revelation concerns the extreme points of $\text{MPC}(H)$. Recall, for each extreme point F , there are countably many, disjoint *bi-pooling intervals* $[x_1, x_2]$, such that F pools $[x_1, x_2]$ to at most two atoms and I_F lies strictly below I_H on the interior (x_1, x_2) . Outside

the union of the bi-pooling intervals, I_F coincides with I_H . A non-trivial interval $[a, b]$ is *fully revealing* if I_F coincides with I_H on $[a, b]$.

Note carefully that full revelation intervals are non-trivial by definition. An extreme point need not admit any full revelation intervals.

Definition 6. An extreme point F of $\text{MPC}(H)$ admits *incomplete full revelation* if there is a non-degenerate bi-pooling interval $[x_1, x_2]$ of F and $\delta > 0$ satisfying at least one of the following:

- (1) F admits two atoms on (x_1, x_2) , both $[x_1 - \delta, x_1]$ and $[x_2, x_2 + \delta]$ are full revelation intervals, and $0 < x_1 < x_2 < 1$ holds.
- (2) F admits exactly one atom on (x_1, x_2) , and at least one of the following holds:
 - (a) Both $[x_1 - \delta, x_1]$ and $[x_2, x_2 + \delta]$ are full revelation intervals, and $0 < x_1 < x_2 < 1$ holds.
 - (b) $[x_1 - \delta, x_1]$ is a full revelation interval, and $0 < x_1 < x_2 = 1$ holds.
 - (c) $[x_2, x_2 + \delta]$ is a full revelation interval, and $0 = x_1 < x_2 < 1$ holds.

In words, F fully reveals the state on non-degenerate intervals to the left and right of a bi-pooling interval (cases (1) and (2a)). Cases (2b) and (2c) are as (2a), except that one neighboring interval is the boundary of the state space.

OB.2 Profitable Restrictions

Say full delegation is not optimal if there does not exist optimal $F \in \text{BR}(H)$.

There exists an extreme point of $\text{MPC}(H)$ that is designer-preferred among $\text{BR}(H)$. We refer to such an extreme point as a *DEF* (**d**esigner-**o**ptimal **e**xtr**e**me **p**oint **u**nder **f**ull **d**elegation). We provide sufficient conditions such that a given DEF is not an optimal IC experiment; accordingly, full delegation is not optimal.

Theorem 4. *Let the designer's payoff u_D be continuously differentiable and strictly convex, let the experimenter's payoff u_E be continuously differentiable, and let the prior H admit a strictly positive density. If the experimenter has locally strict curvature preferences and there exists a DEF with incomplete full revelation, then full delegation is not optimal.*

Proof of Theorem 4. Let F be a DEF with incomplete full information. Let $\delta > 0$ and $[x_1, x_2]$ be as in the definition of incomplete full information. Let y_1 and y_2 be the atoms in $[x_1, x_2]$ (possibly $y_1 = y_2$).

Find a price function p as in Corollary 1 of [Dworczak and Martini \(2019\)](#), i.e. p is convex and continuous, and it holds $p \geq G$, $\text{supp } F \subseteq \{m \in [0, 1]: p(m) = G(m)\}$ and $\int p dF = \int p dH$.

Since $[x_1, x_2]$ is bi-pooling, Corollary 3 of [Arieli et al. \(2023\)](#) tells us that p is affine on $[x_1, x_2]$.

We distinguish the cases of [Definition 6](#) concerning the form of $[x_1, x_2]$. In all cases, the proof strategy is to use [Definition 5](#) to find a price function \tilde{p} that touches G close to points at which p equals G . Using the price-theoretic characterization of IC, we obtain a particular IC experiment \tilde{F} whose support is close to the support of F . For the particularly constructed experiment, we show the designer is strictly better off than under F , which completes the proof.

Case 1. F admits two atoms, $y_1 < y_2$, on (x_1, x_2) , both $[x_1 - \delta, x_1]$ and $[x_2, x_2 + \delta]$ are full revelation intervals, and $0 < x_1 < x_2 < 1$ holds.

Since p is affine on $[x_1, x_2]$ and $y_1, y_2 \in (x_1, x_2)$, both y_1 and y_2 must be in the interior of \mathcal{I}_{ccv} (in view of [Definition 5](#)). Also, since $I_F = I_H$ on $[x_1 - \delta, x_1] \cup [x_2, x_2 + \delta]$, the functions p and G agree on this set; in particular, G is convex on this set. Thus, $[x_1 - \delta, x_1] \cup [x_2, x_2 + \delta] \subseteq \mathcal{I}_{\text{cvx}}$. Consequently, the tuple (x_1, y_1, y_2, x_2) meets the hypothesis of [Definition 5](#). Let $\tilde{T} = (\tilde{x}_1, \tilde{y}_1, \tilde{y}_2, \tilde{x}_2)$ be as in [Definition 5](#); for all $r \in (0, 1]$, let $\tilde{p}(\cdot, r)$ be a supporting price function.

Throughout, we take r sufficiently close to 0 such that $\tilde{x}_1(r) \in [x_1 - \delta, x_1]$ and $\tilde{x}_2(r) \in [x_2, x_2 + \delta]$, which is possible by continuity of \tilde{T} .

The function $\tilde{p}(\cdot, r)$ obtained from [Definition 5](#) has domain $[\tilde{x}_1(r), \tilde{x}_2(r)]$. We extend to $[0, 1]$ by setting $\tilde{p}(m, r) = p(m)$ for all $m \notin [\tilde{x}_1(r), \tilde{x}_2(r)]$. The resulting function clearly lies above G . One may verify that $\tilde{p}(\cdot, r)$ is convex by observing that $\tilde{p}(\cdot, r)$ lies above G on $[\tilde{x}_1(r), \tilde{x}_2(r)]$, while $\tilde{x}_1(r)$ and $\tilde{x}_2(r)$, respectively, lie in the intervals $[x_1 - \delta, x_1]$ and $[x_2, x_2 + \delta]$, respectively, on which p coincides with G .

We claim $\tilde{y}'_1(0) > 0$ and $\tilde{y}'_2(0) < 0$. Indeed, for small r we have $\tilde{y}_1(r), \tilde{y}_2(r) \in \mathcal{I}_{\text{ccv}}$ since y_1 and y_2 are in the interior of \mathcal{I}_{ccv} . The points $(y_1, G(y_1))$ and $(y_2, G(y_2))$ in \mathbb{R}^2 lie on the graph of an affine function with slope $g(y_1) = g(y_2)$. Thus, for $\tilde{T}(r)$ to be supported by a convex price function, concavity of G around y_1 and y_2 requires $\tilde{y}_1(r) \geq y_1$ and $\tilde{y}_2(r) \leq y_2$ for sufficiently small r . Since \tilde{y}_1 and \tilde{y}_2 are assumed to be continuously differentiable with non-zero derivatives at $r = 0$, we conclude $\tilde{y}'_1(0) > 0$ and $\tilde{y}'_2(0) < 0$.

[Figure 7](#) schematically depicts the assumed and derived properties of p and $\tilde{p}(\cdot, r)$ for some fixed r .

Let q be the constant slope of I_F on (y_1, y_2) , i.e. $q = F(y_1)$. Since $[x_1, x_2]$ is bi-pooling with atoms y_1 and y_2 , both $i \in \{1, 2\}$ satisfy

$$I_H(x_i) + H(x_i)(y_i - x_i) = I_F(y_i) + q(y_i - y_i),$$

where the right side of course simply equals $I_F(y_i)$. The function $t \mapsto I_H(t) + H(t)(y_i - t)$

has a non-zero derivative at $t = x_i$. Therefore, for sufficiently small r the Inverse Function Theorem yields $\tilde{t}_i(r)$ such that

$$I_H(\tilde{t}_i(r)) + H(\tilde{t}_i(r))(\tilde{y}_i(r) - \tilde{t}_i) = I_F(y_i) + q(\tilde{y}_i(r) - y_i), \quad (10)$$

Moreover, $\tilde{t}_i(r) \rightarrow x_i$ as $r \rightarrow 0$, and $\tilde{t}_i(r)$ is differentiable with respect to r , with

$$\tilde{t}'_i(0) = \tilde{y}'_i(0) \frac{q - H(x_i)}{h(x_i)(y_i - x_i)}.$$

We have $\tilde{t}'_1(0) > 0$ and $\tilde{t}'_2(0) < 0$ since $\tilde{y}'_1(0) > 0$ and $\tilde{y}'_2(0) < 0$ and $x_1 < y_1$ and $y_2 < x_2$.²⁴ Finally, note that since I_F is affine on $[y_1, y_2]$, the right side of (10) simply equals $I_F(\tilde{y}_i(r))$, for all r .

Using $(\tilde{t}_1(r), \tilde{x}_1(r), \tilde{t}_2(r), \tilde{x}_2(r)) \rightarrow (x_1, x_1, x_2, x_2)$ as $r \rightarrow 0$, an application of the Intermediate Value Theorem shows that for r sufficiently small there exist $\tilde{s}_1(r) \leq \tilde{x}_1(r)$ and $\tilde{s}_2(r) \geq \tilde{x}_2(r)$ such that

$$\int_{[\tilde{s}_1(r), \tilde{t}_1(r)]} (\tilde{x}_1(r) - \omega) dH(\omega) = \int_{[\tilde{t}_2(r), \tilde{s}_2(r)]} (\tilde{x}_2(r) - \omega) dH(\omega) = 0, \quad (11)$$

i.e. $\tilde{x}_1(r)$ is the expectation with respect to H conditional on $[\tilde{s}_1, \tilde{t}_1(r)]$, and analogously for $\tilde{x}_2(r)$. We have $\tilde{s}_1(r) \rightarrow x_1$ and $\tilde{s}_2(r) \rightarrow x_2$ as $r \rightarrow 0$. The equations (11) imply

$$\forall i \in \{1, 2\}, \quad I_H(\tilde{s}_i(r)) + H(\tilde{s}_i(r))(\tilde{x}_i(r) - \tilde{s}_i(r)) = I_H(\tilde{t}_i(r)) + H(\tilde{t}_i(r))(\tilde{x}_i(r) - \tilde{t}_i(r)).$$

In what follows, fix r sufficiently small such that $\tilde{s}_1(r), \tilde{t}_1(r), \tilde{t}_2(r)$, and $\tilde{s}_2(r)$ are well-defined, and $x_1 - \delta < \tilde{s}_1(r) < \tilde{x}_1(r) < \tilde{t}_1(r) < \tilde{y}_1(r) < \tilde{y}_2(r) < \tilde{t}_2(r) < \tilde{x}_2(r) < \tilde{s}_2(r) < x_2 + \delta$.

We define an experiment \tilde{F}_r via its integrated CDF, $I_{\tilde{F}_r}$. For all $m \in [0, 1]$, let

$$I_{\tilde{F}_r}(m) = \begin{cases} I_F(m) & \text{if } m \notin [\tilde{s}_1(r), \tilde{s}_2(r)], \\ I_H(\tilde{s}_1(r)) + H(\tilde{s}_1(r))(m - \tilde{s}_1(r)) & \text{if } m \in [\tilde{s}_1(r), \tilde{x}_1(r)], \\ I_H(\tilde{t}_1(r)) + H(\tilde{t}_1(r))(m - \tilde{t}_1(r)) & \text{if } m \in (\tilde{x}_1(r), \tilde{y}_1(r)], \\ I_F(m) & \text{if } m \in (\tilde{y}_1(r), \tilde{y}_2(r)], \\ I_H(\tilde{t}_2(r)) + H(\tilde{t}_2(r))(m - \tilde{t}_2(r)) & \text{if } m \in (\tilde{y}_2(r), \tilde{x}_2(r)], \\ I_H(\tilde{s}_2(r)) + H(\tilde{s}_2(r))(m - \tilde{s}_2(r)) & \text{if } m \in (\tilde{x}_2(r), \tilde{s}_2(r)]. \end{cases}$$

The choice of $(\tilde{s}_1(r), \tilde{t}_1(r), \tilde{t}_2(r), \tilde{s}_2(r))$ in (10) and (11) ensures that \tilde{F}_r is an experiment, where we recall that I_F is affine on (y_1, y_2) and given by the right side of (10). Lemma 2 implies that \tilde{F}_r is IC since $\text{supp } \tilde{F}_r \subseteq \{m \in [0, 1] : \tilde{p}(m, r) = G(m)\}$.

We show that the designer strictly prefers \tilde{F}_r to F for r sufficiently small. The designer's

²⁴Also, $H(x_1) < q = F(y_1) < H(x_2)$ since the slope of I_F is $F(y_1)$ on (y_1, y_2) , while it is $H(x_i)$ at x_i , for both i .

utility from \tilde{F}_r is given by

$$\int_{x_1-\delta}^{\tilde{s}_1(r)} u_D(\omega) dH(\omega) + (H(\tilde{t}_1(r)) - H(\tilde{s}_1(r)))u_D(\tilde{x}_1(r)) + (q - H(\tilde{t}_1(r)))u_D(\tilde{y}_1(r)) \quad (12)$$

$$+ (H(\tilde{t}_2(r)) - q)u_D(\tilde{y}_2(r)) + (H(\tilde{s}_2(r)) - H(\tilde{t}_2(r)))u_D(\tilde{x}_2(r)) + \int_{\tilde{s}_2(r)}^{x_2+\delta} u_D(\omega) dH(\omega) \quad (13)$$

$$+ \int_{[0,1] \setminus [x_1-\delta, x_2+\delta]} u_D(m) dF(m).$$

The designer's utility from F is given by

$$\int_{x_1-\delta}^{x_1} u_D(\omega) dH(\omega) + (q - H(x_1))u_D(y_1) \quad (14)$$

$$+ (H(x_2) - q)u_D(y_2) + \int_{x_2}^{x_2+\delta} u_D(\omega) dH(\omega) \quad (15)$$

$$+ \int_{[0,1] \setminus [x_1-\delta, x_2+\delta]} u_D(m) dF(m).$$

We claim that for r sufficiently small, the sum in (12) is strictly larger than the sum in (14), and the sum in (13) is strictly larger than the sum in (15). The arguments are similar, so we show only the former. Dividing by r , the difference between (12) and (14) is

$$\frac{(u_D(\tilde{y}_1(r)) - u_D(y_1))}{r}(q - H(x_1)) + (u_D(y_1) - u_D(\tilde{x}_1(r)))\frac{(H(x_1) - H(\tilde{t}_1(r)))}{r} \quad (16)$$

$$+ \frac{1}{r} \int_{[\tilde{s}_1(r), x_1]} (u_D(\tilde{x}_1(r)) - u_D(\omega)) dH(\omega) \quad (17)$$

As $r \rightarrow 0$, the sum of the two terms in (16) converges to

$$\begin{aligned} & \tilde{y}'_1(0)(u'_D(y_1)(q - H(x_1)) - (u_D(y_1) - u_D(x_1))h(x_1)\tilde{t}'_1(0)) \\ & = \tilde{y}'_1(0) \left(u'_D(y_1) - \frac{u_D(y_1) - u_D(x_1)}{y_1 - x_1} \right) (q - H(x_1)). \end{aligned}$$

This limit is strictly positive since u_D is strictly convex and since $\tilde{y}'_1(0) > 0$ holds. Turning to the integral in (17), we argue that this integral vanishes. We have

$$\begin{aligned} & \frac{1}{r} \left| \int_{[\tilde{s}_1(r), x_1]} (u_D(\tilde{x}_1(r)) - u_D(\omega)) dH(\omega) \right| \\ & \leq \frac{1}{r} |(u_D(\tilde{x}_1(r)) - u_D(x_1))(H(x_1) - H(\tilde{s}_1(r)))| + \frac{1}{r} \left| \int_{[\tilde{s}_1(r), x_1]} (u_D(x_1) - u_D(\omega)) dH(\omega) \right|. \end{aligned}$$

The first term in this upper bound, $\frac{1}{r}(u_D(\tilde{x}_1(r)) - u_D(x_1))(H(x_1) - H(\tilde{s}_1(r)))$, vanishes as $r \rightarrow 0$ since $\tilde{x}_1(r)$ is differentiable at $r = 0$, and since $\tilde{s}_1(r) \rightarrow x_1$ as $r \rightarrow 0$. Consider the

second term. Find $L > 0$ such that $|u'_D| \leq L$. We then have

$$\begin{aligned}
& \frac{1}{r} \left| \int_{[\tilde{s}_1(r), x_1]} (u_D(x_1) - u_D(\omega)) dH(\omega) \right| \\
& \leq \frac{1}{r} \int_{[\tilde{s}_1(r), x_1]} |(u_D(x_1) - u_D(\omega))| dH(\omega) \\
& = \frac{1}{r} \int_{[\tilde{s}_1(r), x_1]} \left| \int_{[\omega, x_1]} u'_D(r) dr \right| dH(\omega) \\
& \leq \frac{L}{r} \int_{[\tilde{s}_1(r), x_1]} (x_1 - \omega) dH(\omega) \\
& = \frac{L}{r} (x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r))) + \frac{L}{r} \int_{[\tilde{s}_1(r), x_1]} (\tilde{x}_1(r) - \omega) dH(\omega) \\
& = \frac{L}{r} (x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r))) + \frac{L}{r} \int_{[x_1, \tilde{t}_1(r)]} (\omega - \tilde{x}_1(r)) dH(\omega) \quad (\text{by (11)}) \\
& \leq \frac{L}{r} (x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r))) + \frac{L}{r} (\tilde{t}_1(r) - \tilde{x}_1(r))(H(\tilde{t}_1(r)) - H(x_1))
\end{aligned}$$

Now, $\frac{1}{r}(x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r)))$ vanishes as $r \rightarrow 0$ since \tilde{x}_1 is differentiable at $r = 0$, and since $\tilde{s}_1(r) \rightarrow x_1$ as $r \rightarrow 0$. Moreover, $\frac{1}{r}(\tilde{t}_1(r) - \tilde{x}_1(r))(H(\tilde{t}_1(r)) - H(x_1))$ vanishes since $\tilde{t}_1(r)$ is differentiable at $r = 0$, and since $\tilde{t}_1(r)$ and $\tilde{x}_1(r)$ both converge to x_1 as $r \rightarrow 0$. Thus, (17) vanishes. Thus, the sum of (16) and (17) is strictly positive for r sufficiently close to 0.

Case 2. F admits exactly one atom y_1 on (x_1, x_2) , both $[x_1 - \delta, x_1]$ and $[x_2, x_2 + \delta]$ are full revelation intervals, and $0 < x_1 < x_2 < 1$ holds.

Similarly to the previous case, we have $y_1 \in \text{int } \mathcal{I}_{\text{ccv}}$, and $[x_1 - \delta, x_1] \cup [x_2, x_2 + \delta] \subseteq \mathcal{I}_{\text{cvx}}$. Thus, the tuple (x_1, y_1, y_1, x_2) meets the hypothesis of Definition 5. Let $\tilde{T} = (\tilde{x}_1, \tilde{y}_1, \tilde{y}_2, \tilde{x}_2)$ be as postulated by Definition 5. For all $r \in (0, 1]$, let $\tilde{p}(\cdot, r)$ be a supporting price function. The component $\tilde{y}_2(r)$ shall be irrelevant for our argument.

Recall that \tilde{y}_1 is continuously differentiable with non-zero derivative at $r = 0$. Without loss, let $\tilde{y}'_1(0) > 0$, the other case being similar.

Throughout, take r sufficiently small such that $\tilde{y}_1(r) > y_1$ and $\tilde{x}_1(r) \in [x_1 - \delta, x_1]$ and $\tilde{x}_2(r) \in [x_2, x_2 + \delta]$, where we recall that $\tilde{x}_1(r) \leq x_1$ and $\tilde{x}_2(r) \geq x_2$ follow directly from Definition 5.

Figure 8 schematically depicts the assumed and derived properties of p and $\tilde{p}(\cdot, r)$ for some fixed r .

As in the proof for the previous case, we extend $\tilde{p}(\cdot, r)$ to $[0, 1]$ by setting $\tilde{p}(m, r) = p(m)$ for all $m \notin [\tilde{x}_1(r), \tilde{x}_2(r)]$. The resulting function is continuous, convex, and lies above G .

Similar to the previous case, we next construct three thresholds $\tilde{s}_1(r), \tilde{t}_1(r), \tilde{s}_2(r)$. (The

reader is invited to think of the fourth threshold $\tilde{t}_2(r)$ from the previous case as being fixed to x_2 .) Note, it holds

$$I_H(x_1) + H(x_1)(y_1 - x_1) = I_H(x_2) + H(x_2)(y_1 - x_2),$$

and the function $t \mapsto I_H(t) + H(t)(y_1 - t)$ has a strictly positive derivative at x_1 . Therefore, an application of the Inverse Function Theorem shows that for r sufficiently small there exists $\tilde{t}_1(r)$ such that

$$\begin{aligned} I_H(\tilde{t}_1(r)) + H(\tilde{t}_1(r))(\tilde{y}_1(r) - \tilde{t}_1(r)) &= I_H(x_2) + H(x_2)(\tilde{y}_1(r) - x_2) \\ \Leftrightarrow \int_{[\tilde{t}_1(r), x_2]} (\tilde{y}_1(r) - \omega) dH(\omega) &= 0. \end{aligned}$$

Moreover, $\tilde{t}_1(r) \rightarrow x_1$ as $r \rightarrow 0$, and $\tilde{t}_1(r)$ is differentiable with respect to r , with

$$\tilde{t}'_1(0) = \tilde{y}'_1(0) \frac{H(x_2) - H(x_1)}{h(x_1)(y_1 - x_1)}.$$

Note $\tilde{t}'_1(0) > 0$ since $\tilde{y}'_1(0) > 0$ and $x_1 < y_1 < x_2$.

Using $(\tilde{x}_1(r), \tilde{t}_1(r), \tilde{x}_2(r)) \rightarrow (x_1, x_1, x_2)$ as $r \rightarrow 0$ and $\tilde{x}_1(r) \leq x_1 < \tilde{t}_1(r)$ and $x_2 \leq \tilde{x}_2(r)$, an application of the Intermediate Value Theorem shows that for r sufficiently small there exist $\tilde{s}_1(r)$ and $\tilde{s}_2(r)$ such that

$$\int_{[\tilde{s}_1(r), \tilde{t}_1(r)]} (\tilde{x}_1(r) - \omega) dH(\omega) = \int_{[x_2, \tilde{s}_2(r)]} (\tilde{x}_2(r) - \omega) dH(\omega) = 0. \quad (18)$$

We have $\tilde{s}_1(r) \leq \tilde{x}_1(r)$ and $\tilde{x}_2(r) \leq \tilde{s}_2(r)$. Moreover, $\tilde{s}_1(r) \rightarrow x_1$ and $\tilde{s}_2(r) \rightarrow x_2$ as $r \rightarrow 0$. The equations (18) are equivalent to

$$\begin{aligned} I_H(\tilde{s}_1(r)) + H(\tilde{s}_1(r))(\tilde{x}_1(r) - \tilde{s}_1(r)) &= I_H(\tilde{t}_1(r)) + H(\tilde{t}_1(r))(\tilde{x}_1(r) - \tilde{t}_1(r)), \\ I_H(x_2) + H(x_2)(\tilde{x}_2(r) - x_2) &= I_H(\tilde{s}_2(r)) + H(\tilde{s}_2(r))(\tilde{x}_1(r) - \tilde{s}_2(r)) \end{aligned}$$

Define the experiment \tilde{F}_r via its integrated CDF as follows:

$$\forall m \in [0, 1] \quad I_{\tilde{F}_r}(m) = \begin{cases} I_F(m) & \text{if } m \notin [\tilde{s}_1(r), \tilde{s}_2(r)], \\ I_H(\tilde{s}_1(r)) + H(\tilde{s}_1(r))(m - \tilde{s}_1(r)) & \text{if } m \in [\tilde{s}_1(r), \tilde{x}_1(r)], \\ I_H(\tilde{t}_1(r)) + H(\tilde{t}_1(r))(m - \tilde{t}_1(r)) & \text{if } m \in (\tilde{x}_1(r), \tilde{y}_1(r)], \\ I_H(x_2) + H(x_2)(m - x_2) & \text{if } m \in (\tilde{y}_1(r), \tilde{x}_2(r)], \\ I_H(\tilde{s}_2(r)) + H(\tilde{s}_2(r))(m - \tilde{s}_2(r)) & \text{if } m \in (\tilde{x}_2(r), \tilde{s}_2(r)]. \end{cases}$$

The choice of $\tilde{t}_1(r)$, $\tilde{s}_1(r)$, and $\tilde{s}_2(r)$ ensure that \tilde{F}_r is an experiment. [Lemma 2](#) implies that \tilde{F}_r is IC since $\text{supp } \tilde{F}_r \subseteq \{m \in [0, 1]: \tilde{p}(m, r) = G(m)\}$. We show the designer is strictly better off under \tilde{F}_r than F for r sufficiently small.

The designer's utility from \tilde{F}_r , denoted $U_D(\tilde{F}_r)$, is given by

$$U_D(\tilde{F}_r) = (H(\tilde{t}_1(r)) - H(\tilde{s}_1(r)))u_D(\tilde{x}_1(r)) + (H(x_2) - H(\tilde{t}_1(r)))u_D(\tilde{y}_1(r)) \\ + (H(\tilde{s}_2(r)) - H(x_2))u_D(\tilde{x}_2(r)) + \int_{[0,1] \setminus [\tilde{s}_1(r), \tilde{s}_2(r)]} u_D(m) dF(m)$$

The designer's utility from F , denoted $U_D(F)$, is given by

$$U_D(F) = (H(x_2) - H(x_1))u_D(y_1) + \int_{[\tilde{s}_1(r), x_1]} u_D(\omega) dH(\omega) \\ + \int_{[x_2, \tilde{s}_2(r)]} u_D(\omega) dH(\omega) + \int_{[0,1] \setminus [\tilde{s}_1(r), \tilde{s}_2(r)]} u_D(m) dF(m).$$

The difference $\frac{1}{r}(U_D(\tilde{F}_r) - U_D(F))$ equals

$$(H(x_2) - H(x_1)) \frac{u_D(\tilde{y}_1(r)) - u_D(y_1)}{r} + \frac{(H(x_1) - H(\tilde{t}_1(r)))}{r} (u_D(\tilde{y}_1(r)) - u_D(\tilde{x}_1(r))) \quad (19)$$

$$+ \frac{1}{r} \int_{[\tilde{s}_1(r), x_1]} (u_D(\tilde{x}_1(r)) - u_D(\omega)) dH(\omega) \quad (20)$$

$$+ \frac{1}{r} \int_{[x_2, \tilde{s}_2(r)]} (u_D(\tilde{x}_2(r)) - u_D(\omega)) dH(\omega). \quad (21)$$

We argue that, as $r \rightarrow 0$, the sum in (19) converges to a strictly positive term, whereas both (20) and (21) vanish.

The limit of (19) as $r \rightarrow 0$ equals

$$(H(x_2) - H(x_1))\tilde{y}'_1(0) \left(u'_D(y_1) - \frac{u_D(y_1) - u_D(x_1)}{y_1 - x_1} \right).$$

This limit is strictly positive since u_D is strictly convex and $\tilde{y}'_1(0) > 0$.

Next, the integral in (20) vanishes; the argument is analogous to the argument that in the previous **Case 1** the integral (17) vanishes. We have

$$\frac{1}{r} \left| \int_{[\tilde{s}_1(r), x_1]} (u_D(\tilde{x}_1(r)) - u_D(\omega)) dH(\omega) \right| \\ \leq \frac{1}{r} |(u_D(\tilde{x}_1(r)) - u_D(x_1))(H(x_1) - H(\tilde{s}_1(r)))| + \frac{1}{r} \left| \int_{[\tilde{s}_1(r), x_1]} (u_D(x_1) - u_D(\omega)) dH(\omega) \right|.$$

The first term in this upper bound, $\frac{1}{r}(u_D(\tilde{x}_1(r)) - u_D(x_1))(H(x_1) - H(\tilde{s}_1(r)))$, vanishes as $r \rightarrow 0$ since $\tilde{x}_1(r)$ is differentiable at $r = 0$, and since $\tilde{s}_1(r) \rightarrow x_1$ as $r \rightarrow 0$. Consider the second term. Find $L > 0$ such that $|u'_D| \leq L$. We have

$$\frac{1}{r} \left| \int_{[\tilde{s}_1(r), x_1]} (u_D(x_1) - u_D(\omega)) dH(\omega) \right|$$

$$\begin{aligned}
&\leq \frac{1}{r} \int_{[\tilde{s}_1(r), x_1]} |u_D(x_1) - u_D(\omega)| dH(\omega) \\
&= \frac{1}{r} \int_{[\tilde{s}_1(r), x_1]} \left| \int_{[\omega, x_1]} u'_D(r) dr \right| dH(\omega) \\
&\leq \frac{L}{r} \int_{[\tilde{s}_1(r), x_1]} (x_1 - \omega) dH(\omega) \\
&= \frac{L}{r} (x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r))) + \frac{L}{r} \int_{[\tilde{s}_1(r), x_1]} (\tilde{x}_1(r) - \omega) dH(\omega) \\
&= \frac{L}{r} (x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r))) + \frac{L}{r} \int_{[x_1, \tilde{t}_1(r)]} (\omega - \tilde{x}_1(r)) dH(\omega) \quad (\text{by (18)}) \\
&\leq \frac{L}{r} (x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r))) + \frac{L}{r} (\tilde{t}_1(r) - \tilde{x}_1(r))(H(\tilde{t}_1(r)) - H(x_1))
\end{aligned}$$

Now, $\frac{1}{r}(x_1 - \tilde{x}_1(r))(H(x_1) - H(\tilde{s}_1(r)))$ vanishes as $r \rightarrow 0$ since \tilde{x}_1 is differentiable at $r = 0$, and since $\tilde{s}_1(r) \rightarrow x_1$ as $r \rightarrow 0$. Moreover, $\frac{1}{r}(\tilde{t}_1(r) - \tilde{x}_1(r))(H(\tilde{t}_1(r)) - H(x_1))$ vanishes since $\tilde{t}_1(r)$ is differentiable at $r = 0$, and since $\tilde{t}_1(r)$ and $\tilde{x}_1(r)$ both converge to x_1 as $r \rightarrow 0$. Thus, (20) vanishes.

Finally, the integral (21) vanishes; the argument is simpler than but analogous to the argument that (20) vanishes. We have

$$\begin{aligned}
|(21)| &= \frac{1}{r} \left| \int_{[x_2, \tilde{s}_2(r)]} (u_D(\tilde{x}_2(r)) - u_D(\omega)) dH(\omega) \right| \\
&\leq \frac{1}{r} |(u_D(\tilde{x}_2(r)) - u_D(x_2))(H(\tilde{s}_2(r)) - H(x_2))| + \frac{1}{r} \left| \int_{[x_2, \tilde{s}_2(r)]} (u_D(x_2) - u_D(\omega)) dH(\omega) \right|.
\end{aligned}$$

The first term in this upper bound, $\frac{1}{r}(u_D(\tilde{x}_2(r)) - u_D(x_2))(H(\tilde{s}_2(r)) - H(x_2))$, vanishes as $r \rightarrow 0$ since $\tilde{x}_2(r)$ is differentiable at $r = 0$, and since $\tilde{s}_2(r) \rightarrow x_2$ as $r \rightarrow 0$. Consider the second term. Find $L > 0$ such that $|u'_D| \leq L$. We have

$$\begin{aligned}
\frac{1}{r} \left| \int_{[x_2, \tilde{s}_2(r)]} (u_D(x_2) - u_D(\omega)) dH(\omega) \right| &\leq \frac{1}{r} \int_{[x_2, \tilde{s}_2(r)]} |(u_D(x_2) - u_D(\omega))| dH(\omega) \\
&= \frac{1}{r} \int_{[x_2, \tilde{s}_2(r)]} \left| - \int_{[x_2, \omega]} u'_D(r) dr \right| dH(\omega) \\
&\leq \frac{L}{r} \int_{[x_2, \tilde{s}_2(r)]} (\omega - x_2) dH(\omega) \\
&= \frac{L}{r} \int_{[x_2, \tilde{s}_2(r)]} (\hat{x}_2(r) - x_2) dH(\omega) \quad (\text{by (18)}) \\
&= \frac{L}{r} (\hat{x}_2(r) - x_2)(H(\tilde{s}_2(r)) - H(x_2)),
\end{aligned}$$

Now, $\frac{L}{r}(\hat{x}_2(r) - x_2)(H(\tilde{s}_2(r)) - H(x_2))$ vanishes as $r \rightarrow 0$ since \hat{x}_2 is differentiable at $r = 0$ and since $\tilde{s}_2(r) \rightarrow x_2$ as $r \rightarrow 0$. Thus, (21) vanishes.

Case 3. F admits exactly one atom y_1 on (x_1, x_2) , the interval $[x_1 - \delta, x_1]$ is a full revelation interval, and $0 < x_1 < x_2 = 1$ holds.

The proof is similar to the proof of [Theorem 2](#), and hence we shall be brief. Abbreviate $y = y_1$ and $x = x_1$. As before, $y \in \text{int } \mathcal{I}_{\text{ccv}}$, and $[x - \delta, x] \subseteq \mathcal{I}_{\text{cvx}}$. In this case, the price p is affine on $[x, 1]$; in particular, $G(x) + g(y)(y - x) = G(y)$. For \tilde{y} converging to y from above, consider the affine function $t \mapsto G(\tilde{y}) + g(\tilde{y})(t - \tilde{y})$. For \tilde{y} close to y , this affine function intersects G at a point \tilde{x} in $[x - \delta, x]$ since $y \in \text{int } \mathcal{I}_{\text{ccv}}$, and $[x - \delta, x] \subseteq \mathcal{I}_{\text{cvx}}$. Consider the function \tilde{p} that coincides with p up to \tilde{x} , then equals the affine function up to \tilde{y} , and finally continues with a sufficiently steep slope such that \tilde{p} is convex and lies above G on $[\tilde{y}, 1]$. The function \tilde{p} is convex and lies above G on all of $[0, 1]$. Now find \tilde{t} such that $\tilde{y} = \mathbb{E}_H[\omega \mid \omega \in [\tilde{t}, 1]]$ and $\tilde{x} = \mathbb{E}_H[\omega \mid \omega \in [\tilde{s}, \tilde{t}]]$, which is possible for \tilde{y} sufficiently close to y . Finally, let \tilde{F} be the experiment that coincides with F up to \tilde{s} , then pools H on $[\tilde{s}, \tilde{t}]$ to \tilde{x} , and pools H on $[\tilde{t}, 1]$ to \tilde{y} . For \tilde{y} sufficiently close to y , the point \tilde{s} is in $[x - \delta, x]$, implying that \tilde{F} is supported on points m such that $\tilde{p}(m) = G(m)$. In particular \tilde{F} is IC. A calculation similar to the one in the proof of [Theorem 2](#) shows that the designer is strictly better off under \tilde{F} than under F .

Case 4. F admits exactly one atom y_1 on (x_1, x_2) , the interval $[x_2, x_2 + \delta]$ is a full revelation interval, and $0 = x_1 < x_2 < 1$ holds. Analogous to the previous case. \square

OC Extreme Point Representation

In this appendix, we provide a general extreme point representation of IC experiments. Let G satisfy part (i) of the regularity condition of [Dworczak and Martini \(2019\)](#), as in [Section 7.2](#); in particular, [Lemma 2](#) applies.

Let ν be a Borel probability measure on $\text{MPC}(H)$, where $\text{MPC}(H)$ has the L^1 -norm. Let $F \in \text{MPC}(H)$. Say ν represents F if $V(F) = \int V(\Phi) d\nu(\Phi)$ holds for all continuous linear functionals $V: \text{MPC}(H) \rightarrow \mathbb{R}$. Say ν is supported on a measurable set $\mathcal{F} \subseteq \text{MPC}(H)$ if $\nu(\mathcal{F}) = 1$.

Theorem 5. For all IC experiments F there exists a Borel probability measure ν that represents F and is supported on the set of incentive-compatible extreme points of $\text{MPC}(H)$.²⁵

²⁵The set of incentive-compatible extreme points of $\text{MPC}(H)$ is measurable since IC experiments form a closed set (see [Appendix A.1](#)) while the extreme points form a G_δ -set ([Aliprantis and Border \(2006, Lemma 7.63\)](#)).

The content of [Theorem 5](#) is that the extreme points can be chosen to be IC, and that they are extreme points of the entire set $\text{MPC}(H)$.

[Theorem 5](#) does *not* imply that the set of IC experiments is convex. Indeed, it is not generally convex: for S -shaped preferences of the experimenter, IC experiments can have at most one atom in the concave part of G ([Lemma 3](#)), implying that a proper mixture of two distinct double censorship experiments is not IC.

Proof of Theorem 5. Let F be IC. By [Lemma 2](#), there exists a continuous, convex function $p: [0, 1] \rightarrow \mathbb{R}$ such that $p \geq G$ and $\text{supp} F \subseteq \{m \in [0, 1]: p(m) = G(m)\}$. By Choquet's Representation Theorem (e.g. [Kleiner et al. \(2021, Proposition 1\)](#)), there exists a measure ν that represents F and is supported on the extreme points of $\text{MPC}(H)$.

We argue that ν -almost all Φ are IC. We use [Lemma 2](#). Let $K = \{m \in [0, 1]: p(m) = G(m)\}$ and $O = [0, 1] \setminus K$. The set K is closed and O is open (in $[0, 1]$) since $p - G \geq 0$ holds and since $p - G$ is lower semicontinuous.

Let γ be a continuous function such that $\gamma(m) > 0$ for all $m \in O$ and $\gamma(m) = 0$ for all $m \in K$; e.g., let $\gamma(m) = \min_{m' \in K} |m - m'|$, where the minimum is well-defined and continuous by closedness of K and Berge's Maximum Theorem. Thus, the linear functional $\Phi \mapsto \int \gamma(m) d\Phi(m)$ is continuous. Since ν represents F and $\text{supp} F \subseteq \{m \in [0, 1]: p(m) = G(m)\} = K$, we have $0 = \int \int_O \gamma(m) d\Phi(m) d\nu(\Phi)$. Since γ is strictly positive on O , it follows that ν -almost all Φ assign mass 0 to the set O . Since $[0, 1] \setminus O = K = \{m \in [0, 1]: p(m) = G(m)\}$ and since K is closed, it follows that ν -almost all Φ satisfy $\text{supp} \Phi \subseteq \{m \in [0, 1]: p(m) = G(m)\}$. Invoking [Lemma 2](#) with the price function p , it follows that ν -almost all Φ are IC. \square