# Coarse Information Design*

Qianjun Lyu
*Universidad Carlos III de Madrid*

Wing Suen
*University of Hong Kong*

Yimeng Zhang
*Universidad Carlos III de Madrid*

October 10, 2026

*Abstract:* We study an information design problem with continuous state and limited signal resources. Under convex and S-shaped value functions, the optimal information structure is interval-partitional and exhibits a dual expectations property: each induced signal is the conditional mean (taken under the prior density) of each interval; and each interval cutoff is the barycenter (taken under the value function curvature) of the interval formed by neighboring signals. This property enables an examination into which part of the state space is more finely partitioned. The analysis can be extended to general value functions and adapted to study coarse mechanism design.

*Keywords*: constrained information design, limited signal resources, dual expectations, scrutiny, coarse nonlinear pricing

*JEL Classification*: D81, D82, D83

*We are especially grateful to Gregorio Curello for his continuous support and insightful discussions. We thank Sarah Auster, Yunus Aybas, Dirk Bergemann, Florian Brandl, Francesc Dilmé , Michael Greinecker, Li Hao, Ángel Hernando-Veciana, Ilwoo Hwang, Mathijs Janssen, Andreas Kleiner, Jin Li, Benny Moldovanu, Paula Onuchic, Ran Shorrer, Ali Shourideh, Aloysius Siow, Peter Sørensen, Junze Sun, Dezső Szalay, Lina Uhe, Kun Zhang, Chen Zhao, and various seminar and conference audiences for valuable suggestions and comments. We have benefited tremendously from the constructive feedback of anonymous referees and wish to express our special thanks to co-editor Gabriel Carroll, whose meticulous review significantly improved the clarity and rigor of this work and uncovered an error in an earlier version of this paper. Lyu acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866. Zhang acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2047/1-390685813.

## 1. Introduction

Coarse information is ubiquitous. Teachers evaluate students by letter grades or on a pass/fail basis; online platforms often convey product reviews through coarse rating systems; and regulatory policies disclose the efficiency and safety level of different products via finite categories. These coarse information mechanisms can emerge due to limited cognitive and memory capacities, imperfect communication channels, technological constraints on the measurement instruments, or simply for the sake of convenience. In pain medicine, the intensity of pain naturally falls into a continuum, but it would be almost impossible for patients to fully describe their subjective feeling or convert it into a real number on a continuous scale. Instead, doctors rely on instruments such as the Numerical Rating Scale, from 0 to 10, based on patients' self-reported pain level to reach a diagnosis. In an organizational context, individuals at different levels of the hierarchy may not have the time or the expertise to digest complicated information from each other if the communication involves exhaustive details. In practice, they typically adopt language protocols that are less precise but more comprehensible to facilitate decision-making. Furthermore, because different individuals in the organization are making different decisions, the optimal way to coarsen information is tailored to different needs. An executive summary of a research report written for the head of engineering, for example, should be quite different from a summary of the same report written for the CEO. In this paper, we take the coarse nature of information structure as given and study how to design information optimally subject to such a constraint.

Formally, we study a class of information design problems where the uncertain state $\theta$ is continuously distributed on $[0,1]$ and is payoff relevant only through its expectation. The (information) designer—or the sender—commits to a Blackwell experiment $(\pi, \Sigma)$, where $\Sigma$ is the signal space and $\pi(\sigma|\theta)$ is the probability of obtaining signal realization $\sigma \in \Sigma$ conditional on state $\theta$. The receiver updates her belief based on the realized signal and then chooses an action optimally. We introduce a constraint on the "signal resources" such that $\Sigma$ is restricted to be a finite set with a cardinality of at most $N$. Our model allows sender and receiver to be the same person. In that case, an unconstrained information design problem would be trivial as the optimal information structure is fully informative. While in our model, given that the state is continuous and the signals are finite, a fully informative experiment is infeasible. In other words, a coarse experiment necessarily involves pooling across different states. How to allocate

the limited signal resources then becomes a relevant economic question—which parts of the state space should receive *closer scrutiny* in the experiment relative to other parts?

To better motivate our results, in this introduction, unless otherwise specified, we illustrate our findings under scenarios in which the sender's and receiver's preferences are sufficiently aligned such that more information is strictly beneficial (i.e., the sender's value function $u$ as a function of the induced posterior mean is convex). In this case, the insights from our coarse information design problem are orthogonal to those prevalent in the unconstrained information design literature, where coarsening arises from the designer's strategic incentive to pool states when the value function has a non-convex region.

Consider an *interval-partitional* information structure, where an experiment divides the state space into a finite number of subintervals and reveals which subinterval the state belongs to. We say that a subinterval of the state space receives "closer scrutiny" than others if the width of this subinterval is smaller than that of others. In Figure 1, we represent an experiment by the set of cutoffs that defines the partition. Under experiment $A$, the state space in the neighborhood of 0.7 receives the closest scrutiny. In contrast, experiment $B$ explores the state space near 0.3 with the closest scrutiny. We refer to the subinterval that receives the closest scrutiny as the "center of scrutiny." In both experiments, the intervals become progressively wider, farther away from the center of scrutiny.

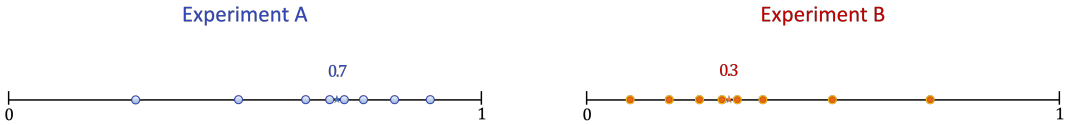



Figure 1: Experiment $A$ gives close scrutiny to the state space near 0.7; experiment $B$ gives close scrutiny to the state space near 0.3.

Naturally, the designer should give closer scrutiny to states that are more likely to happen (according to the prior distribution), because having the receiver make informed decisions under those states carries a greater weight in his ex ante expected utility. This insight is developed in other applications such as quantization theory (Gray and Neuhoff, 1998) or preference evolution (Robson, 2001; Netzer, 2009). On the other hand, the designer should also give closer scrutiny to those states where local information is more valuable (i.e., where the value function has greater curvature, measured

by the second derivative $u''$ of the value function). The latter concern has received less attention in the existing literature, let alone the connection between these two forces.

Our first main result characterizes the designer's optimal interval-partitional information structure and establishes a novel dual relation between the prior density $f$ and the curvature of the value function $u''$. For each interval-partitional experiment, represented by the cutoffs $\{s_k\}_{k=0}^N$, the signal $x_k$ induced within the interval $(s_{k-1}, s_k)$ equals the expected value of this interval, computed using the conditional density of the prior distribution, $f(\cdot)/(F(s_k) - F(s_{k-1}))$. Under the *optimal* interval partition, each cutoff point $s_k$ equals the "expected value" of the interval formed by two adjacent signals $(x_k, x_{k+1})$, computed using the conditional "density", $u''(\cdot)/(u'(x_{k+1}) - u'(x_k))$. We call such a property of the optimal experiment *dual expectations*.

The dual expectations property enables us to characterize the optimal experiment as the fixed point of a monotone equation system. The second key finding highlights an additional feature of the optimal interval partition: In a *logconcave* environment—where both the prior density $f$ and the curvature $u''$ are logconcave—the widths of the subintervals are *single-dipped*. Specifically, there is a center of scrutiny—a subinterval that receives the closest scrutiny. As we move away from this center, the level of scrutiny diminishes for states located farther away (as illustrated in Figure 1). This pattern of information structure is commonly observed in various mandatory energy-efficiency rating standards, such as Energy Performance Certificates for residential buildings in Austria, France, Germany, and Italy. [1] We leverage the dual expectations property to facilitate comparative statics analysis. When either the prior density $f$ or the curvature function $u''$ adopts a likelihood-ratio increase, all the interval cutoff points $\{s_k\}_{k=1}^{N-1}$ and the posterior means $\{x_k\}_{k=1}^N$ will shift to the right. Furthermore, if either $f$ or $u''$ becomes less variable according to the *uniform conditional variability order* (Whitt, 1985), then both the interval cutoff points and the posterior means become more compressed, in the sense that there exists an $n^*$ such that $s_k$ shift to the right for all $k < n^*$ and $s_k$ shifts to the left for all $k > n^*$.

With convex value functions, our framework has a natural application to mechanism design where the agent's payoff is linear in her private information and the contract is constrained to be finite. In Section 4, we illustrate how to transform a standard nonlinear pricing problem with a finite menu into a coarse information design problem.

[1] For specific guidelines, see Austria (page 5), France (page 29), Germany, and Italy (Table 2).

Most of our results are transferable to study the design and properties of the optimal finite menu.

In Sections 5 and 6, we discuss S-shaped value functions and more general value functions to address scenarios where the optimal unconstrained information structure involves pooling (i.e., when $u'' < 0$ for some regions of the state space). With S-shaped value functions, the optimal coarse information structure is interval-partitional. Moreover, the dual expectations characterization, the single-dipped property, as well as the comparative statics results with respect to the change of the curvature function $u''$, remain true. In particular, the dual expectations property is modified with respect to a signed measure—the optimal cutoff point $s_k$ equals the *barycenter* of the interval formed by two adjacent signals $(x_k, x_{k+1})$ under the signed measure $u'$. These results rely on the fact that only the last subinterval of the optimal coarse information structure contains the concave region of the state space. With more general value functions, interval-partitional information structures need not be optimal. However, within this class of information structures, we show that the dual expectations property remains valid, incorporating the potential usage of a bi-tangency condition when the barycenter is not well defined.

We provide further discussions of our model in Section 7. Section 7.1 discusses the scenario where there are only finite actions and the number of messages is smaller than the number of actions. Section 7.2 uses numerical examples beyond logconcave environments to demonstrate the robustness of our conclusions regarding the center of scrutiny. In Section 7.3 we establish connections between our model and cheap talk.

**Related literature.** Our paper contributes to the literature on information design with a continuous state space. We adopt the approach introduced by Gentzkow and Kamenica (2016), which represents an information structure by the integral of its cumulative distribution function. Kolotilin (2018) and Dworczak and Martini (2019) use alternative approaches to studying this problem. Kleiner et al. (2021) and Arieli et al. (2023) show that the optimal unconstrained information structure exhibits a bi-pooling property. Curello and Sinander (2023) study the comparative statics of mean-measurable persuasion problems and identify the conditions under which a sender with "more convex" value function will design a more informative signal structure. The bulk of the Bayesian persuasion literature starting from Kamenica and Gentzkow (2011) and Rayo and Segal (2010) is concerned with the strategic use of pooling: how the sender

designs an experiment to "concavify" his value function by pooling information to influence the receiver's action. When the signal space is constrained to be finite, pooling becomes necessary even when the value function is convex. Our paper examines how to effectively pool information in such environments.

The constraint imposed by coarse information on the operation of markets is discussed by Wilson (1989), McAfee (2002), Hoppe et al. (2010) and Kos (2012). Dow (1991) considers a sequential search problem for a decision maker whose memory is represented by a partition of the set of possible past prices. Harbaugh and Rasmusen (2018) and Ostrovsky and Schwarz (2010) analyze coarse grading from an information design perspective, while Crémer et al. (2007) study coarse language code under a discrete environment where the loss function only depends on the number of states that are pooled together in one message.

Recently, the literature on information design has begun to explore the limitation of communication and information channels. Mensch (2021); Onuchic and Ray (2023); Kolotilin et al. (2025) study information design subject to monotone information structures. Aybas and Turkel (2025) examine an information design problem under a bounded signal space. They characterize the highest achievable sender payoff and analyze how it changes with the cardinality of the signal space. Their paper explores finite state and finite action space, and allows the sender's payoff to depend arbitrarily on beliefs. In contrast, we focus specifically on mean-measurable persuasion problems and provide a general characterization for the optimal information structure.

Our paper is related to Hopenhayn and Saeedi (2024), which studies the optimal design of coarse rating systems that partition a continuum of sellers with heterogeneous qualities into a finite number of quality groups in a competitive market. They characterize the social planner's optimal rating scheme in market equilibrium, compare optimal rating schemes under different shapes of supply function, and develop comparative statics for rating thresholds with respect to the prior distribution of seller quality under the assumption of linear supply. In contrast to our setting, because equilibrium supply is endogenous, their designer's objective function (total welfare) is generally nonlinear in the distribution of posterior quality except for the linear supply case. Tian (2022) uses "cell functions" (a mapping from subintervals to payoffs) as primitives to study the optimal interval partition problem. He focuses on submodular cell functions, and shows that the optimal interval cutoffs shift up when the prior distribution shifts up.

We use interim value functions as primitives and characterize the optimal information structure for general value functions. Moreover, the dual relation between value function curvature and prior density allows us to develop comparative statics with respect to changes in payoff functions.

The coarse information design problem also resembles quantization in information theory (Gray and Neuhoff, 1998; Mease and Nair, 2006). Early contributions in the quantization literature (Graf and Luschgy, 2000) typically maximize the expected within-group "similarity" based on a distance metric, e.g., minimizing the expected distance between the realized values and their cluster centroids, which in general, cannot be transformed into a well-defined interim convex value function over a domain of posterior means. More recently, a growing literature (Banerjee et al., 2005; Boissonnat et al., 2010; Fischer, 2010; Boutoille and Pagès, 2025) has generalized quantization and clustering problems using Bregman divergences. We show that while information design and Bregman quantization originate from different objectives, the optimal solutions coincide. We discuss this connection in depth in Section 3.

## 2. Model

Consider an information design model where the payoff-relevant state $\theta$ is drawn from a prior distribution $F$ on $[0,1]$. We assume that $F$ admits a continuous density function $f$ with full support. A designer commits to an information structure: a signal space $\Sigma$ and a mapping $\pi : [0,1] \to \Delta(\Sigma)$ from the state to a distribution over signals, which induces a random posterior. We focus on models where the designer's interim payoff depends on the induced posterior belief only through the posterior mean,[2] and denote the interim value function by $u : [0,1] \to \mathbb{R}$. Throughout the paper, we maintain the assumption that $u$ is twice continuously differentiable and $u''$ is not equal to zero almost everywhere. In Section 7.1, we consider value functions that are piecewise affine, a situation that often arises when the underlying action space is finite. Since the only relevant information is the posterior mean, it is without loss of generality to assume

[2] As Dworczak and Martini (2019) point out, the posterior-mean dependent setup is satisfied when the receiver's optimal action only depends on the expected state and the sender's preference over action is linear on the state. Specifically, the receiver's utility function can be expressed as $y_1(a) + y_2(a)\theta + y_3(\theta)$ and the designer's utility function can be represented by $z_1(a) + z_2(a)\theta$, where $a$ is an element from a compact set $A$ and $y_1, y_2, y_3, z_1, z_2$ are upper semi-continuous functions. Ivanov (2021) observes that if the designer's utility function includes a third term dependent solely on the state $z_3(\theta)$, the optimal experiment remains unchanged.

that the realized signal is the posterior mean itself. From now on, we will primarily work with the induced distribution of posterior means, denoted by $G$.

We follow Gentzkow and Kamenica (2016) by representing an information structure as the integral of the induced distribution function of posterior means. Let $F_0$ be the degenerate distribution that puts probability mass one on the prior mean of $F$. For any distribution $G$ on $[0,1]$, define $I_G(x) := \int_0^x G(t)\,\mathrm{d}t$ and call it the *integral distribution* of $G$. By Strassen's (1965) theorem, $G$ can be induced by some signal structure if and only if $I_{F_0} \leq I_G \leq I_F$; namely, $G$ must be a mean-preserving contraction of $F$.

The main ingredient of our model is the introduction of a capacity constraint on the signal space $\Sigma$. In particular, we require $\Sigma$ to contain no more than $N$ elements. Consequently, the distribution $G$ of posterior means can only have a finite support. We can then write the designer's optimization program as:

$$\max_{G\in\triangle([0,1])} \int_0^1 u(x)\,\mathrm{d}G(x) \tag{1}$$

$$\text{s.t.} \qquad I_{F_0} \leq I_G \leq I_F, \tag{MPC}$$

$$|\text{supp}\,(G)| \leq N. \tag{D}$$

A distribution with finite support has an integral distribution which is increasing, convex and piecewise affine, with kinks at every element in the support of the distribution. Let ICPA denote the set of such integral distribution functions defined on $[0,1]$ that satisfy the mean-preserving contraction constraint (MPC). For every $I \in$ ICPA, define the set of kink points of $I$ by $\mathscr{K}_I := \{x \in (0,1) : I'(x^-) \neq I'(x^+)\}$.

We now transform the objective function (1) into an explicit integral of an ICPA function:

$$\int_0^1 u(x)\,\mathrm{d}G(x) = u(1) - u'(1)I_F(1) + \int_0^1 u''(x)I_G(x)\,\mathrm{d}x$$

The first two terms are constants. Hence, the original problem can be rewritten as:

$$\max_{I_G\in\text{ICPA}} \int_0^1 u''(x)I_G(x)\,\mathrm{d}x \tag{2}$$

$$\text{s.t.} \qquad \left|\mathscr{K}_{I_G}\right| \leq N. \tag{D}$$

The objective function (2) represents the "signed weighted" area under the integral distribution function, where the signed weights are given by the *curvature* of the value function, $u''$. Intuitively, the level of an integral distribution $I_G$ measures how informative the corresponding information structure is locally. A higher $I_G$ is closer to the full-information integral distribution $I_F$. It yields a higher payoff when providing local information is very rewarding, that is, when associated with higher positive $u''$. Similarly, a lower $I_G$ is closer to the null-information integral distribution $I_{F_0}$, and is more preferable to the designer if $u''$ is negative. Obviously, if $u$ is convex within some region, a local upward shift of $I_G$ increases the designer's expected payoff, provided that such a shift does not exceed $I_F$. Moreover, a clockwise rotation of a certain segment of $I_G$ at the point when $u$ switches from convex to concave is beneficial, provided that $I_G$ remains in the ICPA class.

Throughout the paper, we repeatedly visit the following class of applications to illustrate the results.

**Example 1** (Purchase Decision)**.** The state $\theta \in [0,1]$ is the value of a good, and the price of the good is $p$. The decision maker chooses to buy the good ($a = 1$) or not ($a = 0$). She first observes a signal from some experiment to learn about $\theta$. After the signal is observed but before the purchase decision is made, a cost shock $\eta$ is realized, where $\eta$ is distributed with density $h$ on $[\underline{\eta}, \overline{\eta}]$ with $\underline{\eta} \leq -p$, which can be interpreted as either the actual cost of consuming the good or the opportunity cost of outside option. When her posterior expectation of the state is $x$, she buys the good if and only if $x - p - \eta \geq 0$. Her interim utility function is her expected consumer surplus, $u(x) = \int_{\underline{\eta}}^{x-p} (x - p - \eta) h(\eta) \, \mathrm{d}\eta$, which is a convex function with curvature $u''(x) = h(x-p)$. On the other hand, the seller wants to maximize the probability of sale. Hence, his interim payoff is the expected revenue $\hat{u}(x) = pH(x-p)$, with curvature $\hat{u}''(x) = ph'(x-p)$. The example encapsulates two sets of models: when the *buyer* designs the information structure, the framework corresponds to the model where a single agent designs a coarse experiment to make a better-informed decision in the future, pioneered by Dow (1991); when the *seller* designs the information structure, it corresponds to a problem of persuading a privately informed receiver (Kolotilin et al., 2017). The former case features convex value functions and is discussed in Section 3, while the latter case features S-shaped value functions when the cost shock follows a unimodal density and is discussed in Section 5.

A natural class of coarse information structures is the set of interval-partitional information structures. That is, we partition the state space $[0,1]$ into $N$ subintervals, each of which sends a distinct signal. Interval partitions are widely observed in contexts such as certification or grading schemes, where signals naturally correspond to intervals over an underlying quality or performance scale. They are easier to implement than general signal structures that may require randomization or involve pooling states from disparate parts of the state space. Motivated by these considerations, in the main text of this paper, we restrict our attention to this particular class of information structures. In most of this paper, the solution to this problem also solves the more general problem without such a restriction. We will be clear when this issue arises.

For an interval-partitional information structure $G$ defined by cutoff points, $0 = s_0 < s_1 < \ldots < s_N = 1$, its integral representation is illustrated in Figure 2. The tangency points between $I_F$ and $I_G$ correspond to these cutoff points because $G(s_k) = I'_G(s_k) = I'_F(s_k) = F(s_k)$ for all $k = 0, \ldots, N$. Moreover each kink point $x_k$ in $I_G$ between two adjacent tangency points $s_{k-1}$ and $s_k$ represents the posterior mean of the state over the corresponding subinterval $(s_{k-1}, s_k)$.

**Lemma 1.** *For any $I_G \in$ ICPA, if $x_k$ is the only kink point between two adjacent tangency points $s_{k-1}$ and $s_k$, then $x_k = \mathbb{E}_F[\theta \mid \theta \in (s_{k-1}, s_k)]$.*

*Proof.* The kink point $x_k$ satisfies the equation,

$$I_F(s_{k-1}) + F(s_{k-1})(x_k - s_{k-1}) = I_F(s_k) + F(s_k)(x_k - s_k).$$

Solving this equation gives:

$$x_k = \frac{\int_{s_{k-1}}^{s_k} \theta \, \mathrm{d}F(\theta)}{\int_{s_{k-1}}^{s_k} \mathrm{d}F(\theta)}.$$

□

In fact, for any $N$-interval partition $G$, it is effectively the only information structure to induce the corresponding distribution of posterior means.

## 3. Convex Value Functions

In this section, we discuss convex value functions, where sender and receiver's preferences are sufficiently aligned such that more information is strictly beneficial. The

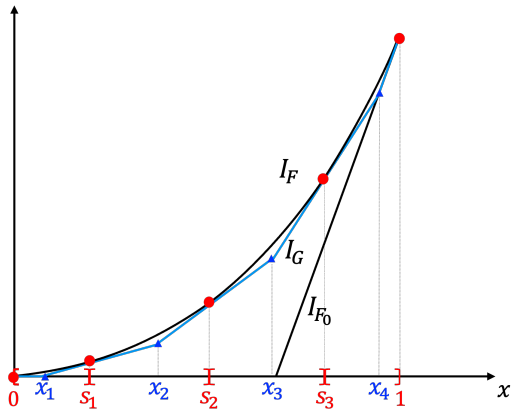



Figure 2: The integral distribution function of an interval partition with 4 signals.

finiteness constraint raises an important question regarding how to pool information efficiently when the number of disposable "signal resources" is limited. Moreover, this class of value functions incorporates many natural applications, such as a single-agent decision problem or a sender-receiver game with quadratic utilities.

Upon normalization, it is without loss of generality in this case to treat $u''$ as a density function that captures the relative value of information provision. Our main characterization of the optimal interval partition features a dual relation between the prior density $f$ and the curvature density $u''$. For any $0 \leq a < b \leq 1$, let

$$\phi(a,b) := \mathbb{E}_F[t \mid t \in (a,b)],$$
$$\mu(a,b) := \mathbb{E}_{u'}[t \mid t \in (a,b)].$$

Here, $\mathbb{E}_{u'}$ is the conditional expectation operator using $u''(\cdot)/(u'(b)-u'(a))$ as the conditional density function. When $u$ is strictly convex, this is a valid density of full support, and the "conditional expectation" function $\mu(\cdot)$ is well defined.

**Theorem 1.** *Suppose $u$ is strictly convex. The optimal $N$-interval partition $G$ exists and is characterized by $\{s_k\}_{k=0}^N$ and $\{x_k\}_{k=1}^N$ satisfying:*

$$x_k = \phi(s_{k-1}, s_k) \qquad \text{for } k = 1,\dots,N; \tag{CE-$F$}$$
$$s_k = \mu(x_k, x_{k+1}) \qquad \text{for } k = 1,\dots,N-1. \tag{CE-$u'$}$$

*Proof. Existence:* The designer's objective function can be viewed as a function of $N-1$ interval-cutoffs. It is continuous and the feasible set of cutoffs is compact. It follows

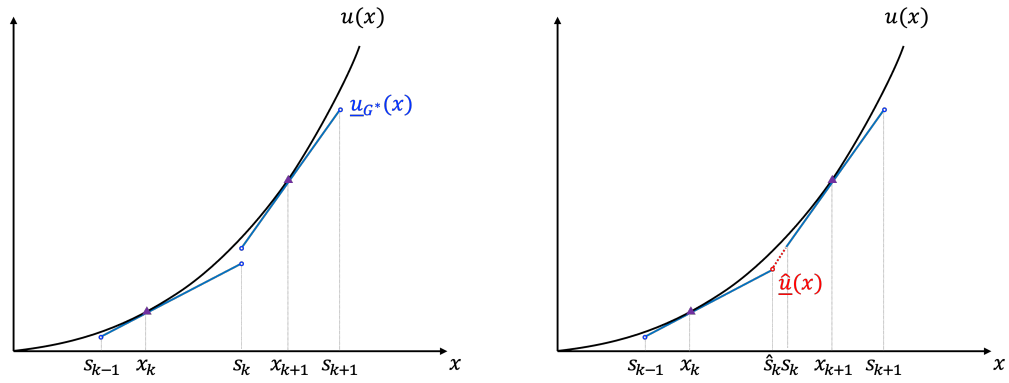



Figure 3: Suboptimality of discontinuous minorant function

that an optimal solution exists.

*Characterization:* First, (CE-$F$) must hold by definition of an interval-partitional information structure. To prove (CE-$u'$), for any interval-partitional structure $G$, we define the *minorant function* $\underline{u}_G$ on each (open) segment as follows:

$$\underline{u}_G(x) := \sum_{k=1}^{N} \Big[u(x_k) + u'(x_k)(x - x_k)\Big] \mathbb{I}_{(s_{k-1}, s_k)}(x). \tag{3}$$

On each subinterval $(s_{k-1}, s_k)$, $\underline{u}_G(\cdot)$ is the subgradient of $u(\cdot)$ at the posterior mean of that subinterval. By construction, a minorant function is piecewise affine with increasing slopes.

Under an optimal interval partition $G^*$, the minorant function must be *continuous* at the interval cutoff points. Suppose to the contrary that $\underline{u}_{G^*}(s_k^-) \neq \underline{u}_{G^*}(s_k^+)$. In particular, suppose $\underline{u}_{G^*}(s_k^-) < \underline{u}_{G^*}(s_k^+)$ (see left panel of Figure 3). Then we consider another function,

$$\underline{\hat{u}}(x) = \begin{cases} \max\{u(x_k) + u'(x_k)(x - x_k),\, u(x_{k+1}) + u'(x_{k+1})(x - x_{k+1})\} & \text{if } x \in (s_{k-1}, s_{k+1}), \\ \underline{u}_{G^*}(x) & \text{otherwise.} \end{cases}$$

By construction $\underline{\hat{u}}(x)$ is continuous at $x = \hat{s}_k$, where $\hat{s}_k$ satisfies $u(x_k) + u'(x_k)(\hat{s}_k - x_k) = u(x_{k+1}) + u'(x_{k+1})(\hat{s}_k - x_{k+1})$. Given $\underline{u}_{G^*}(s_k^-) < \underline{u}_{G^*}(s_k^+)$, it is immediate that $\hat{s}_k \in (s_{k-1}, s_k)$ (see right panel of Figure 3). Moreover, $\underline{\hat{u}}(x) \geq \underline{u}_{G^*}(x)$ for all $x$ in each open segment, and the inequality holds strictly for $x \in (\hat{s}_k, s_k)$.

Now consider an interval-partitional structure $\hat{G}$, which is characterized by the cutoffs $\{s_0, \ldots, s_{k-1}, \hat{s}_k, s_{k+1}, \ldots, s_N\}$. The following inequalities show that $\hat{G}$ generates strictly higher expected payoff than $G^*$.

$$\begin{aligned}\int_0^1 u(x)\,\mathrm{d}G^*(x) &= \int_0^1 \underline{u}_{G^*}(x)\,\mathrm{d}G^*(x) = \int_0^1 \underline{u}_{G^*}(x)\,\mathrm{d}F(x)\\ &< \int_0^1 \underline{\hat{u}}(x)\,\mathrm{d}F(x) = \int_0^1 \underline{\hat{u}}(x)\,\mathrm{d}\hat{G}(x) \le \int_0^1 u(x)\,\mathrm{d}\hat{G}(x).\end{aligned}$$

The first equality follows from the fact that the information structure $G^*$ puts mass points of each subinterval $(s_{k-1}, s_k)$ on the corresponding posterior mean $x_k$. The second equality comes from the linearity of $\underline{u}_{G^*}$ in each subinterval. Though the values of $\underline{u}_G(x)$ at the interval cutoffs are not assigned at this stage, this equality remains true because $F$ puts zero measure on this countable set. The first inequality on the second line follows because $\underline{\hat{u}}$ is everywhere above $\underline{u}_{G^*}$. The subsequent equality again comes from the fact that $\underline{\hat{u}}$ is piecewise affine. Finally the last inequality is due to the fact that $u$ is everywhere above $\underline{\hat{u}}$. This string of inequalities demonstrates that if $\underline{u}_{G^*}$ is not continuous at any cutoff point $s_k$, then there exists another information structure $\hat{G}$ that outperforms $G^*$—a contradiction. The analogous argument could be formulated when $\underline{u}_{G^*}(s_k^-) > \underline{u}_{G^*}(s_k^+)$, in which case $\hat{s}_k \in (s_k, s_{k+1})$.

We can therefore complete the minorant function at each $s_k$ by letting $\underline{u}_{G^*}(s_k) = \underline{u}_{G^*}(s_k^-) = \underline{u}_{G^*}(s_k^+)$. The continuity of $\underline{u}_{G^*}$ implies that the completed minorant function under optimal information structure must be a convex, piecewise affine function. Finally, if we regard $u$ as the "integral distribution" of $u'$, then $u$ bears a similar relationship to $\underline{u}_{G^*}$ as $I_F$ bears to $I_G$. Consequently, Lemma 1 implies that $s_k = \mathbb{E}_{u'}[\theta \mid \theta \in (x_k, x_{k+1})]$ for $k = 1, \ldots, N-1$. □

Under an interval-partitional information structure, each signal $x_k$ is the posterior expectation of the state conditional on the interval $(s_{k-1}, s_k)$. Theorem 1 states that under the *optimal* interval-partitional information structure, each interval cutoff $s_k$ must be the expectation of the random variable that follows distribution $u'$, conditional on the interval $(x_k, x_{k+1})$.[3]

To better understand the result, first note that due to the linear structure of payoff functions, every affine segment of the minorant function $\underline{u}_G$ on the subinterval $(s_{k-1}, s_k)$

[3]We use $u'$ here as a shorthand to stand for the distribution function $(u'(\cdot) - u'(0))/(u'(1) - u'(0))$.

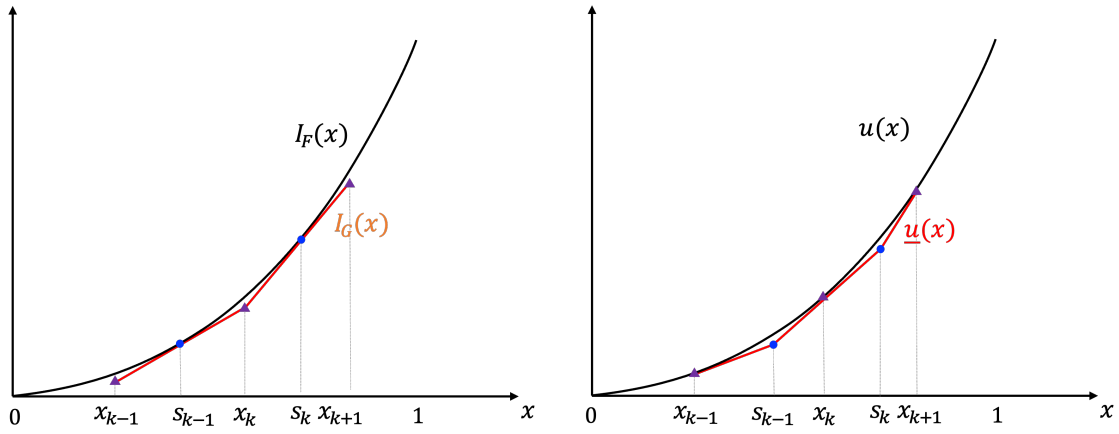



Figure 4: Dual expectations.

represents the designer's payoff when an action is taken upon the realization of the signal $x_k$.[4] Therefore, every two adjacent subgradients must intersect at the cutoff point $s_k$, for otherwise one can always construct a new information structure that produces a set of uniformly higher pieces of subgradients, yielding a higher payoff to the designer. To put it differently, the designer should be indifferent between categorizing the cutoff state $s_k$ into either the upper subinterval or the lower subinterval under the optimal information structure. Figure 4 puts together the minorant function and the integral distribution of an interval partition. The minorant function $\underline{u}$ exhibits a geometric relationship with $u$ akin to the relationship between $I_G$ and $I_F$, which leads to the property of dual expectations.

**Remark.** When the value function $u$ is convex, our geometric representation of dual expectations coincides with the solution of a Bregman quantization problem.[5] Specifically, for any two points $p, q \in [0,1]$, define the Bregman divergence induced by $u$ to measure the distortion as

$$D_u(p|q) = u(p) - u(q) - u'(q)(p-q).$$

[4] In the case of single-agent information acquisition problems, the piece of subgradient is exactly the designer's payoff at each state by the envelope theorem (Milgrom and Segal, 2002). When the designer and the receiver do not share perfectly aligned interests, due to the linear payoff structure, the subgradient generates the same expected payoff on each partitioned subinterval as the designer's actual payoff function.

[5] We thank a referee for drawing our attention to this literature and providing many helpful comments.

The $N$-Bregman quantization seeks to minimize the expected Bregman distortion, namely,

$$\min_{q_1,\ldots,q_N} \int \min\{D_u(p|q_1),\ldots,D_u(p|q_N)\}\,\mathrm{d}F(p);$$

or equivalently,

$$\max_{q_1,\ldots,q_N} \int \max\left\{u(q_1)+u'(q_1)(p-q_1),\ldots,u(q_N)+u'(q_N)(p-q_N)\right\}\,\mathrm{d}F(p).$$

Fix a set of centroids $\{q_k\}_{k=1}^N$, the integrand of the original minimization problem induces a Bregman Voronoi diagram (Boissonnat et al., 2010), where cells are separated by Bregman bisectors, pinned down by intersections of affine tangent functions. As a result, the integrand of the equivalent maximization problem is a continuous, piecewise affine function, arising from the upper envelope of $N$ affine functions. At the Bregman quantization optimum, each centroid $\{q_k\}_{k=1}^N$ must coincide with the conditional expectation of its associated Bregman Voronoi cell (Banerjee et al., 2005, Proposition 1). On the other hand, in coarse information design, the signals $\{x_k\}_{k=1}^N$ are constrained to be the conditional means of partitioned subintervals by construction whereas the continuity of the minorant function emerges as an optimality condition rather than a built-in geometric property. Nevertheless, the solutions to these two problems coincide and such geometric equivalence implies that the characterization of the optimal coarse information structure can be reached via Lloyd's algorithm adapted to Bregman divergence (see Banerjee et al. (2005)), and all the results in this Section apply to the uni-dimensional Bregman quantization.

The characterization of (CE-$F$) and (CE-$u'$) can extend to general value functions where $u'$ is a general signed measure with a well-behaved Radon-Nikodym derivative (with respect to Lebesgue measure) and the conditional "expectations" are calculated accordingly. We will provide more details in Sections 5 and 6.

Theorem 1 gives the necessary condition for the optimal $N$-interval partition. The next result identifies environments under which the equation system has a unique solution, and therefore the necessary condition is also sufficient. Logconcave priors are commonly used in information economics and are satisfied in many classes of probability distributions. In contrast, logconcavity of $u''$ concerns an endogenous object—the curvature of the value function. In binary decision problems where the receiver has

private information, the logconcavity of $u''$ is equivalent to the prior distribution of the receiver's private signal being logconcave, as illustrated in the purchase decision example (Example 1). Later we discuss other examples where the curvature can also be logconcave.

**Proposition 1.** *If both $f$ and $u''$ are logconcave, the solution to the equation system (CE-F) and (CE-$u'$) is unique and characterizes the optimal $N$-interval partition.*

The proof relies on the following property of logconcave densities. Because we will further exploit this property for subsequent results, we state it here for easier reference. Obviously, a similar property as stated in Lemma 2 holds for the $\mu(\cdot)$ function if $u''$ is logconcave. When both $f$ and $u''$ are logconcave, we refer to that as a *logconcave environment*.

**Lemma 2** (Mease and Nair (2006); Szalay (2012))**.** *If $f$ is logconcave and has full support, then for any $0 \le a < b \le 1$ and any $\varepsilon_1, \varepsilon_2 \ge 0$ such that $[a+\varepsilon_1, b+\varepsilon_2] \subseteq [0,1]$,*

$$\phi(a+\varepsilon_1, b+\varepsilon_2) \le \phi(a,b) + \max\{\varepsilon_1, \varepsilon_2\},$$

*with strict inequality if $\varepsilon_1 \neq \varepsilon_2$.*

*Proof of Proposition 1.* The equation system (CE-$F$) and (CE-$u'$) is a monotone mapping. By Tarski's theorem, there exists a largest and a smallest fixed point (denoted $(x_1', s_1', \ldots, x_N')$ and $(x_1'', s_1'', \ldots, x_N'')$, respectively). If these two fixed points do not coincide, the first index where these two vectors differ may be either odd or even. If it is odd and the index is not $N$, let the corresponding elements be $x_j' > x_j''$. Because $s_{j-1}' = s_{j-1}''$, Lemma 2 implies that $s_j' - s_j'' > x_j' - x_j''$. Iterating this argument forward leads to $x_N' - x_N'' > s_{N-1}' - s_{N-1}''$. But $x_N' - x_N'' = \phi(s_{N-1}', 1) - \phi(s_{N-1}'', 1) < s_{N-1}' - s_{N-1}''$ by Lemma 2, a contradiction. Meanwhile, if $x_N' > x_N''$ are the first elements that differ, then the contradiction trivially holds. If the first index where the two fixed points differ is even, let the corresponding elements be $s_k' > s_k''$. A parallel argument again leads to $x_N' - x_N'' > s_{N-1}' - s_{N-1}''$, a contradiction. Since we already showed that the optimal solution exists and there is only one candidate solution that satisfies the necessary condition for optimality, this unique solution must be the optimal solution. [6] □

[6] Recently, Boutoille and Pagès (2025) prove a similar uniqueness result for optimal Bregman quantization.

When the value function is convex, the optimal $N$-interval partition characterized above also solves the general information design problem (2) without restricting to interval-partitional structures. The designer can gain a higher payoff by raising the informativeness of the experiment (i.e., by raising $I_G$) whenever $u''$ is positive. Since it is always feasible to raise $I_G$ if $G$ is not interval-partitional, it is without loss of generality to focus on interval-partitional structures. Moreover, the "information-maximizing" designer should always fully utilize the "signal resources" by setting $|\mathscr{K}_{I_G}| = N$.

**Proposition 2.** *Suppose $u$ is convex. The optimal information structure $G$ is an interval-partitional structure with $|\mathscr{K}_{I_G}| = N$.*

The optimality of interval partition is also shown in Ivanov (2021). For completeness, we provide a simple proof in the Appendix based on the representation of integral distributions. Specifically, for any integral distribution $I_G$ that is not piecewise tangent to $I_F$ or that has less than $N$ kinks, we construct an alternative $I_{G'}$ with $N$ kinks, induced by an interval-partitional information structure, such that $I_{G'}$ lies everywhere weakly above $I_G$. This implies that the experiment $G$ is a mean-preserving contraction of $G'$. The argument implies that the optimality of interval partitions holds for any convex value function $u$, even if $u$ is not strictly convex or not differentiable.

### 3.1. Allocating scarce signal resources

If a region of the state space is finely partitioned into more subintervals, then the experiment reveals finer details about that region of the state space as the receiver is more likely to distinguish across signal realizations. Given an interval partition, we say a subinterval receives *closer scrutiny* than another if the width of this subinterval is smaller than that of another. Fixing the number of signals, if some region of the state space receives closer scrutiny, then some other region of the state space will receive less scrutiny. A natural question is: Which part of the state space should receive closer scrutiny in the optimal experiment?

In the simplest case, both $f$ and $u''$ are uniform. Then $\phi(a,b) = \mu(a,b) = (a+b)/2$. Equations (CE-$F$) and (CE-$u'$) imply that the state space $[0,1]$ is divided into $N$ equal-sized intervals with $s_k = k/N$ for $k = 0,\dots,N$; and the induced posterior means are evenly spaced, with $x_k = (2k-1)/(2N)$ for $k = 1,\dots,N$. The optimal experiment gives every part of the state space equal scrutiny.

To develop more general lessons, denote the width of the $k$-th interval $(s_{k-1}, s_k)$

by $w_k = s_k - s_{k-1}$ for $k = 1, \ldots, N$, and the distance between two adjacent posterior means by $d_k = x_{k+1} - x_k$ for $k = 1, \ldots, N-1$. Let $\{\Delta_i\}_{i=1}^{2N-1}$ be the interleaved sequence $\{w_1, d_1, w_2, \ldots, d_{N-1}, w_N\}$, where $\Delta_{2k-1} = w_k$ and $\Delta_{2k} = d_k$. See Figure 5. We say that the sequence $\{\Delta_i\}_{i=1}^{2N-1}$ is *single-dipped* if it is decreasing then increasing or if it is monotone.

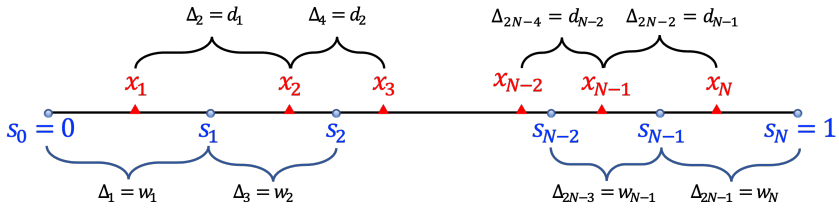



Figure 5: The interleaved sequence $\{\Delta_i\}$.

**Theorem 2.** *In a logconcave environment,* $\{\Delta_i\}_{i=1}^{2N-1}$ *is single-dipped.*

*Proof.* It suffices to show that $\Delta_i \geq \Delta_{i-1}$ implies $\Delta_{i+1} \geq \Delta_i$ for any $i = 2, \ldots, 2N-2$. There are two cases to consider.

Suppose $i$ is even, and $d_k \geq w_k$ but $w_{k+1} < d_k$ for some $k < N-1$. Then

$$x_{k+1} = \phi(s_k, s_{k+1}) = \phi(s_{k-1} + w_k, s_k + w_{k+1}) < \phi(s_{k-1} + d_k, s_k + d_k) \leq x_k + d_k = x_{k+1}.$$

The strict inequality obtains because $\phi$ is strictly monotone in each argument. The second inequality comes from Lemma 2. This is a contradiction. For $k = N-1$ , we have

$$x_N = \phi(s_{N-1}, s_N) = \phi(s_{N-2} + w_{N-1}, s_{N-1} + w_N) \leq \phi(s_{N-2} + d_{N-1}, s_{N-1} + w_N) < x_{N-1} + d_{N-1} = x_N.$$

In this case, the strict inequality follows from Lemma 2. Contradiction. Suppose $i$ is odd, and $w_k \geq d_{k-1}$ but $d_k < w_k$. Applying a similar argument above, we obtain

$$s_k = \mu(x_k, x_{k+1}) = \mu(x_{k-1} + d_{k-1}, x_k + d_k) < \mu(x_{k-1} + w_k, x_k + w_k) \leq s_{k-1} + w_k = s_k,$$

which is also a contradiction. □

An immediate consequence of Theorem 2 is that both the sequence of widths of subintervals, $\{w_k\}_{k=1}^N$, and the sequence of distances between adjacent posterior means,

$\{d_k\}_{k=1}^{N-1}$, are single-dipped in a logconcave environment. The part of the state space where $w_i$ attains the minimum is the region that receives the closest scrutiny in the optimal experiment. The result that the widths and distances are single-dipped implies there is a center of scrutiny—the subinterval with the minimum width. Moreover, the optimal experiment pays less and less scrutiny to states that are farther and farther away from this center.

Define the index $i^*$ such that $w_{i^*} = \min\{w_1, \ldots, w_N\}$. The subinterval $(s_{i^*-1}, s_{i^*})$ is the center of scrutiny of the optimal information structure. The location of this center generally reflects the designer's two concerns: (1) revealing finer information about the states that are more likely to occur (higher density $f$); and (2) focusing on states with higher local value of information (greater curvature $u''$). Since the logconcavity of $f$ and $u''$ implies that these functions are single-peaked (Dharmadhikari and Joag-dev, 1988), the two ends of the state space have lower probability mass and smaller curvature when both distributions have interior modes. Intuitively, if we fix one of the concerns, such as assuming a uniform prior distribution, the scrutiny center is likely to be near the mode of the curvature. The following result confirms this intuition and extends it to a more general environment where $f$ and $u''$ are single-peaked with possibly different modes, and they need not be logconcave.

**Proposition 3.** *(i) If $u''$ is a constant and $f$ is single-peaked with mode $m_f$ in the $j$-th interval, then $i^* \in \{j-1, j, j+1\}$.*

*(ii) If $f$ is a constant and $u''$ is single-peaked with mode $m_u$ in the $k$-th interval, then $i^* \in \{k-1, k, k+1\}$.*

*(iii) If $f$ is single-peaked with mode $m_f$ in the $j$-th interval and $u''$ is single-peaked with mode $m_u$ in the $k$-th interval and $k \leq j$, then $i^* \in \{k-1, \ldots, j+1\}$.*

**Example 2.** A risk-neutral agent with unknown productivity $\theta \in [0,1]$ chooses effort level $a \geq 0$ to maximize the expected wage net of effort cost $C(a)$. Before exerting effort, the agent designs a coarse information structure to learn about his productivity. Output is given by $a\theta$, and the wage scheme is linear in output with a fixed piece rate $\beta \in (0,1)$. Thus, the agent's utility is $\beta a\theta - C(a)$. The prior distribution of $\theta$ is single-peaked at $m_f = 0.5$. Let $a^*(x)$ be the agent's chosen level of effort when the posterior mean of $\theta$ is $x$. The agent's interim utility is $u(x) = \beta a^*(x)x - C(a^*(x))$. Suppose the cost function takes the iso-elastic form, $C(a) = a^\gamma/\gamma$, with $\gamma > 1$. Then, the agent's effort supply also takes the iso-elastic form, $a^*(x) = (\beta x)^{1/(\gamma-1)}$. By the

envelope theorem, $u'(x) = \beta a^*(x)$, and therefore

$$u''(x) = \frac{\beta^2}{\gamma - 1}(\beta x)^{\frac{2-\gamma}{\gamma-1}}.$$

Now consider two cases. When $\gamma > 2$, $u''(x)$ is strictly decreasing in $x$, so it is single-peaked with a mode at $m_u = 0$. Intuitively, when the marginal cost of effort rises sharply, the agent's optimal effort becomes relatively unresponsive to expected productivity at higher values of $x$. As a result, the agent's payoff is not very sensitive to $x$ in the upper part of the productivity distribution, making it less valuable to finely differentiate high-productivity states. In this case, the optimal information structure for the agent will resemble Experiment B in Figure 1, with a center of scrutiny near the lower part of the productivity distribution. Note that in this case, $u''(x)$ is not logconcave, but the location of scrutiny center is still described by Proposition 3. On the other hand, when $\gamma \in (1, 2)$, $u''(x)$ is strictly increasing in $x$ and is also logconcave. In this case, the mode of curvature shifts to $m_u = 1$. Here, the marginal cost of effort is less steep at the top, so effort is more sensitive to the change in productivity at the upper end. Consequently, the agent's utility is more sensitive to posterior means in the upper range, and the optimal information structure will resemble Experiment A in Figure 1, with a center of scrutiny near the upper part of the productivity distribution.

### 3.2. Comparative statics

In this section, we present some comparative statics results as we vary the prior distribution $f$ and the curvature of value function $u''$.

**Definition 1.** A function $\hat{f}$ *dominates* $f$ in likelihood ratio order (written as $\hat{f} \succeq_{lr} f$) if $\hat{f}(a)f(b) \leq \hat{f}(b)f(a)$ for any $0 \leq a \leq b \leq 1$.

If both $f$ and $\hat{f}$ are probability density functions with common support, then the above definition implies that $\hat{f}(\cdot)/f(\cdot)$ is increasing. Definition 1 does not require $f$ or $\hat{f}$ to be positive, which is useful later on when we discuss comparative statics for S-shaped value functions (where the curvature function $u''$ is single-crossing).

**Proposition 4.** *In a logconcave environment, if either (a) the prior density changes from $f$ to $\hat{f}$, with $\hat{f} \succeq_{lr} f$; or (b) the value function changes from $u$ to $\hat{u}$, with $\hat{u}'' \succeq_{lr} u''$, then all the interval cutoff points $\{s_k\}_{k=1}^{N-1}$ and all the induced posterior means $\{x_k\}_{k=1}^{N}$ will increase.*

*Proof.* We prove the case where $f$ changes to $\hat{f}$. The proof for the change of value function is analogous. A likelihood ratio increase in $f$ raises the conditional mean $\phi(a,b)$ for any given $a$ and $b$. In a logconcave environment the optimal $(x_1,\ldots,x_N,s_1,\ldots,s_{N-1})$ is the unique fixed point of $\Gamma$, the mapping corresponding to the right-hand-side of the equation system (CE-$F$) and (CE-$u'$). Since a likelihood ratio increase from $f$ to $\hat{f}$ raises $\Gamma$, and $\Gamma$ is monotone, a standard result in monotone comparative statics (Topkis, 1998, Corollary 2.5.2) establishes that any fixed point under $\hat{f}$ is larger than the unique (and hence smallest) fixed point under $f$. □

Proposition 4 holds as long as the solution to the system of equations (CE-$F$) and (CE-$u'$) is unique either in the original environment or in the new environment. The comparative statics result with regard to the change of prior distribution is intuitive. A higher distribution (in the likelihood ratio order) means that the state is more likely to fall in the upper part of the state space. Naturally, the optimal experiment will induce more signals that reflect higher states.[7]

Part (b) of Proposition 4 is more novel to the literature and deserves some additional discussion. For two value functions $u$ and $\hat{u}$, an increasing ratio of their corresponding curvatures $\hat{u}''/u''$ is equivalent to the condition that the marginal value function $\hat{u}'$ is *more convex* than $u'$, i.e., there exists a strictly increasing and convex function $\psi$ such that $\hat{u}'(x)=\psi(u'(x))$. This condition implies that the designer's value function under $\hat{u}$ on average changes faster in the upper region. Therefore, the designer has incentives to learn higher states more precisely, as information is more valuable in higher states.

**Example 1a (continued).** We now revisit the purchase decision example through the lens of Proposition 4. The buyer and the seller differ in the relative value of information across different regions of the state space. The buyer's value function curvature is $h(x-p)$, which represents the ex ante probability that she is indifferent between making the purchase or not when the posterior mean is $x$. She values information most around the region where she is most likely to be indifferent (when consumer surplus from trade $x-p$ is near the mode of the cost shock). The seller, however, has a value function curvature $ph'(x-p)$. He only cares about the probability of trade; hence information

[7] Tian (2022) shows this part of the result in a more general environment where uniqueness is not required. Hopenhayn and Saeedi (2024) establish this part of result in the context of implementable rating thresholds in market equilibrium under linear supply or Cournot competition. Szalay (2012), Chen and Gordon (2015), Deimen and Szalay (2023) and Smith et al. (2021) make similar observations in the context of cheap talk models and social learning models.

is most valuable to him when the probability that the buyer is on the margin between buying and not buying changes fastest. When $h$ is an increasing function, the seller's value function is convex. Moreover, if $h$ is logconcave, then $\hat{u}''/u'' = ph'(x-p)/h(x-p)$ is decreasing in $x$. That is, the buyer's value of information increases faster with $x$ than does the seller's. Suppose the state $\theta$ follows some distribution with log-concave density function, then our result implies that the partitional cutoffs $\{s_k\}_{k=1}^{N-1}$ and induced posterior means $\{x_k\}_{k=1}^{N}$ under the buyer-optimal information structure are uniformly higher than those under the seller-optimal information structure.

Next, we investigate how the optimal information structure changes when either $f$ or $u''$ becomes less variable.

**Definition 2.** A function $\hat{f}$ is *uniformly less variable* than $f$ (written as $f \succeq_{uv} \hat{f}$) if there exists some $p \in (0,1)$ such that $\hat{f}(a)f(b) \leq \hat{f}(b)f(a)$ for all $0 \leq a \leq b \leq p$ and $\hat{f}(a)f(b) \geq \hat{f}(b)f(a)$ for all $p \leq a \leq b \leq 1$.

Again, we adopt this formulation to facilitate analysis in later sections where single-crossing functions are involved. When $\hat{f}$ and $f$ are probability density functions, the above definition reduces to $\hat{f}(\cdot)/f(\cdot)$ being unimodal. If these two distributions are not (first-order) stochastically ordered, then $\hat{f}$ is said to be *uniformly less variable than* $f$ (Whitt, 1985; Shaked and Shanthikumar, 2007).[8] Intuitively, when either the density or the curvature becomes less variable, scrutiny tends to increase near the region where $f$ or $u''$ approaches the peak. The next result formalizes the reasoning.

**Proposition 5.** *In a logconcave environment, if either (a) the prior density changes from $f$ to $\hat{f}$, with $f \succeq_{uv} \hat{f}$; or (b) the value function changes from $u$ to $\hat{u}$, with $u'' \succeq_{uv} \hat{u}''$, then the new sequence of optimal signals and interval cutoffs, $\{\hat{x}_1, \hat{s}_1, \ldots, \hat{s}_{N-1}, \hat{x}_N\}$, crosses the original sequence $\{x_1, s_1, \ldots, s_{N-1}, x_N\}$ at most once and from above.*

*Proof.* We prove the proposition for (a); the proof for (b) is similar. Suppose $\hat{f}/f$ reaches a peak at $p \in (0,1)$, and let $k^*$ be the integer such that $\hat{s}_{k^*} \leq p < \hat{s}_{k^*+1}$. Define the sequence,

$$\{\delta_i\}_{i=1}^{2N-1} := \{x_1 - \hat{x}_1, s_1 - \hat{s}_1, \ldots, s_{N-1} - \hat{s}_{N-1}, x_N - \hat{x}_N\}.$$

[8] Uniform variability order does not require $f$ and $\hat{f}$ to have the same mean. For example, a normal distribution with a smaller variance is uniformly less variable than another normal distribution with a larger variance, regardless of the values of the respective means.

To prove this proposition, it suffices to show that the two subsequences, $\{\delta_i\}_{i=1}^{2k^*+1}$ and $\{\delta_i\}_{i=2k^*+1}^{2N-1}$, are both single-crossing from below.

Let $j$ be the first nonnegative term in the subsequence $\{\delta_i\}_{i=1}^{2k^*+1}$. We want to show that $\delta_i \geq 0$ implies $\delta_{i+1} \geq 0$ and $\delta_i > 0$ implies $\delta_{i+1} > 0$ for $i = j, \ldots, 2k^*$.

Suppose $j$ is odd and $j > 1$; that is, $\delta_j = x_k - \hat{x}_k \geq 0$ and $\delta_{j-1} = s_{k-1} - \hat{s}_{k-1} < 0$ for some $k \in \{2, \ldots, k^*\}$. Let $\hat{\phi}(\cdot)$ be the conditional mean function under $\hat{f}$. We have

$$\begin{aligned} x_k - \hat{x}_k = \phi(s_{k-1}, s_k) - \hat{\phi}(\hat{s}_{k-1}, \hat{s}_k) &\leq \phi(s_{k-1}, s_k) - \phi(\hat{s}_{k-1}, \hat{s}_k) \\ &< \phi(\hat{s}_{k-1}, \max\{s_k, \hat{s}_k\}) - \phi(\hat{s}_{k-1}, \hat{s}_k) \\ &\leq \max\{0, s_k - \hat{s}_k\}, \end{aligned}$$

where the first inequality follows because $\hat{f}(\cdot)/f(\cdot)$ is increasing on $[0, p]$, and the second inequality follows because $\phi(\cdot, \cdot)$ is strictly increasing in both arguments. The last inequality comes from Lemma 2. Since $\delta_j = x_k - \hat{x}_k$ is nonnegative, the sequence of inequalities implies $\delta_{j+1} = s_k - \hat{s}_k > x_k - \hat{x}_k = \delta_j \geq 0$. Carrying this argument forward,

$$s_k - \hat{s}_k = \mu(x_k, x_{k+1}) - \mu(\hat{x}_k, \hat{x}_{k+1}) \leq \max\{x_k - \hat{x}_k, \max\{x_{k+1} - \hat{x}_{k+1}, 0\}\}.$$

Since $s_k - \hat{s}_k > x_k - \hat{x}_k \geq 0$ and therefore $x_k - \hat{x}_k \neq x_{k+1} - \hat{x}_{k+1} > 0$, the above inequality must hold strictly by Lemma 2, which implies $\delta_{j+2} = x_{k+1} - \hat{x}_{k+1} > s_k - \hat{s}_k = \delta_{j+1} > 0$. Similarly,

$$\begin{aligned} x_{k+1} - \hat{x}_{k+1} = \phi(s_k, s_{k+1}) - \hat{\phi}(\hat{s}_k, \hat{s}_{k+1}) &\leq \phi(s_k, s_{k+1}) - \phi(\hat{s}_k, \hat{s}_{k+1}) \\ &\leq \max\{s_k - \hat{s}_k, \max\{s_{k+1} - \hat{s}_{k+1}, 0\}\}. \end{aligned}$$

Since $x_{k+1} - \hat{x}_{k+1} > s_k - \hat{s}_k > 0$ and therefore $s_k - \hat{s}_k \neq s_{k+1} - \hat{s}_{k+1} > 0$, again the last inequality above must hold strictly by Lemma 2, which further implies $s_{k+1} - \hat{s}_{k+1} > x_{k+1} - \hat{x}_{k+1} > 0$. Continue this line of argument inductively, we show that $\{\delta_j, \ldots, \delta_{2k^*+1}\}$ is positive and strictly increasing.[9]

Suppose $j = 1$, then $x_1 \geq \hat{x}_1$ and $s_0 = \hat{s}_0 = 0$. If $x_1 > \hat{x}_1$, We have

$$0 < x_1 - \hat{x}_1 = \phi(0, s_1) - \hat{\phi}(0, \hat{s}_1) \leq \phi(0, s_1) - \phi(0, \hat{s}_1) \leq \max\{0, s_1 - \hat{s}_1\},$$

[9]Notice that the proof for the boundary case $x_{k^*+1} - \hat{x}_{k^*+1} = \delta_{2k^*+1} > \delta_{2k^*} = s_{k^*} - \hat{s}_{k^*}$ is obtained from the preceding $\mu$-comparison.

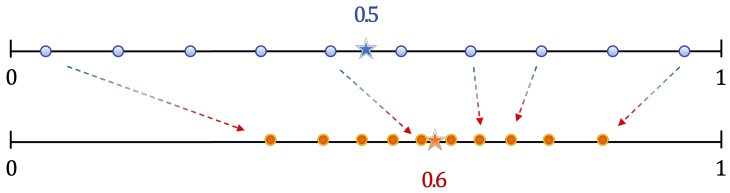



Figure 6: Comparative statics under uniform variability order.

which implies that $s_1 - \hat{s}_1 > x_1 - \hat{x}_1 > 0$. Repeating the same argument, $\{\delta_1, \delta_2, \ldots, \delta_{2k^*+1}\}$ is positive and strictly increasing. If $x_1 = \hat{x}_1$, then $s_1 \geq \hat{s}_1$. If the inequality is strict, we can resort to the same argument as above and show that $\{\delta_2, \ldots, \delta_{2k^*+1}\}$ is positive and strictly increasing. Otherwise, all inequalities hold as equalities while we proceed inductively, and the subsequence $\{\delta_2, \ldots, \delta_{2k^*+1}\}$ is identically zero.

Similar reasoning applies to the case where $j$ is even. Thus, we establish that $\{\delta_i\}_{i=1}^{2k^*+1}$ is single-crossing from below.

For the subsequence $\{\delta_i\}_{i=2k^*+1}^{2N-1}$, the single-crossing property is equivalent to the single-crossing property of the following:

$$\{\hat{x}_N - x_N, \hat{s}_{N-1} - s_{N-1}, \ldots, \hat{s}_{k^*+1} - s_{k^*+1}, \hat{x}_{k^*+1} - x_{k^*+1}\}.$$

Note that $f(\cdot)/\hat{f}(\cdot)$ is increasing on $[p, 1]$. We apply a symmetric argument to prove the sequence listed above is single-crossing from below. Combining the single-crossing property of $\{\delta_i\}_{i=1}^{2k^*+1}$ and $\{\delta_i\}_{i=2k^*+1}^{2N-1}$ establishes the desired result. □

An implication of Proposition 5 is that $\{\hat{s}_k\}_{k=1}^{N-1}$ crosses $\{s_k\}_{k=1}^{N-1}$ at most once and from above. If they indeed cross, there exists an $n^*$ such that all cutoffs to the left of $s_{n^*}$ shift up and all cutoffs to the right of $s_{n^*}$ shift down. See Figure 6 for an illustration. This figure shows that a uniformly less variable density will cause the sequence of optimal partitional cutoffs to be more compressed. Moreover, because $s_1$ increases and $s_{N-1}$ decreases, the widths of the intervals other than the outermost ones must become narrower on average. In this sense, the designer is giving more scrutiny to the middle part of the state distribution when either $f$ or $u''$ becomes uniformly less variable. Intuitively, a less variable $f$ or $u''$ means either the likelihood or the relative information value is now more compressed toward the mode. Naturally the designer should respond by compressing the interval partition to be finer around the mode as well.

## 4. Coarse Mechanism Design

Our characterization of dual expectations has implications beyond persuasion models. In particular, we can adapt the minorant function approach to study mechanism design problems where the agent's payoff is linear in her private type and where the contract space is constrained to be finite. We illustrate this with a simple model of nonlinear pricing with a finite menu.[10] Our main objective is to demonstrate the transferability of our analytical framework to an analogous coarse mechanism design problem.

Consider the classic nonlinear pricing model of Mussa and Rosen (1978). A seller supplies products to a continuum of buyers with unit demand. Each buyer has private information about her preference type $\theta \in [0,1]$, distributed according to $F$ with a continuous density function $f$ with full support. The seller designs a menu that specifies a set of quality and price pairs $\{(q_i, p_i)\}_{i\in\Sigma}$. A buyer with type $\theta$ who chooses a specific pair $(q,p)$ from the menu receives the payoff $\theta v(q) - p$. Seller's production cost of supplying quality $q \in [0,\bar{q}]$ is $c(q)$. For simplicity, we assume $v$ and $c$ are continuous on $[0,\bar{q}]$ and twice continuously differentiable on $(0,\bar{q})$ with $v' > 0$, $v'' < 0$, $c' > 0$, $c'' \geq 0$ on $(0,\bar{q})$, and $v(0) = c(0) = 0$. Additionally, we impose the constraint that the seller can only offer menus with a finite number of options, i.e., $|\Sigma| \leq N$. Throughout our discussion, we focus on the profit-maximizing menu for the seller; the analysis can be readily adapted to study welfare maximization by a planner.

By the revelation principle, we consider direct mechanisms, $\{q(\theta), p(\theta)\}$, where $q(\cdot)$ and $p(\cdot)$ are step functions because the menu is constrained to be finite. In fact, we can further restrict our attention to (weakly) increasing allocation rule $q(\theta)$, as any such rule is implementable by adjusting the transfers $p(\theta)$. For each incentive compatible allocation $q(\theta)$, the seller's profit takes the well-known formula,

$$\int_0^1 \left[ v\left(q\left(\theta\right)\right)\left(\theta - \frac{1-F\left(\theta\right)}{f\left(\theta\right)}\right) - c\left(q(\theta)\right)\right] \mathrm{d}F\left(\theta\right). \tag{4}$$

Define $\varphi := \theta - (1-F(\theta))/f(\theta)$ to be the virtual valuation. We assume that virtual valuation is strictly increasing in type. Since the unconstrained optimal allocation is to

[10] Wong (2014) uses the first-order approach to derive results for this problem. He mainly focuses on the marginal benefit from increasing the size of the menu, and discusses the approximation property when $N$ grows large. Bergemann et al. (2021) study a linear-quadratic version of this model using the quantization approach. Bergemann et al. (2026) show the optimality of endogenously coarse menu when the seller can jointly design the information structure and selling mechanism.

provide $q = 0$ to all types with non-positive virtual valuations, we focus on the allocation for those types with positive virtual valuations, and use $J$ to represent the truncated distribution of virtual valuations on $[0,1]$ (suitably normalized so that $J(0) = 0$ and $J(1) = 1$). We maintain the assumption that $J$ admits a continuous density function on $[0,1]$ with full support.

With slight abuse of notation, let $q(\varphi)$ represent the quality allocation to virtual type $\varphi$ in a finite menu, and define $\underline{\pi}(\varphi) := \varphi v(q(\varphi)) - c(q(\varphi))$ correspondingly. Formally, the coarse mechanism design problem is the following:[11]

$$\begin{aligned} \max_{q(\cdot)} \quad & \int_0^1 \underline{\pi}(\varphi)\,\mathrm{d}J(\varphi) && \text{(CMD)} \\ \text{s.t.} \quad & q(\cdot) \text{ is nondecreasing,} && \text{(Monotonicity)} \\ & q([0,1]) \text{ has at most } N+1 \text{ elements,} && \text{(Finite Menu)} \\ & q(0) = 0. && \text{(IR)} \end{aligned}$$

The program is essentially to pick a monotone, finite allocation rule $q(\varphi)$ to maximize the expected virtual surplus. Exclusion of consumers is captured by $q(\varphi) = 0$. Therefore an $N$-option menu is equivalent to an $N+1$-valued allocation rule. Since the allocation rule has to be monotone, the optimal menu partitions the (virtual) type space into $N+1$ subintervals with a sequence of cutoff types, $0 < s_1 < \ldots < s_{N+1} = 1$. Within each type segment $(s_{k-1}, s_k)$ where $k = 2, \ldots, N+1$, the allocation assigns $q_k$ uniformly. Consumers with virtual types below $s_1$ are excluded and are assigned $q_1 = 0$. To leverage the dual expectations property in this coarse mechanism design problem, let

$$\pi(\varphi) := \max_{q \in [0,\bar{q}]} \varphi v(q) - c(q) \tag{5}$$

be the pointwise maximal virtual surplus on $\varphi \in [0,1]$ and $q^*(\varphi)$ be the corresponding optimal quality. We assume that $q^*(\varphi) \in (0,\bar{q})$ for every $\varphi \in (0,1)$. Then by our assumptions on $v$ and $c$, the maximizer is unique and for $\varphi \in (0,1)$, is characterized by the first-order condition $\varphi v'(q^*(\varphi)) = c'(q^*(\varphi))$. Consequently, $\pi$ is continuously differentiable on $[0,1]$ and twice-continuously differentiable on $(0,1)$. Because the pointwise profit is linear in the consumer's virtual type, $\pi(\varphi)$ is also strictly increasing and strictly convex on $[0,1]$ with $\pi'' > 0$ on $(0,1)$. Moreover, for any $0 \le a < b \le 1$,

[11] In the program, we exclude negative virtual types as it is optimal to assign $q = 0$ to them.

we define[12]

$$\phi^m(a,b) := \mathbb{E}_J[t \mid t \in (a,b)],$$
$$\mu^m(a,b) := \mathbb{E}_{\pi'}[t \mid t \in (a,b)].$$

**Proposition 6.** *The solution to program (CMD) satisfies*

$$q(\varphi) = \begin{cases} 0 & \text{if } \varphi \le s_1, \\ q^*(x_k) & \text{if } \varphi \in (s_{k-1}, s_k], \quad \text{for } k = 2, \dots, N+1; \end{cases}$$

*where* $\{s_k\}_{k=1}^{N+1}$ *and* $\{x_k\}_{k=1}^{N+1}$ *satisfy:*

$$x_k = \phi^m(s_{k-1}, s_k) \qquad \text{for } k = 2, \dots, N+1, \tag{CE-$J$}$$
$$s_k = \mu^m(x_k, x_{k+1}) \qquad \text{for } k = 1, \dots, N; \tag{CE-$\pi'$}$$

*with* $x_1 = 0$ *and* $s_{N+1} = 1$.

*Proof.* By definition, $\underline{\pi}(\varphi)$ is weakly below $\pi(\varphi)$ for any $\varphi \in [0,1]$. Our goal is to show that the optimal $q(\varphi)$ induces a virtual surplus function $\underline{\pi}(\varphi)$ that shares the same structure as the minorant function $\underline{u}$ identified in the corresponding coarse information design problem, except for the first segment $(0, s_1)$ where there is no trade. Any incentive-compatible $N+1$-quality allocation rule $q(\varphi)$ would induce some $\underline{\pi}$ which is piecewise affine with $N+1$ pieces, including a horizontal piece in the first segment. We show that, under the optimal $q(\varphi)$, the corresponding $\underline{\pi}$ satisfies the following two properties: (a) it is continuous at $s_k$ for all $k = 1, \dots, N$ and thereby is also convex; and (b) on each type segment $(s_{k-1}, s_k)$ with $k = 2, \dots, N+1$, it is tangent to $\pi$ at the conditional expectation $x_k$ of the corresponding segment.

For part (a), suppose the surplus function $\underline{\pi}$ induced by an allocation $q(\cdot)$ is discontinuous at some cutoff point $s_k$, as depicted in the left panel of Figure 7. Then we can construct another incentive-compatible allocation $\hat{q}(\cdot)$, which supplies higher quality to consumers in $(\hat{s}_k, s_k)$ while the allocations to other consumers are unchanged. The corresponding $\underline{\hat{\pi}}$ function (shown by the blue dotted line for $\varphi \in (\hat{s}_k, s_k)$ in Figure 7) for this modified allocation is everywhere higher than $\underline{\pi}$ for the original allocation. This

[12] Again by a scaling argument and by the convexity of $\pi$, it is without loss of generality to normalize its derivative $\pi'$ to be a cumulative distribution function.

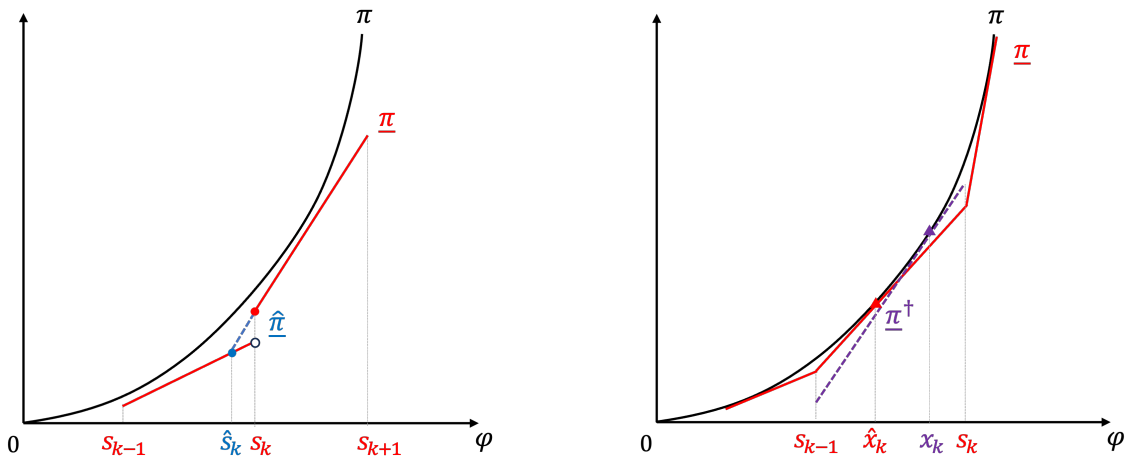



Figure 7: Properties of optimal $\underline{\pi}$: it is continuous (left panel) and is tangent to $\pi$ at $x_k = \phi^m(s_{k-1}, s_k)$ (right panel).

would yield a strictly higher profit, hence a contradiction. The convexity of the optimal $\underline{\pi}$ then follows directly from continuity and from the fact that the slope of $\underline{\pi}$ is equal to $v(q(\varphi))$, where $q(\cdot)$ must be nondecreasing for incentive compatibility.

For part (b), first observe that each piece of $\underline{\pi}$ must be tangent to $\pi$ at some interior point, for otherwise the seller can adjust the allocation to raise $\underline{\pi}$ and increase profit. It remains to be proven that the optimal allocation for the segment $(s_{k-1}, s_k)$ for $k = 2, ..., N+1$ must equal the unconstrained optimal allocation for the average virtual type of that segment, i.e., $q_k = q^*(x_k)$ for $k = 2, \ldots, N+1$. Suppose not, and instead $\underline{\pi}$ is tangent to $\pi$ at some point $\hat{x}_k \neq x_k$. Then $\underline{\pi}(\varphi) = \pi(\hat{x}_k) + \pi'(\hat{x}_k)(\varphi - \hat{x}_k)$ (shown as the red line in the right panel of Figure 7). We can construct another allocation $q^\dagger(\cdot)$ such that $q_k^\dagger = q^*(x_k)$ with the corresponding surplus function $\underline{\pi}^\dagger$ tangent to $\pi$ at $x_k$ (shown as the purple dotted piece), and is otherwise equal to the original $q$. Then,

$$\begin{aligned}\int_{s_{k-1}}^{s_k} \underline{\pi}(\varphi)\,\mathrm{d}J(\varphi) &= [J(s_k) - J(s_{k-1})]\left[\pi(\hat{x}_k) + \pi'(\hat{x}_k)(x_k - \hat{x}_k)\right] \\ &< [J(s_k) - J(s_{k-1})]\,\pi(x_k) = \int_{s_{k-1}}^{s_k} \underline{\pi}^\dagger\,\mathrm{d}J(\varphi),\end{aligned}$$

where the inequality follows from the strict convexity of $\pi$. This contradicts the optimality of the allocation corresponding to the original $\underline{\pi}$ function. Therefore $q_k = q^*(x_k)$ must hold for $k = 2, \ldots, N+1$. Furthermore, the optimal quality allocation $\underline{\pi}(\varphi) = 0$

for $\varphi \in [0, s_1]$ follows from (IR) and the continuity of $\underline{\pi}$ at $s_1$. We can then let $x_1 = 0$.

Consequently, the virtual surplus function $\underline{\pi}$ under the optimal quality allocation must be a convex and piecewise affine function. We can again use Lemma 1 to conclude that (CE-$\pi'$) must hold. □

Our analysis implies that the optimal $\underline{\pi}$ exhibits the same property as the minorant function that we introduce before. For $k = 2, \ldots, N+1$, the optimal menu allocates goods with quality $q_k = q^*(x_k)$ to consumers with virtual types belonging to the segment $(s_{k-1}, s_k]$, where $x_k$ is the expected type of this segment. Consumers with virtual types below $s_1$ are excluded and receive the outside option. Moreover, for $k = 1, \ldots, N$, since the interval cutoff $s_k$ is a kink-point of $\underline{\pi}$, we have $s_k$ as the expected value of the interval $(x_k, x_{k+1})$ under the measure $\pi'$. Hence, the dual expectations characterization is still valid for the finite-menu nonlinear pricing problem.

It is worth noting that the coarse mechanism design problems and the coarse information design problems are not isomorphic to each other. The equation system (CE-$F$) and (CE-$u'$) and the equation system (CE-$J$) and (CE-$\pi'$) have different initial (boundary) conditions: in coarse information design, the boundary conditions are $s_0 = 0$ and $s_{N+1} = 1$ if there are $N+1$ signals, whereas in coarse mechanism design, the boundary conditions are $x_1 = 0$ and $s_{N+1} = 1$. Nonetheless, the general properties of the equation system (CE-$F$) and (CE-$u'$) only rely on the dual expectations system, not the boundary conditions. Therefore all results in Section 3 continue to hold in the equation system (CE-$J$) and (CE-$\pi'$), with minor modification of the relevant sequence when needed.[13]

**Example 3.** Suppose $v(q) = q^\beta$ and $c(q) = \gamma q$ for $\beta \in (\frac{1}{2}, 1)$. The unconstrained optimal allocation for the seller is $q^*(\varphi) = (\beta\varphi/\gamma)^{1/(1-\beta)}$, and the "value function" is $\pi(\varphi) = (1-\beta)(\gamma/\beta)(\beta\varphi/\gamma)^{1/(1-\beta)}$. Let the value of the good change to $\hat{v}(q) = q^{\hat{\beta}}$, with $\hat{\beta} > \beta$. Note that $\hat{v}(q) > v(q)$ for all $q > 1$.[14] The value function $\hat{\pi}(\varphi)$ corresponding to the higher valuation of quality satisfies the property that $\hat{\pi}''/\pi''$ is increasing. If the density of virtual valuation is logconcave on $[0, 1]$, this setting is a logconcave environment. By the same logic leading to Proposition 4, all the interval cutoff points will move to the right when the benefit increases from $v(q)$ to $\hat{v}(q)$. In particular, we

[13] For Theorem 2 (single-dipped property), the relevant sequence in the statement becomes $\{d_1, w_2, \ldots, d_N, w_{N+1}\}$. For Proposition 3, we redefine $w_{i^*} = \min\{w_2, \ldots, w_{N+1}\}$. For Proposition 5 (comparative statics with respect to uniform variability order), the relevant sequence becomes $\{\hat{s}_1 - s_1, \hat{x}_2 - x_2, \ldots, \hat{s}_N - s_N, \hat{x}_{N+1} - x_{N+1}\}$.

[14] The seller's optimal menu will offer $q_k > 1$ for all $k > 1$ when the marginal cost $\gamma$ is sufficiently low.

have $\hat{s}_1 > s_1$. This means that the seller's optimal menu will exclude a larger set of consumers, despite the fact that the high-quality good has become more valuable to all consumers. Intuitively, an increase in $\beta$ not only implies higher valuation for quality, but also leads to greater importance to screen the high virtual value buyers. When the screening device is constrained, the optimal finite menu naturally involves more discrimination towards high-end customers, leading to the exclusion of more low-end customers. This is a novel feature for finite menu design, as the coverage of customers in an unconstrained problem only depends on the distribution of virtual value and will not change when $v$ changes.

## 5. S-shaped Value Functions

The analysis so far is based on the premise that the value function $u$ is convex, under which the designer wishes to transmit as much information as possible. In this section, we extend our analysis to S-shaped value functions, where the designer faces the tension between providing information on the convex region and suppressing it on the concave region.

Formally, a function $u$ is *S-shaped* if there exists some inflection point $t \in (0,1)$ such that $u$ is strictly convex on $(0,t)$ and strictly concave on $(t,1)$. In other words, $u''$ is single-crossing from above.[15] S-shaped value functions have been extensively studied in persuasion literature because they capture a range of economic applications. The unconstrained optimal information structure $G^*$ features "upper-censorship," namely the optimal experiment without the finiteness constraint will reveal full information on $[0,s^*)$ and coarsen information by pooling all states in $(s^*,1]$ for some $s^* < t$ (Kolotilin et al., 2022). In this section, we rule out the uninteresting case where the unconstrained solution reveals no information.

Under S-shaped value functions, the dual expectations characterization, the single-dipped property, as well as the comparative statics results with respect to the change of value function, remain true when the relevant conditions are properly modified. In particular, when the value function $u$ is not convex, $u''$ can take negative values and may not be a valid density. Nevertheless, it can be interpreted as the "density" corresponding to a *signed measure*. For any $0 \leq a < b \leq 1$, define its *barycenter* under signed measure

[15] With obvious modification, the analysis in this section also extends to the case where $u''$ is single-crossing from below.

$u'$ by

$$\mu(a,b) := \frac{\int_a^b \theta u''(\theta)\,\mathrm{d}\theta}{\int_a^b u''(\theta)\,\mathrm{d}\theta}, \tag{6}$$

whenever $\int_a^b u''(\theta)\,\mathrm{d}\theta \neq 0$. We can therefore interpret $\mu(a,b)$ in equation (CE-$u'$) as the barycenter of the interval $(a,b)$ under signed measure $u'$.

**Proposition 7.** *Suppose $u$ is S-shaped. The optimal information structure $G$ is an interval-partitional structure with $|\mathscr{K}_{I_G}| = N$. It satisfies the equation system (CE-$F$) and (CE-$u'$), which is well-defined at the optimum. Moreover, the optimal information structure $G$ is less informative than the unconstrained one $G^*$; in particular, $s_{N-1} < s^*$.*

*Sketch of proof.* The proof of the optimality of interval-partitional structure follows a similar logic as the proof of Proposition 2: raising $I_G$ increases the value of the objective function (2) whenever $u'' > 0$. Suppose $I_G$ corresponds to the optimal information structure and there is a segment of $I_G$ that is not tangent to $I_F$ (so that $G$ is not interval-partitional). If $u''$ is not entirely positive on this segment (otherwise, the argument is similar), then we consider an alternative information structure $I_{G'}$ which is a clockwise rotation of $I_G$, with $I_{G'} > I_G$ to the left of the inflection point $t$ on the same segment and $I_{G'} < I_G$ to the right of the inflection point on the same segment. This alternative information structure will increase informativeness locally in the convex region and reduce informativeness in the concave region and raise the value of (2). The proof for the remaining statements are provided in the online appendix. □

In Section 6, we make an observation in Proposition 9 and 10(ii) that, for any value functions, whenever the optimal coarse information is interval-partitional, the minorant function is (continuous and) convex. Continuity implies that an optimal interval partition for S-shaped value functions must satisfy (CE-$F$) and (CE-$u'$), and convexity implies that (CE-$u'$) is well-defined at the optimum. To see that $s_{N-1} < s^*$, in an information design problem without finiteness constraint, the optimal censorship threshold $s^*$ lies in the convex region, and it solves the designer's local trade-off between providing full information around the marginal state $s^*$ in the convex region and pooling it into the concave region. With the finiteness constraint, suppose the interval cutoff of the last interval is $s_{N-1} = s^*$, then the marginal gain from information provision around the same state diminishes since now $s^*$ cannot be perfectly revealed, but is instead pooled with the lower interval $(s_{N-2}, s^*)$. As a result, equalization at the margin requires the

designer optimally pools more of the convex region with the concave region to diminish the marginal gains from censorship at the upper end. The optimal cutoff point under $G$ satisfies $s_{N-1} < s^*$. This implies that $G$ is less Blackwell-informative than $G^*$.

The main challenge in generalizing the remaining results from the case of convex value functions arises because the key logconcavity property (Lemma 2) no longer applies when the curvature $u''$ is single-crossing. Nevertheless, a one-sided version of Lemma 2 still holds.

**Lemma 3.** *Suppose $u''$ is logconcave on $(0,t)$. Then for any $a < a+\varepsilon < t < b$ such that $u'(b) - u'(a+\varepsilon) > 0$, we have*

$$\mu(a+\varepsilon, b) < \mu(a,b) + \varepsilon.$$

This weaker condition, together with Lemma 2, turns out to be sufficient for most of the remaining results because, under the optimal interval-partitional information structure, all interval cutoffs $\{s_k\}_{k=1}^{N-1}$ and the posterior means $\{x_k\}_{k=1}^{N-1}$ remain within the convex region of the value function. The only complication is that the last posterior mean $x_N$ may lie in either the convex region or in the concave region, and therefore requires a case-by-case verification. To maintain focus of the main text, we put the proof of Lemma 3 and the proof of the following proposition in an online appendix.

**Proposition 8.** *Suppose $u$ is S-shaped, $f$ is logconcave, and $u''$ is logconcave on $(0,t)$.*

*(i)* *$\{\Delta_i\}_{i=1}^{2N-1}$ is single-dipped.*

*(ii)* *The optimal interval-partitional information structure is unique.*

*(iii)* *If the value function changes from $u$ to $\hat{u}$, with $\hat{u}'' \succeq_{lr} u''$ and $\hat{u}''$ is single-crossing at $\hat{t}$ and logconcave on $(0,\hat{t})$, then all interval cutoffs and all induced posterior means will increase.*

*(iv)* *If the value function changes from $u$ to $\hat{u}$, with $u'' \succeq_{uv} \hat{u}''$ where $\hat{u}''$ is single-crossing at $\hat{t}$, then the new sequence of optimal posterior means and interval cutoffs, $\{\hat{x}_1, \hat{s}_1, \ldots, \hat{s}_{N-1}, \hat{x}_N\}$, crosses the original sequence $\{x_1, s_1, \ldots, s_{N-1}, x_N\}$ at most once and from above.*

Part (i) of the proposition extends the single-dipped property to S-shaped value functions. Intuitively, the single-dipped property under convex value functions means

that the information designer puts less scrutiny on the two ends of the state space. When the value function is S-shaped, the last subinterval $(s_{N-1}, 1)$ covers the concave region and therefore should deserve even less scrutiny. Hence, the single-dipped property preserves.

We prove part (ii) of Proposition 8 by showing that there is only one solution to the equation system (CE-$F$) and (CE-$u'$) which satisfies the requirement that all cutoffs lie in the convex region. There may exist other solutions to the equation system in which some cutoffs lie in the concave region, but such solutions will not be optimal because Proposition 7 implies that $s_{N-1} < s^* < t$ in an optimal information structure.

Parts (iii) and (iv) mirror the comparative statics results in Propositions 4(b) and 5(b), respectively. Note that we allow $u''$ and $\hat{u}''$ to have different inflection points, $t$ and $\hat{t}$. In Proposition 8(iii), $\hat{u}'' \succeq_{lr} u''$ implies $\hat{t} \geq t$, meaning that $\hat{u}$ has a smaller concave region than $u$. In the extreme case, $\hat{u}$ can be convex (represented by some $\hat{t} > 1$), and the statement remains valid. Also note that the proof of Proposition 8(iii) does not use the fact that $u''$ is logconcave on $(0, t)$. In other words, for Proposition 8(iii) to apply, we only require the dominating value function curvature to be logconcave while positive; the corresponding value function curvature which is likelihood-ratio dominated need not be logconcave while positive. In Proposition 8(iv), $u'' \succeq_{uv} \hat{u}''$ implies either $t \leq \hat{t} \leq p$ or $p \leq \hat{t} \leq t$. If both inflection points are to the left of $p$, then $\hat{u}$ has a smaller concave region than $u$. Otherwise, if both inflection points are to the right of $p$, then $\hat{u}$ has a larger concave region than $u$.

**Example 1b (continued).** In the purchase decision example, when the density $h$ of cost shock is single-peaked, the seller's value function is S-shaped, with curvature $ph'$. Because the buyer's value function curvature $h$ is logconcave, and $h \succeq_{lr} ph'$, Proposition 8(iii) applies. This means that the conclusion of Example 1a remains valid when the seller's value function is S-shaped. Intuitively, a falling tail of the cost shock density $h$ implies that better information about higher states has diminishing returns, as the probability that the consumer will be on the margin between buying and not buying falls when mean quality is already high. Compared to the previous case (Example 1a), where the seller's value function is convex and the density $h$ is rising, the seller has even less incentive than does the buyer to provide finer information about higher states.

Nevertheless, Proposition 4(a) and 5(a) no longer extend to S-shaped value function. This reflects the asymmetric role played by the prior density and the curvature

of the value function once the latter becomes a signed measure. To see where the issue arises, consider a binary-signal example and a likelihood-ratio increase in the prior. Suppose that under the initial prior, the optimal upper signal $x_2$ already lies above the inflection point $t$. A likelihood-ratio increase in prior shifts the upper signal further to the right and hence farther into the concave region. Since the barycenter associated with the signed measure is decreasing in its upper endpoint $x_2$ whenever $x_2 > t$, such a change can lead to a lower cutoff $\hat{s}_1 < s_1$ in order for the dual expectation condition to continue to hold. Thus, unlike in the convex case, the change in prior need not move the signals and cutoffs in the same direction.[16]

## 6. General Value Functions

The coarse information design problem with convex or S-shaped value functions is relatively tractable because interval-partitional information structures are optimal. With general value functions, it is possible that the optimal information structure entails information policies that send more than one signal within a partitional subinterval, and the analysis would become more involved. In this section, we restrict our discussion to interval-partitional information structures, and provide a generalization of the dual expectations property. Moreover, when the interval partition is indeed optimal among all possible coarse information structures, we establish a general convexity property about the minorant function.

**Proposition 9.** *For any general value function $u$, an optimal interval-partitional information structure with $N$ intervals, characterized by $\{s_k\}_{k=0}^N$ and $\{x_k\}_{k=1}^N$, must satisfy:*

*(i)* $x_k = \phi(s_{k-1}, s_k)$, *for* $k = 1, \ldots, N$;

*(ii)* $u(x_k) + u'(x_k)(s_k - x_k) = u(x_{k+1}) + u'(x_{k+1})(s_k - x_{k+1})$, *for* $k = 1, \ldots, N-1$.

The second part of Proposition 9 establishes the continuity of the minorant function, which further generalizes the dual expectations property in the following sense. Suppose the solution of the system of equations satisfies $u'(x_{k+1}) \neq u'(x_k)$, we can restate the continuity of minorant function as $s_k = \mu(x_k, x_{k+1})$, where $\mu(\cdot)$ is given by equation (6) that defines the barycenter under the signed measure $u'$. Thus, the dual expectations property continues to hold as the barycenter is well defined. On the other hand,

[16] For example, take $u'' = \frac{1}{2} - x$. Let $f(x) = 1$ and $\hat{f}(x) = x + \frac{1}{2}$. The solutions to both systems of equations are unique and we have $\hat{x}_2 > x_2$ but $\hat{s}_1 < s_1$.

if the solution happens to be $u'(x_{k+1}) = u'(x_k)$, then the continuity of the minorant function implies that

$$u'(x_k) = \frac{u(x_{k+1}) - u(x_k)}{x_{k+1} - x_k} = u'(x_{k+1}).$$

In this case, the solution coincides with a bi-tangent line, i.e., the minorant function is affine on $[s_{k-1}, s_{k+1}]$. In either case, the continuity condition would return an equation, contributing to a system of equations of $2N-1$ equations for $2N-1$ unknowns.

For the next result, define for $x \in [x_1, x_N)$ the function,

$$\overline{u}_G(x) := \sum_{k=1}^{N-1} \left[ u(x_k) + \frac{u(x_{k+1}) - u(x_k)}{x_{k+1} - x_k}(x - x_k) \right] \mathbb{I}_{[x_k, x_{k+1})}(x),$$

where $(x_1, \ldots, x_N)$ are the induced posterior means under interval-partitional information structure $G$. The graph of $\overline{u}_G$ is piecewise affine and connects the adjacent posterior means induced by $G$. In the case that $u$ is convex, it is obvious that both $\overline{u}_G$ and $\underline{u}_G$ are convex, and on $[x_1, x_N)$ they satisfy:

$$\overline{u}_G(x) \geq u(x) \geq \underline{u}_G(x).$$

For general value functions $u$, the minorant function $\underline{u}_G$ need not be everywhere below $u$. Nevertheless, the following properties hold.

**Proposition 10.** *For any general value function u, suppose an interval-partitional information structure G with N subintervals is optimal among all coarse information structures. Then*

*(i)* $\overline{u}_G(x) \geq u(x)$ *for all* $x \in [x_1, x_N)$*, and* $\overline{u}_G$ *is convex.*

*(ii) The minorant function* $\underline{u}_G$ *is convex. Moreover, if the value function is convex or S-shaped, the minorant function has strictly increasing slopes for each affine segment, i.e.,* $u'(x_{k+1}) > u'(x_k)$ *for* $k = 1, \ldots, N-1$.

Proposition 10(ii) indicates a general convexity property about the minorant function provided that the interval partition is indeed optimal among all possible coarse information structures. It is in contrast to Kolotilin et al. (2025), where the optimal interval disclosure under M-shaped (concave-convex-concave) value functions can induce

a concave minorant function. The reason is that they restrict attention to monotone deterministic information structures.

On the other hand, extending the properties and comparative statics results to general value functions could be difficult, since the equation system under general value functions can be non-monotone. Consequently, it requires extensive case-by-case discussion on whether each signal falls into convex or concave regions, in contrast to the S-shaped case, where only the last signal could fall into the concave region. Moreover, the logconcavity property identified in Lemma 2 and Lemma 3, which play an essential role in previous cases, cannot be easily adapted to general value functions.

## 7. Discussion

### 7.1. Convex piecewise affine value functions

When the set of actions is finite, the value function is piecewise affine. In this section, we study the case where the value function is convex and piecewise affine. The main challenge here is that the minorant function corresponding to an interval partition now may not be unique because the subgradient at any kink of the value function is not unique. Nevertheless, we show that this concern is not an issue. An important implication is that the optimal interval partition will not generate a posterior mean that keeps the information designer indifferent between two adjacent actions (where a kink arises).

Suppose the value function $u$ is convex and has $M$ kinks. Then it can be represented by

$$u(x) = \sum_{m=0}^{M} \big[u(\upsilon_m) + u'(\upsilon_m)(x - \upsilon_m)\big] \mathbb{I}_{[\upsilon_m, \upsilon_{m+1}]}(x),$$

where $\{\upsilon_m\}_{m=0}^{M+1}$ includes all the kink points, and the end points $\upsilon_0 = 0$, $\upsilon_{M+1} = 1$. Throughout this section, we use $u'(\cdot)$ to represent the right derivative whenever the left derivative is not equal to the right. As before, we invoke the normalization that $u'(\upsilon_0) = 0$ and $u'(\upsilon_M) = 1$. To avoid triviality, we assume that the number of signals $N$ satisfies $M \geq N \geq 2$. The next result shows that the posterior means of the optimal interval partition cannot fall into the set of kink points of $u$.

**Lemma 4.** *Under any optimal interval partition $G^*$ for convex piecewise affine value functions, $\{x_i\}_{i=1}^{N} \cap \{\upsilon_m\}_{m=1}^{M} = \emptyset$.*

To understand the intuition, consider the case where the information designer is also the decision maker. He seeks to maximize the value of information. However, the value of information is minimized when the decision maker is indifferent between actions—which occurs at the kink points of $u$. Therefore, if a subinterval yields a posterior mean that lies at a kink, the designer can always locally adjust the interval cutoffs to produce a different posterior mean, thereby obtaining a strictly higher information value.

With Lemma 4, we can uniquely pin down the minorant function in the same way as it is defined in equation (3), where $u'(x_i)$ is just the derivative, since the left derivative coincides with the right derivative. Then we can apply exactly the same argument as in the smooth case to argue that, under the optimal interval partition, $\underline{u}_{G^*}(x)$ must be continuous. Hence every $\{s_i\}_{i=1}^{N-1}$ must be a meeting point of two affine segments of the value function. Let $\mathscr{C}$ be the set of points where any two affine pieces meet with each other:

$$\mathscr{C} = \left\{x \in [0,1] : u(v_i) + u'(v_i)(x - v_i) = u(v_j) + u'(v_j)(x - v_j) \text{ for } 0 \leq i < j \leq M\right\}.$$

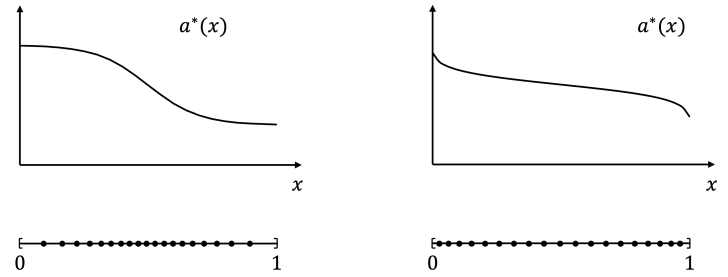
**Corollary 1.** *Under any optimal information structure $G^*$ for convex piecewise affine value functions, $\{s_i\}_{i=1}^{N-1} \subseteq \mathscr{C}$.*

Under any optimal information structure $G^*$, all interval cutoffs $\{s_i\}_{i=1}^{N-1}$ are selected from the set $\mathscr{C}$. Hence, the number of information structures that we can search over to find the optimal one is finite.

### 7.2. Bimodal curvature and double-dipped interval widths

A logconcave environment ensures both $f$ and $u''$ are single-peaked, which in turn implies that the interval widths in the optimal information structure are single-dipped. Whether the value function curvature $u''$ is single-peaked or not depends on the nature of the underlying decision problem. However, one of the basic insights of this paper—namely, the optimal information structure pays more scrutiny to the part of the state space where value function curvature is relatively high—does not depend on the logconcavity assumption.

To illustrate, suppose a decision maker chooses to consume $a$ units of a good to maximize her expected utility $\mathrm{E}[B(a) - \theta a]$, where $B(a)$ is the intrinsic benefit from consumption, and $\theta$ represents per-unit cost. Her indirect value function will have curvature $u''(x) = -\mathrm{d}a^*(x)/\mathrm{d}x$, where $a^*(x)$ is the "demand" for this good when the ex-




(a) demand function for logconcave $u''$ (b) demand function for bimodal $u''$

Figure 8: The demand function in panel (a) is $a^*(x) = \frac{1}{1+e^{10x-5}} + \frac{1}{2}$; it is most cost-sensitive near the middle part of the state distribution. The demand function in panel (b) is $a^*(x) = \log \frac{0.99-0.98x}{0.01+0.98x} + 5$; it is most cost-sensitive near the two ends of the state distribution. The line segments at the bottom of these two panels show the corresponding optimal partitioning of the state space.

pected cost is $x$. The shape of the demand function $a^*(\cdot)$ generally depends on $B'^{-1}(\cdot)$, the inverse of the marginal benefit function. Panel (a) of Figure 8 shows a demand function that takes a logit form. In this panel, demand is most sensitive to cost at intermediate levels of $x$. If the density $f$ of the state is uniform, then the center of scrutiny is near 0.5. Panel (b) shows a different possibility, where demand is more cost-sensitive for values of $x$ near 0 and 1. The value function curvature $u''$ corresponding to this demand pattern is bimodal. The corresponding optimal information structure exhibits interval widths which are double-dipped: the decision maker pays closest scrutiny to states near 0 and 1 because her action responds more to the state at the two ends of the state distribution. Finally, if the demand function is linear, then $u''$ will be constant. With uniform prior density $f$, linear demand will lead to an optimal information structure in which all the intervals are evenly spaced.

### 7.3. Connection with cheap talk

Our model of coarse information design bears some resemblance to the standard cheap talk model of Crawford and Sobel (1982), where information coarsening arises endogenously due to incentive compatibility constraints. In Crawford and Sobel (1982), to verify an interval-partitional information structure is an equilibrium strategy, we only need to verify that the cutoff types are indifferent between two adjacent actions taken by the receiver under two adjacent posterior means. For example, suppose that the sender's payoff is $-(a-\kappa_1(\theta-\kappa_0))^2$ and the receiver's payoff is $-(a-\theta)^2$, where $a$

is the action taken by the receiver, and $\kappa_1 \geq 1$, $\kappa_0 \in [0,1]$. Then, the cheap-talk equilibrium, given a fixed number of signals $N$, is determined by the following system of equations:

$$\begin{aligned} x_k &= \phi(s_{k-1}, s_k) \qquad \text{for } k = 1, \ldots, N; \\ s_k &= \frac{x_k + x_{k+1}}{2\kappa_1} + \kappa_0 \qquad \text{for } k = 1, \ldots, N-1. \end{aligned}$$

The second set of equations are the indifference conditions for the cutoff types. They are generally different from the optimality conditions (CE-$u'$) for the coarse information design problem.[17] Nevertheless this system of equations is still a monotone mapping satisfying Lemma 2 if the prior density $f$ is logconcave. This means that some of our results should align with the equilibrium characterization in this cheap talk model, such as the existence of a center of scrutiny. The insight driving this result in the cheap talk model, however, differs from ours. Specifically, the sender's bias is smaller when $\theta$ is close to $\kappa_0$ than when it is farther away (Gordon, 2010; Deimen and Szalay, 2019, 2023). Hence, credibility is easier to obtain for states close to $\kappa_0$ and the intervals are narrower in that part of the state space. In contrast, in our coarse information design problem, the center of scrutiny is mainly determined by the value function curvature and the prior density of the state. With quadratic payoffs, the value function curvature $u''(x)$ of the sender is a constant that does not depend on $x$. Hence, the values of the parameters $\kappa_1$ and $\kappa_0$ have no effect on the optimal information structure. Indeed, in this example, the sender-optimal coarse experiment with $N$ signals is identical to the receiver-optimal coarse experiment with the same number of signals, because the value function curvature of the receiver is also independent of $x$.

Recent developments in cheap-talk games have explored the optimal information design by the sender. This literature assumes that an uninformed sender commits to an experiment, and sends an unverifiable message about the experimental outcome. When the loss function is quadratic with $\kappa_1 = 1$, the sender's problem can be transformed into a problem similar to ours, with an additional set of truth-telling constraints that require the distance between every two adjacent messages to be at least $2\kappa_0$, and with

[17] When the sender and receiver have common interests, i.e., $\kappa_1 = 1$ and $\kappa_0 = 0$, the cheap talk equilibrium for a fixed number of messages coincides with our coarse information design problem with a quadratic value function. Jäger et al. (2011) study a multidimensional cheap talk problem with common interests in depth.

the number of messages $N$ being endogenously determined through this constraint. As elaborated in Kreutzkamp and Lou (2025), this additional constraint makes bi-pooling a possible candidate solution despite the quadratic value function. In our model, $N$ is exogenously fixed and there are no truth-telling constraints. Hence, even with the same payoff structures, the optimal experiment in our setup has an interval-partitional structure.

## 8. Conclusion

The literature on Bayesian persuasion often assumes no restrictions on the set of experiments that can be chosen, making the information design problem trivial when the value function is convex. One approach that has been taken to relax this assumption is to introduce a posterior-based cost of acquiring information that depends on the experiment chosen (Caplin and Dean, 2015; Matějka, 2015; Bloedel and Segal, 2021; Ravid et al., 2022; Matysková and Montes, 2023); see Denti (2022) and Denti et al. (2022) for a critical assessment of that approach. In this paper, we adopt an alternative track by imposing a finiteness constraint on the signal space to reflect the coarseness of information structures. We take information coarsening as given and consider the best way of doing it. This approach naturally leads to a research agenda that investigates which parts of the state space deserve most attention or scrutiny in information acquisition decisions. Our paper only takes a first stab at this research agenda by considering a simple environment with one-dimensional state where the belief about the state matters only through the mean. Many questions regarding sequential information acquisition, competitive and complementary constrained information provision, or coarse information design under more varied payoff or informational environments remain to be explored.

## Appendix

**Proof of Lemma 2.** Take any $\varepsilon \in [0, 1-b]$. We can write

$$\phi(a+\varepsilon, b+\varepsilon) = \int_{a+\varepsilon}^{b+\varepsilon} x \frac{f(x)}{F(b+\varepsilon)-F(a+\varepsilon)}\,\mathrm{d}x = \int_{a}^{b} (x'+\varepsilon)\frac{f(x'+\varepsilon)}{F(b+\varepsilon)-F(a+\varepsilon)}\,\mathrm{d}x',$$
$$\phi(a,b)+\varepsilon = \int_{a}^{b} (x'+\varepsilon)\frac{f(x')}{F(b)-F(a)}\,\mathrm{d}x'.$$

The difference is

$$\int_{a}^{b} x\left(\frac{f(x+\varepsilon)}{F(b+\varepsilon)-F(a+\varepsilon)} - \frac{f(x)}{F(b)-F(a)}\right)\mathrm{d}x.$$

Due to the logconcavity of $f$, the first density is dominated by the second in the likelihood ratio order. Hence, the difference is nonpositive: $\phi(a+\varepsilon, b+\varepsilon) \le \phi(a,b)+\varepsilon$.

To see the inequality when $\varepsilon_1 \neq \varepsilon_2$, consider two cases. Case 1: $b+\varepsilon_1 \le 1$. Since $\phi(\cdot,\cdot)$ is strictly increasing in each argument,

$$\phi(a+\varepsilon_1, b+\varepsilon_2) \le \phi(a+\max\{\varepsilon_1,\varepsilon_2\}, b+\max\{\varepsilon_1,\varepsilon_2\}) \le \phi(a,b)+\max\{\varepsilon_1,\varepsilon_2\},$$

with strict inequality if $\varepsilon_1 \neq \varepsilon_2$.

Case 2: $b+\varepsilon_1 > 1$. In particular, it implies $\varepsilon_1 > \varepsilon_2$ and $a < 1-\varepsilon_1 < b$. Then

$$\phi(a+\varepsilon_1, b+\varepsilon_2) \le \phi(a+\varepsilon_1, 1) = \phi(a+\varepsilon_1, 1-\varepsilon_1+\varepsilon_1) \le \phi(a, 1-\varepsilon_1)+\varepsilon_1 < \phi(a,b)+\varepsilon_1.$$

□

**Proof of Proposition 2.** The objective function (2) is the $u''$-weighted area below $I_G$. Since $u''$ is nonnegative, an information structure $G'$ performs better than another information structure $G$ if $I_{G'} \ge I_G$. Consider an arbitrary induced integral distribution $I_G$ with a segment strictly below $I_F$ that does not contain a tangency point. We can construct an alternative integral distribution $I_{G'}$ by pointwise increasing this segment until it crosses the adjacent linear pieces (see the left panel of Figure 9). By construction, $I_{G'}$ lies above $I_G$ and remains below $I_F$. Moreover, the number of induced posterior means remains unchanged. Thus, the information structure associated with $I_{G'}$ is feasible and yields a strictly higher value of the objective function. It follows that under the optimal

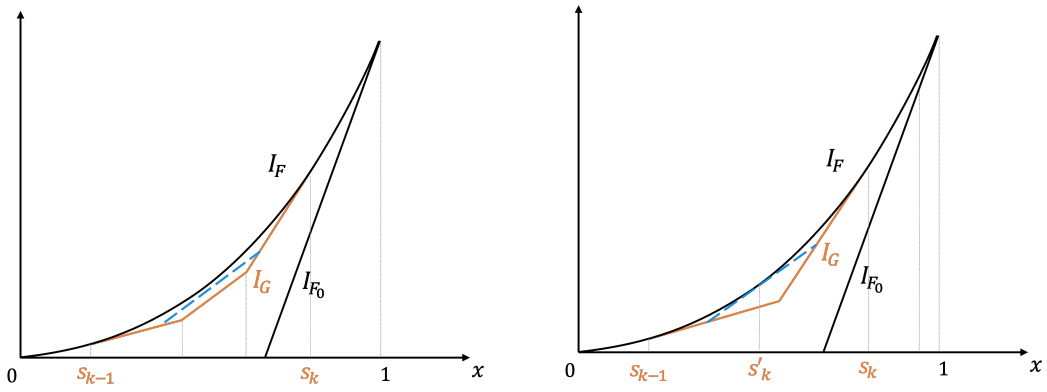



Figure 9: Illustration of Proposition 2

information structure, every segment of $I_G$ must be tangent to $I_F$. By Lemma 1, such an $I_G$ can be implemented via an interval-partitional information structure.

Next, suppose the optimal interval partition induces $N' < N$ posteriors. Then the sender can divide some interval $[s_{k-1}, s_k)$ into two pieces $[s_{k-1}, s'_k) \cup [s'_k, s_k)$ (see the right panel of Figure 9). This new integral distribution with an additional kink point is everywhere higher than the original integral distribution and will produce a higher value of the objective function when $u$ is convex. □

**Proof of Proposition 3.** We prove (iii) first. The functions $\phi(\cdot)$ and $\mu(\cdot)$ are conditional expectations. Therefore, $\phi(a,b) \geq (a+b)/2$ and $\mu(a,b) \geq (a+b)/2$ for any $a < b \leq \min\{m_f, m_u\}$, because both $f$ and $u''$ are increasing in the relevant region. An induction argument establishes that $w_1 \geq w_2 \geq \ldots \geq w_{k-1}$. Similarly, $\phi(a,b) \leq (a+b)/2$ and $\mu(a,b) \leq (a+b)/2$ for any $\max\{m_f, m_u\} \leq a < b$, because both $f$ and $u''$ are decreasing in the relevant region. An induction argument establishes that $w_N \geq w_{N-1} \geq w_{j+1}$. Together, they imply that there exists an $i^* \in \{k-1, \ldots, j+1\}$ such that the sequence of widths attains a minimum at $w_{i^*}$.

A constant $u''$ is a special case of single-peaked $u''$. We can set the mode of $u''$ at $m_f$, and part (i) follows. Similarly, if $f$ is a constant, we can set its mode at $m_u$, and part (ii) follows. □

**Proof of Proposition 9.** Part (i) follows from the definition of an interval-partitional

information structure. Under an interval-partitional information structure $G$,

$$I_G(x) = \begin{cases} 0 & \text{if } x \in [0, x_1) \\ I_F(s_k) + F(s_k)(x - s_k) & \text{if } x \in [x_k, x_{k+1}) \text{ for } k = 1, \ldots, N-1 \\ I_F(1) + (x-1) & \text{if } x \in [x_N, 1]. \end{cases}$$

Therefore, the objective function (2) can be written as a function of the interval cutoffs,

$$\sum_{k=1}^{N-1} \int_{x_k}^{x_{k+1}} u''(x) \left(I_F(s_k) + F(s_k)(x - s_k)\right) \mathrm{d}x + \int_{x_N}^{1} u''(x) \left(I_F(1) + (x-1)\right) \mathrm{d}x,$$

where $x_k = \phi(s_{k-1}, s_k)$. The first-order condition with respect to each interval cutoff $s_k$ leads to,

$$\int_{x_k}^{x_{k+1}} u''(x)(x - s_k)\, \mathrm{d}x = 0, \quad k = 1, \ldots, N-1.$$

Upon integration-by-parts, the above condition leads to the continuity of the minorant function:

$$u(x_k) + u'(x_k)(s_k - x_k) = u(x_{k+1}) + u'(x_{k+1})(s_k - x_{k+1}), \quad k = 1, \ldots, N-1.$$

□

**Proof of Proposition 10.** We first prove that $\overline{u}_G(x) \geq u(x)$ in (i). Let

$$L(a, b) := \frac{u(b) - u(a)}{b - a}.$$

Suppose to the contrary that under the optimal interval partition $G$, there exists $\hat{x} \in (x_k, x_{k+1})$ such that $u(\hat{x}) > u(x_k) + L(x_k, x_{k+1})(\hat{x} - x_k)$. This would imply $L(x_k, \hat{x}) > L(x_k, x_{k+1}) > L(\hat{x}, x_{k+1})$. We then construct another information structure that is strictly better than $G$, which would contradict its optimality.

*Case 1:* $\hat{x} > \phi(s_{k-1}, s_{k+1})$. Consider an information structure $\hat{G}$ such that in the interval $(s_{k-1}, s_{k+1})$, we let the experiment induce a pair of posterior means, $x_k$ or $\hat{x}$. Because $s_{k-1} < x_k < \phi(s_{k-1}, s_{k+1}) < \hat{x} < x_{k+1} < s_{k+1}$, by Lemma 1 of Arieli et al. (2023), such an experiment constitute a bi-pooling policy and is feasible. The information structure $\hat{G}$

is otherwise identical to $G$ for states outside $(s_{k-1}, s_{k+1})$. Observe that

$$\int_{s_{k-1}}^{s_{k+1}} u(x)\,\mathrm{d}G = (F(s_{k+1}) - F(s_{k-1}))\,[u(x_k) + L(x_k, x_{k+1})(\phi(s_{k-1}, s_{k+1}) - x_k)]$$
$$< (F(s_{k+1}) - F(s_{k-1}))\,[u(x_k) + L(x_k, \hat{x})(\phi(s_{k-1}, s_{k+1}) - x_k)] = \int_{s_{k-1}}^{s_{k+1}} u(x)\,\mathrm{d}\hat{G}.$$

Thus $\hat{G}$ is strictly better than $G$.

*Case 2:* $\hat{x} < \phi(s_{k-1}, s_{k+1})$. Consider an information structure $\hat{G}$ such that in the interval $(s_{k-1}, s_{k+1})$, the experiment will induce a pair of posterior means, $\hat{x}$ and $x_{k+1}$. Because $x_k < \hat{x} < \phi(s_{k-1}, s_{k+1}) < x_{k+1}$, this new information structure $\hat{G}$ is also feasible. With a similar argument as above, $\hat{G}$ gives a strictly higher payoff than $G$.

*Case 3:* $\hat{x} = \phi(s_{k-1}, s_{k+1})$. Because $u$ is continuous, we can always find a nearby $x'$ such that $x' > \phi(s_{k-1}, s_{k+1})$ and $u(x') > u(x_k) + L(x_k, x_{k+1})(x' - x_k)$. Therefore, we can adopt the same argument as case 1 above to construct a better information structure.

To prove part (ii), note that $u(x_k) + u'(x_k)(s_k - x_k) = u(x_{k+1}) + u'(x_{k+1})(s_k - x_{k+1})$ for $k = 1, \ldots, N-1$ can be re-arranged into:

$$L(x_k, x_{k+1}) = \frac{s_k - x_k}{x_{k+1} - x_k} u'(x_k) + \frac{x_{k+1} - s_k}{x_{k+1} - x_k} u'(x_{k+1}).$$

Thus, $L(x_k, x_{k+1})$ must lie between $u'(x_k)$ and $u'(x_{k+1})$. Suppose $u'(x_k) > u'(x_{k+1})$, then we would have $u'(x_k) > L(x_k, x_{k+1}) > u'(x_{k+1})$. But this would imply $u(x) > u(x_k) + L(x_k, x_{k+1})(x - x_k)$ for $x$ sufficiently close to (and larger than) $x_k$, which contradicts part (i). This shows that $u'(x_k) \leq u'(x_{k+1})$ for all $k = 1, \ldots, N-1$ under the optimal interval partition.

When the value function is convex, it is obvious that $u'(x_k) < u'(x_{k+1})$ holds. Now we argue that the same strict inequality holds when the value function is S-shaped. Suppose $u'(x_k) = u'(x_{k+1})$, then $u'(x_k) = L(x_k, x_{k+1}) = u'(x_{k+1})$, which further implies that $u(x_{k+1}) = u(x_k) + u'(x_k)(x_{k+1} - x_k)$ (bi-tangency). Note that $u''$ is single-crossing from above; therefore $u'(x) > u'(x_k)$ for all $x \in (x_k, x_{k+1})$. This implies that

$$u(x_{k+1}) = u(x_k) + \int_{x_k}^{x_{k+1}} u'(x)\,\mathrm{d}x > u(x_k) + \int_{x_k}^{x_{k+1}} u'(x_k)\,\mathrm{d}x = u(x_k) + u'(x_k)(x_{k+1} - x_k).$$

This contradicts the bi-tangency condition, and we have established part (ii).

Finally, the proof of part (ii) shows that $L(x_k, x_{k+1})$ must lie between $u'(x_k)$ and $u'(x_{k+1})$. Since we have established that $u'(x_k)$ is nondecreasing in $k$, $L(x_k, x_{k+1})$ must also be nondecreasing in $k$, and therefore $\overline{u}_G$ is convex. This completes the proof of part (i). □

**Proof of Lemma 4.** Suppose to the contrary that $x_i = v_m$ for some $i < N$. (The case where $i = N$ can be proved in the same way). Let $\underline{u}_i$ and $\overline{u}_i$ be the subgradient lines of $u$ at $x_i$, with slopes equal to the *left derivative* and *right derivative*, respectively, at that point, i.e., $\underline{u}_i(x) := u(x_i) + u'(v_{m-1})(x - x_i) = u(v_m) + u'(v_{m-1})(x - v_m)$ and $\overline{u}_i(x) := u(x_i) + u'(v_m)(x - x_i) = u(v_m) + u'(v_m)(x - v_m)$. Similarly, let $\underline{u}_{i+1}$ and $\overline{u}_{i+1}$ be the subgradient lines of $u$ at $x_{i+1}$, with slopes equal to the *left derivative* and *right derivative* at that point, i.e., $\underline{u}_{i+1}(x) := u(x_{i+1}) + u'(x_{i+1}^-)(x - x_{i+1})$ and $\overline{u}_{i+1}(x) := u(x_{i+1}) + u'(x_{i+1}^+)(x - x_{i+1})$. Note that $x_{i+1} \geq v_{m+1}$, because otherwise either $x_i$ or $x_{i+1}$ is redundant. Let $\{s_i\}_{i=1}^{N-1}$ be the interval cutoff points under the optimal information structure $G^*$. There are two cases to consider.

*Case 1:* $\overline{u}_i(s_i) > \overline{u}_{i+1}(s_i)$. There exists some $\varepsilon > 0$ such that $s_i + \varepsilon < x_{i+1}$ and $\overline{u}_i(x) > \overline{u}_{i+1}(x)$ for all $x \in [s_i, s_i + \varepsilon]$. Consider another information structure $\hat{G}$, which is identical to $G^*$ except that the cutoff point $s_i$ is replaced by $s_i + \varepsilon$. Thus, $\hat{x}_i = \phi(s_{i-1}, s_i + \varepsilon) > x_i$ and $\hat{x}_{i+1} = \phi(s_i + \varepsilon, s_{i+1}) > x_{i+1}$. Moreover, pick $\varepsilon$ small enough such that $u(\hat{x}_i) = \overline{u}_i(\hat{x}_i)$ and $u(\hat{x}_{i+1}) = \overline{u}_{i+1}(\hat{x}_{i+1})$. We want to show that $\hat{G}$ gives a higher payoff than $G$. To see this,

$$\begin{aligned}
\int_{s_{i-1}}^{s_{i+1}} u(x)\,\mathrm{d}\hat{G}(x) &= \int_{s_{i-1}}^{s_i+\varepsilon} \overline{u}_i(x)\,\mathrm{d}\hat{G}(x) + \int_{s_i+\varepsilon}^{s_{i+1}} \overline{u}_{i+1}(x)\,\mathrm{d}\hat{G}(x) \\
&= \int_{s_{i-1}}^{s_i+\varepsilon} \overline{u}_i(x)\,\mathrm{d}F(x) + \int_{s_i+\varepsilon}^{s_{i+1}} \overline{u}_{i+1}(x)\,\mathrm{d}F(x) \\
&> \int_{s_{i-1}}^{s_i} \overline{u}_i(x)\,\mathrm{d}F(x) + \int_{s_i}^{s_{i+1}} \overline{u}_{i+1}(x)\,\mathrm{d}F(x) = \int_{s_{i-1}}^{s_{i+1}} u(x)\,\mathrm{d}G^*(x).
\end{aligned}$$

The inequality follows from the premise that $\overline{u}_i > \overline{u}_{i+1}$ on $[s_i, s_i + \varepsilon]$. Since $G^*$ and $\hat{G}$ yields the same expected payoff for all states outside $[s_{i-1}, s_{i+1}]$, $G^*$ is strictly suboptimal.

*Case 2:* $\underline{u}_i(s_i) < \overline{u}_{i+1}(s_i)$. There exists some $\varepsilon > 0$ such that $s_i - \varepsilon > x_i$ and $\underline{u}_i(x) < \overline{u}_{i+1}(x)$ for all $x \in [s_i - \varepsilon, s_i]$. Consider another information structure $\hat{G}$, which is identical to $G^*$ except that $s_i$ is replaced by $s_i - \varepsilon$. Thus, $\hat{x}_i = \phi(s_{i-1}, s_i - \varepsilon) < x_i$ and

$\hat{x}_{i+1} = \phi(s_i - \varepsilon, s_{i+1}) < x_{i+1}$. Moreover, pick $\varepsilon$ small enough such that $u(\hat{x}_i) = \underline{u}_i(\hat{x}_i)$ and $u(\hat{x}_{i+1}) = \underline{u}_{i+1}(\hat{x}_{i+1})$. We now show that $\hat{G}$ gives a higher payoff than $G^*$,

$$\begin{aligned}
\int_{s_{i-1}}^{s_{i+1}} u(x)\,\mathrm{d}\hat{G} &= \int_{s_{i-1}}^{s_i-\varepsilon} \underline{u}_i(x)\,\mathrm{d}\hat{G}(x) + \int_{s_i-\varepsilon}^{s_{i+1}} \underline{u}_{i+1}(x)\,\mathrm{d}\hat{G}(x) \\
&\geq \int_{s_{i-1}}^{s_i-\varepsilon} \underline{u}_i(x)\,\mathrm{d}\hat{G}(x) + \int_{s_i-\varepsilon}^{s_{i+1}} \overline{u}_{i+1}(x)\,\mathrm{d}\hat{G}(x) \\
&= \int_{s_{i-1}}^{s_i-\varepsilon} \underline{u}_i(x)\,\mathrm{d}F(x) + \int_{s_i-\varepsilon}^{s_{i+1}} \overline{u}_{i+1}(x)\,\mathrm{d}F(x) \\
&> \int_{s_{i-1}}^{s_i} \underline{u}_i(x)\,\mathrm{d}F(x) + \int_{s_i}^{s_{i+1}} \overline{u}_{i+1}(x)\,\mathrm{d}F(x) = \int_{s_{i-1}}^{s_{i+1}} u(x)\,\mathrm{d}G^*(x).
\end{aligned}$$

The first inequality on the second line comes from $\hat{x}_{i+1} < x_{i+1}$ and from the fact that $u(x)$ is piecewise affine and convex. It is strict if and only if $x_{i+1}$ is one of the kink points, i.e. $x_{i+1} \in \{v_k\}_{k>m}$. The second inequality follows from the premise that $\underline{u}_i < \overline{u}_{i+1}$ on $[s_i - \varepsilon, s_i]$. Hence, we conclude that the information structure $\hat{G}$ yields a higher expected payoff than $G^*$.

By construction, $\overline{u}_i > \underline{u}_i$ on $(x_i, 1]$ and, in particular, at $s_i$. Therefore at least one of these two cases (i.e., $\overline{u}_i(s_i) > \overline{u}_{i+1}(s_i)$ or $\underline{u}_i(s_i) < \overline{u}_{i+1}(s_i)$) must hold. This concludes the proof of the suboptimality of $G^*$. □

## References


**Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita**, "Optimal persuasion via bi-pooling," *Theoretical Economics*, 2023, *18* (1), 15–36.

**Aybas, Yunus C. and Eray Turkel**, "Persuasion with coarse communication," 2025. Working paper, Texas A&M University.

**Banerjee, Arindam, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh**, "Clustering with Bregman divergences," *Journal of Machine Learning Research*, 2005, *6* (58), 1705–1749.

**Bergemann, Dirk, Edmund M. Yeh, and Jinkun Zhang**, "Nonlinear pricing with finite information," *Games and Economic Behavior*, 2021, *130*, 62–84.

**_ , Tibor Heumann, and Stephen Morris**, "Screening with persuasion," *Journal of Political Economy*, 2026, *134* (2), 570–625.

**Bloedel, Alexander W. and Ilya Segal**, "Persuading a rationally inattentive agent," 2021. Working paper, Stanford University.

**Boissonnat, Jean-Daniel, Frank Nielsen, and Richard Nock**, "Bregman Voronoi diagrams," *Discrete & Computational Geometry*, 2010, *44* (2), 281–307.

**Boutoille, Guillaume and Gilles Pagès**, "Optimal Bregman quantization: Existence and uniqueness of optimal quantizers revisited," 2025. Working paper. `https://arxiv.org/abs/2506.01746`.

**Caplin, Andrew and Mark Dean**, "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

**Chen, Ying and Sidartha Gordon**, "Information transmission in nested sender-receiver games," *Economic Theory*, 2015, *58* (3), 543–569.

**Crawford, Vincent P. and Joel Sobel**, "Strategic information transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

**Crémer, Jacques, Luis Garicano, and Andrea Prat**, "Language and the theory of the firm," *Quarterly Journal of Economics*, 2007, *122* (1), 373–407.

**Curello, Gregorio and Ludvig Sinander**, "The comparative statics of persuasion," 2023. Working paper, University of Bonn.

**Deimen, Inga and Dezső Szalay**, "Delegated expertise, authority, and communication," *American Economic Review*, 2019, *109* (4), 1349–1374.

**_ and _** , "Communication in the shadow of catastrophe," 2023. Working paper, University of Bonn.

**Denti, Tommaso**, "Posterior separable cost of information," *American Economic Review*, 2022, *112* (10), 3215–3259.

**_ , Massimo Marinacci, and Aldo Rustichini**, "Experimental cost of information," *American Economic Review*, 2022, *112* (9), 3106–3123.

**Dharmadhikari, Sudhakar and Kumar Joag-dev**, *Unimodality, Convexity, and Applications*, San Diego, CA: Academic Press, 1988.

**Dow, James**, "Search decisions with limited memory," *Review of Economic Studies*, 1991, *58* (1), 1–14.

**Dworczak, Piotr and Giorgio Martini**, "The simple economics of optimal persuasion," *Journal of Political Economy*, 2019, *127* (5), 1993–2048.

**Fischer, Aurélie**, "Quantization and clustering with Bregman divergences," *Journal of Multivariate Analysis*, 2010, *101* (9), 2207–2221.

**Gentzkow, Matthew and Emir Kamenica**, "A Rothschild-Stiglitz approach to Bayesian persuasion," *American Economic Review*, 2016, *106* (5), 597–601.

**Gordon, Sidartha**, "On infinite cheap talk equilibria," 2010. Working paper, University of Montreal.

**Graf, Siegfried and Harald Luschgy**, *Foundations of Quantization for Probability Distributions*, Berlin: Springer, 2000.

**Gray, Robert M. and David L. Neuhoff**, "Quantization," *IEEE Transactions on Information Theory*, 1998, *44* (6), 2325–2383.

**Harbaugh, Rick and Eric Rasmusen**, "Coarse grades: Informing the public by withholding information," *American Economic Journal: Microeconomics*, 2018, *10* (1), 210–235.

**Hopenhayn, Hugo and Maryam Saeedi**, "Optimal simple ratings," 2024. Working paper, University of California, Los Angeles.

**Hoppe, Heidrun C., Benny Moldovanu, and Emre Ozdenoren**, "Coarse matching with incomplete information," *Economic Theory*, 2010, *47* (1), 75–104.

**Ivanov, Maxim**, "Optimal monotone signals in Bayesian persuasion mechanisms," *Economic Theory*, 2021, *72* (3), 955–1000.

**Jäger, Gerhard, Lars P. Metzger, and Frank Riedel**, "Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signal," *Games and Economic Behavior*, 2011, *73* (2), 517–537.

**Kamenica, Emir and Matthew Gentzkow**, "Bayesian persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

**Kleiner, Andreas, Benny Moldovanu, and Philipp Strack**, "Extreme points and majorization: Economic applications," *Econometrica*, 2021, *89* (4), 1557–1593.

**Kolotilin, Anton**, "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 2018, *13* (2), 607–635.

**_ , Hongyi Li, and Andriy Zapechelnyuk**, "On monotone persuasion," 2025. SSRN Working paper.

**_ , Tymofiy Mylovanov, and Andriy Zapechelnyuk**, "Censorship as optimal persuasion," *Theoretical Economics*, 2022, *17* (2), 561–585.

**_ , _ , _ , and Ming Li**, "Persuasion of a privately informed receiver," *Econometrica*, 2017, *85* (6), 1949–1964.

**Kos, Nenad**, "Communication and efficiency in auctions," *Games and Economic Behavior*, 2012, *75* (1), 233–249.

**Kreutzkamp, Sophie and Yichuan Lou**, "Persuasion without ex-post commitment," *Journal of Economic Theory*, 2025, *228*, 106058.

**Matějka, Filip**, "Rigid pricing and rationally inattentive consumer," *Journal of Economic Theory*, 2015, *158* (B), 656–678.

**Matysková, Ludmila and Alfonso Montes**, "Bayesian persuasion with costly information acquisition," *Journal of Economic Theory*, 2023, *223*, 105678.

**McAfee, R. Preston**, "Coarse matching," *Econometrica*, 2002, *70* (5), 2025–2034.

**Mease, David and Vijayan N. Nair**, "Unique optimal partitions of distributions and connections to hazard rates and stochastic ordering," *Statistica Sinica*, 2006, *16* (4), 1299–1312.

**Mensch, Jeffrey**, "Monotone persuasion," *Games and Economic Behavior*, 2021, *130*, 521–542.

**Milgrom, Paul and Ilya Segal**, "Envelope theorems for arbitrary choice sets," *Econometrica*, 2002, *70* (2), 583–601.

**Mussa, Michael and Sherwin Rosen**, "Monopoly and product quality," *Journal of Economic Theory*, 1978, *18* (2), 301–317.

**Netzer, Nick**, "Evolution of time preferences and attitudes toward risk," *American Economic Review*, 2009, *99* (3), 937–955.

**Onuchic, Paula and Debraj Ray**, "Conveying value via categories," *Theoretical Economics*, 2023, *18* (4), 1407–1439.

**Ostrovsky, Michael and Michael Schwarz**, "Information disclosure and unraveling in matching markets," *American Economic Journal: Microeconomics*, 2010, *2* (2), 34–63.

**Ravid, Doron, Anne-Katrin Roesler, and Balázs Szentes**, "Learning before trading: On the inefficiency of ignoring free information," *Journal of Political Economy*, 2022, *130* (2), 346–387.

**Rayo, Luis and Ilya Segal**, "Optimal information disclosure," *Journal of Political Economy*, 2010, *118* (5), 949–987.

**Robson, Arthur J.**, "The biological basis of economic behavior," *Journal of Economic Literature*, 2001, *39* (1), 11–33.

**Shaked, Moshe and J. George Shanthikumar**, *Stochastic Orders*, New York, NY: Springer, 2007.

**Smith, Lones, Peter Norman Sørensen, and Jianrong Tian**, "Informational herding, optimal experimentation, and contrarianism," *Review of Economic Studies*, 2021, *88* (5), 2527–2554.

**Strassen, Volker**, "The existence of probability measures with given marginals," *Annals of Mathematical Statistics*, 1965, *36* (2), 423–439.

**Szalay, Dezső**, "Strategic information transmission and stochastic orders," 2012. Working paper, University of Bonn.

**Tian, Jianrong**, "Optimal interval division," *Economic Journal*, 2022, *132* (641), 424–435.

**Topkis, Donald M.**, *Supermodularity and Complementarity*, Princeton, NJ: Princeton University Press, 1998.

**Whitt, Ward**, "Uniform conditional variability ordering of probability distributions," *Journal of Applied Probability*, 1985, *22* (3), 619–633.

**Wilson, Robert**, "Efficient and competitive rationing," *Econometrica*, 1989, *57* (1), 1–40.

**Wong, Adam Chi Leung**, "The choice of the number of varieties: Justifying simple mechanisms," *Journal of Mathematical Economics*, 2014, *54*, 7–21.

[This page is intentionally left blank.]

## Online Appendix: Proof of results for S-shaped value functions

***Proof of Proposition 7.*** To see that an optimal interval partition for S-shaped value functions must satisfy (CE-$F$) and (CE-$u'$), Proposition 9 in Section 6 establishes that, for general value functions, an optimal interval-partitional information structure must satisfy (CE-$F$) and (CE-$u'$), whenever the latter is well defined. Moreover, under S-shaped value functions, given the optimality of interval partitions, Proposition 10(ii) shows that the induced posterior means of an optimal interval partition must satisfy $u'(x_{k+1}) > u'(x_k)$ for $k = 1, \ldots, N-1$. This ensures that the barycenter (6) for the signed measure $u'$ is well defined.

We now establish the last part of the proposition. Because the optimal information structure $G^*$ in the unconstrained problem is perfectly revealing for states below $s^*$ and pools all states above $s^*$, to show that the optimal coarse information structure $G$ is a garbling of $G^*$, it suffices to show that $s_{N-1} < s^*$. The proof for $s_{N-1} < s^*$ follows from Claims 1 and 2 below. These two claims rely on properties of an optimal interval partition established in Propositions 9 and 10 for general value functions. Since the proof sketch has already established that an optimal coarse information structure $G$ for S-shaped value functions is indeed interval-partitional, the restrictions imposed on $G$ by Proposition 10 apply. □

**Claim 1.** *Under an optimal $N$-interval partition $G$, $s_{N-1} < t$.*

*Proof.* Suppose $s_{N-1} \geq t$. It implies $x_N > t$. This in turn implies $x_{N-1} < t$; for otherwise the line segment connecting $u(x_{N-1})$ and $u(x_N)$ is not above $u(x)$, contradicting Proposition 10(i). By concavity of $u$ on $[t, x_N]$ and the fact that $u(x_N) = \underline{u}_G(x_N)$, we have $u(t) < u(x_N) + u'(x_N)(t - x_N) \leq \underline{u}_G(t)$, where the last inequality is based on the convexity of the minorant function from Proposition 10(ii) and the supposition $t \leq s_{N-1}$. On the other hand, by convexity of $u$ on $[x_{N-1}, t]$ and the fact that $u(x_{N-1}) = \underline{u}_G(x_{N-1})$, we have $u(t) > \underline{u}_G(t)$. This contradicts the continuity of $\underline{u}_G(\cdot)$ from Proposition 9. □

**Claim 2.** *Under an optimal $N$-interval partition $G$, $s_{N-1} < s^*$.*

*Proof.* Let $x^* = \phi(s^*, 1)$. By Theorem 1 in Dworczak and Martini (2019), $s^* < t < x^*$. Let $\underline{u}^*(x) = u(x^*) + u'(x^*)(x - x^*)$. Dworczak and Martini (2019) show that $\underline{u}^*(x) \geq u(x)$ on $[s^*, 1]$, with equality at $s^*$ and $x^*$.

Now suppose $s_{N-1} \geq s^*$, and therefore $x_N \geq x^*$. Consequently, $x_N \geq x^* > t$. The concavity of $u$ on $(t,1]$ and the fact that $u(x_N) = \underline{u}_G(x_N)$ imply that $\underline{u}_G(x^*) \geq u(x^*) = \underline{u}^*(x^*)$, and that the slopes satisfy $\underline{u}'_G(x) = u'(x_N) \leq u'(x^*) = \underline{u}^{*\prime}(x)$ for all $x \in (s_{N-1}, 1]$. Therefore,

$$\underline{u}_G(s_{N-1}^+) \geq \underline{u}^*(s_{N-1}).$$

On the other hand, by Claim 1, $s_{N-1} < t$. Therefore,

$$\underline{u}^*(s_{N-1}) \geq u(s_{N-1}) > \underline{u}_G(s_{N-1}^-) = u(x_{N-1}) + u'(x_{N-1})(s_{N-1}^- - x_{N-1}),$$

where the last inequality comes from the fact $u$ is convex on $[0, s_{N-1}]$ and $u(x_{N-1}) = \underline{u}_G(x_{N-1})$. The two displayed inequalities contradict the continuity of the minorant function $\underline{u}_G$ required by Proposition 9. □

**Proof of Lemma 3.** We show that $\frac{\partial \mu(a,b)}{\partial a} < 1$ for all $a \in (0,t)$ such that $u'(b) - u'(a) > 0$. Note that

$$\frac{\partial \mu(a,b)}{\partial a} = \frac{u''(a)\int_a^b [u'(b) - u'(x)]\,\mathrm{d}x}{(u'(b) - u'(a))^2}.$$

Let $c \in (0,t)$ be the point such that $u'(c) = u'(b)$. Since $u'$ is quasiconcave, we have $u'(b) - u'(x) < 0$ for all $x \in (c,b)$. Therefore,

$$\begin{aligned}
\frac{\partial \mu}{\partial a} &< \frac{u''(a)\int_a^c [u'(b) - u'(x)]\,\mathrm{d}x}{(u'(b) - u'(a))^2} \\
&\leq \frac{1}{u'(b) - u'(a)} \int_a^c \frac{u''(x)}{u'(b) - u'(x)} \big(u'(b) - u'(x)\big)\,\mathrm{d}x \\
&= \frac{1}{u'(b) - u'(a)} \big(u'(c) - u'(a)\big) = 1.
\end{aligned}$$

The second inequality above comes from logconcavity of $u''$ on the convex region (and hence $\frac{u''(x)}{u'(b) - u'(x)} = \frac{u''(x)}{u'(c) - u'(x)}$ is increasing in $x$), and from the fact that $u'(b) - u'(x) > 0$ for all $x \in [a,c)$. □

Before proving Proposition 8, we make several claims below. Define the barycenter for $u'$ as $\mu(a,b) := \frac{\int_a^b \theta u''(\theta)\,\mathrm{d}\theta}{\int_a^b u''(\theta)\,\mathrm{d}\theta}$, whenever the denominator is nonzero. Similarly, define the barycenter for $\hat{u}'$ as $\hat{\mu}(a,b) := \frac{\int_a^b \theta \hat{u}''(\theta)\,\mathrm{d}\theta}{\int_a^b \hat{u}''(\theta)\,\mathrm{d}\theta}$, whenever the denominator is nonzero.

**Claim 3.** *For any $0 < a < t < b \leq 1$ such that $u'(b) - u'(a) > 0$, $\frac{\partial \mu(a,b)}{\partial b} < 0$.*

*Proof.* The derivative of $\mu$ with respect to $b$ is

$$\frac{\partial \mu(a,b)}{\partial b} = \frac{u''(b)\int_a^b [u'(x)-u'(a)]\,\mathrm{d}x}{[u'(b)-u'(a)]^2}.$$

This expression is negative because $u''(b) < 0$; moreover $u'(b) > u'(a)$ implies $u'(x) > u'(a)$ for all $x \in (a,b]$. □

**Claim 4.** *Suppose $\{s_k\}_{k=0}^N$ with $\{x_k\}_{k=1}^N$ solve the equation system (CE-F) and (CE-u′) with $s_{N-1} < t$. Then $u'(x_k) > u'(x_{k-1})$ for all $k = 2,\ldots,N$.*

*Proof.* It suffices to check $u'(x_N) > u'(x_{N-1})$, since the claim trivially holds for all $k = 2,\ldots,N-1$ as both $x_k$ and $x_{k-1}$ lie in the convex region for such $k$. Notice that if $x_N \le t$, the claim again trivially holds; therefore it suffices to prove the inequality for $x_N > t > s_{N-1}$. Suppose $u'(x_N) \le u'(x_{N-1})$ instead. Then

$$\begin{aligned} u(t) &> u(x_{N-1}) + u'(x_{N-1})(t - x_{N-1}) \\ &= u(x_{N-1}) + u'(x_{N-1})(s_{N-1} - x_{N-1}) + u'(x_{N-1})(t - s_{N-1}) \\ &\ge u(x_N) + u'(x_N)(s_{N-1} - x_N) + u'(x_N)(t - s_{N-1}) \\ &= u(x_N) + u'(x_N)(t - x_N) > u(t), \end{aligned}$$

a contradiction. The first inequality utilizes the convexity of $u$ on $(x_{N-1}, t)$. The second inequality follows from that $u(x_{N-1})+u'(x_{N-1})(s_{N-1}-x_{N-1}) = u(x_N)+u'(x_N)(s_{N-1}-x_N)$, $t > s_{N-1}$ and $u'(x_N) \le u'(x_{N-1})$. The last inequality utilizes the concavity of $u$ on $(t, x_N)$. □

**Claim 5.** *Suppose $\hat{u}'' \succeq_{lr} u''$. For any $a \in [0,t)$ and $b > a$,*

*(i) $u'(b)-u'(a) > 0$ implies that $\hat{u}'(b)-\hat{u}'(a) > 0$;*

*(ii) $u'(b)-u'(a) > 0$ implies $\hat{\mu}(a,b) \ge \mu(a,b)$.*

*Proof.* The statements are trivially true if $\hat{u}$ is convex. Hence, we focus on the case where $\hat{u}$ is S-shaped. To show part (i), $\hat{u}'' \succeq_{lr} u''$ implies $\frac{\hat{u}''(s)}{\hat{u}''(a)} \ge \frac{u''(s)}{u''(a)}$ for all $s \ge a$, and hence

$$\frac{\hat{u}'(b)-\hat{u}'(a)}{\hat{u}''(a)} = \int_a^b \frac{\hat{u}''(s)}{\hat{u}''(a)}\,\mathrm{d}s \ge \int_a^b \frac{u''(s)}{u''(a)}\,\mathrm{d}s = \frac{u'(b)-u'(a)}{u''(a)}.$$

Since $a < t \leq \hat{t}$, we have $u''(a) > 0$ and $\hat{u}''(a) > 0$. Therefore, $u'(b) - u'(a) > 0$ implies $\hat{u}'(b) - \hat{u}'(a) > 0$.

Now we show part (ii). If $b < t$, then the inequality follows directly from the property of likelihood ratio order for positive density functions. If $b \in [t, \hat{t}]$, then $\mu(a,b) \leq \mu(a,t) \leq \hat{\mu}(a,t) \leq \hat{\mu}(a,b)$. If $b > \hat{t}$, and suppose to the contrary that $\mu(a,b) - \hat{\mu}(a,b) > 0$, then because $\mu(a,\cdot) - \hat{\mu}(a,\cdot)$ is continuous, there must exist some $q \in [\hat{t}, b)$ such that $\mu(a,q) - \hat{\mu}(a,q) = 0$ and crosses 0 from below. That is,

$$\frac{u''(q)}{u'(q) - u'(a)}(q - \mu(a,q)) = \left.\frac{\partial \mu(a,x)}{\partial x}\right|_{x=q} > \left.\frac{\partial \hat{\mu}(a,x)}{\partial x}\right|_{x=q} = \frac{\hat{u}''(q)}{\hat{u}'(q) - \hat{u}'(a)}(q - \hat{\mu}(a,q)).$$

Since $t - \mu(a,t) > 0$ and $\mu(a,\cdot)$ is decreasing on $[t,b)$, we have $q - \hat{\mu}(a,q) = q - \mu(a,q) > 0$. Meanwhile, $b > q > \hat{t} \geq t$ implies $\hat{u}''(q) < 0$, $u''(q) < 0$, $u'(q) > u'(b) > u'(a)$, and $\hat{u}'(q) > \hat{u}'(b) > \hat{u}'(a)$. Rearranging terms, we obtain:

$$\frac{\hat{u}''(q)}{u''(q)} > \frac{\int_a^q \hat{u}''(s)\,\mathrm{d}s}{\int_a^q u''(s)\,\mathrm{d}s} = \frac{\hat{u}''(a)\int_a^q \frac{\hat{u}''(s)}{\hat{u}''(a)}\,\mathrm{d}s}{u''(a)\int_a^q \frac{u''(s)}{u''(a)}\,\mathrm{d}s} \geq \frac{\hat{u}''(a)}{u''(a)},$$

where the last inequality comes from the definition of likelihood ratio order and that $a < t \leq \hat{t}$. Yet this contradicts $\hat{u}'' \succeq_{lr} u''$. □

**Proof of Proposition 8(i).** The proof of Theorem 2 shows that $\Delta_i \geq \Delta_{i-1}$ implies $\Delta_{i+1} \geq \Delta_i$ within the convex region, namely, for all $i = 2, \ldots, 2N-4$. For S-shaped value function, it remains to be shown that $\Delta_i \geq \Delta_{i-1}$ implies $\Delta_{i+1} \geq \Delta_i$ for $i = 2N-3$ and $i = 2N-2$.

For $i = 2N-3$, this is equivalent to showing that $w_{N-1} \geq d_{N-2}$ implies $d_{N-1} \geq w_{N-1}$. Suppose instead $d_{N-1} < w_{N-1}$. Then

$$\begin{aligned}
s_{N-1} = \mu(x_{N-1}, x_N) &= \mu(x_{N-2} + d_{N-2}, x_{N-1} + d_{N-1}) \\
&\leq \mu\left(x_{N-2} + d_{N-2}, x_{N-1} + \min\{d_{N-1}, t - x_{N-1}\}\right) \\
&\leq \mu\left(x_{N-2} + w_{N-1}, x_{N-1} + \min\{d_{N-1}, t - x_{N-1}\}\right) \\
&< \mu(x_{N-2}, x_{N-1}) + \max\left\{w_{N-1}, \min\{d_{N-1}, t - x_{N-1}\}\right\} \\
&= s_{N-2} + w_{N-1} = s_{N-1},
\end{aligned}$$

which is a contradiction. The first inequality in the above follows from Claim 3, where

the condition for this claim to apply is established by Proposition 7 and Proposition 10(ii). The last strict inequality follows from Lemma 2 and the fact that $\min\{d_{N-1}, t - x_{N-1}\} \leq d_{N-1} < w_{N-1}$.

For $i = 2N-2$, we want to show that $d_{N-1} \geq w_{N-1}$ implies $w_N \geq d_{N-1}$. Suppose we have $w_N < d_{N-1}$ instead. Then,

$$\begin{aligned} x_N = \phi(s_{N-1}, s_N) &= \phi(s_{N-2} + w_{N-1}, s_{N-1} + w_N) \\ &\leq \phi(s_{N-2} + d_{N-1}, s_{N-1} + w_N) \\ &< \phi(s_{N-2}, s_{N-1}) + \max\{d_{N-1}, w_N\} = x_{N-1} + d_{N-1} = x_N, \end{aligned}$$

which is a contradiction. □

**Proof of Proposition 8(ii).** We establish part (ii) by proving a stronger result that there is only one solution to the equation system (CE-$F$) and (CE-$u'$) with the property that $s_{N-1} < t$. To prove this stronger result, we cannot invoke the optimality of interval partition for S-shaped value functions and Proposition 10(ii). Instead we rely on Claim 4, which establishes that any solution to (CE-$F$) and (CE-$u'$) conditional on $s_{N-1} < t$ would generate a minorant function with strictly increasing slopes.

By Claim 4, the barycenters specified in (CE-$F$) and (CE-$u'$) are well defined when $s_{N-1} < t$. We prove the uniqueness of the solution to this equation system subject to $s_{N-1} < t$. Suppose both $\{s'_k\}_{k=0}^N$ with $\{x'_k\}_{k=1}^N$ and $\{s''_k\}_{k=0}^N$ with $\{x''_k\}_{k=1}^N$ solve the equation system (CE-$F$) and (CE-$u'$) with $s'_{N-1} \leq s''_{N-1} < t$. Remove the last two equations, $x_N = \phi(s_{N-1}, 1)$ and $s_{N-1} = \mu(x_{N-1}, x_N)$, from the original equation system and consider the remaining equations:

$$\begin{aligned} x_k &= \phi(s_{k-1}, s_k) \quad \text{for } k = 1, \ldots, N-1 \\ s_k &= \mu(x_k, x_{k+1}) \quad \text{for } k = 1, \ldots, N-2, \end{aligned}$$

with $s_0 = 0$ and $s_{N-1} = s'_{N-1}$. By the same argument in Proposition 1, $\{x'_k\}_{k=1}^{N-1}$ and $\{s'_k\}_{k=1}^{N-2}$ must be the unique solution to the reduced equation system. Similarly, $\{x''_k\}_{k=1}^{N-1}$ and $\{s''_k\}_{k=1}^{N-2}$ must be the unique solution to the reduced equation system with $s_0 = 0$ and $s_{N-1} = s''_{N-1}$.

Now, suppose $s''_{N-1} = s'_{N-1}$. This would imply that $(x'_1, \ldots, x'_{N-1}) = (x''_1, \ldots, x''_{N-1})$. Moreover, $x'_N = \phi(s'_{N-1}, 1) = \phi(s''_{N-1}, 1) = x''_N$. Thus, these two solutions would be

identical. In order to have two distinct solutions, it must hold that $s''_{N-1} > s'_{N-1}$, and hence $x''_N > x'_N$. Corollary 2.5.2 in Topkis (1998) implies that $(x''_1, \ldots, x''_{N-1})$ is (elementwise) larger than $(x'_1, \ldots, x'_{N-1})$. Moreover, we must have $s''_{N-1} - s'_{N-1} > x''_{N-1} - x'_{N-1}$. To see this, suppose instead $s''_{N-1} - s'_{N-1} \leq x''_{N-1} - x'_{N-1}$. Then Lemma 2 would imply $s''_{N-2} - s'_{N-2} \geq x''_{N-1} - x'_{N-1}$. Iterating the argument back leads to $x''_1 - x'_1 \geq s''_1 - s'_1$, which contradicts Lemma 2. Therefore, we can restate the inequality,

$$s''_{N-1} - s'_{N-1} = \mu(x''_{N-1}, x''_N) - \mu(x'_{N-1}, x'_N) > x''_{N-1} - x'_{N-1}.$$

On the other hand, applying Lemma 2 again gives

$$x''_N - x'_N = \phi(s''_{N-1}, 1) - \phi(s'_{N-1}, 1) < s''_{N-1} - s'_{N-1}.$$

Now consider three possibilities. (a) If $t \geq x''_N > x'_N$, then by Lemma 2

$$s''_{N-1} - s'_{N-1} = \mu(x''_{N-1}, x''_N) - \mu(x'_{N-1}, x'_N) \leq \max\{x''_{N-1} - x'_{N-1}, x''_N - x'_N\},$$

contradicting the two inequalities above. (b) If $x''_N > t \geq x'_N$, then

$$\begin{aligned} s''_{N-1} - s'_{N-1} &= \mu(x''_{N-1}, x''_N) - \mu(x'_{N-1}, x'_N) \\ &< \mu(x''_{N-1}, t) - \mu(x'_{N-1}, x'_N) \leq \max\{x''_{N-1} - x'_{N-1}, t - x'_N\} \\ &\leq t - x'_N < x''_N - x'_N, \end{aligned}$$

where the first inequality in the second line follows from Claim 3, which is applicable since $u'(x''_N) > u'(x''_{N-1})$ (by Claim 4), and it in turn implies $u'(x) > u'(x''_{N-1})$ for all $x \in (t, x''_N)$. The first inequality in the last line comes from that $s''_{N-1} - s'_{N-1} > x''_{N-1} - x'_{N-1}$. But then the final inequality contradicts $s''_{N-1} - s'_{N-1} > x''_N - x'_N$. (c) If $x''_N > x'_N > t$, then

$$s''_{N-1} - s'_{N-1} = \mu(x''_{N-1}, x''_N) - \mu(x'_{N-1}, x'_N) \leq \mu(x''_{N-1}, x'_N) - \mu(x'_{N-1}, x'_N) \leq x''_{N-1} - x'_{N-1}.$$

The first inequality above follows from Claim 3 (and Claim 4), and the last inequality comes from Lemma 3 (which applies because $u'(x'_N) - u'(x''_{N-1}) > u'(x''_N) - u'(x''_{N-1}) > 0$ by Claim 4). Yet this again leads to a contradiction. Therefore, we conclude that $s''_{N-1} = s'_{N-1}$, and consequently the two solutions must be identical. □

**Proof of Proposition 8(iii).** Let $\hat{u}'' \succeq_{lr} u''$. Suppose that $\hat{u}''$ is single-crossing at $\hat{t}$ and

logconcave on $(0, \hat{t})$, and that $f$ is logconcave. Let the optimal interval cutoffs under $u$ and $\hat{u}$ be $\{s_k\}_{k=1}^{N-1}$ and $\{\hat{s}_k\}_{k=1}^{N-1}$, respectively.

*Step 1:* We prove $\hat{s}_{N-1} \geq s_{N-1}$. Suppose to the contrary that $\hat{s}_{N-1} < s_{N-1}$. Consequently, $\hat{x}_N = \phi(\hat{s}_{N-1}, 1) < \phi(s_{N-1}, 1) = x_N$. First, we argue that $s_{N-1} - \hat{s}_{N-1} < x_{N-1} - \hat{x}_{N-1}$. We discuss three cases separately: (a) $\hat{t} \leq \hat{x}_N < x_N$; (b) $\hat{x}_N < \hat{t} < x_N$; and (c) $\hat{x}_N < x_N \leq \hat{t}$.

In case (a), we have

$$\begin{aligned} s_{N-1} = \mu(x_{N-1}, x_N) &\leq \hat{\mu}(x_{N-1}, x_N) \\ &< \hat{\mu}(x_{N-1}, \hat{x}_N) \\ &\leq \hat{\mu}\left(\max\{x_{N-1}, \hat{x}_{N-1}\}, \hat{x}_N\right) \\ &= \hat{\mu}\left(\hat{x}_{N-1} + \max\{x_{N-1} - \hat{x}_{N-1}, 0\}, \hat{x}_N\right) \\ &\leq \hat{\mu}(\hat{x}_{N-1}, \hat{x}_N) + \max\{x_{N-1} - \hat{x}_{N-1}, 0\} \\ &= \hat{s}_{N-1} + \max\{x_{N-1} - \hat{x}_{N-1}, 0\}. \end{aligned}$$

The first inequality comes from Claim 5, where the premise of the claim is established by Proposition 7 and Proposition 10(ii). The second inequality follows from the fact that $\hat{\mu}(x_{N-1}, \cdot)$ is decreasing on $(\hat{t}, 1)$ by Claim 3 and $\hat{x}_N < x_N$. The third inequality holds because $\hat{\mu}(\cdot, \hat{x}_N)$ is increasing on $(0, \hat{x}_{N-1})$ since $\hat{u}'(\hat{x}_N) - \hat{u}'(\hat{x}_{N-1}) > 0$. The fourth inequality invokes Lemma 3. Given that $s_{N-1} - \hat{s}_{N-1} > 0$, we obtain $s_{N-1} - \hat{s}_{N-1} < x_{N-1} - \hat{x}_{N-1}$.

In case (b),

$$\begin{aligned} s_{N-1} = \mu(x_{N-1}, x_N) &\leq \hat{\mu}(x_{N-1}, x_N) \\ &< \hat{\mu}(x_{N-1}, \hat{t}) \\ &\leq \hat{\mu}\left(\max\{x_{N-1}, \hat{x}_{N-1}\}, \hat{t}\right) \\ &= \hat{\mu}\left(\hat{x}_{N-1} + \max\{x_{N-1} - \hat{x}_{N-1}, 0\}, \hat{x}_N + \hat{t} - \hat{x}_N\right) \\ &\leq \hat{\mu}(\hat{x}_{N-1}, \hat{x}_N) + \max\left\{\max\{x_{N-1} - \hat{x}_{N-1}, 0\}, \hat{t} - \hat{x}_N\right\} \\ &= \hat{s}_{N-1} + \max\{x_{N-1} - \hat{x}_{N-1}, \hat{t} - \hat{x}_N\}. \end{aligned}$$

Because $f$ is logconcave, we have $s_{N-1} - \hat{s}_{N-1} > x_N - \hat{x}_N > \hat{t} - \hat{x}_N$, where the second inequality comes from the premise of case (b). Therefore, we reach the same conclusion that $s_{N-1} - \hat{s}_{N-1} < x_{N-1} - \hat{x}_{N-1}$.

In case (c),

$$\begin{aligned} s_{N-1} = \mu(x_{N-1}, x_N) &\le \hat{\mu}(x_{N-1}, x_N) \\ &\le \hat{\mu}\left(\max\{x_{N-1}, \hat{x}_{N-1}\}, x_N\right) \\ &= \hat{\mu}\left(\hat{x}_{N-1} + \max\{x_{N-1} - \hat{x}_{N-1}, 0\}, \hat{x}_N + x_N - \hat{x}_N\right) \\ &\le \hat{\mu}(\hat{x}_{N-1}, \hat{x}_N) + \max\left\{\max\{x_{N-1} - \hat{x}_{N-1}, 0\}, x_N - \hat{x}_N\right\} \\ &= \hat{s}_{N-1} + \max\{x_{N-1} - \hat{x}_{N-1}, x_N - \hat{x}_N\}. \end{aligned}$$

Since $s_{N-1} - \hat{s}_{N-1} > x_N - \hat{x}_N$, we obtain $s_{N-1} - \hat{s}_{N-1} \le x_{N-1} - \hat{x}_{N-1}$. In fact, the latter inequality must hold strictly, for otherwise this would violate Lemma 2 due to $x_N - \hat{x}_N \neq x_{N-1} - \hat{x}_{N-1} > 0$.

Having shown that $\hat{s}_{N-1} < s_{N-1}$ implies $s_{N-1} - \hat{s}_{N-1} < x_{N-1} - \hat{x}_{N-1}$ in all these three cases, by logconcavity of $f$ it further implies:

$$s_{N-2} - \hat{s}_{N-2} > x_{N-1} - \hat{x}_{N-1} > s_{N-1} - \hat{s}_{N-1}.$$

Once in the convex region, we can apply the similar argument inductively and get $x_k - \hat{x}_k > s_k - \hat{s}_k$, for $k = 1, \ldots, N-1$. In particular, $x_1 - \hat{x}_1 > s_1 - \hat{s}_1$, which contradicts the property of logconcavity of $f$ from Lemma 2. Therefore, it must hold that $\hat{s}_{N-1} \ge s_{N-1}$.

*Step 2:* We prove $\hat{x}_{N-1} \ge x_{N-1}$. Suppose to the contrary that $\hat{x}_{N-1} < x_{N-1}$. By logconcavity of $f$, we must have $s_{N-2} - \hat{s}_{N-2} > x_{N-1} - \hat{x}_{N-1} > 0$. On the other hand,

$$\begin{aligned} s_{N-2} = \mu(x_{N-2}, x_{N-1}) &\le \hat{\mu}(x_{N-2}, x_{N-1}) \\ &\le \hat{\mu}\left(\max\{x_{N-2}, \hat{x}_{N-2}\}, x_{N-1}\right) \\ &= \hat{\mu}\left(\hat{x}_{N-2} + \max\{x_{N-2} - \hat{x}_{N-2}, 0\}, \hat{x}_{N-1} + x_{N-1} - \hat{x}_{N-1}\right) \\ &\le \hat{\mu}\left(\hat{x}_{N-2}, \hat{x}_{N-1}\right) + \max\left\{\max\{x_{N-2} - \hat{x}_{N-2}, 0\}, x_{N-1} - \hat{x}_{N-1}\right\} \\ &= \hat{s}_{N-2} + \max\{x_{N-2} - \hat{x}_{N-2}, x_{N-1} - \hat{x}_{N-1}\}. \end{aligned}$$

Hence, $s_{N-2} - \hat{s}_{N-2} \le x_{N-2} - \hat{x}_{N-2}$. This inequality holds strictly, i.e., $s_{N-2} - \hat{s}_{N-2} < x_{N-2} - \hat{x}_{N-2}$, for the same reason as in case (c) of Step 1. Similar to Step 1, we then conduct this argument inductively and obtain $x_1 - \hat{x}_1 > s_1 - \hat{s}_1$, which contradicts the logconcavity of $f$. Therefore, we must have $\hat{x}_{N-1} \ge x_{N-1}$.

*Step 3:* Repeat the similar argument inductively, we prove that $\hat{x}_k \ge x_k$ for all $k = 1, \ldots, N$, and $\hat{s}_k \ge s_k$ for all $k = 1, \ldots, N-1$. □

**Proof of Proposition 8(iv).** We prove the result for a change of value function. The case for a change of prior distribution is similar.

Define the sequence,

$$\{\delta_i\}_{i=1}^{2N-1} := \{x_1 - \hat{x}_1, s_1 - \hat{s}_1, x_2 - \hat{x}_2, \ldots, s_{N-1} - \hat{s}_{N-1}, x_N - \hat{x}_N\}.$$

Note that $\delta_{2N-2} \geq (>)\ 0$ implies $\delta_{2N-1} \geq (>)\ 0$, as $s_{N-1} \geq (>)\ \hat{s}_{N-1}$ implies $x_N \geq (>)\ \hat{x}_N$. Therefore, to establish the proposition, it suffices to show that the subsequence, $\{\delta_i\}_{i=1}^{2N-2}$, is single-crossing from below.

Also note that $u'' \succeq_{uv} \hat{u}''$ implies either $t \leq \hat{t} \leq p$ or $p \leq \hat{t} \leq t$. Suppose $t \leq \hat{t} \leq p$. Then $x_{N-1} < s_{N-1} \leq t \leq p$ and $\hat{x}_{N-1} < \hat{s}_{N-1} \leq \hat{t} \leq p$. This means that all the relevant terms involved in the sequence $\{\delta_i\}_{i=1}^{2N-2}$ are in the convex region. We can use the same argument as in Proposition 5(b) to show that this sequence is single-crossing from below.

From now on, we focus on the case where $p \leq \hat{t} \leq t$. There are four cases to consider.

*Case 1:* $\max\{\hat{x}_{N-1}, x_{N-1}\} \leq p$. If $x_N \leq p$, then the argument for Proposition 5(b) directly applies without modification. Hence, we focus on the case with $x_{N-1} \leq p < x_N$. In this case $k^* = N-1$ is the index such that $x_{k^*} \leq p < x_{k^*+1}$. We can apply the argument in Proposition 5(b) to conclude that the sequence $\{\delta_i\}_{i=1}^{2k^*}$ is single-crossing from below.

*Case 2:* $x_{N-1} \leq p \leq \hat{x}_{N-1}$ with $x_{N-1} < \hat{x}_{N-1}$. From the argument that leads to Proposition 5(b), we know that if $\delta_j \geq 0$ for some $j \leq 2N-4$, then all $\{\delta_i\}_{i=j}^{2N-3}$ will be nonnegative, contradicting the premise of Case 2, i.e., $\delta_{2N-3} = x_{N-1} - \hat{x}_{N-1} < 0$. This shows that all terms except the last term of the sequence $\{\delta_i\}_{i=1}^{2N-2}$ are negative under Case 2, and thus the sequence must be single-crossing from below.

*Case 3:* $\hat{x}_{N-1} \leq p \leq x_{N-1}$ with $\hat{x}_{N-1} < x_{N-1}$. The subsequence $\{\delta_i\}_{i=1}^{2N-4}$ is single-crossing from below by the same reasoning in Proposition 5(b), because $s_{N-2} < t$ and $\hat{s}_{N-2} < \hat{t} \leq t$ (the subsequence is in the convex region). Moreover, $\delta_{2N-4} \geq 0$ leading to $\delta_{2N-3} > 0$ is automatically satisfied by the premise that $\hat{x}_{N-1} < x_{N-1}$. It remains to show that $s_{N-1} - \hat{s}_{N-1} > 0$.

Consider two subcases: (a) $s_{N-2} \leq \hat{s}_{N-2}$; and (b) $s_{N-2} > \hat{s}_{N-2}$.

In subcase (a), it is trivially true that $s_{N-1} > \hat{s}_{N-1}$, for otherwise $x_{N-1} = \phi(s_{N-2}, s_{N-1}) \leq \phi(\hat{s}_{N-2}, \hat{s}_{N-1}) = \hat{x}_{N-1}$, contradicting the premise of Case 3.

In subcase (b), the proof of $s_{N-1} - \hat{s}_{N-1} > 0$ contains two steps.

*Step 1*. We want to show that $s_{N-2} - \hat{s}_{N-2} > x_{N-2} - \hat{x}_{N-2}$ implies that $s_{N-1} > \hat{s}_{N-1}$.

We first observe that $s_{N-2} - \hat{s}_{N-2} > x_{N-2} - \hat{x}_{N-2}$ implies $x_{N-1} - \hat{x}_{N-1} > s_{N-2} - \hat{s}_{N-2}$. To see this,

$$\begin{aligned}
s_{N-2} &= \mu(x_{N-2}, x_{N-1}) \\
&\leq \mu\left(\hat{x}_{N-2} + \max\{x_{N-2} - \hat{x}_{N-2}, 0\}, \hat{x}_{N-1} + x_{N-1} - \hat{x}_{N-1}\right) \\
&\leq \mu(\hat{x}_{N-2}, \hat{x}_{N-1}) + \max\{x_{N-2} - \hat{x}_{N-2}, x_{N-1} - \hat{x}_{N-1}\} \\
&\leq \hat{\mu}(\hat{x}_{N-2}, \hat{x}_{N-1}) + \max\{x_{N-2} - \hat{x}_{N-2}, x_{N-1} - \hat{x}_{N-1}\} \\
&= \hat{s}_{N-2} + \max\{x_{N-2} - \hat{x}_{N-2}, x_{N-1} - \hat{x}_{N-1}\}.
\end{aligned}$$

The third inequality is valid since $\hat{x}_{N-2} < \hat{x}_{N-1} \leq p$. From the sequence of inequalities, if $s_{N-2} - \hat{s}_{N-2} > x_{N-2} - \hat{x}_{N-2}$, we can conclude that $x_{N-1} - \hat{x}_{N-1} > s_{N-2} - \hat{s}_{N-2}$. Consequently, $s_{N-1} - \hat{s}_{N-1} > x_{N-1} - \hat{x}_{N-1} > 0$ holds by logconcavity of $f$.

*Step 2*. We want to show that step 1's premise $s_{N-2} - \hat{s}_{N-2} > x_{N-2} - \hat{x}_{N-2}$ holds.

Let $\delta_j$ be the first nonnegative term in $\{\delta_i\}_{i=1}^{2N-4}$. If $j = 2N-4$, then the desired inequality holds trivially since $s_{N-2} - \hat{s}_{N-2} \geq 0 > x_{N-2} - \hat{x}_{N-2}$. Therefore, we study the remaining cases where $j \leq 2N-5$. Similar to Proposition 5(b), let $k^* \geq 1$ be the integer such that $x_{k^*} \leq p < x_{k^*+1}$. Based on the argument for Proposition 5(b), for $\{\delta_i\}_{i=j}^{2k^*}$, if nonempty, $\delta_i \geq (>)\, 0$ implies $\delta_{i+1} \geq (>)\, \delta_i$ for $i = j, \ldots, 2k^* - 1$.

Suppose $k^* = N-2$, and by the premise $s_{N-2} - \hat{s}_{N-2} > 0$, the desired inequality trivially holds since $s_{N-2} - \hat{s}_{N-2} = s_{k^*} - \hat{s}_{k^*} = \delta_{2k^*} > \delta_{2k^*-1} = x_{k^*} - \hat{x}_{k^*} = x_{N-2} - \hat{x}_{N-2} \geq 0$ where the last inequality comes from the first non-negative term has an index $j \leq 2N-5$.

Suppose $k^* \leq N-3$. To show the desired inequality, we show that $\{\delta_i\}_{i=2k^*}^{2N-4}$ is strictly increasing. Note that by the premise of case 3, $\hat{x}_{k^*} < \hat{x}_{k^*+1} \leq \hat{x}_{N-2} < \hat{x}_{N-1} \leq p < x_{k^*+1} < x_{N-1}$.

First, we prove $\delta_{2k^*+1} > \delta_{2k^*}$, i.e. $x_{k^*+1} - \hat{x}_{k^*+1} > s_{k^*} - \hat{s}_{k^*}$. Suppose $s_{k^*} - \hat{s}_{k^*} \leq 0$. Then the inequality is automatically true as $x_{k^*+1} - \hat{x}_{k^*+1} > 0$. On the other hand, suppose

$s_{k^*} - \hat{s}_{k^*} > 0$. Then

$$\begin{aligned}
s_{k^*} &= \mu(x_{k^*}, x_{k^*+1}) \\
&\leq \mu\left(\hat{x}_{k^*} + \max\{x_{k^*} - \hat{x}_{k^*}, 0\}, \hat{x}_{k^*+1} + x_{k^*+1} - \hat{x}_{k^*+1}\right) \\
&\leq \mu(\hat{x}_{k^*}, \hat{x}_{k^*+1}) + \max\{x_{k^*} - \hat{x}_{k^*}, x_{k^*+1} - \hat{x}_{k^*+1}\} \\
&\leq \hat{\mu}(\hat{x}_{k^*}, \hat{x}_{k^*+1}) + \max\{x_{k^*} - \hat{x}_{k^*}, x_{k^*+1} - \hat{x}_{k^*+1}\} \\
&= \hat{s}_{k^*} + \max\{x_{k^*} - \hat{x}_{k^*}, x_{k^*+1} - \hat{x}_{k^*+1}\}.
\end{aligned}$$

The first inequality is by monotonicity of $\mu(\cdot, x_{k^*+1})$, the second inequality is by Lemma 2, and the third inequality follows from the fact that $\hat{u}'' \succeq_{lr} u''$ on $(0, p)$. Moreover, either $x_{k^*} - \hat{x}_{k^*} > 0$, then $s_{k^*} - \hat{s}_{k^*} > x_{k^*} - \hat{x}_{k^*}$ by $\{\delta_i\}_{i=j}^{2k^*}$ being strictly increasing, and hence $x_{k^*+1} - \hat{x}_{k^*+1} > s_{k^*} - \hat{s}_{k^*}$; or $x_{k^*} - \hat{x}_{k^*} \leq 0$, then we again have $x_{k^*+1} - \hat{x}_{k^*+1} > s_{k^*} - \hat{s}_{k^*}$.

Now we proceed by showing that if $\delta_m > \delta_{m-1}$, then $\delta_{m+1} > \delta_m$ for all $m = 2k^* + 1, \ldots, 2N - 5$.

Suppose $m$ is odd, let $\delta_m = x_l - \hat{x}_l$ for some $k^* + 1 \leq l < N - 1$ such that $x_l - \hat{x}_l > s_{l-1} - \hat{s}_{l-1}$. Since $s_l > p > \hat{s}_l$, we have

$$\begin{aligned}
x_l &= \phi(s_{l-1}, s_l) \\
&\leq \phi\left(\hat{s}_{l-1} + \max\{s_{l-1} - \hat{s}_{l-1}, 0\}, \hat{s}_l + s_l - \hat{s}_l\right) \\
&\leq \phi(\hat{s}_{l-1}, \hat{s}_l) + \max\{s_{l-1} - \hat{s}_{l-1}, s_l - \hat{s}_l\} \\
&= \hat{x}_l + \max\{s_{l-1} - \hat{s}_{l-1}, s_l - \hat{s}_l\}.
\end{aligned}$$

Since $x_l - \hat{x}_l > s_{l-1} - \hat{s}_{l-1}$, it must hold that $s_l - \hat{s}_l > x_l - \hat{x}_l$.

Suppose $m$ is even, let $\delta_m = s_l - \hat{s}_l$ for some $k^* + 1 \leq l < N - 2$ such that $s_l - \hat{s}_l > x_l - \hat{x}_l$. Then

$$\begin{aligned}
s_l &= \mu(x_l, x_{l+1}) \\
&= \mu(\hat{x}_l + x_l - \hat{x}_l, \hat{x}_{l+1} + x_{l+1} - \hat{x}_{l+1}) \\
&\leq \mu(\hat{x}_l, \hat{x}_{l+1}) + \max\{x_l - \hat{x}_l, x_{l+1} - \hat{x}_{l+1}\} \\
&\leq \hat{\mu}(\hat{x}_l, \hat{x}_{l+1}) + \max\{x_l - \hat{x}_l, x_{l+1} - \hat{x}_{l+1}\} \\
&= \hat{s}_l + \max\{x_l - \hat{x}_l, x_{l+1} - \hat{x}_{l+1}\}.
\end{aligned}$$

Since $s_l - \hat{s}_l > x_l - \hat{x}_l$, it must hold that $x_{l+1} - \hat{x}_{l+1} > s_l - \hat{s}_l$. Therefore, we conclude

that $\{\delta_i\}_{i=2k^*}^{2N-4}$ is strictly increasing.

*Case 4:* $p \leq \min\{\hat{x}_{N-1}, x_{N-1}\}$. Again, the subsequence $\{\delta_i\}_{i=1}^{2N-4}$ is single-crossing from below by the same argument in Proposition 5(b). It remains to be shown that $s_{N-1} < (\leq)\, \hat{s}_{N-1}$ (and thereby $x_N < (\leq)\, \hat{x}_N$) implies $x_{N-1} < (\leq)\, \hat{x}_{N-1}$. Observe that $p \leq \min\{\hat{x}_{N-1}, x_{N-1}\}$ and Claim 5(ii) imply $\hat{s}_{N-1} = \hat{\mu}(\hat{x}_{N-1}, \hat{x}_N) \leq \mu(\hat{x}_{N-1}, \hat{x}_N)$.

Suppose $s_{N-1} < \hat{s}_{N-1}$. If $\hat{x}_N \geq t$, then

$$\begin{aligned}
\hat{s}_{N-1} \leq \mu(\hat{x}_{N-1}, \hat{x}_N) &\leq \mu\left(\hat{x}_{N-1}, \max\{x_N, t\}\right) \\
&\leq \mu\left(\max\{\hat{x}_{N-1}, x_{N-1}\}, \max\{x_N, t\}\right) \\
&= \mu\left(x_{N-1} + \max\{\hat{x}_{N-1} - x_{N-1}, 0\}, x_N + \max\{0, t - x_N\}\right) \\
&\leq s_{N-1} + \max\left\{\hat{x}_{N-1} - x_{N-1}, 0, t - x_N\right\} \\
&= s_{N-1} + \max\{0, \hat{x}_{N-1} - x_{N-1}\}.
\end{aligned}$$

The second inequality follows from the monotonicity of $\mu(\hat{x}_{N-1}, \cdot)$ from Claim 3, which is applicable because $\hat{u}'(\hat{x}_N) - \hat{u}'(\hat{x}_{N-1}) > 0$, $u'' \succeq_{lr} \hat{u}''$ on $(p, 1)$, and by Claim 5(i), $u'(\hat{x}_N) - u'(\hat{x}_{N-1}) > 0$. The third inequality comes from the following observation: if $x_{N-1} > \hat{x}_{N-1}$, then because $u'(\max\{x_N, t\}) - u'(x_{N-1}) > 0$, $\mu(\hat{x}_{N-1}, \max\{x_N, t\}) < \mu(x_{N-1}, \max\{x_N, t\})$. The fourth inequality holds by Lemma 2 if $t \geq x_N$, or by Lemma 3 if $t < x_N$. The last equality holds because $\hat{s}_{N-1} - s_{N-1} > \hat{x}_N - x_N \geq t - x_N$ if $t > x_N$; otherwise if $t \leq x_N$, the equality is trivially true. Hence, it must be the case that $\hat{x}_{N-1} - x_{N-1} > 0$, as desired.

If $t > \hat{x}_N > x_N$, then

$$\begin{aligned}
\hat{s}_{N-1} \leq \mu(\hat{x}_{N-1}, \hat{x}_N) &\leq \mu\left(\max\{\hat{x}_{N-1}, x_{N-1}\}, x_N + \hat{x}_N - x_N\right) \\
&= \mu\left(x_{N-1} + \max\{\hat{x}_{N-1} - x_{N-1}, 0\}, x_N + \hat{x}_N - x_N\right) \\
&\leq s_{N-1} + \max\{\hat{x}_{N-1} - x_{N-1}, \hat{x}_N - x_N\}.
\end{aligned}$$

By logconcavity, we have $\hat{x}_N - x_N < \hat{s}_{N-1} - s_{N-1}$. So it must be the case that $\hat{x}_{N-1} - x_{N-1} > \hat{s}_{N-1} - s_{N-1} > 0$.

On the other hand, suppose $s_{N-1} = \hat{s}_{N-1}$, and thereby $x_N = \hat{x}_N$. Then we have

$$\begin{aligned}
\hat{s}_{N-1} \leq \mu(\hat{x}_{N-1}, \hat{x}_N) &= \mu(\hat{x}_{N-1}, x_N) \\
&\leq \mu(\max\{\hat{x}_{N-1}, x_{N-1}\}, x_N) \\
&\leq \mu\left(x_{N-1} + \max\{\hat{x}_{N-1} - x_{N-1}, 0\}, x_N\right) \\
&\leq s_{N-1} + \max\{0, \hat{x}_{N-1} - x_{N-1}\}.
\end{aligned}$$

If $\hat{x}_{N-1} < x_{N-1}$, then the second inequality must be strict, a contradiction. Hence, it must be that $\hat{x}_{N-1} \geq x_{N-1}$. □