ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 314

Coarse Information Design

Qianjun Lyu Wing Suen Yimeng Zhang

June 2024 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Coarse Information Design*

Qianjun Lyu
University of Bonn

Wing Suen
University of Hong Kong

Yimeng Zhang
University of Hong Kong

May 24, 2024

Abstract: We study an information design problem with continuous state and discrete signal space. Under convex 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 conditional mean (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: dual expectations, scrutiny, S-shaped value function, coarse nonlinear pricing, energy efficiency ratings

JEL Classification: D81, D82, D83

*We are especially grateful to Gregorio Curello for his continuous support and insightful discussions. We also thank Sarah Auster, Yunus Aybas, Dirk Bergemann, Michael Greinecker, Li Hao, Ilwoo Hwang, Mathijs Janssen, Andreas Kleiner, Benny Moldovanu, Paula Onuchic, Ali Shourideh, Aloysius Siow, Peter Sørensen, Junze Sun, Dezső Szalay, Kun Zhang, Chen Zhao, and participants at University of Bonn, Chinese University of Hong Kong, National Taiwan University, National University of Singapore, University of Washington, the 2023 Stony Brook International Conference on Game Theory, EWMES 2023 and VIEE for helpful suggestion and comments on this topic. Lyu acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2126/1-390838866.

1. Introduction

Information is critical in decision making, whether in a single-agent environment or in a sender-receiver game where interests are not perfectly aligned. Oftentimes, information takes a discrete form. For example, teachers evaluate students by letter grades or on a pass/fail basis; Michelin Guide ranks the recommended restaurants up to three stars for excellence; online platforms often convey product reviews through coarse rating systems; and regulatory policies disclose efficiency and safety level of different products via discrete categories. In this paper, we take the discrete nature of information structure as given and study how to design information optimally subject to such a constraint.

Coarse information emerges naturally 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, for example, 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.

Formally we study a class of information design problems where the uncertain state θ 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 (π, Σ), where Σ is the signal space and $\pi(\sigma|\theta)$ is the probability of obtaining signal realization $\sigma \in \Sigma$ conditional on state θ. The receiver—who may be the designer herself—updates her belief based on the realized signal and then chooses an action optimally. The constraint we introduce in this paper is that Σ is restricted to be a finite set with a cardinality of at most N. Given that the state is continuous and the signals are finite, an experiment necessarily involves pooling across different states. How to allocate the limited "signal resources" then becomes a nontrivial question.

From a design perspective, an economically relevant question emerges: Which parts of the state space should receive *closer scrutiny* in the experiment relative to other parts? In this introduction, unless otherwise specified, we illustrate our findings under scenarios in which the sender 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). The insights extend to a broader class of value functions.

To elaborate on the notion of scrutiny, 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."

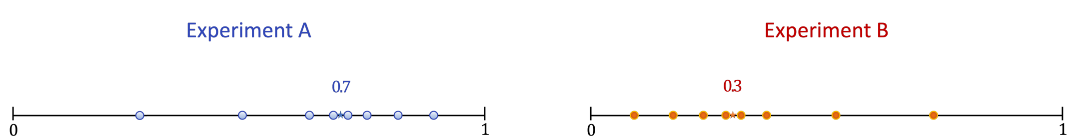

Figure 1: Experiment A gives close scrutiny to the state space near 0.7; experiment B gives close scrutiny to the state space near 0.3.

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, 1988) or preference evolution (Robson, 2001; Netzer, 2009). On the other hand, the designer should also give closer scrutiny to those states where the 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 in Section 3 characterizes the designer's optimal 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* information structure, 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 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 system of equations. The second key finding highlights an additional feature of the optimal information structure: 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 gradually diminishes for states located farther away (as illustrated in Figure 1). Such a pattern of information structure is commonly observed in various (mandatory) energy-efficiency rating standards, such as Energy Performance Certificates in housing and automobile markets worldwide.

We then leverage this dual expectations property to facilitate the comparative statics analysis with respect to both the prior distribution and the value function. 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, when either f or u'' becomes less variable according to the *uniform conditional variability order* (Whitt, 1985), then both the interval cutoff points $\{s_k\}_{k=1}^{N-1}$ and the posterior means $\{x_k\}_{k=1}^N$ become more compressed, in the sense that there exists an n^* such that s_k and x_k shift to the right for all $k < n^*$ and s_k and x_k shifts to the left for all $k > n^*$.

With convex value function, our framework has a natural application to mechanism design where the agent's payoff is linear in his 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.

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

In Section 5 and 6, we extend our analysis to S-shaped value functions and more general value functions to address scenarios where information can have negative value (i.e., $u'' < 0$ for some regions of the state space). With S-shaped value functions, the optimal experiment is still interval-partitional and therefore our dual expectations property continues to hold with slight modifications. With more general value functions, we show that the dual expectations property remains valid, incorporating the potential usage of a bi-pooling information structure through a bi-tangency condition similar to Dworczak and Martini (2019) and Arieli et al. (2023).

We provide further discussions of our model in Section 7. Section 7.1 studies the design of energy efficiency ratings system as an application of our model. 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 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. Recently, Curello and Sinander (2023) study the comparative statics of general linear persuasion problems. They 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 about the strategic use of pooling: how the sender designs an experiment to "concavify" her value function by strategically pooling information in a way 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 studies a constrained information design problem and examines how to effectively pool information in such environment.

The constraint imposed by coarse information on the operation of markets is discussed by Wilson (1989), McAfee (2002) and Hoppe et al. (2010). 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. Our paper generalizes the insight into a broader class of design problems. Harbaugh and Rasmusen (2018) study how coarse grading can increase information by raising the incentive to get costly certification; while Ostrovsky and Schwarz (2010) analyze coarse grading from an information design perspective. Lipman (2009) points out the efficiency loss from the vagueness of language, and Crémer et al. (2007) study the optimal 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. They discuss the implications of coarse language for the theory of the firm and organizational hierarchy. Applying the toolkit of Voronoi diagram, Jäger et al. (2011) study the behavior and stability of the optimal language by a distance-based criterion.

Recently, the literature on information design has begun to explore the limitation of communication and information channels. Onuchic and Ray (2023) and Mensch (2021) study information design subject to monotone information structures. Aybas and Turkel (2022) 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,[1] and allows the sender's payoff to depend arbitrarily on beliefs. In contrast, we focus specifically on linear persuasion problems and provide a general characterization for the optimal information structure.[2]

Furthermore, our paper is related to Hopenhayn and Saeedi (2022), which studies the optimal coarse ratings system that partitions a continuum of sellers with different qualities into a finite number of quality groups in a competitive environment. They characterize the social planner's optimal rating scheme and compare the rating schemes under different shapes of supply functions.[3] 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

[1] Le Treust and Tomala (2019) and Doval and Skreta (2022) develop powerful machinery to study persuasion problems with general constraints under finite states.

[2] Ivanov (2021) also addresses linear persuasion problems, but he allows the sender's payoff to depend on the information structure non-strategically, i.e., a non-strategic receiver acts upon both the realized signal and the information structure.

[3] Their model reduces to a linear persuasion problem only when the supply function is linear, in which case, the value function has a uniform curvature.

cutoffs shift up when the prior distribution shifts up without characterizing the solution. Instead, we use interim value functions as primitives and characterize the optimal information structure for general value functions.[4] Moreover, the dual relation between value function curvature and prior density allows us to develop comparative statics with respect to the payoffs.

Similar finite constraints are also studied in mechanism design problems with transfers. Bergemann et al. (2012, 2021) and Wong (2014) introduce the discreteness constraint in a standard nonlinear pricing problem and discuss the payoff bound of the finite menu and asymptotic convergence rate. In Section 4, we establish a connection between these two problems. Recently, Bergemann et al. (2022) study the role of information design in a classic nonlinear pricing problem, demonstrating that the unconstrained optimal information structure is endogenously coarse.

This paper is related to the classic cheap talk model of Crawford and Sobel (1982), where information coarsening emerges endogenously from incentive compatibility. Comparative statics in cheap talk models regarding the prior distribution have been explored in Szalay (2012), Chen and Gordon (2015), and Deimen and Szalay (2023). Recently, the literature on cheap talk begins to explore the sender's optimal design of information using the majorization approach (Ivanov, 2010; Lou, 2023; Kreutzkamp, 2023). In these models, the optimal unconstrained information structure may adopt a bi-pooling policy even under convex value functions. In Section 7.3, we discuss the connection between our research and the cheap talk game more thoroughly.

The coarse information design problem addressed in this paper resembles quantization in information theory (Gray and Neuhoff, 1988; Mease and Nair, 2006). These two classes of problems are not nested. The main difference is that quantization algorithms typically maximize the expected "similarity" within a group based on a distance metric, e.g., minimizing the expected distance between the realized values and the centroid within a cluster, which in general, cannot be transformed into a well-defined interim value function with a domain of posterior means. One exception is when the loss function is quadratic,[5] in which case the quantization problem becomes equivalent to our problem with uniform value function curvature.

[4]Though we consider a linear environment, the two problems do not nest each other as a general value function can violate single-crossing and therefore submodularity of cell functions .

[5]The least-squares criterion, or mean-squared-error model is the workhorse in the quantization literature. For a classical algorithm, see Lloyd (1982).

2. Model

Consider an information design model where the payoff-relevant state θ 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 Σ 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,[6] 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 can be partitioned into finitely many interval regions such that u is either convex or concave on each interval.[7] Since the only relevant information is the posterior mean, it is without loss of generality to assume 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 the finiteness constraints on the signal space Σ. In particular, we require Σ to contain no more than N elements. Consequently, the distribution G of posterior means can only have a finite support.

[6]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.

[7]Henceforth, we refer to the class of value functions that satisfy such conditions as *regular* value functions, as introduced by Dworczak and Martini (2019).

We can then write the designer's optimization program as:

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

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

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

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

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

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

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

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 case I_F. Hence, it yields a higher payoff if providing local information is very rewarding, namely, associated with higher positive u''. Similarly, a lower I_G is closer to the null information case I_{F_0}, and is more preferable to the designer if u'' is negative. Obviously, if u is convex within some region, a locally upward shift of I_G increases the designer's expected payoff, provided that such 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 ICPL class.

2.1. Preliminaries

In this section, we elaborate on how to construct an integral distribution I_G with finitely many kinks $|\mathscr{K}_{I_G}| \leq N$ and its connection to the implied signal structure.

Consider the following class of information structures. Partition the state space into $K \leq N$ subintervals with a sequence of cutoff points, $0 = s_0 < s_1 < \ldots < s_K = 1$. Within each subinterval (s_{k-1}, s_k) for $k \in \{1, \ldots, K\}$,[8] there are $J_k \geq 1$ signals induced. Denote the induced posterior means by $\{x_k^j\}_{j=1}^{J_k}$, such that $\sum_{k=1}^{K} J_k \leq N$. Following Kleiner et al. (2021) and Arieli et al. (2023), we call this class of information structures *J-pooling* policies if $\max\{J_1, \ldots, J_K\} = J$. Since K can be equal to 1, any discrete information structure satisfying (D) is a J-pooling policy for some value of J. In the special case of $J = 2$, we call it a *bi-pooling* structure;[9] and if $J = 1$, we call it an *interval-partitional* structure. The following lemma identifies the corresponding J-pooling information structure for every ICPL function.

Lemma 1. *For any $I_G \in$ ICPL, the corresponding distribution G can be induced by a J-pooling policy where*

(i) $\{s_k\}_{k=0}^{K}$ are the points where I_G is tangent to I_F;

(ii) for each $k = 1, \ldots, K$, $\{x_k^j\}_{j=1}^{J_k}$ are the kink points of $\mathscr{K}_{I_G}$ that lie in (s_{k-1}, s_k).

The tangency points between I_G and I_F are the partition cutoffs and the kink points of I_G are the induced posterior means under a J-pooling policy. For any J-pooling policy with $J > 1$, some segments of the corresponding I_G curve is strictly below I_F. For an interval-partitional structure, on the other hand, every segment of I_G has exactly one tangency point with I_F. See Figure 2.

An important implication of Lemma 1 is that if I_G has only one kink point between two adjacent tangency points s_{k-1} and s_k, then the kink point must be the posterior mean conditional on the subinterval (s_{k-1}, s_k). For example, this property holds trivially for I_{F_0}, where the two adjacent tangency points are 0 and 1, and the only kink point is the prior mean $\mathbb{E}_F[\theta]$. The following corollary suggests that this simple property holds generally for any *subinterval*.

[8]Throughout the paper, we always use the open subinterval to characterize the partition for consistency. It is innocuous when the distribution is atomless and can be modified to study distributions with atoms.

[9]Arieli et al. (2023) show that any bi-pooling policy over an interval can alternatively be implemented by some pure nested-interval policy and by some potentially mixed FOSD-ranked monotone policy.

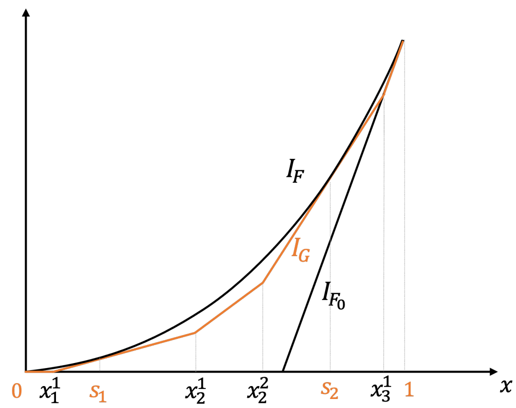

Figure 2: The integral distribution function I_G is tangent to I_F at $\{s_0, s_1, s_2, s_3\}$, where $s_0 = 0$ and $s_3 = 1$. The J-pooling policy derived from I_G sends only one signal each in the first and third intervals. If $\theta \in (s_1, s_2)$, the information structure generates two distinct signals, inducing posterior mean x_2^1 or x_2^2.

Corollary 1. *For any $I_G \in$ ICPL, if $J_k = 1$ for some k, then $x_k^1 = \mathbb{E}_F[\theta \mid \theta \in (s_k, s_{k+1})]$.*

The next result establishes the existence of solution and shows that it suffices to consider a much smaller subset of ICPL functions, namely, the bi-pooling information structure. The proof is in the spirit of Kleiner et al. (2021) and Arieli et al. (2023). Some minor modification is required, because although the objective function is still a linear functional, the set of distributions that satisfy the discreteness constraint (D) is not convex. It remains to show that the corresponding feasible set of I_G is compact and the objective function in (2) is continuous, which ensures the existence of an optimal solution.[10]

Lemma 2. *For any feasible J-pooling policy with $J \geq 3$, there exists a feasible bi-pooling policy that produces a (weakly) higher value of the objective function (2). Moreover, the maximum of (2) can be attained by some bi-pooling policy.*

3. Convex Value Functions

In this section, we characterize the optimal information design for convex value functions, where sender and receiver's preferences are sufficiently aligned such that more

[10]We thank Yunus Aybas and Eray Turkel for sharing an earlier version of their working paper that provides detailed proof of the existence of the solution. Although they impose the affine-closedness assumption on the value function, their proof applies to general regular value functions.

information is strictly beneficial. The discreteness constraint raises an important question regarding how to pool information efficiently when the number of disposable "signal resources" is budgeted. 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. It can also be adapted to study contract design with quasi-linear preferences and finiteness constraint on the contract space.

Naturally, with convex value functions, the designer can increase her payoff by raising the informativeness of the experiment (i.e., by raising I_G). Consequently, it is without loss of generality to focus on interval-partitional structures.[11] Moreover the optimal experiment should always fully utilize her "signal resources" by setting $|\mathscr{K}_{I_G}| = N$. In other words, the "information-maximizing" designer always exhausts his signal budget.

Lemma 3. *Suppose u is convex. The optimal information structure G can be implemented by an interval-partitional structure with $|\mathscr{K}_{I_G}| = N$.*

We provide a simple proof in the Appendix, using the representation of integral distribution. Note that Lemma 3 remains valid for *any* convex function u even when it is non-differentiable.

For any $0 \leq a < b \leq 1$, let

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

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" $\mu(\cdot,\cdot)$ is well-defined. Our main characterization of the optimal information structure features a *dual relation* between the prior distribution f and the curvature distribution u''.

[11] The optimality of interval-partitional policies with a finite signal space and a convex value function is well understood in the literature. Ivanov (2021) proves this in an equivalent setting as ours. Specifically, for fixed marginal distribution of the signals, he shows that any information structure that is not interval-partitional is majorized by an interval-partitional information structure. Optimality of interval-partitional structures is also observed in other applications, such as Dow (1991) in a search model, Wilson (1989) and McAfee (2002) in matching models, and Sørensen (1996) and Smith et al. (2021) in social learning models.

Theorem 1. *Suppose u is strictly convex. The optimal information structure G, characterized by $\{s_k\}_{k=0}^N$ and $\{x_k\}_{k=1}^N$, must satisfy:*

$$x_k = \phi(s_{k-1}, s_k) \qquad \textit{for } k = 1, \dots, N; \tag{CE-F}$$

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

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

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

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, for any interval structure G, the corresponding $\underline{u}_G(x)$ is piecewise affine with increasing slopes.

Under an optimal interval partition G^*, the minorant function must be continuous at the interval cutoff points, i.e., $\underline{u}_{G^*}(s_k^-) = \underline{u}_{G^*}(s_k^+)$ for all $k = 1, \dots, N-1$. Suppose to the contrary, say, $\underline{u}_{G^*}(s_k^-) < \underline{u}_{G^*}(s_k^+)$, as in the left panel of Figure 3. Then we can construct another function,

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

Let $\hat{s}_k$ be the solution that $u(x_k) + u'(x_k)(x - x_k) = u(x_{k+1}) + u'(x_{k+1})(x - x_{k+1})$. It is immediate that $\hat{s}_k \in (s_{k-1}, s_k)$, as depicted in the right panel of Figure 3 where the two pieces meet. By construction $\underline{\hat{u}}(x)$ is continuous at $x = \hat{s}_k$. 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 another interval structure $\hat{G}$, characterized by the partitional cutoffs $\{s_0, \dots, s_{k-1}, \hat{s}_k, s_{k+1}, \dots, s_N\}$. The following inequalities show that this interval structure

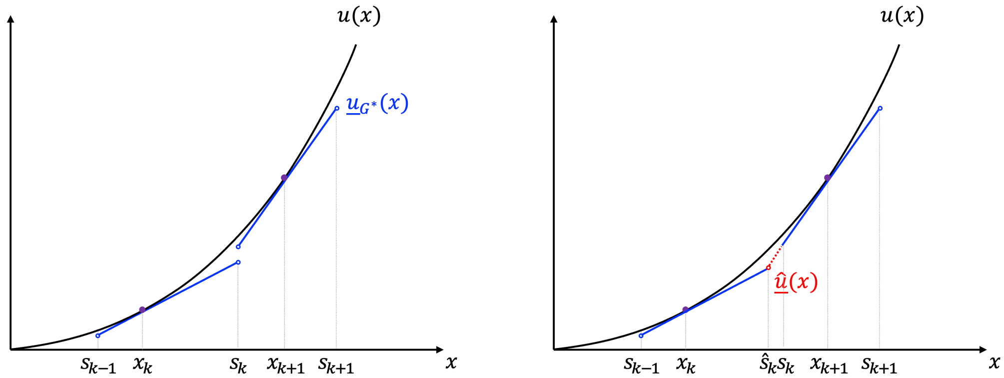

Figure 3: Suboptimality of discontinuous minorant function

$\hat{G}$ generates higher expected payoff than G^*.

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

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.[12] The inequality on the second line follows because $\underline{\hat{u}}$ is everywhere above $\underline{u}_{G^*}$. The subsequent equality again comes from the piecewise linearity of $\underline{\hat{u}}$. 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.

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 function" of the "distribution" u', then u bears a similar relationship to $\underline{u}_{G^*}$ as I_F bears to I_G. Consequently, Corollary 1 implies that

[12] Though the values of $\underline{u}_G(x)$ at the interval cutoffs are not assigned at this stage, this equality remains true because F put zero measure on this countable set.

$s_k = \mathbb{E}_{u'}[\theta \mid \theta \in (x_k, x_{k+1})]$ for all $k = 1, \ldots, N-1$. □

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 conditional on the interval (x_k, x_{k+1}) with respect to the distribution u'.[13]

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), is representative of the designer's payoff when an action is taken upon the realization of the signal x_k.[14] Therefore, under the optimal interval partition, 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 from categorizing the cutoff state s_k into either the upper subinterval or the lower subinterval under the optimal information structure.

Thus we can identify a *minorant function* $\underline{u}$,[15] which is continuous, convex and piecewise linear such that $\underline{u}$ is tangent to the value function u at the support of optimal information structure $\{x_k\}_{k=1}^{N}$. Moreover, the set of kink points of $\underline{u}$ are the cutoffs of the optimal interval partition structure $\{s_k\}_{k=1}^{N-1}$. Figure 4 puts together the minorant function and the integral distribution of an interval partitions. The minorant function $\underline{u}$ exhibits a geometric relationship with u akin to the relationship between I_G and I_F,

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

[14] Specifically, 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, the subgradient does not necessarily capture the designer's actual state-wise payoff. Nevertheless, due to the linear payoff structure, it generates the same expected payoff on each partitioned subinterval as the designer's actual payoff function.

[15] The minorant function for the optimal interval partition $\underline{u}_{G^*}$ is reminiscent of the "price function" introduced by Dworczak and Martini (2019) and Kolotilin (2018). One difference is that the minorant function we obtain is everywhere below the value function, as opposed to the price function. This is because the existence and the continuity of the price function are guaranteed by the strong duality in the unconstrained information design problem, while in our problem, the discreteness constraint breaks this strong duality. Nevertheless, one can view the minorant function as the dual payoff (c-transform in Kantorovich duality) characterized using the techniques from optimal transport (Galichon, 2016; Smolin and Yamashita, 2022; Dworczak and Kolotilin, 2023) if the action space is restricted to the set of actions induced under the optimal interval partition. It is also related to the virtual utility specified in Mensch (2021).

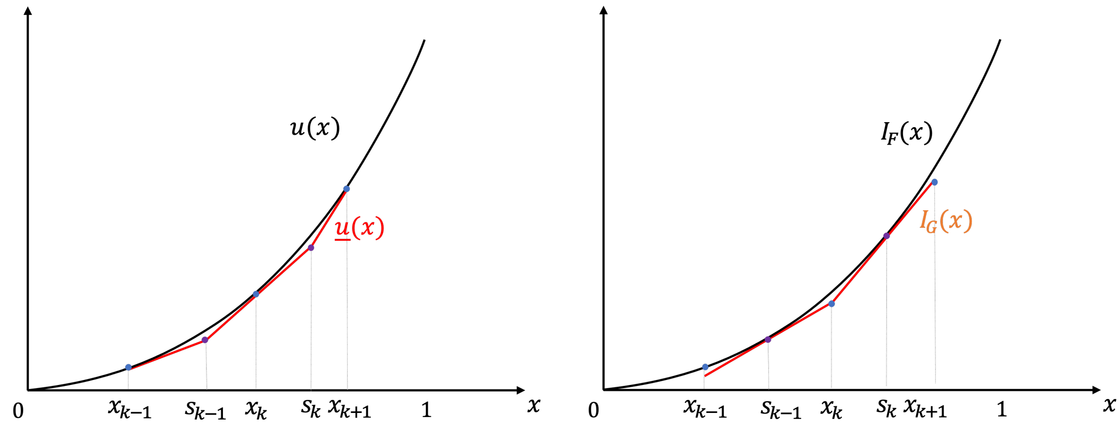

Figure 4: Dual expectations.

which leads to the property of dual expectations.

In fact, 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, provided that the constrained-optimal information structure is indeed interval-partitional (see Section 5 and Section 6).

Theorem 1 provides a necessary condition for the optimal information structure. The next result identifies environments under which the equation system has a unique solution and therefore becomes also sufficient. Notice that logconcave distributions contain a wide range of distributions and are widely adopted in information economics.

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 information structure.*

The proof relies on the following property of logconcave densities.[16] Because we will further exploit this property for subsequent results, we state it here for easier reference.

Lemma 4 (Mease and Nair (2006); Szalay (2012)). *If f is logconcave, then for any $0 \le a < b \le 1$ and any $\varepsilon \ge 0$ such that $[a+\varepsilon, b+\varepsilon] \subseteq [0,1]$,*

$$\phi(a+\varepsilon, b+\varepsilon) \le \phi(a,b) + \varepsilon. \tag{4}$$

[16]The role of this property is explored in other applications, for example, quantization models (Mease and Nair, 2006) and cheap talk models (Szalay, 2012; Deimen and Szalay, 2019).

Obviously, a similar property as stated in Lemma 4 holds for the $\mu(\cdot,\cdot)$ function if u'' is logconcave. When both f and u'' are logconcave, we refer to that as a *logconcave environment* in the following analysis.

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 structure, we say a subinterval of the state space receives *closer scrutiny* than another if the width of this subinterval is smaller than that of another subinterval. 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 $k = 1,\dots,N$, and the distance between two adjacent posterior means by $d_k = x_{k+1} - x_k$ for $k = 1,\dots,N-1$. Let $\{\Delta_i\}_{i=1}^{2N-1}$ be the interleaved sequence $\{w_1, d_1, w_2, \dots, d_{N-1}, w_N\}$. 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.

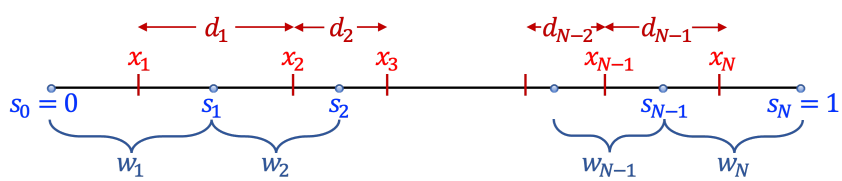

Figure 5: The interleaved sequence of Δ_i.

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 $i = 2,\dots,2N-2$. There are two cases to consider.

If i is even, then starting with the hypothesis that $d_k \geq w_k$ for some $k \leq N-1$, we want to show $w_{k+1} \geq d_k$. Suppose to the contrary that $d_k > w_{k+1}$. 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 ϕ is strictly monotone in each argument. The second inequality comes from Lemma 4. This is a contradiction.

If i is odd, then starting with the hypothesis that $w_k \geq d_{k-1}$ for some $k = 2, \ldots, N-1$, we want to show $d_k \geq w_k$. Suppose to the contrary that $w_k > d_k$. Applying the 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 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—in the sense that 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, 1998), the two ends of the state space have lower probability mass and smaller curvature. 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.[17] The following result confirms this intuition and extends it to a more general environment

[17] It does not necessarily contain the mode because u'' can be highly asymmetric about the mode, e.g., very flat on one side and very steep on the other side.

where the two functions have different modes.

Proposition 2. *(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 $i^* \in \{k-1, \ldots, j+1\}$.

To elaborate further, we provide an example below and will refer to it in later sections.

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 designs an experiment to acquire information about θ. After the signals are observed but before the purchase decision is made, a cost shock η is realized, where η is distributed with density h on $[\underline{\eta}, \overline{\eta}]$, where $\underline{\eta} \leq -p$. 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 $u''(x) = h(x-p)$.

If h is uniform and the prior density f has mode $m_f = 0.7$, then Proposition 2(i) suggests that the optimal experiment will give close scrutiny to states near 0.7, because such states have greater probability of occurring. The optimal experiment will resemble Experiment A in Figure 1 shown in the introduction. On the other hand, if f is uniform and h is single-peaked with mode 0 and $p = 0.3$, then u'' is single-peaked at $m_u = 0.3$. The purchase decision is particularly sensitive to belief when the posterior mean is near 0.3. The optimal experiment will resemble Experiment B in Figure 1. When both f and u'' are single-peaked, Proposition 2(iii) suggests that the optimal experiment will balance these two considerations. The exact location of the center of scrutiny will depend on details of f and h, but it is unlikely to be outside the interval $(0.3, 0.7)$.[18]

3.2. Comparative statics

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

[18]Proposition 2(iii) does not exclude the possibility that $i^* = k-1$ or $i^* = j+1$, in which case the center of scrutiny will fall outside $(0.3, 0.7)$. This may occur, for example, if f is almost flat and h is highly asymmetric near its peak.

Proposition 3. *In a logconcave environment, if either (a) the prior density changes from f to $\hat{f}$, with $\hat{f}(\cdot)/f(\cdot)$ being increasing; or (b) the value function changes from u to $\hat{u}$, with $\hat{u}''(\cdot)/u''(\cdot)$ being increasing, then all the interval cutoff points $\{s_k\}_{k=1}^{N-1}$ will increase and all the induced posterior means $\{x_k\}_{k=1}^{N}$ will also increase.*

Proof. We prove the case where f changes to $\hat{f}$. The proof for the change of value function is analogous. Since F increases according to the likelihood ratio order, the conditional distribution F on any interval $(a,b) \subseteq [0,1]$ also increases, which implies that the conditional mean $\phi(a,b)$ becomes higher for any given a and b. In a logconcave environment the optimal $(x_1,\dots,x_N,s_1,\dots,s_{N-1})$ is the unique fixed point of Γ, the mapping corresponding to the right-hand-side of the equation system (CE-F) and (CE-u'). Since a likelihood ratio increase in F raises Γ, and Γ is monotone, a standard result in monotone comparative statics (Topkis, 1998) establishes that any fixed point under $\hat{f}$ is larger than the unique fixed point under f. □

The proposition 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.[19]

Part (b) of Proposition 3 is more novel to 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 an increasing and convex function ψ such that $\hat{u}'(x) = \psi(u'(x))$.[20] This condition implies that the designer's value function under 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.

We may revisit the purchase decision example through the lens of Proposition 3.

[19] Tian (2022) shows this part of the result in a more general environment where uniqueness is not required. 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.

[20] Hopenhayn and Saeedi (2022) touch upon this intuition by comparing the solutions when the value function changes from a linear function to a convex function. Nevertheless, the two results do not nest each other as their model can only be treated as a linear persuasion problem when the supply function is linear.

Suppose it is the *seller* of the good rather than the consumer who is designing the information structure. The price of the good is fixed at p and the seller wants to maximize the probability of making a sale. His interim value function is $\hat{u}(x) = pH(x-p)$, and therefore $\hat{u}''(x) = ph'(x-p)$. If the density h is nondecreasing, then $\hat{u}$ is convex. Furthermore, if h is logconcave, then $\hat{u}''(x)/u''(x) = ph'(x-p)/h(x-p)$ is decreasing in x. Our result shows that the induced posterior means $\{\hat{x}_k\}_{k=1}^N$ under the seller-optimal information structure are lower than the ones under the buyer-optimal information structure.[21] Intuitively, the interests of the buyer and the seller are not aligned: the buyer cares about consumer surplus, which can be sensitive to high-value realizations as she is likely to consume the good in such states, while the seller cares about the probability of trade, which is less sensitive to high states as the value of the good already well exceeds the price. The misalignment in constrained signal allocation driven by different u'' is not captured in the standard unconstrained information design problem as the unconstrained solution always entails full disclosure.

Next, we investigate how the optimal information structure changes when either f or u'' becomes less variable. We say that a distribution with density $\hat{f}$ is *uniformly less variable than* one with density f if $\hat{f}(\cdot)/f(\cdot)$ is unimodal (Whitt, 1985; Shaked and Shanthikumar, 2007).[22] 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, while the means of the two distributions are not necessarily the same.

Proposition 4. *In a logconcave environment, if either (a) $\hat{f}$ is uniformly less variable than f; or (b) $\hat{u}''$ is uniformly less variable than u'', then the corresponding 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 $s_{k^*} \leq p < s_{k^*+1}$. Define

[21] Later we show in Section 5 that this conclusion remains valid when h is logconcave but nonmonotone, which causes the seller's value function to become S-shaped.

[22] Define $S(\cdot)$ be the number of sign changes for a function. Whitt (1985) shows that $\hat{f}$ is uniformly less variable than f if and only if the sign change counter $S(f - c\hat{f}) \leq 2$ for all $c > 0$, with equality for $c = 1$; and in the case of equality, the sign sequence is $+, -, +$. This is a stronger notion than the standard variability ordering (convex order), as it implies that standard variability ordering is preserved conditional on any subset of the state space.

the sequence,

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

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

Let j be the first non-negative term in the subsequence $\{\delta_j\}_{j=1}^{2k^*+1}$. We want to show that $\delta_j \geq 0$ implies $\delta_{j+1} \geq 0$. Suppose j is odd; that is, $\delta_j = x_k - \hat{x}_k \geq 0$ and $s_{k-1}-\hat{s}_{k-1} < 0$ for some $k \in \{2, \ldots, k^*\}$ (and if $j=1$, then $k=1$ and $s_{k-1}-\hat{s}_{k-1}=0$). Then, letting $\hat{\phi}(\cdot)$ be the conditional mean function under $\hat{f}$, we have

$$x_k - \hat{x}_k = \phi(s_{k-1}, s_k) - \hat{\phi}(\hat{s}_{k-1}, \hat{s}_k) \leq \phi(s_{k-1}, s_k) - \phi(\hat{s}_{k-1}, \hat{s}_k) \leq \max\{s_{k-1}-\hat{s}_{k-1}, s_k - \hat{s}_k\},$$

where the first inequality follows because $\hat{f}(\cdot)/f(\cdot)$ is increasing on $[0,p]$, and the last inequality follows from Lemma 4. Since $s_{k-1}-\hat{s}_{k-1} \leq 0$, the above inequality implies $\delta_{j+1} = s_k - \hat{s}_k \geq \delta_j \geq 0$. Similarly,

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

Since $s_k - \hat{s}_k \geq x_k - \hat{x}_k$, the above inequality implies $\delta_{j+2} = x_{k+1} - \hat{x}_{k+1} \geq \delta_{j+1} \geq 0$. By induction, $\{\delta_{j+1}, \ldots, \delta_{2k^*+1}\}$ are all non-negative. Similar reasoning applies to the case where j is even. Thus, we establish that $\{\delta_j\}_{j=1}^{2k^*+1}$ is single-crossing from below.

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

$$\{\hat{x}_N - x_N, \hat{s}_{N-1} - s_{N-1}, \hat{x}_{N-1} - x_{N-1}, \ldots \hat{x}_{k^*+2} - x_{k^*+2}, \hat{s}_{k^*+1} - s_{k^*+1}, \hat{x}_{k^*+1} - x_{k^*+1}\}.$$

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 $\{\delta_j\}_{j=1}^{2k^*+1}$ and $\{\delta_j\}_{j=2k^*+1}^{2N-1}$ establishes the desired result. □

An implication of Proposition 4 is that $\{\hat{x}_k\}_{k=1}^{N}$ crosses $\{x_k\}_{k=1}^{N}$ at most once and from above.[23] If they indeed cross, there exists a n^* such that x_k increases for all $k < n^*$

[23]Proposition 4 does not exclude the possibility that these sequences do not cross. For example, if $\hat{f}(\cdot)/f(\cdot)$ is unimodal with a peak p close to 1, then changing the density from f to $\hat{f}$ is almost like a likelihood ratio increase in the prior distribution, and it is possible that all interval cutoffs and induced

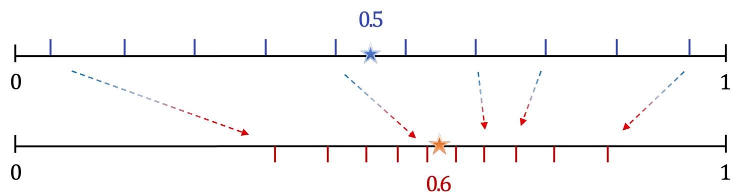

Figure 6: Comparative statics under uniform variability order.

and decreases for all $k \geq n^*$. See Figure 6 for an illustration. This figure shows that a uniformly less variable density will cause the induced posterior means to be more compressed.[24]

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 his 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 (Bergemann et al. (2012, 2021) and Wong (2014)).[25] 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 his preference type $\theta \in [0,1]$, distributed according to F with a continuous density function f. 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 θ 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 is $c(q)$. For simplicity, we assume v and c are twice differentiable with $v' > 0$, $v'' < 0$, $c' > 0$, $c'' > 0$, 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., $|\Sigma| \leq N$, as introduced in Bergemann et al. (2012, 2021) and Wong (2014).[26] Throughout our discussion, we

posterior means increase.

[24]Deimen and Szalay (2023) explore the implications of uniform variability order in cheap talk applications.

[25]Bolton and Dewatripont (2005) provides several applications in contract theory that inherits similar structures, e.g. credit rationing, optimal taxation and monopoly regulation.

[26]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

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 can focus on 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,

$$\int_0^1 \left[v\left(q\left(\theta\right)\right)\left(\theta - \frac{1-F\left(\theta\right)}{f\left(\theta\right)}\right) - c\left(q(\theta)\right)\right] \mathrm{d}F\left(\theta\right). \tag{5}$$

Define $\varphi := \theta - (1-F(\theta))/f(\theta)$ to be the virtual valuation, and let H be its associated distribution on $[0,1]$. Throughout the section, we assume that virtual valuation is strictly increasing in type and H admits a continuous density function h. Henceforth, we directly work with virtual valuation.

With slight abuse of notation, let $q(\varphi)$ represent the quality allocation to virtual type φ 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:

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

The program is essentially to pick a monotone, finite allocation rule $q(\varphi)$ to maximize the expected virtual surplus.[27] 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, $0 = s_0 < s_1 < \ldots < s_{N+1} = 1$. Within each type segment (s_{k-1}, s_k) where $k = 1, \ldots, N+1$, the allocation assigns q_k uniformly.

To leverage the dual expectations property in this coarse mechanism design problem,

when N grows large. Bergemann et al. (2021) study a linear-quadratic version of this model using the quantization approach.

[27] Exclusion of consumers is captured by $q = 0$. Therefore an N-option menu is equivalent to an $N+1$-valued allocation rule.

let

$$\pi(\varphi) := \max_{q \geq 0} \ \varphi v(q) - c(q), \tag{6}$$

be the pointwise maximal virtual surplus on $\varphi \in [0,1]$, and let $q^*(\varphi)$ be the unconstrained optimal allocation to (6). Because the pointwise profit is linear in the consumer's virtual type, $\pi(\varphi)$ is increasing and convex on $[0,1]$. Moreover, for any $0 \leq a < b \leq 1$, we define[28]

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

Proposition 5. *The solution to program (CMD), characterized by* $\{s_k\}_{k=0}^{N+1}$ *and* $\{q_k\}_{k=1}^{N+1}$*, must satisfy:*

$$x_k = \phi^m(s_{k-1}, s_k) \qquad \text{for } k = 2, \ldots, N+1; \tag{CE-H}$$
$$s_k = \mu^m(x_k, x_{k+1}) \qquad \text{for } k = 1, \ldots, N; \tag{CE-π'}$$
$$q_k = q^*(x_k) \qquad \text{for } k = 2, \ldots, N+1. \tag{Quality}$$

with $x_1 = 0$ *and* $q_1 = 0$.

Proof. By definition, $\underline{\pi}(\varphi)$ is weakly below $\pi(\varphi)$ for any $\varphi \in [0,1]$. Our goal is to show that given π and H, the optimal $q(\varphi)$ in the coarse mechanism design problem 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 $k = 1, \ldots, N$ and thereby is also convex; and (b) on each type segment (s_{k-1}, s_k) with $k = 2, \ldots, N+1$, it is tangent to π at the conditional expectation of the corresponding segment x_k.

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

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

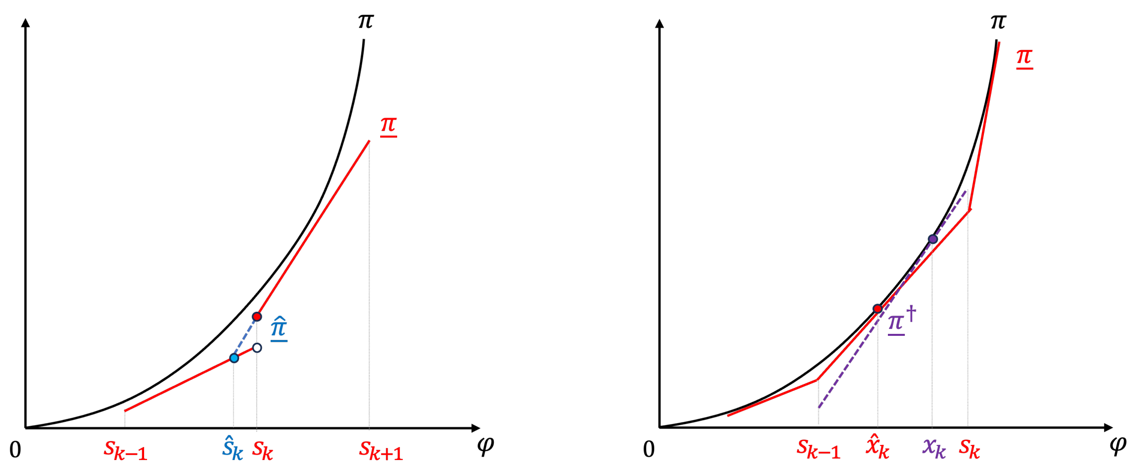

Figure 7: Properties of optimal $\underline{\pi}$.

construct another incentive-compatible allocation $\hat{q}(\cdot)$, which supplies higher quality to consumers in $(\hat{s}_k, s_k)$ while the allocation to other consumers are unchanged. The corresponding $\hat{\underline{\pi}}$ function (shown by the blue dotted line for $\varphi \in (\hat{s}_k, s_k)$ in the left panel of Figure 7) for this modified allocation is everywhere higher than $\underline{\pi}$ for the original allocation. This 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 π at some interior point, for otherwise the seller can adjust the allocation to raise $\underline{\pi}$ and increase the profit. Hence, it remains to prove that the optimal allocation for the segment (s_{k-1}, s_k) must equal the optimal allocation for the average virtual type of that segment, i.e., $q_k = q^*(x_k)$. Suppose not, and instead $\underline{\pi}$ is tangent to π 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 π at x_k (shown as the purple dotted piece), and is otherwise equal to the original q. Then,

$$\int_{s_{k-1}}^{s_k} \underline{\pi}(\varphi)\,\mathrm{d}H(\varphi) = [H(s_k) - H(s_{k-1})]\big[\pi(\hat{x}_k) + \pi'(\hat{x}_k)(x_k - \hat{x}_k)\big]$$

$$< [H(s_k) - H(s_{k-1})]\,\pi(x_k) = \int_{s_{k-1}}^{s_k} \underline{\pi}^\dagger\,\mathrm{d}H(\varphi),$$

where the inequality follows from the strict convexity of π. This contradicts the optimality of the allocation corresponding to the original $\underline{\pi}$ function. Therefore $q_k = q^*(x_k)$ must hold.

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

Overall, 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 $\varphi \in (0, s_1)$, together with those with negative φ, 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 π'. Hence, the dual expectations characterization is still valid for the finite-menu nonlinear pricing problem, and the remaining results in Section 3 continue to hold (where the prior distribution is interpreted as the distribution of virtual type).

Example 2. Suppose $v(q) = q^\beta$ and $c(q) = \gamma q$ for $\beta < 1$. Then the unconstrained optimal allocation for the seller is $q^*(\varphi) = (\beta\varphi/\gamma)^{1/(1-\beta)}$, and the "value function" for this problem 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 $q > 1$.[29] The value function $\hat{\pi}(\varphi)$ corresponding to the higher valuation of quality satisfies the property that $\hat{\pi}''(\cdot)/\pi''(\cdot)$ 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 3, all the interval cutoff points will move to the right when the benefit increases from $v(q)$ to $\hat{v}(q)$. In particular, we 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 good has become more valuable to all consumers. Intuitively, an increase in β not only implies higher valuation for quality, but also means 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. Note that this is a novel feature for finite menu design, as the coverage

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

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 interim 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.

A function u is *S-shaped* if there exists some $\hat{x} \in (0,1)$ such that u is convex on $(0,\hat{x})$ and concave on $(\hat{x},1)$. In other words, u'' is single-crossing from above.[30] S-shaped value functions have been extensively studied in persuasion literature because they capture a range of economic applications. The unconstrained optimal information structure features "upper-censorship," namely the optimal experiment without the discreteness constraint will reveal full information on $[0,s^*)$ and coarsen information by pooling all states in $(s^*,1]$ for some $s^* \leq \hat{x}$ (Kolotilin et al., 2022).[31] The optimal unconstrained information structure G^* has an integral distribution I_{G^*} that coincides with I_F everywhere to the left of s^*; then becomes linear with slope $F(s^*)$ to the right of s^* until it intersects with I_{F_0}, and then coincides with I_{F_0} to the right of the intersection point. See the blue solid curve in Figure 8(b).

Proposition 6. *Suppose u is S-shaped and the optimal information structures with and without the discreteness constraint (D) are G and G^*, respectively, where G^* features "upper-censorship" with $s^* \in (0,\hat{x}]$. Then,*

(i) G has an interval-partitional structure with $|\mathcal{K}_{I_G}| = N$;

(ii) G is less informative than G^*.

Proof. (i) Suppose the optimal solution is a bi-pooling policy G such that two posterior means x_k^1 and x_k^2 are induced within the interval (s_{k-1},s_k). If $\hat{x} \geq x_k^2$ or $\hat{x} \leq x_k^1$, we can use the same argument as in the proof of Lemma 3 to show that the policy is suboptimal. If $\hat{x} \in (x_k^1, x_k^2)$, we can construct another integral distribution function $\hat{I}$ by the following procedure. First, clockwise rotate the piece of I_G on (x_k^1, x_k^2) around the point $(\hat{x}, I_G(\hat{x}))$.

[30] The analysis in this section also extends to the case where u'' is single-crossing from below.

[31] It is possible that under some prior distributions, the optimal information structure contains no information, i.e., $s^* = 0$.

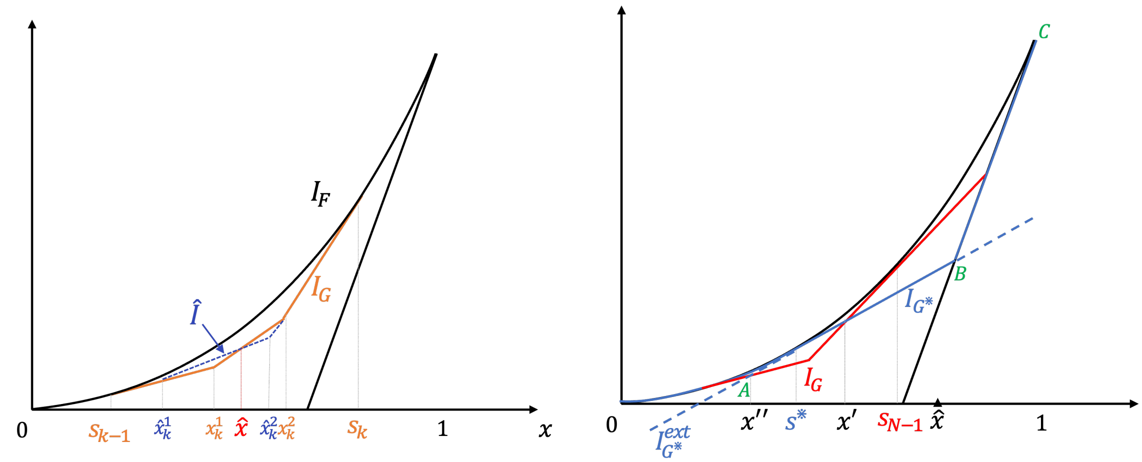

Figure 8: In the left panel, a clockwise rotation of the integral distribution function around the inflection point $\hat{x}$ would raise the value of the objective function. In the right panel, the solid blue curve I_{G^*} represents the optimal unconstrained information structure. If I_G is a feasible discrete information structure with $s_{N-1} > s^*$, another feasible information structure $\hat{I}$ which coincides with I_G to the left of x'' and follows the line segments ABC to the right of x'' will increase the value of the objective function.

Then we extend the new piece until it hits I_G at $\hat{x}_k^1 < x_k^1$ and hits the extended line of the next piece of I_G at $\hat{x}_k^2 > x_k^2$. See Figure 8(a) for a graphical illustration. By construction, $\hat{I} > I_G$ on $(\hat{x}_k^1, \hat{x})$ and $\hat{I} < I_G$ on $(\hat{x}, \hat{x}_k^2)$. Since $u''(x) > 0$ on $(\hat{x}_k^1, \hat{x})$ and $u''(x) < 0$ on $(\hat{x}, \hat{x}_k^2)$, we must have $\int_0^1 u''(x) I_G(x) dx < \int_0^1 u''(x) \hat{I}(x) \mathrm{d}x$, a contradiction. The argument for $|\mathscr{K}_{I_G}| = N$ is similar to the convex case and therefore omitted here.

(ii) Because I_{G^*} coincides with I_F to the left of s^*, and G has an interval-partitional structure by part (i), G is less Blackwell-informative than G^* if and only if $s_{N-1} \leq s^*$. Note that s_{N-1} cannot exceed $\hat{x}$, for otherwise we could rotate I_G clockwise at $\hat{x}$ to increase the value of the objective function (2).

Suppose G is not less informative than G^*, i.e., $s_{N-1} \in (s^*, \hat{x}]$. Note that $I_G(s_{N-1}) = I_F(s_{N-1}) > I_{G^*}(s_{N-1})$ and $I_G(s^*) \leq I_F(s^*) = I_{G^*}(s^*)$. Therefore there exists $x' \in [s^*, s_{N-1})$ such that I_G crosses I_{G^*} from below at x'. Moreover, let $I_{G^*}^{\text{ext}}(x) := I_{G^*}(s^*) - F(s^*)(s^* - x)$ for $x \in [0,1]$ to be the extrapolation of the linear segment of I_{G^*} (see Figure 8(b)). Then we have $I_G(0) > I_{G^*}^{\text{ext}}(0)$ and $I_G(s^*) \leq I_{G^*}^{\text{ext}}(s^*)$. Therefore there exists $x'' \in (0, s^*]$ such that $I_{G^*}^{\text{ext}}$ crosses I_G from below at x''.

Consider an alternative informal structure with integral distribution $\hat{I}$ given by:

$$\hat{I}(x) = \begin{cases} I_G(x) & \text{if } x < x'' \\ I_{G^*}^{\text{ext}}(x) & \text{if } x \in [x'', x'] \\ I_{G^*}(x) & \text{if } x > x'. \end{cases}$$

Note that $\hat{I} \in \text{ICPL}$ with $|\mathcal{K}_{\hat{I}}| \le |\mathcal{K}_{I_G}|$. Thus $\hat{I}$ is a feasible discrete information structure. By construction, $\hat{I}(x) > I_G(x)$ for $x \in [x'', x']$. Since $x' < s_{N-1} \le \hat{x}$, we have $u'' > 0$ for $x \in [x'', x']$. This gives

$$\int_0^{x'} u''(x)\hat{I}(x)\,\mathrm{d}x > \int_0^{x'} u''(x)I_G(x)\,\mathrm{d}x.$$

Moreover, another information structure $\tilde{I}$ which coincides with I_{G^*} on $[0, x')$ and coincides with I_G on $[x', 1]$ is feasible in the unconstrained information design problem without the discreteness constraint. Therefore,

$$\int_{x'}^1 u''(x)\hat{I}(x)\,\mathrm{d}x = \int_{x'}^1 u''(x)I_{G^*}(x)\,\mathrm{d}x \ge \int_{x'}^1 u''(x)\tilde{I}(x)\,\mathrm{d}x = \int_{x'}^1 u''(x)I_G(x)\,\mathrm{d}x,$$

where the inequality follows due to revealed preference. Combining these two inequalities shows that $\hat{I}$ is strictly better than I_G, a contradiction. □

The result that the optimal experiment has an interval-partitional structure even when u is S-shaped follows a similar logic as in the proof of Lemma 3. If the information structure exhibits bi-pooling in a subinterval that contains $\hat{x}$, rotating the integral function clockwise at $\hat{x}$ will increase payoff because u'' is positive to the left of $\hat{x}$ and negative to the right of $\hat{x}$. Part (ii) of Proposition 6 shows that I_G must be everywhere below I_{G^*}; therefore the optimal discrete information structure is less informative than the optimal unconstrained experiment in the sense of Blackwell (1953).

Once we establish the optimality of interval-partitional structure, we can apply our dual expectations characterization to solve the problem. One subtlety is that since now u'' is negative on some region, the object $\mu(\cdot)$ in (CE-u') is the barycenter instead of the conditional expectation. We provide a more thorough discussion on this point in the next section. Another issue is that in our discussion of the purchase decision example in

Section 3.2, we assume that the value function $\hat{u}$ for the seller is convex. If the density h of the cost shock is logconcave but not nondecreasing, then $\hat{u}''(x) = ph'(x-p)$ is single-crossing from above, and so $\hat{u}$ is an S-shaped value function. For such value function, we need to provide an additional argument to establish Proposition 3. We offer such an argument in the appendix. Our earlier conclusion that the induced posterior means under the seller-optimal information structure are lower than the ones under the buyer-optimal structure remains valid even when the seller's value function $\hat{u}$ is S-shaped.

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,[32] it is possible that the optimal information structure entails bi-pooling policies, and the analysis would become more involved. Nevertheless, in this section, we show that the dual expectations property between interval cutoffs and induced posterior means still holds, albeit in a modified way.

When the value function u is not convex, u'' may be negative and therefore cannot be a valid density. Nevertheless, it can be interpreted as the "density" corresponding to a *signed measure*. For some measurable set A under signed measure u', its *barycenter* is given by

$$\frac{\int_A \theta u''(\theta)\,\mathrm{d}\theta}{\int_A u''(\theta)\,\mathrm{d}\theta}.$$

We can therefore interpret $\mu(a,b)$ in equation (CE-u') as the barycenter of set $A=(a,b)$ under signed measure u'. When the optimal information structure under general value function is interval-partitional, it is still characterized by the system of equations (CE-F) and (CE-u') stated in Proposition 1. The dual expectations property continues to hold, with "expectations" interpreted as "barycenter."

The case where the information structure may not be interval-partitional is illustrated in Figure 9. The information structure represented by $\underline{u}(x)$ in this figure partitions the state space into three intervals: $(0,s_1)$, (s_1,s_2) and $(s_2,1)$.[33] The first interval

[32]General value functions can arise in persuasion mechanisms where the designer's information disclosure is contingent on the receiver's private type (Kolotilin et al., 2022), and delegation model where the principal's equilibrium payoff can be formulated as a linear functional (Kleiner, 2022). Our framework can be accommodated to study those models when the action space is finite.

[33]For general value functions, $\underline{u}(x)$ is not necessarily everywhere below or above the value function.

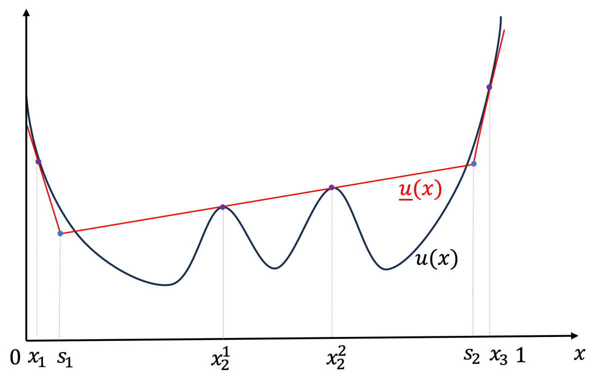

Figure 9: Optimal information structure for a general value function.

induces signal x_1 and the third interval induces signal x_3. The second interval induces two possible signals, x_2^1 and x_2^2. If bi-pooling is optimal on (s_1, s_2), then $\underline{u}$ must be tangent to the value function u at the two induced posterior means. Such bi-tangency requirement pins down x_2^1 and x_2^2. For given x_2^1 and x_2^2, our characterization for the remaining posterior means and cutoff points works as follows: the cutoffs s_1 and s_2 are the respective barycenters of (x_1, x_2^1) and (x_2^2, x_3) under the signed measure u', and x_1 and x_3 are the conditional means of $(0, s_1)$ and $(s_2, 1)$, respectively, under the prior distribution. Moreover, one can check the feasibility of such bi-pooling information structure by verifying whether the interval (s_1, s_2) so obtained is compatible with the two given signals x_1^2 and x_2^2. The necessary and sufficient condition for such feasibility are described by Lemma 1 in Arieli et al. (2023).

Proposition 7. *Suppose an information structure $I_G \in ICPL$ is optimal, with interval cutoff points $s_0 < \ldots < s_K$ such that I_G is tangent to I_F at these points.*

(i) If $J_k = 1$, then $x_k = \mathbb{E}_F[\theta | \theta \in (s_{k-1}, s_k)]$.

(ii) If $J_k = 2$, then $x_k^1 < x_k^2$ and they lie on an affine line bi-tangent to $u(x)$.

(iii) Each cutoff s_k is the barycenter of (y_k, y_{k+1}) under the signed measure u' where,

$$y_k = \begin{cases} x_k & \text{if } J_k = 1 \\ x_k^2 & \text{if } J_k = 2; \end{cases} \qquad y_{k+1} = \begin{cases} x_{k+1} & \text{if } J_{k+1} = 1 \\ x_{k+1}^1 & \text{if } J_{k+1} = 2. \end{cases}$$

Proposition 7 only gives necessary conditions for an optimal information structure. Nevertheless, the set of information structures that satisfies those conditions is finite.

In principle, an optimal solution can be derived by comparing the associated payoffs.

7. Discussion

7.1. An application: design of energy efficiency ratings

Providing accessible information on the energy efficiency of electrical appliances or automobiles can assist households in estimating the costs associated with energy usage, enabling them to make informed decisions. However, energy consumption often involves externalities, meaning that the sole objective of enhancing households' welfare may not fully align with broader policy objectives that encompass environmental and other considerations. The development of suitable energy efficiency ratings must take into account these diverse concerns.

Consider the rating design for gasoline vehicles as an example. Let θ be a random variable which is uniformly distributed on $[\underline{\theta}, 1]$, where $\underline{\theta} \in (0, 1)$. It represents parameters such as fuel economy (e.g., fuel usage per 100 kilometers) or CO_2 emissions.

Suppose a household drives a car for a miles. Total energy consumption would be $a\theta$. The unit cost of fuel consumption is p. The household has a quasi-linear preference for money and CARA utility from car usage,

$$U(a, \theta) = 1 - e^{-a} - pa\theta.$$

Given an estimation of the expected energy efficiency x, the household's optimal usage is $a(x) = -\log(px)$. Thus, its indirect value function is $u(x) = 1 - e^{-a(x)} - pa(x)x$, which is convex. Now consider a government that not only cares about the household's welfare but also aims to reduce total energy consumption for environmental reasons or to decrease private car usage for traffic management. Its indirect value function is $\hat{u}(x) = u(x) - \lambda_1 a(x)x - \lambda_2 a(x)$, where $\lambda_1, \lambda_2 > 0$ measure the intensity of the two respective objectives. Note that $\hat{u}$ is concave to the left of $\lambda_2/(\lambda_1 + p)$ and is convex to the right of it.

Our numerical solution for the government's optimal rating design is depicted in the second row of Figure 10. It suggests that the width of the first class (representing the most efficient vehicles) is wider than that of the second one. Because the government's value function is concave near $\underline{\theta}$ (assuming $\underline{\theta} < \lambda_2/(\lambda_1 + p)$), the first interval of the optimal rating scheme is relatively wide. Intuitively, when energy performance is very

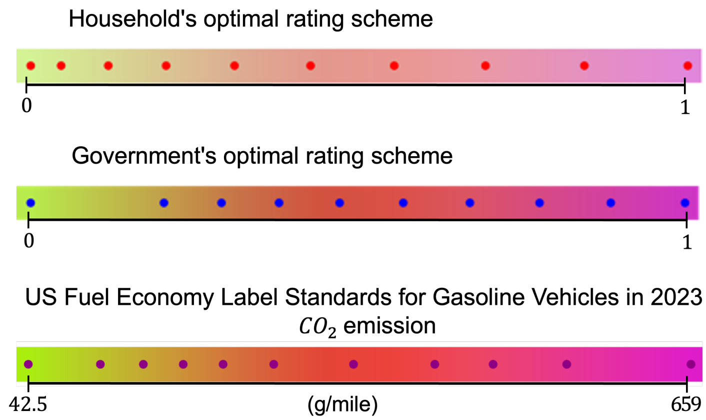

Figure 10: Rating schemes. The first and second panels are the optimal rating schemes for the household and for the government calculated based on the parameters $\underline{\theta} = 0.01$, $p = 0.1$, $\lambda_1 = 0.3$, $\lambda_2 = 0.05$, and uniform prior distribution. The third panel shows the CO_2 emissions standard for vehicles in the U.S. in 2023 (Source: `https://www.fueleconomy.gov/feg/label/learn-more-gasoline-label.shtml`).

efficient, usage is very sensitive to efficiency (as $\mathrm{d}y(x)/\mathrm{d}x = -p/x$ is large in absolute value when x is small). Thus, a coarser interval around the smallest values of θ can help achieve the government's objectives of reducing usage and energy consumption. This observation is consistent with the rating standards of fuel economy labels for gasoline vehicles in the U.S. (the bottom panel of Figure 10).[34] A similar pattern can also be found in the ratings of electrical devices such as washing machines and the energy ratings scheme for residential buildings in the EU.

7.2. Beyond logconcave environments

The conditions (CE-F) and (CE-u') for interval cutoffs and induced posterior means are necessary but not sufficient for optimality. When the solution to that equation system is unique, the necessary conditions are also sufficient. A logconcave environment ensures uniqueness of the solution to (CE-F) and (CE-u') and facilitates comparative statics analysis.[35] In this subsection, we provide numerical examples of situations where the curvature functions or the prior densities are not logconcave or not even single-peaked.

[34]Though we use a uniform prior distribution for the numerical examples, the observation that the first interval is coarser is robust across different years in 2017–2024 (the U.S. Environmental Protection Agency changes the rating scheme every year as the underlying distribution of efficiency varies across years).

[35]When there are multiple solutions to (CE-F) and (CE-u'), the optimal solution need not be the largest or the smallest fixed point, thus rendering standard monotone comparative statics difficult.

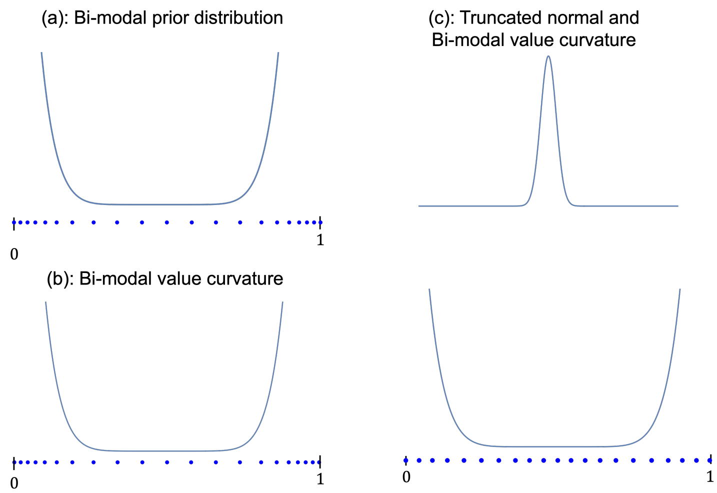

Figure 11: Numerical solutions for different prior distributions and value functions. In (a), $f(x) = 2048(x-0.5)^2$ and $u''(x) = 2$. In (b), $f(x) = 1$ and $u''(x) = (x-0.5)^8$. In (c), the prior distribution is truncated Normal on $[0,1]$ with parameters $(0.5, 0.03)$, and $u''(x) = (x-0.5)^8$.

In these examples the uniqueness of the solution to (CE-F) and (CE-u') can be numerically verified.[36] Furthermore, the main message in our analysis—the decision maker gives closer scrutiny to states that are more likely to occur or states with higher instrumental value of information—remains robust.

In Figure 11(a), we use a bi-modal density function with modes at 0 and 1, while the value function is quadratic. The dotted line at the bottom of each panel plots the interval cutoffs of the optimal information structure. The decision maker gives the closest scrutiny to those states around 0 and 1 as they are most likely to be realized. In 11(b), we consider the uniform prior distribution, while u'' is bi-modal with modes at 0 and 1. Symmetrically, the optimal information structure gives close scrutiny to states around 0 and 1 as they have more impact over the decision maker's payoffs. In 11(c), u'' is still bi-modal, but the prior distribution is a truncated normal distribution with a single mode at 0.5. In this case, the optimal information structure needs to compromise between the two forces, and gives close scrutiny to states around 0, 0.5 and 1. The widths of successive intervals are not single-dipped in these examples.

[36] Starting with initial values near 1, an iterated solution to (CE-F) and (CE-u') converges to the largest fixed point; and starting with initial values near 0, it converges to the smallest fixed point. The solution is unique if the largest fixed point and the smallest fixed point coincide.

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. Suppose, for example, that the sender's utility is $-(y-\kappa_1\theta)^2$ and the receiver's utility is $-(y-\theta)^2$, where $\kappa_1 > 1$. Then, the cheap-talk equilibrium, given a fixed number of signals N, is determined by the following system of equations:

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

The second set of equations are the indifference conditions for the cutoff types. They are different from the optimality conditions (CE-u') for the coarse information design problem; in general, the dual expectations property does not hold in cheap talk models.[37] Nevertheless this system of equations is still a monotone contraction mapping 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. However, the insight driving this result in the cheap talk model differs from ours. Specifically, with $\kappa_1 > 1$, the sender's bias is smaller when θ is small in magnitude than when it is large. Hence, credibility is easier to generate in smaller states. Consequently, the intervals are narrower for states closer to 0 and become wider farther away from 0 (Gordon, 2010; Deimen and Szalay, 2019, 2023). 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. As with quadratic utilities, u'' is a constant regardless of κ_1. Hence, κ_1 has no effect on the optimal information structure in our model.[38]

Recent developments in cheap-talk games have also explored the optimal informa-

[37] The dual expectations property comes from the continuity of the minorant function at the interval cutoffs. That is, the cutoff types are indifferent in terms of the virtual utility rather than the actual utility.

[38] 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.

tion design by the sender. This literature assumes that an uninformed sender commits to an experiment, and sends an unverifiable message about the experimental outcome. In a uniform-quadratic setting with constant bias κ_0 (Ivanov, 2010; Lou, 2023; Kreutzkamp, 2023), 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 the number of messages N being endogenously determined through this constraint. As elaborated in Lou (2023) and Kreutzkamp (2023), 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 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-separable cost of acquiring information that depends on the experiment chosen (Caplin and Dean, 2015; Matějka, 2015; Bloedel and Segal, 2021; Ravid et al., 2022); see Denti (2022) for a critical assessment of that approach. In this paper, we adopt an alternative track by imposing a discreteness 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 1. Fix some $I_G \in$ ICPL with a finite number of kinks. It can only have a finite number of tangency points to I_F. Because I_G is piecewise linear and I_F is strictly convex, $I_G(\cdot) < I_F(\cdot)$ whenever they are not tangent. Enumerate the set of tangency points in increasing order and let $\{s_k\}_{k=0}^K = \{x \in [0,1] : I_F(x) = I_G(x)\}$, with $s_0 = 0$ and $s_K = 1$.

Allocate all the kink points of I_G into the interval $[s_{k-1}, s_k)$ that contains them. This process is well-defined as every kinks point must lie in the interior of $[s_{k-1}, s_k)$ for some $k = 1, \ldots, K$. Within each interval $[s_{k-1}, s_k)$, enumerate the kinks point in increasing order to get the sequence $\{x_k^j\}_{j=1}^{J_k}$, where J_k is the number of kink points contained in $[s_{k-1}, s_k)$.

Now we claim that the J-pooling policy with $\{s_k\}_{k=0}^K$ and $\{\{x_k^j\}_{j=1}^{J_k}\}_{k=1}^K$ can indeed induce a distribution G that corresponds to the integral distribution function I_G. It suffices to show that for every $k = 1, \ldots, K$, the conditional distribution of G restricted on $[s_{k-1}, s_k)$, denoted by $G|_{[s_{k-1}, s_k)}$, can be indeed induced by that of F on the same interval. We need to check two consistency conditions: for every $k = 1, \cdots, K$, (a) $G(s_k) = F(s_k)$; and (b) $I_{F|_{[s_{k-1},s_k)}} \geq I_{G|_{[s_{k-1},s_k)}} \geq I_{F|^0_{[s_{k-1},s_k)}}$, where $F|^0_{[s_{k-1},s_k)}$ is the degenerate distribution on $[s_{k-1}, s_k)$ that puts probability mass one at the conditional mean $x_k = \mathbb{E}_F[\theta \mid \theta \in [s_{k-1}, s_k)]$.

Part (a) is easy to check as the tangency condition implies that $F(s_k) = I_F'(s_k) = I_G'(s_k) = G(s_k)$ for $k = 0, \ldots, K$. To verify (b), for any $x \in [s_{k-1}, s_k)$ we write:

$$I_G|_{[s_{k-1},s_k)}(x) = \int_{s_{k-1}}^{x} \frac{G(t) - G(s_{k-1})}{G(s_k) - G(s_{k-1})} \, \mathrm{d}t = \int_{s_{k-1}}^{x} \frac{G(t) - F(s_{k-1})}{F(s_k) - F(s_{k-1})} \, \mathrm{d}t. \tag{7}$$

It is immediate that the first inequality of (b) holds if and only if $I_F \geq I_G$ on (s_{k-1}, s_k). For the second inequality, note that the integral function corresponding to the degenerate distribution is

$$I_{F|^0_{(s_{k-1},s_k)}}(x) = \begin{cases} 0 & \text{if } x \in (s_{k-1}, x_k) \\ x - x_k & \text{if } x \in [x_k, s_k). \end{cases} \tag{8}$$

It is obvious that the second inequality of (b) holds for $x \in [s_{k-1}, x_k)$. For $x \in [x_k, s_k)$, we multiply both (7) and (8) by $F(s_k)-F(s_{k-1})$ and the difference reduces to:

$$\begin{aligned}
&I_G(x)-I_G(s_{k-1})-(x-s_{k-1})F(s_{k-1})-(x-x_k)(F(s_k)-F(s_{k-1}))\\
&=[I_G(s_k)+(x_k-s_k)F(s_k)-I_G(s_{k-1})-(x_k-s_{k-1})F(s_{k-1})]+[I_G(x)-I_G(s_k)-(x-s_k)F(s_k)]\\
&=[I_F(s_k)+(x_k-s_k)F(s_k)-I_F(s_{k-1})-(x_k-s_{k-1})F(s_{k-1})]+\left[I_G(x)-I_G(s_k)-(x-s_k)I'_G(s_k)\right]\\
&=I_G(x)-I_G(s_k)-(x-s_k)I'_G(s_k)\geq 0,
\end{aligned}$$

where the first term in brackets in the third line is equal to 0 by definition of x_k as the conditional mean on $[s_{k-1}, s_k)$, and the last inequality follows because I_G is convex. This establishes (b), and therefore by Strassen's (1965) theorem the conditional distribution of G restricted on $[s_{k-1}, s_k)$ can be indeed induced by the prior on the same interval. □

Proof of Lemma 2. Suppose an optimal information structure G induces $J_k \geq 3$ signals between (s_{k-1}, s_k). By Lemma 1, the corresponding I_G exhibits $J_k \geq 3$ kinks between (s_{k-1}, s_k). Then one can construct two functions $\overline{I}, \underline{I} \in$ ICPL with the following properties: (a) $I_F > \overline{I} > I_G > \underline{I} > I_{F_0}$ on (x_k^1, x_k^3), and $\overline{I} = I_G = \underline{I}$ for all $x \notin (x_k^1, x_k^3)$; (2) $\mathscr{K}_{\overline{I}} = \mathscr{K}_{\underline{I}} = \mathscr{K}_{I_G}$; and (c) $I_G = \lambda \overline{I} + (1-\lambda)\underline{I}$ for some $\lambda \in (0,1)$ (see Figure 12). Such construction is feasible without violating (MPC) because in this region, I_G is strictly below I_F and we can always slightly adjust the slopes of each piece. Since the objective function in program (2) is a linear functional over the set of feasible ICPL functions, I_G must be sub-optimal.

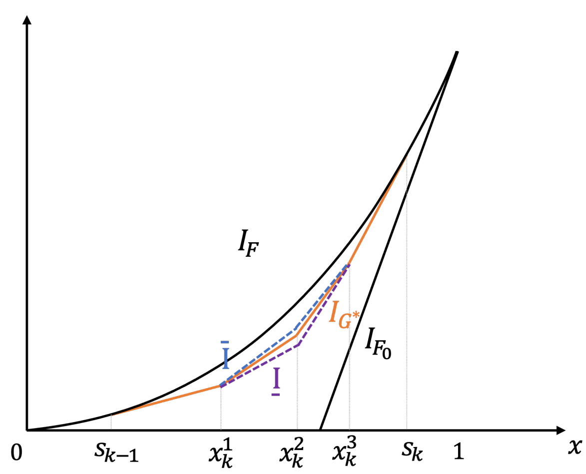

Figure 12

The proof for the existence of optimal signal structure directly comes from Theorem

4 in Appendix C, Aybas and Turkel (2021). In particular, it follows from the compactness of the feasible ICPL functions and continuity of the objective function in (2) with sup-norm. □

Proof of Proposition 1. Let $\Gamma : [0,1]^{2N-1} \to [0,1]^{2N-1}$ be the mapping corresponding to the right-hand-side of (CE-F) and (CE-u'). Since Γ is monotone, Tarski's theorem implies that a largest (denoted $z' = (x_1', s_1', \dots, s_{N-1}', x_N')$) and a smallest (denoted $z'' = (x_1'', s_1'', \dots, s_{N-1}'', x_N'')$) fixed point exist. Since $z' \geq z''$, we must have $x_1' - x_1'' \geq 0$. If $x_1' - x_1'' = 0$, then the fact that $x_1 = \phi(0, s_1)$ and $\phi(\cdot)$ is strictly increasing implies that $s_1' - s_1'' = 0$. Since $s_1 = \mu(x_1, x_2)$ and $\mu(\cdot)$ is strictly increasing, this further implies that $x_2' - x_2'' = 0$. Iterating this argument then shows that $z' = z''$. If $x_1' - x_1'' > 0$, the fact that $\phi(\cdot)$ is strictly increasing and satisfies property (4) stated in Lemma 4 below when f is logconcave implies that $s_1' - s_1'' > x_1' - x_1''$. Since a similar property as stated in Lemma 4 holds for $\mu(\cdot)$ when u'' is logconcave, this further implies that $x_2' - x_2'' > s_1' - s_1''$. Iterating this argument 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 4, a contradiction. This shows that the largest and the smallest fixed points coincide. Because an optimal information structure exists and any candidate solution must be a fixed point of Γ, the unique fixed point of Γ is the optimal solution. □

Proof of Lemma 3. The objective function (2) is the u''-weighted area below I_G. Since u'' is non-negative, an information structure G' performs better than another information structure G if $I_{G'} \geq I_G$. Consider an arbitrary bi-pooling policy G such that two posterior means x_k^1 and x_k^2 are induced within the interval $[s_{k-1}, s_k)$ (see the left panel of Figure 13). We can construct another integral distribution function $I_{G'}$ by fixing the point $(x_k^1, I_G(x_k^1))$ and slightly rotating the piece of I_G on (x_k^1, x_k^2) in the counterclockwise direction until it hits the next piece of I_G on (x_k^2, s_k) (see the blue dashed piece in the left graph). By construction $I_{G'}$ is everywhere above I_G and everywhere below I_F. Moreover, the number of induced posterior means does not change. Therefore the information structure represented by $I_{G'}$ is feasible and produces a higher value of the objective function.

Next, suppose the optimal information structure 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 13). This new integral distribution with an additional kink point

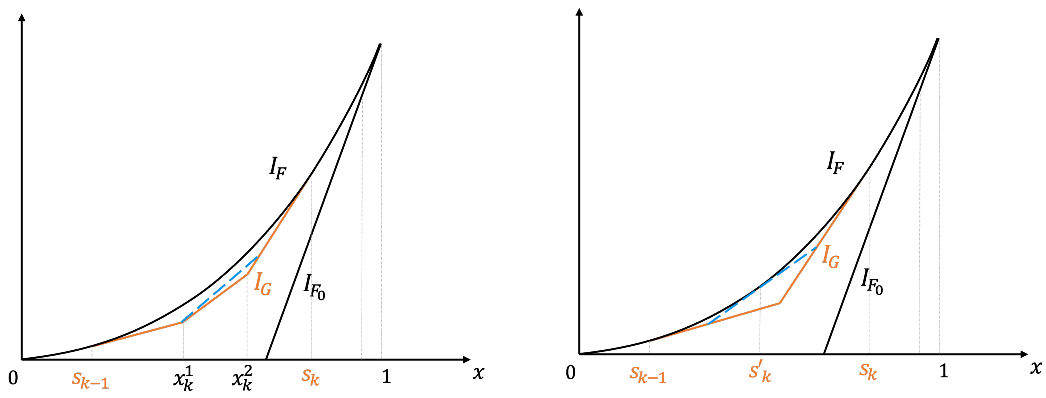

Figure 13: A rotation of I_G is profitable if the original information structure exhibits bi-pooling. Adding one more kink to I_G is profitable if the original information structure uses less than N signals.

is everywhere higher than the original integral distribution, and will produce a higher value of the objective function when u is convex.

□

Proof of Lemma 4. Mease and Nair (2006) and Szalay (2012) already discover this property. Here we present a simple proof with consistent notation for easy reference. Rewrite $\phi(a+\varepsilon, b+\varepsilon)$ and $\phi(a,b)+\varepsilon$ as

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

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. □

Proof of Proposition 2. 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 arbitrarily set the mode of u'' at m_f, and part (i) follows. Similarly, if f is a constant, we can arbitrarily set its mode at m_u, and part (ii) follows. □

Proof of Proposition 3 *(extended to the case when $\hat{u}$ is S-shaped)*. Suppose u'' and f are both logconcave, and let $\hat{u}''(\cdot)/u''(\cdot)$ be decreasing. Suppose $\hat{u}$ is an S-shaped value function, with $\hat{u}''(\cdot)$ changes from positive to negative at $\hat{x} \in (0,1)$. Equations (CE-F) and (CE-u') still characterize the necessary conditions for optimality. We show that $\hat{s}_{N-1} \leq s_{N-1}$.

Take s_{N-1} as fixed at some value $s^\dagger \in (0, \hat{x})$. The right-hand-side of equations (CE-F) for $k = 1, \ldots, N-1$ and (CE-u') for $k = 1, \ldots, N-2$ defines a mapping $\Gamma^\dagger : [0, s^\dagger]^{2N-3} \to [0, s^\dagger]^{2N-3}$. The fixed point of this mapping gives the solution $(x_1, \ldots, x_{N-1}, s_1, \ldots, s_{N-2})$ to this equation system as a function of $s^\dagger$. Because $\Gamma^\dagger$ is monotone and is increasing in $s^\dagger$, the largest fixed point is increasing in $s^\dagger$. Moreover, because $\hat{u}''(\cdot)/u''(\cdot)$ is decreasing and is positive on $[0, s^\dagger]$, monotone comparative statics implies $\hat{x}_k(s^\dagger) \leq x_k(s^\dagger)$ for all $s^\dagger$. Finally, define $x^\dagger(s^\dagger) := \phi(s^\dagger, 1)$, so that equation (CE-$F$) for $k = N$ is satisfied when $x_N = x^\dagger(s^\dagger)$. The remaining optimality condition is $s^\dagger = \mu(x_{N-1}(s^\dagger), x^\dagger(s^\dagger))$. We note that $\gamma(\cdot) := \mu(x_{N-1}(\cdot), x^\dagger(\cdot))$ is an increasing mapping. The optimal s_{N-1} must be a fixed point of γ. Since both $x^\dagger(\cdot)$ and $x_k(\cdot)$ are increasing functions and $\hat{x}_k(\cdot) \leq x_k(\cdot)$ (for $k = 1, \ldots, N-1$), $\hat{s}_{N-1} \leq s_{N-1}$ would imply $\hat{x}_k \leq x_k$ for all k. Similarly, this would also imply $\hat{s}_k \leq s_k$ for all k.

Use $\hat{\mu}(\cdot)$ to denote the conditional mean function using $\hat{u}''$ as density, and $\tilde{\mu}(\cdot)$ to denote the conditional mean function using $\max\{\hat{u}'', 0\}$ as density. We have $\hat{\mu}(\cdot) \leq \tilde{\mu}(\cdot) \leq \mu(\cdot)$. Therefore,

$$\hat{\gamma}(s^\dagger) = \hat{\mu}(\hat{x}_{N-1}(s^\dagger), x^\dagger(s^\dagger)) \leq \mu(x_{N-1}(s^\dagger), x^\dagger(s^\dagger)) = \gamma(s^\dagger),$$

because $\hat{x}_{N-1}(s^\dagger) \leq x_{N-1}(s^\dagger)$. This implies that the largest fixed point of $\hat{\gamma}$ is smaller than the (unique) fixed point of γ, and therefore $\hat{s}_{N-1} \leq s_{N-1}$. □

Proof of Proposition 9. Part (i) is by construction. Part (iii) can be obtained from the same first-order condition as in Proposition 1. To show (ii), notice that the corresponding I_G supported on $[x_k^1, x_k^2]$ is completely below I_F. The optimality of such I_G implies that a local (clockwise or anti-clockwise) rotation at any point $(\tilde{x}, I_G(\tilde{x}))$ where $\tilde{x} \in [x_k^1, x_k^2]$ is not profitable. Denote the piece of I_G on $[x_k^1, x_k^2]$ as $I_G(\tilde{x}) + \beta(x - \tilde{x})$. Then the first-order necessary condition for objective function (2) with respect to β reduces to,

$$\int_{x_k^1}^{x_k^2} (x - \tilde{x}) u''(x)\, \mathrm{d}x = 0, \quad \text{for every } \tilde{x} \in [x_k^1, x_k^2].$$

Integration by part then implies,

$$u(x_k^1) - u'(x_k^1)(x_1^k - \tilde{x}) = u(x_k^2) - u'(x_k^2)(x_2^k - \tilde{x}), \quad \text{for every } \tilde{x} \in [x_k^1, x_k^2].$$

Graphically, this equation means that there exists an affine line that is bi-tangent to $u(\cdot)$ at both x_k^1 and x_k^2. □

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," 2021. Working paper, Stanford University. `https://arxiv.org/abs/1910.13547v4`.

_ and _ , "Persuasion with coarse communication," 2022. Working paper, Stanford University.

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

_ , Ji Shen, Yun Xu, and Edmund M. Yeh, "Mechanism design with limited information," in Rahul Jain and Rajgopal Kannan, eds., *Game Theory for Networks*, Berlin: Springer, 2012, pp. 1–10.

_ , Tibor Heumann, and Stephen Morris, "Screening with persuasion," 2022. Cowles Foundation Discussion Paper No. 2338R.

Blackwell, David, "Equivalent comparisons of experiments," *Annals of Mathematics and Statistics*, 1953, *24* (2), 265–272.

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

Bolton, Patrick and Mathias Dewatripont, *Contract theory*, Cambridge, MA: MIT Press, 2005.

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.

Dharmadhikari, Sudhakar and Kumar Joag-dev, *Unimodality, Convexity, and Applications*, San Diego, LA: Academic Press, 1998.

Doval, Laura and Vasiliki Skreta, "Constrained information design," *Mathematics of Operations Research*, 2022, *forthcoming.*

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

Dworczak, Piotr and Anton Kolotilin, "The persuasion duality," *Theoretical Economics*, 2023, *forthcoming.*

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

Galichon, Alfred, *Optimal transport methods in economics*, Princeton, NJ: Princeton University Press, 2016.

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

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

Gray, Robert M. and David L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, 1988, *44* (6), 2325–2383.

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

Hopenhayn, Hugo and Maryam Saeedi, "Optimal simple ratings," 2022. 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, "Informational control and organizational design," *Journal of Economic Theory*, 2010, *145* (2), 721–751.

—, "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, "Optimal delegation in a multidimensional world," 2022. Working paper, `https://arxiv.org/pdf/2208.11835`.

—, 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.

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

Kreutzkamp, Sophie, "Endogenous information acquisition in cheap-talk games," 2023. Working paper, University of Bonn.

Lipman, Barton L., "Why is language vague?," 2009. Working paper, Boston University.

Lloyd, Stuart P., "Least squares quantization in PCM," *IEEE transactions on information theory*, 1982, *28* (2), 129–137.

Lou, Yichuan, "Sender-optimal learning and credible communication," 2023. Working paper, University of Tokyo.

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

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," *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: 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.

Smolin, Alex and Takuro Yamashita, "Information design in concave games," 2022. Working paper, Toulouse School of Economics.

Sørensen, Peter Norman, "Rational Social learning," 1996. Ph.D. Thesis, Massachusetts Institute of Technology.

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: Princeton University Press, 1998.

Treust, Maël Le and Tristan Tomala, "Persuasion with limited communication capacity," *Journal of Economic Theory*, 2019, *184*, 104940.

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," *Journal of Mathematical Economics*, 2014, *54*, 7–21.