Optimal Information Structures with Information Control

Attila Ambrus[†]
Luke Zhao[‡]
Last Updated: August 22, 2025

Abstract

This paper considers optimal information structures in restricting the information of the Sender in a cheap talk game. Monotone partitional information structures might not obtain optimality, and some signals that the Sender receives are required to be partially garbled together in order to relax the Sender's incentive constraints. We provide a partial characterization of the optimal structure and show that it is from a class of structures that are more general than monotone partitional, but only exhibit certain local nonmonotonicities of the information structure. Our analysis builds on the graphical characterization of feasible information structures in Gentzkow and Kamenica (2016) and shows that this methodology can be used in complex information design problems with large state and action spaces. In an extension we use the same approach to solve problems with capacity constraints on communication.

1 Introduction

There are many economically important contexts in which an uninformed decision maker needs to seek advice from an informed but biased expert before making a decision, such as when a policymaker consults with an advisor with specialized knowledge or when corporate headquarters gather information from a particular organizational unit. Starting with the seminal work of Crawford and Sobel (1982), there is a large literature investigating such interactions focusing on the case when communication between the parties takes the form of cheap talk: an informed Sender sending a cheap talk message to a Receiver who subsequently chooses an action.[1] Most of this literature assumes that the information of the Sender is exogenously given. However, there are many settings in which the Receiver can restrict the information that the Sender can learn before the communication takes place. For example, organizational headquarters can put restrictions on what processes the head of an organizational unit can monitor and keep records on. Or a regulatory agency collecting information on demand conditions from firms at a market can restrict the type of data the firms are allowed to collect on their consumers. More recently, Fischer and Stocken (2001) and Ivanov (2010b) showed that such reduction of the quality of the information

[1]See also Green and Stokey (2007) the first version of which was written around the same time.

of the Sender can be beneficial for the Receiver (and indeed to both parties from an ex ante point of view). The main intuition is that coarser information can ease the incentive constraints of the Sender and hence can facilitate credible communication of more information.[2]

The main focus of the current paper is characterizing the optimal way of restricting the information of the Sender. Ivanov (2010b) characterizes the optimal information structure in the uniform-quadratic specification of the Crawford and Sobel model subject to the assumption that the optimal information structure is monotone partitional, but leaves the question open whether restricting attention to such information structures results in suboptimality. We show that indeed it is, even in the uniform-quadratic specification of the cheap talk model. Our leading example is a range of bias parameters of the Sender for which the optimal partitional structure involves partitioning the state space (the unit interval in $\mathbb{R}$) to three equal size cells and the Sender only learning which cell the state falls in, or equivalently only learning the conditional expectation of the state given the cell where the state is. For biases in the range we consider in this example, there is no partitional structure with more than three cells for which it is incentive compatible for the Sender to truthfully reveal the observed cell for all of the possible cell realizations. Inevitably, when partitioning the state space into four or more cells, conditional expectations of at least two neighboring cells would be close enough to each other such that the biased Sender would have an incentive to claim that the state is from the neighboring cell, not from the one that the Sender observed. On the other hand, it is possible to construct nonpartitional information structures with four signal realizations, each signal associated with different conditional expectations of the state, such that the Sender can truthfully reveal his information to the Receiver without violating incentive compatibility constraints. These structures resemble a partitional structure with four information cells, but with some garbling of the information from the two middle cells: with some small probability the signal of the Sender being the second highest one even when the true state is from the third highest cell, and vice versa. Relative to a partitional structure restricted on the same states, this garbling pulls the conditional expectations associated with the two signals closer to each other, away from the conditional expectations associated with the extreme signals. This in turn makes it possible to add two other information cells at the extremes without violating incentive compatibility. The trade-off associated with this is that while a higher number of signals is credibly conveyed to the Receiver, the quality of some of the signals is reduced by the garbling in the middle, but for a range of biases the garbled information structure with more signals yields higher payoffs.

For an example how such a garbling of the Sender's information can be welfare improving, consider search committee in an academic department in a setting where a job candidate can only

[2]Relatedly, Austen-Smith (1994) demonstrates that making information acquisition costly for the Receiver can improve the efficiency of communication, and Ivanov (2010a) and Ambrus et al. (2013) show that communication via strategic Mediators, where the information of a Mediator is endogenously coarse relative to the Sender, can have similar effects.

be evaluated by the committee member who is from the same field as the candidate. Assume that candidates can be lumped into four categories: Unacceptable, Mediocre, Good and Excellent, and ideally the committee would like to learn from the expert committee member which category the candidate falls into. However, the committee member is biased towards her own field, and has an incentive to claim that a candidate is Excellent even when she knows that the candidate is only Good. In this case, instead of providing information to the committee member that lets her accurately identify a candidate's category, it might be better to provide information that garbles the Mediocre and Good categories to some extent: in particular one which implies that with positive probability some Mediocre candidates seem Good to the committee member. This makes the committee member less positive about candidates that seem Good to her, and with enough garbling the committee member does not have incentive to pretend that she observed an Excellent signal instead of a merely Good one (assuming that her bias is not too large). The upside is that the committee can learn the true information of the expert member, while the downside is that the expert's information is of lower quality by identifying some Mediocre candidates as Good and vice versa. But if the required garbling is not too much, it improves welfare relative to an alternative information structure that allocates candidates only into three ranked categories and accurately indicates it to the expert member to which category the candidate belongs.

Our main result is a partial characterization of the optimal information structure for quadratic preferences and general state distributions.[3] Just like in Ivanov (2010b), we can restrict attention to information structures in which the signals are just conditional expectations of the state that are far away from each other such that the Sender is willing to truthfully convey them to the Receiver (implying a finite upper bound on the number of signals, determined by the magnitude of the Sender's bias). We show that the optimal information structure always takes the following form. The state space is partitioned into a finite number of interval regions, and there are two types of regions. The first one is standard information partition cells in which the same signal is generated from every state of the cell and this signal is not generated at any state from outside the cell. The second one is interval regions in which there are two different signals are generated, in a non-monotone fashion, that is the region cannot be divided into two subintervals such that each signal is only generated in one subinterval. Moreover, both these signals are only generated within the region. Therefore the type of garbling in optimal structures is limited to garbling information between two neighboring signals. It cannot involve more complicated non-monotonicities garbling three or more signals. A special case of this class of information structures is the standard partitional structure in which each region is an information cell generating only one signal. We also show that the lowest and highest region in an optimal information structure always has to be

[3]We are motivated by examples in which the information designer is the Receiver, but with quadratic preferences the optimal information structures are the same for the Sender and the Receiver: they are the ones that minimize the conditional variance of the state around implemented actions.

an information cell generating only one signal, and the Sender's incentive compatibility constraints must bind around a garbling region.

The partial characterization result can also be used to exactly characterize the optimal information structure for concrete specifications of the cheap talk game. We do this for the uniform-quadratic specification of the game for biases that are not too small. We show that regions of parameter values alternate for which the optimal information structure is monotone partitional and for which the optimal information structure requires regions with garbled information.

Our analysis builds on the graphical characterization of all feasible information structures in terms of integrated cumulative distribution functions of conditional expectations (posterior means) in Gentzkow and Kamenica (2016), which in turn combines insights from Blackwell (1953) and Rothschild and Stiglitz (1970). In our setting with a finite number of signals (implied by the incentive compatibility conditions), these cumulative distribution functions are piecewise linear, making it easier to work with them. The analysis shows that the graphical characterization of Gentzkow and Kamenica can help solving information design problems with large state and action spaces in more general settings than that they considered – it is not necessary to restrict the Sender's preferences over the Receiver's actions to be state-independent. Also, certain graphical features of integrated cumulative distribution functions can be associated with the features of information structures, allowing the graphical method to be utilized when the number of posterior means is larger than what Gentzkow and Kamenica originally shown.

We further demonstrate the usefulness of working with integrated cumulative distribution functions by showing how it can be used in solving information design problems in which instead of incentive compatibility constraints on the side of the Receiver, participants face exogenous technological constraints, either in the form of an upper bound on the number of messages that can be used in communication, or an upper bound on the entropy of the information structure (motivated by having a budget constraint on communication in a setting where more informative structures are more costly).

A paper using similar techniques as ours, in a different context that combines information design and voluntary disclosure, is Shishkin (2024). In particular it investigates a setting in which Sender is trying to persuade Receiver to accept a project, and the signal realization produces a hard evidence with some probability. The Sender then can decide whether to disclose this hard evidence. Our setting differs from this in many aspects, one being that Receiver has a much richer action space, and we demonstrate that the techniques introduced in Gentzkow and Kamenica (2016) can be used in such settings too. More generally, our work builds on insights from the literature on Bayesian persuasion starting with Rayo and Segal (2010) and Kamenica and Gentzkow (2011). In the latter literature Kolotilin (2018), Dworczak and Martini (2019), Ivanov (2021) and Mensch (2021) identify sufficient conditions for the optimality of monotone partitional information structures.

A recent paper investigating a similar framework as our benchmark model, except featuring costly information acquisition, is Kreutzkamp (2023). It derives similar results to the current paper, but using completely different techniques, from the literature on extreme points and majorization,[4] instead of working with the graphical representetion of feasible information structures as the current paper. Kreutzkamp and Lou (2024) further build on these techniques in a variation of the model in which the Sender, after choosing an information structure, can cheat either by manipulating the realization of the signal or lie about the realization. Further relatedly, Lou (2023) investigates a setting in which a principal delegates decision-making to an expert but can both control the expert's information and restrict the set of actions that can be chosen and shows that under some asssumptions the optimal information structure is monotone partitional. The current paper does not consider delegation and assumes cheap talk communication following the information design phase.

Another paper investigating optimal information design in a cheap talk setting is Krähmer (2021). The main difference between the approaches is that the latter paper allows the principal to design information structures that give him information about the realization of a random variable that the agent doesn't learn. Equivalently, he can randomize between information signals for the expert in a way that the expert does not know the scope of information she can possibly learn, but the principal does. We instead focus on situations in which the expert has substantially superior knowledge and expertise relative to the principal, and assume that it is not feasible for the principal to randomize the information structure in a way that provides him extra information relative to the expert. That is, the outcome of any randomization in the mechanism is either not observed directly by either party, or if it is observed by the principal then it is also observed by the expert: the latter, using her superior knowledge of the environment, fully recognizes what information structure got randomly selected. In Krähmer (2021), the principal can use his superior information to cross-check reports from the agent, which allows him to achieve the first best, full revelation of information.[5] In contrast, in our model information design improves over communication with unconstrained information on the side of the expert, but the principal's payoff is bounded away from the first best.

2 The Model

There are two players, a Sender and a Receiver. The state of the world is stochastic, distributed according to a commonly known prior distribution F on support $\Theta = [0, 1]$. Suppose that F has a continuous density function f bounded away from 0. Let μ be the prior mean.

The timeline of the game is as follows. At the beginning of the game, the Receiver first chooses

[4] See Kleiner et al. (2021) for an introduction to these techniques.

[5] See also Watson (1996) for a similar construction.

an information structure $\pi : \theta \to \Delta(S)$, where S is the set of possible signal realizations. The true state of the world θ then realizes according to F, and a signal realization is sent to the Sender according to $\pi(\cdot \mid \theta)$. The Sender then send a message $m \in M$ to the Receiver, where M is the set of possible messages. After receiving m, the Receiver then chooses an action $y \in \mathbb{R}$. That is, the information control game is a classic cheap talk game, with an additional information design layer at the beginning. Compared with the classic cheap talk model, this additional information design layer now grants some power to the Receiver to control the information.

Given the state θ and the action a, the Sender and the Receiver's utilities are quadratic,

$$U^S(y, \theta, b) = -[y - (\theta + b)]^2 \quad \text{and} \quad U^R(y, \theta) = -(y - \theta)^2,$$

where b represents the Sender's state-independent bias in the perspective of the Receiver. Without loss of generality, we assume that $b > 0$. As Lou (2023) pointed out, and as we will show in details later, such preferences are equivalent to preferences that are linear in states. For example, we can replace the Receiver's utility as

$$V^R(y, \theta) = 2\theta y - y^2.$$

This equivalence offers special convenience when considering the information design problem (Gentzkow and Kamenica, 2016). The key implication is that we can consider the distribution of posterior means in the information design phase, instead of the complete distribution of posterior beliefs. This reduces the dimension of posteriors to be considered and in turn simplifies the problem.

The equilibrium concept used is perfect Bayesian equilibrium (PBE). In general, a Sender's strategy is a map from the set of signal realizations S to the set of probability distributions $\Delta(M)$. A Receiver's strategy consists of two parts, an information structure π, and a map from the set of messages M to the set of probability distributions $\Delta(Y)$ on the set possible actions Y. But we can vastly reduce the generality of this problem due to the quadratic preferences. First, there is no need to consider mixed actions. For any posterior belief induced by a signal realization, the Receiver has a unique optimal action given the quadratic preference. Second, as Ivanov (2010b) shows, a revelation-principle type of argument establishes that we can without loss of generality consider truth-telling equilibria in which the Sender truthfully reveals her posterior mean after receiving a signal realization. Given that the Sender reports the truth, the Receiver then should always choose the posterior mean reported as the optimal action. For this reason, it is also equivalent to consider the Sender as the information designer in this game. One can consider the Sender recommending actions to the Receiver by truthfully reporting the posterior means, and the Receiver is willing to follow the recommendations.

In order for the Sender to truthfully report her posterior mean, the information structure must be finite, meaning that it can generate at most finitely many posterior means. To see this, suppose that x and x' are two of the posterior means generated under π. In order for the Sender who gets the posterior mean x not to report x', it must be the case that

$$|x - x'| \geqslant 2b.$$

Since x and x' are arbitrarily chosen, this means there must be at least $2b$ distance between any two posterior means, which further implies that one can only fit finitely many posterior means in $[0, 1]$. Therefore, it suffices to consider information structures that induce finitely many posterior means. For a given bias b, let $N_{\max}(b) = \lceil 1/2b \rceil$ be the maximum number of posterior means that can be fitted within $[0, 1]$. For an information structure π that induces N posterior means, let the posterior means be ordered

$$0 < x_1 < \cdots < x_N < 1.$$

Although the Receiver can use as many signal realizations as possible to induces these posterior means, it suffices for the Receiver to use N signal realizations. In particular, the Receiver can combine the signal realizations that induces the same posterior means into a single new signal realization. Let s_i be the signal realization that induces x_i. When looking for the optimal information structure, we can restrict attention to structures satisfying the incentive compatibility constraints in terms of posterior means

$$x_i - x_{i-1} \geqslant 2b, \quad i = 2, \cdots, N, \tag{IC}$$

that maximizes the ex-ante expected utility of the Receiver

$$\mathbb{E}U^R(\pi) = \sum_{i=1}^{N} \int_{\Theta} -(x_i - \theta)^2 f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta. \tag{1}$$

Notice that truthful reporting by the Sender, and the optimal action choice of the Receiver given the truthful report, are already incorporated in (1).

Before we start the formal analysis, note that our goal is to find a Receiver-optimal information structure in terms of ex-ante payoffs. Given the continuous state space, there are some technicalities that need to be mentioned. For one thing, for an information strcuture π, one can change $\pi(s_i \mid \theta)$ for some $\theta \in \Theta_o$ without changing the ex-ante payoffs as long as Θ_o has measure 0. It should be understood later that if we claim an optimal structure must have certain features, we have ignored such alternations. For another, in a similar fashion, one can add more signal realizations that have zero probability to be realized into an information strcuture without changing the ex-ante payoffs. For this reason, we assume that in all information structures considered below, all signal realizations can be realized with positive total probabilities.

3 Mixing Signal Realizations: An Example

Crawford and Sobel (1982) has shown that the optimal structure in the canonical cheap talk model is a partition structure. If this is still the case in an information control model, the analysis can be vastly simplified. Nevertheless, in this section, we provide a simple example to show that even with a simple prior such as $\mathrm{Uniform}[0,1]$, a partition structure may not be optimal.

As a preparation, we first rewrite the Receiver's ex-ante expected payoff. We can write (1) as

$$\begin{aligned}
\mathbb{E}U^R(\pi) &= \sum_{i=1}^{N}\int_{\Theta}\left(2x_i\theta - x_i^2 - \theta^2\right) f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta \\
&= \sum_{i=1}^{N}\int_{\Theta}\left(2x_i\theta - x_i^2\right) f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta - \sum_{i=1}^{N}\int_{\Theta}\theta^2 f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta \\
&= \sum_{i=1}^{N}\int_{\Theta}\left(2x_i\theta - x_i^2\right) f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta - \int_{\Theta}\theta^2 f(\theta)\mathrm{d}\theta.
\end{aligned}$$

Observe that in the last line, the second term does not depend on the information structure, so that it can be omitted when we only aim to compare different structures. This also implies what actually matters is the $2x_i\theta - x_i^2$ part.

Consider the integral that depends on information structures. The first term is

$$\int_{\Theta} 2\theta x_i f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta = 2x_i w_i \int_{\Theta}\theta\frac{f(\theta)\pi(s_i \mid \theta)}{w_i}\mathrm{d}\theta = 2w_i x_i^2.$$

where

$$w_i \equiv \int_0^1 \pi(s_i \mid \theta) f(\theta)\mathrm{d}\theta$$

is the overall probability, or the *weight* of signal realization s_i is sent according to π.

The second term is

$$\int_{\Theta} -x_i^2 f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta = -x_i^2\int_{\Theta} f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta = -w_i x_i^2.$$

Substitute these results back to (1),

$$\mathbb{E}U^R(\pi) = \sum_{i=1}^{N} w_i x_i^2. \tag{2}$$

The Receiver's goal is to choose an information structure that maximizes (2).

We now consider an example that shows mixing between different signal realizations in some states can improve the Receiver's ex-ante expected payoff. Let F be $\mathrm{Uniform}[0,1]$, and $b = 0.126$. Notice that the bias is just slightly larger than $1/8$, implying that it is impossible to fit four partition cells while maintaining (IC). To see this, suppose x_i are all generated by partition cells $(a_{i-1}, a_i]$, with $a_0 = 0$ and $a_4 = 1$.[6] (IC) implies that

$$x_2 - x_1 = \frac{a_2 + a_1}{2} - \frac{a_1 + a_0}{2} \geqslant 2b,$$

[6]Throughout this paper, following the notation in Crawford and Sobel (1982), we will use a_i to denote partition points.

where the first equality follows from uniform prior. This shows

$$a_2 - a_0 = a_2 \geqslant 4b.$$

Similarly, since $x_4 - x_3 \geqslant 2b$, we also have

$$a_4 - a_2 = 1 - a_2 \geqslant 4b.$$

Together, we have $8b \leqslant 1$, which shows that if $b > 1/8$, a partition structure with four posterior means cannot satisfy all (IC) constraints.

However, if one allows mixing signal realizations in some states, fitting four posterior means is still possible. For instance, we can consider the following structure:

- Partition cells $[0, a_1]$ and $[a_3, 1]$ induce the first and the last posterior means, x_1 and x_4, respectively. Ideally, to fit four posterior means, the Receiver wants to push a_1 closer to 0 and a_3 closer to 1 compared with a partition structure.
- In the middle region $[a_1, a_3]$, the states can send s_2 or s_3, inducing x_2 and x_3, respectively.
- For calculation simplicity, let $x_2 - x_1 = 2b$ and $x_4 - x_3 = 2b$.

This structure is shown in Figure 1.

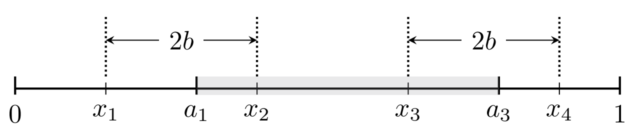


FIGURE 1 A Class of Information Structures with Mixing

One particular information structure that fits the description above is as follows.

- $a_1 = 0.244$, $a_3 = 0.756$, so that $x_1 = 0.122$ and $x_4 = 0.878$.
- For states between 0.244 and 0.4446, s_2 is sent for sure. For states between 0.5554 and 0.756, s_3 is sent for sure.
- For states between 0.4446 and 0.5554, both s_2 and s_3 are sent with positive probability. The probability of sending s_2 decreases linearly from 1 at $\theta = 0.4446$ to 0 at $\theta = 0.5554$.

This structure is visualized in Figure 2. The horizontal axis represents the states, the vertical axis represents the probabilities signals are sent at a state. Later, we will refer to mixing part of a structure similar to the mixing within $[a_1, a_3]$ as *mixing regions*.

We can then use (2) to calculate the Receiver's payoff, which is approximately 0.328. The best partition structure, since four-cell partition is not possible, is the three-cell even-partition structure that induces posterior means $1/6, 1/2, 5/6$ with equal probability. The Receiver's payoff in this case is $35/108 \approx 0.324$. Therefore, the best partition structure is worse than the mixing structure considered above. We will revisit this example in section 6 after introducing some more general results regarding optimal information structures.

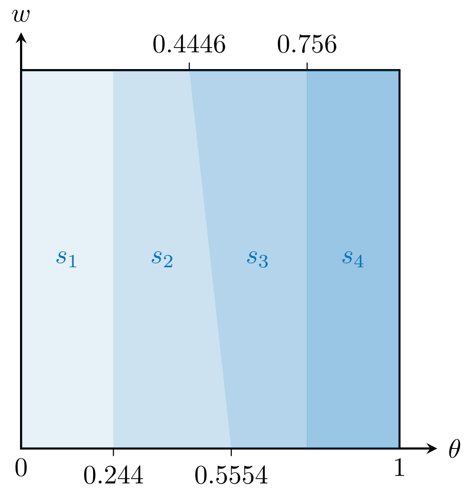


FIGURE 2 An Example Structure

4 The Integrated CDF Framework

The example above shows that it is not sufficient to consider partition structures only. As mixings may lead to higher payoffs, we necessarily need to formally consider the Receiver's optimization problem in its full generality.

4.1 Feasibility

Gentzkow and Kamenica (2016) pointed out that not every distribution of posterior means $\{(x_i, w_i)\}$ such that $\sum w_i x_i = \mu$ can be generated by an information structure and provided a tool – the integrated CDFs – to identify posterior mean distributions that can be induced by information structures. Let G be some potential distribution of posterior means. The integrated CDF is

$$c(x) = \int_0^x G(t)\mathrm{d}t.$$

There are two extreme cases: the most informative information structure that reveals the true state, denoted by $\overline{\pi}$, and the least informative information structure that reveals no information at all, denoted by $\underline{\pi}$. The distribution of the posterior means of the former is just the prior F, while the latter is a degenerate distribution with the singleton support at μ. Therefore, the integrated CDFs are

$$c_{\overline{\pi}}(x) = \int_0^x F(t)\mathrm{d}t \quad \text{and} \quad c_{\underline{\pi}}(x) = \begin{cases} 0 & \text{if } 0 \leqslant x < \mu, \\ x - \mu & \text{if } \mu \leqslant x \geqslant 1. \end{cases}$$

Then, G is feasible if and only if $c(x)$ is a convex function between $c_{\underline{\pi}}$ and $c_{\overline{\pi}}$ everywhere in $[0, 1]$.

In our case, we only need to consider information structures that induce finite posterior means.

Given $\{(x_i, w_i)\}$, the integrated CDF is a $(N+1)$-segment piecewise linear function

$$c_\pi(x) = \begin{cases} 0 & \text{if } 0 \leqslant x < x_1, & \cdots & \text{Segment } 0 \\ w_1 x - w_1 x_1 & \text{if } x_1 \leqslant x < x_2, & \cdots & \text{Segment } 1 \\ \vdots & & & \\ \left(\sum_{j=1}^{i} w_j\right) x - \sum_{j=1}^{i} w_j x_j & \text{if } x_i \leqslant x < x_{i+1}, & \cdots & \text{Segment } i \\ \vdots & & & \\ x - \sum_{j=1}^{N} w_j x_j & \text{if } x_N \leqslant x \leqslant 1, & \cdots & \text{Segment } N. \end{cases} \tag{3}$$

Notice that the segments are numbered starting from 0. Segment 0 is constantly 0 for all integrated CDFs, so that the second segment, Segment 1, is the first "meaningful" segment.[7] Formally, Segment k is the line segment between $(x_k, c_\pi(x_k))$ and $(x_{k+1}, c_\pi(x_{k+1}))$, with $x_0 = 0$ and $x_{N+1} = 1$.

4.2 Connections between the Integrated CDF and the Objective Function

Following the observation that $c_{\overline{\pi}}$ is strictly above $c_{\underline{\pi}}$, it is reasonable to conjecture that the relative locations of the integrated CDF has implications on the informativeness of information structures. This is indeed the case since the integrated CDFs are strongly related to the concept of *mean preserving spread* (See, for example, Machina and Pratt (1997)). Here, by making a "spread", the signal becomes more informative, and at the same time, the integrated CDF become higher. This intuition then lead us to the idea that the area below the integrated CDF may related to the Receiver's payoff.

Proposition 1. The Receiver's expected utility is higher if the area below c_π is larger.

The proof is algebraic and in Appendix A. This is a stronger result than mean-preserving spread since it does not require the integrated CDF of one distribution to be higher than the other pointwise. Notice that this means we can make yet another transformation of the objective function. Starting from (1), apart from the already simple algebraic expression in (2), we can also use the geometric result in Proposition 1 to compare two information structures.

[7] Observe that the last segment, which is just $x - \mu$, is also the same among all integrated CDFs. In other words, the piecewise linear integrated CDF must start and end on $c_{\underline{\pi}}$.

5 Features of Optimal Structures

The integrated CDF framework, in particular, Proposition 1, enables us to obtain some useful insights for characterizing the optimal information structures.

5.1 Graphical Observations

Geometrically, the Receiver's task is to find an $(N+1)$-piece piecewise linear convex function between $c_{\overline{\pi}}$ and $c_{\underline{\pi}}$ that has as large area below as possible, while keeping (IC) by separating each pair of posterior means at least $2b$ away. This naturally means that the Receiver should push the segments up so that they are tangent to $c_{\overline{\pi}}$ – the highest positions possible. A special case is $N_{\max}(b) \leqslant 3$, in which case tangency must be achieved at each segment of the integrated CDF. Formally,

Lemma 2. In an optimal information structure, Segments 1 and $N-1$ must be tangent to $c_{\overline{\pi}}$. If at most two or three posterior means can be induced in all truth-telling information structure, then the integrated CDFs of all optimal information structures must be tangent to $c_{\overline{\pi}}$ at all segments.

Proof. Here we consider the case that there can be at most three posterior means, which shows the idea. Since Segments 0 and 3 are fixed and tangent to $c_{\overline{\pi}}$ at $x = 0$ and $x = 1$, respectively, we only need to consider Segments 1 and 2. Suppose Segment 1 is not tangent to $c_{\overline{\pi}}$, as shown in the left panel of Figure 3. In Figure 3, the opaque lines are $c_{\overline{\pi}}$ (above) and $c_{\underline{\pi}}$ (below), and the solid line is the integrated CDF being examined.

Now consider the following modifications. Move x_1 to the left slightly, as shown in the right panel of Figure 3. For small enough change, the new Segment 1 is still under $c_{\overline{\pi}}$. The IC condition is preserved since the distance between x_1 and x_2 are larger, and finally the area below the integrated CDF is larger, implying that the ex-ante payoff of the Receiver is higher.

Similar argument can be made if Segment 2 is not tangent to $c_{\overline{\pi}}$, in which case one can move x_3 slightly to the right. ■

When more possible posterior means are possible, we cannot claim that all segments are tangent to $c_{\overline{\pi}}$ due to IC constraints. Still, there are some observations we can make regarding segments that are not tangent to $c_{\overline{\pi}}$, if such segments can ever appear in an optimal structure. Consider an optimal information structure $\pi^\star$ with N posterior means. Let $c_{\pi^\star}$ be the corresponding integrated CDF. Moreover, we have the following two observations.

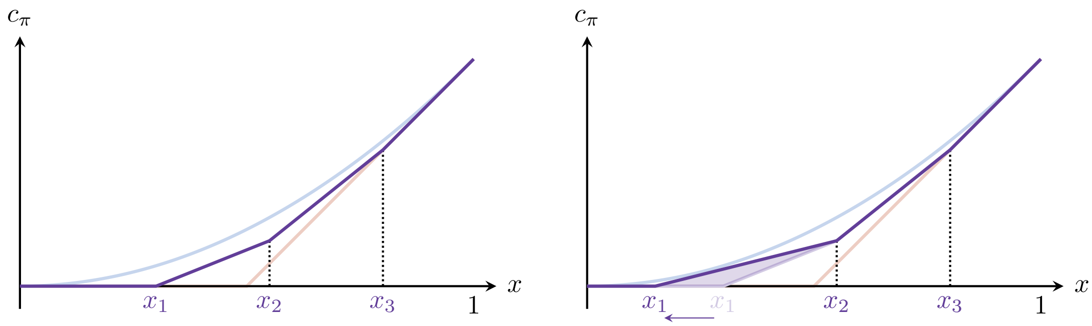


Figure 3 Proof of Lemma 2

Lemma 3. Let π be an optimal structure. $c_{\pi^\star}$ cannot have two consecutive segments not tangent to $c_{\overline{\pi}}$.

Proof. See the left panel of Figure 4. Consider the modification that increases the vertical coordinate of $c_{\pi^\star}$ at x_i slightly. For small increases, the new $c_{\pi^\star}$ is still convex and under $c_{\overline{\pi}}$, yet the area below the new $c_{\pi^\star}$ is strictly larger. Notice that the (IC) condition is still perserved since the location of x_i's are not changed under this modification. ∎

Lemma 4. Let π be an optimal structure. If Segment i in $c_{\pi^\star}$ is not tangent to $c_{\overline{\pi}}$, (IC) must bind at Segment $i-1$ and $i+1$.

Proof. See the right panel of Figure 4. Suppose $x_{i+2} - x_{i+1} > 2b$. Consider the modification that moves x_{i+1} slightly to the right. That is, the modification fixes the locations of x_{i-1}, x_i and x_{i+2} but raises the slope of Segment i, so that the intersection of Segment i and $i+1$ is now higher and more to the right. Since the IC constraint between x_{i+1} and x_{i+2} is not binding, if the move is small, (IC) is not violated anywhere, and the area below the new $c_{\pi^\star}$ is strictly larger. ∎

5.2 Partitions

The results above show that the integrated CDF of optimal structures are repeatedly tangent to the upper bound, $c_{\overline{\pi}}$. This leads to the question: what does tangency mean for an information structure? It turns out that tangency is closely related to a feature of the information structure being quasi-partitional.

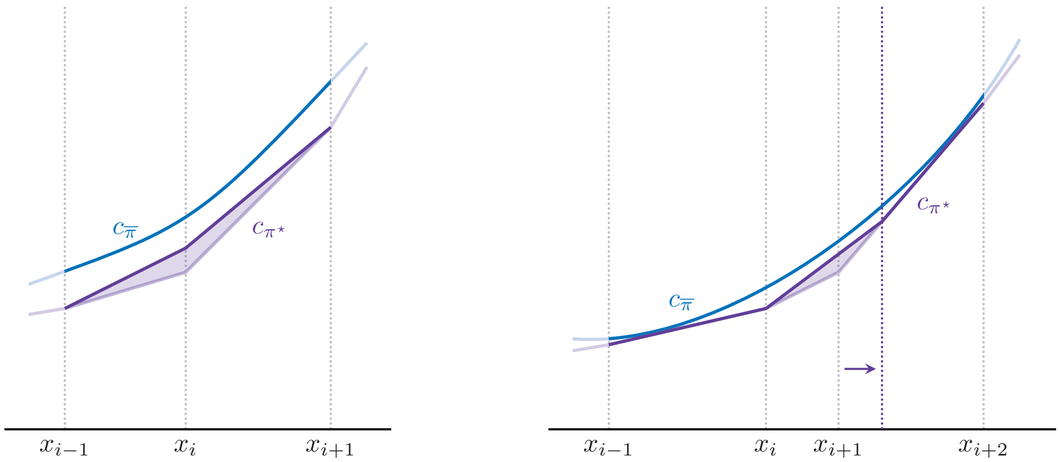


FIGURE 4 Proofs to the Graphical Observations

Proposition 5. Let π be a finite information structure. The following are equivalent:

- Segment i of the integrated CDF of π is tangent to $c_{\overline{\pi}}$ at $(m, c_{\overline{\pi}}(m))$.
- The signal realizations in π never cross over m: a signal realization is sent only from states below m or above m, but never both.

An example is pictured in Figure 5 using the prior $\text{Uniform}[0, 1]$. The corresponding partition structure is the even-partition structure, with the partition point at $1/3$ and $2/3$. The three posterior means are $x_1 = 1/6$, $x_2 = 1/2$ and $x_3 = 5/6$, respectively.

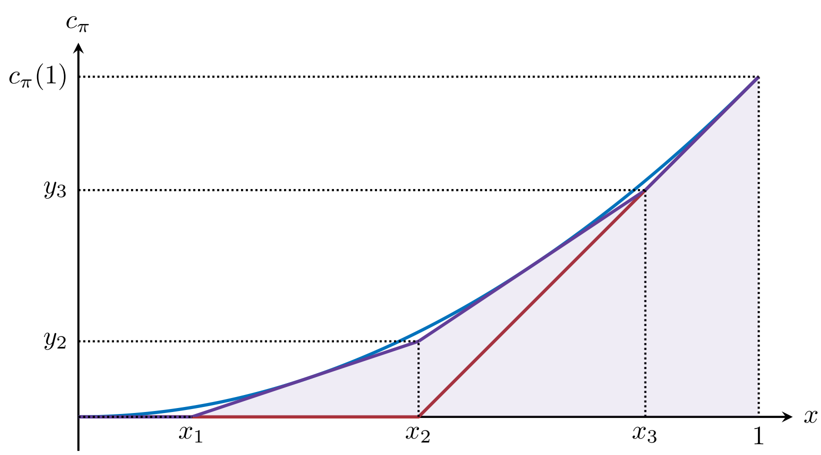


FIGURE 5 An Example with Three Possible Posterior Means

The proof of Proposition 5 is in Appendix A. The intuition is that whenever there is a tangency point $(m, c_{\overline{\pi}(m)})$, one can "cut" the integrated CDF at $x = m$, so that the integrated CDF on $[0, m]$ and the integrated CDF on $[m, 1]$ both become a new qualified integrated CDF that represents the same information structure as before restricted to the corresponding subinterval. Proposition 5 is the bridge between the graphical observations of the integrated CDF and the features of the information structure being represented. Using the graphical observations in Lemma 2, Lemma 3 and Lemma 4, together with Proposition 5, we have the following *partial characterization* of the optimal information structures.

Theorem 6. For any optimal information strcuture of the Receiver:

- There are only two types of components in the structure: partition cells that contain one posterior mean each, or "mixing regions" that contain two posterior means each.
- If the optimal strcuture contains a mixing region that induces x_i and x_{i+1}, then the IC constraints between x_i and x_{i-1}, and between x_{i+2} and x_{i+1} must bind.
- Each posterior mean will not be induced by any state outside its cell or its mixing region.
- The optimal strcuture must start and end with partition cells.

This result states that there are only two types of possible components in an optimal structure and indicates additional features regarding how the structure starts or ends, as well as how the structure behaves around mixing regions. The first feature corresponds to the graphical observation that there are no two consecutive segments of the integrated CDF of an optimal information structure that are both not tangent to $c_{\overline{\pi}}$. For example, if Segment $i-1$ and Segment i are both tangent (at a_i and a_{i+1}, respectively), then x_i is induced by the partition cell $[a_i, a_{i+1}]$. If Segment i of the integrated CDF is not tangent, then Segments $i-1$ and $i+1$ must be tangent, and x_i and x_{i+1} are induced by a mixing region. The second feature also corresponds to the observation that states the IC constraint must bind around a non-tangent segment. The third feature is directly from Proposition 5. The last feature is an implication of Lemma 2.

We should note that the result that IC conditions are binding around mixing regions does not depend on prior distributions (although placing a mixing region in a particular place depends on the prior distribution.). Intuitively, mixing is inferior compared with partition cells by generating less clear signal realizations; that is, the posterior means are less "spread". Without binding IC conditions around it, the Receiver would have some flexibility to reduce the extent of mixing and in turn improve the informativeness of the structure by increasing the extend of spread within the mixing region.

Using Lemma 2 and Theorem 6, we in particular observe that if bias is relatively large and $N_{\max} \leqslant 3$, the optimal structure does not involve mixing.

Corollary 7. If $b \leqslant 1/6$, then the optimal information structure must be a partition structure.

Another implication of Theorem 6, together with the observation that the Receiver's payoff in (2) only depends on $(w_i, x_i)_{i=1}^N$ is as follows.

Corollary 8. For each optimal strcuture, there is a payoff-equivalent structure (hence also optimal) such that a deterministic signal realization is sent at each state.

The proof is again delegated to Appendix A. To illustrate the idea, let's consider a mixing region $[\ell, h]$ that induces x_i and x_{i+1}. Suppose x_i is induced with weight w_i. Pick some $\tilde{\ell} \in [\ell, h]$ and find $\tilde{h}$ such that $\int_{\tilde{\ell}}^{\tilde{h}} f(\theta)\mathrm{d}\theta = w_i$. If we pick $\tilde{\ell} = \ell$, $[\tilde{\ell}, \tilde{h}]$ is the set of the lowest states within $[\ell, h]$ that has weight w_i. Therefore, when sending s_i in $[\tilde{\ell}, \tilde{h}]$, the induced posterior mean $\tilde{x}_i$ must be weakly smaller than x_i. Similarly, if we pick $\tilde{\ell}$ such that $\tilde{h} = h$ and send s_i in $[\tilde{\ell}, \tilde{h}]$, these states are the highest states within $[\ell, h]$that has weight w_i, so that the induced posterior mean $\tilde{x}_i$ must be weakly larger than x_i. In turn, by the intermediate value theorem, there must be some $\tilde{\ell}$ in between such that exactly x_i is induced. Since we only have two signal realizations, what left, $[\ell, \tilde{\ell}] \cup [\tilde{h}, h]$ must have weight w_{i+1} and induces x_{i+1}.

Of course, any structure that induces this distribution will be optimal, and there is no need to use a "linear" probability, or "straight line division" between two signal realizations involved in a mixing. For example, the following structure in Figure 6 is a four-signal structure such that when a state mixes between s_2 and s_3, they are mixed with equal probability. One can verify that this structure is payoff-equivalent to the structure in Figure 2.

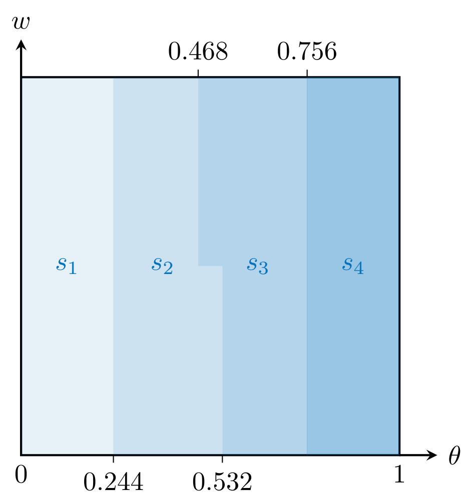


FIGURE 6 Another Example of Four-Signal Optimal Structures

6 Full Characterization of Optimal Information Structure with Uniform Prior

In this section, we show how the partial characterization of the optimal information structures from Theorem 6 can help in finding the optimal structure by revisiting the uniform prior example first introduced in section 3. We will also discuss scenarios where more posterior means (i.e., signal realizations) are possible.

6.1 Four-Posterior-Mean Example Revisited

First, notice that the class of the structures introduced in section 3 indeed satisfy all the features in Theorem 6. Also, since the mixing regions cannot be placed at the beginning or the end of the support, the structure is also the *only* possible class of structure that is optimal and involves mixing.

Since the structure begins and ends with paritition cells, the total probability weights that s_1 and s_4 are sent (or equivalently, that x_1 and x_4 are induced) are

$$w_1 = 2x_1 \quad \text{and} \quad w_4 = 2(1 - x_4).$$

To calculate the other two weights, the following two relations are used:

$$\sum_{i=1}^{4} w_i = 1 \quad \text{and} \quad \sum_{i=1}^{4} w_i x_i = \frac{1}{2}.$$

Substitute w_1 and w_4 back and solve the system of equations to get

$$\begin{aligned} w_2 &= \frac{\frac{1}{2} - x_3 - 2x_1(x_1 - x_3) - 2(1 - x_4)(x_4 - x_3)}{x_2 - x_3}, \\ w_3 &= \frac{\frac{1}{2} - x_2 - 2x_1(x_1 - x_2) - 2(1 - x_4)(x_4 - x_2)}{x_3 - x_2}. \end{aligned}$$

Further substitute $x_2 = x_1 + 2b$ and $x_3 = x_4 - 2b$ to get

$$\begin{aligned} w_2 &= \frac{(2x_4 - 2x_1 - 1)(1 - 2x_1 - 4b)}{2(x_4 - x_1 - 4b)}, \\ w_3 &= \frac{(2x_4 - 2x_1 - 1)(2x_4 - 4b - 1)}{2(x_4 - x_1 - 4b)}. \end{aligned} \tag{4}$$

The algebra here shows that the Receiver in fact only needs to consider two parameters: x_1 and x_4. All other parameters in the optimization can be written as functions of x_1 and x_4. This is due to the fact that the optimal information structure is already largely pinned down by Theorem 6.

The Receiver's optimization problem is to find the optimal four-signal structure described above and to compare it with the optimal three-signal structure, which is the three-cell even partition structure.

Theorem 9. Let $1/8 < b < 1/6$. There exists $\bar{b} \in (1/8, 1/6)$ such that if $b < \bar{b}$, the Receiver's optimal information structures is a four-signal structure that satisfies:

- The incentive compatibility constraint is binding everywhere.
- The first posterior mean is $x_1 = \frac{1}{2} - 3b$.

- Both x_1 and x_4 are induced by partition cells, while x_2 and x_3 are induced by a mixing region.

If $b > \bar{b}$, the optimal structure is the three-cell even-partition structure. Both structures are optimal if $b = \bar{b}$.

The proof uses arguments involving standard Kuhn-Tucker conditions and is in Appendix A. If we set $b = 0.126$, the optimal structure is the one presented in section 3.

It may be helpful to consider why mixing could be useful in more details. For now, let's assume that IC is binding everywhere. For a given pair (x_2, x_3) such that $x_3 = x_2 + 2b$, if it is generated by a symmetric[8] mixing region, then mixing feasibility – x_2 and x_3 can be generated by a mixing region, meaning that x_2 and x_3 cannot be too far away – implies $x_2 - a_1 > b$ and $a_3 - x_3 > b$. This in turn shows that $a_1 - x_1 < b$ and $x_4 - a_3 < b$, so that the first and the last cells are smaller. This shows that (i) the mixing region itself does not save space; instead, it captures more space (size larger than $4b$) than two partition cells (total size equal to $4b$); and (ii) the mixing region fits more signal realizations by shifting the conditional expectations of the state after signals in the mixing region closer to each other, away from the state boundaries.

If $b = 1/8$, the structure is the four-cell even-paritition structure. When b is slightly larger, as shown in Figure 1 or Figure 6 for $b = 0.126$, the optimal four-signal structures are very close to the four-cell even-partition structure. Intuitively, the partition structures with more cells minimize the conditional variances, and the even partitions then guarantee all states are valued equally, which is ideal given the uniform distribution. Therefore, the Receiver prefers the four-cell even-partition structure if possible. Yet when $b > 1/8$, the IC constraints rule out the four-cell even-partition structure, so that the best the Receiver can do is to mimic the four-cell even-partition structure by adding minimal mixing in the middle and slightly squeezing the boundary cells. However, if b is significantly larger and closer to $1/6$, the mixing structures are less resemble to the four-cell even-parititon structure, meaning that the Receiver has to mix heavily in the middle and reduce the size of the the boundary cells by a lot. The mixing and uneveness then lead to worse payoffs than having one fewer signal realizations, and the three-cell even-partition structure indeed performs better.

The same logic applies when b is slightly smaller than $1/8$. In this case, a five-signal structure is available, but it is rather uneven. The Receiver would like to use the four-cell even-partition strcuture instead. As b further decreases, the Receiver would switch to a five-signal structure. This is shown in Figure 7.

Theorem 9 also shows in general how one can use Theorem 6 to solve optimal information

[8] In the sense that $x_2 - a_1 = a_3 - x_3$ and the two posterior means are induced with the same weight. There is no reason to consider asymmetric mixing region given the uniform prior, although we did not impose symmetry when showing Theorem 9. We also did not impose that IC constraints are binding everywhere.

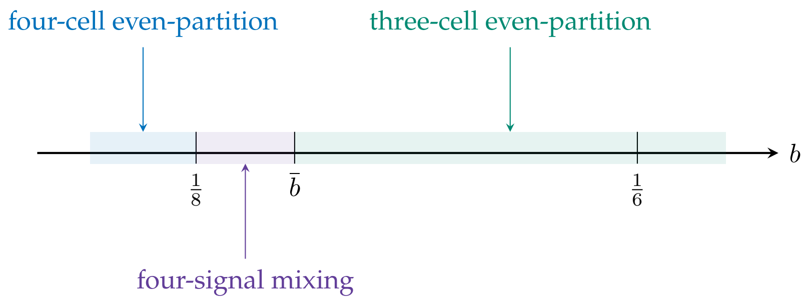


FIGURE 7 Optimal Information Structure of the Receiver for Different Biases

structures. Since the optimal structure can only contains two types of components, the Receiver should first consider the possibilities of numbers and locations of mixing regions (i.e., the classes of structures). In the uniform example, when $N_{\max} = 4$, there is at most 1 mixing region that must locate in the middle. For larger $N_{\max}$, the Receiver could use more mixing regions or place the mixing regions in different locations. The way that Theorem 6 helps simplify the Receiver's problem is by (i) reducing the possibilities of structures that need to be considered, and (ii) reducing the number of choice variables in the optimization problem.

6.2 More Posterior Means

The case with $N_{\max} = 5$. Suppose that $1/10 < b < 1/8$, so that $N_{\max} = 5$. We first note that unlike the the case with $N_{\max} = 4$, the optimal partitional information structure is an uneven partition structure with smaller and larger cells alternating, as shown in Ivanov (2010b).

For general optimal structures, we also need to consider mixing regions. By Theorem 6, as the mixing regions cannot be at the boundaries, there can be at most one mixing region – a region either induces x_2 and x_3 or induces x_3 or x_4. We should note that these two possibilities are symmetric given the uniform prior, so that it is sufficient to consider a mixing region that induces, say, x_2 and x_3. That is,

- x_1 is generated by partition cell $[0, a_1]$, x_4 is generated by partition cell $[a_3, a_4]$, and x_5 is generated by partition cell $[a_4, 1]$.
- x_2 and x_3 is generated in mixing region $[a_1, a_3]$.

An example of such structures is shown in Figure 8, and the states in the shaded area are in the mixing region.

Using the uniform prior,

$$a_1 = 2x_1, \quad a_3 = 2x_4 - 2x_5 + 1, \quad \text{and} \quad a_4 = 2x_5 - 1.$$

Because the IC constraints must bind around the mixing region, $x_2 = x_1 + 2b$ and $x_3 = x_4 - 2b$. In

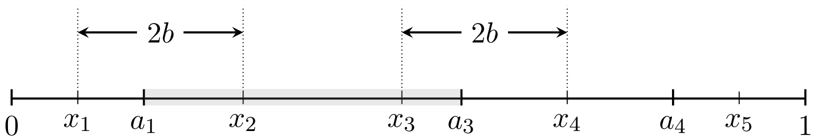


Figure 8 A Five-Signal Example

order for x_2 and x_3 to be able to be generated by an information structure,

$$2x_2 - a_1 \geqslant 2x_3 - a_3,$$

which simplifies to

$$2x_5 - 1 \leqslant 8b. \tag{5}$$

There are also two remaining IC constraints,

$$\begin{aligned} x_5 - x_4 &\geqslant 2b \\ x_4 - x_1 &\geqslant 6b, \end{aligned} \tag{6}$$

where the second equation is the IC constraint $x_3 - x_2 \geqslant 2b$, combined with the two binding constraints. It is also easy to solve the weights,

$$w_1 = 2x_1, \quad w_4 = 4x_5 - 2x_4 - 2, \quad w_5 = 2 - 2x_5.$$

The other two weights are solved by

$$\begin{aligned} w_2 + w_3 &= a_3 - a_1, \\ w_2x_2 + w_3x_3 &= \frac{1}{2}(a_3^2 - a_1^2), \end{aligned}$$

which gives

$$\begin{aligned} w_2 &= -\frac{(4b + 2x_1 - 2x_5 + 1)(2x_1 - 2x_4 + 2x_5 - 1)}{2(4b + x_1 - x_4)}, \\ w_3 &= \frac{(-4b + 2x_4 - 2x_5 + 1)(2x_1 - 2x_4 + 2x_5 - 1)}{2(4b + x_1 - x_4)}. \end{aligned}$$

Notice that the Receiver only has three choice variables: x_1, x_4, and x_5. The Receiver maximizes $\sum_{i=1}^{5} w_i x_i^2$ subject to (5) and (6). Let $\lambda_1, \lambda_2, \lambda_3$ be the three Lagrange multipliers, we can construct the Lagrange function and get

$$\begin{aligned} &x_1 = \frac{1-8b}{2}, \quad x_4 = \frac{1+4b}{2}, \quad x_5 = \frac{1+8b}{2}, \quad \text{and} \\ &\lambda_1 = b - 8b^2, \quad \lambda_2 = 8b(10b-1), \quad \lambda_3 = 4b(10b-1). \end{aligned}$$

Notice that all three Lagrange multipliers are positive given $1/10 < b < 1/8$. This implies all three constraints are binding. In particular, a binding (5) implies the mixing is degenerate, i.e., both x_2 and x_3 are also generated by partition cells. This is the partition structure suggested by Ivanov (2010b). Therefore, the general optimal structure is either this uneven five-cell partition structure when b is closer to $1/10$, or the four-cell even-partition structure when b is closer to $1/8$. We should highlight that with Theorem 6, solving the optimal structure when $1/10 < b < 1/8$ again reduces to a single Lagrange problem with limited number of choice variables.

The case with $N_{\max} = 6$. As the last example, consider $1/12 < b < 1/10$, so that the Receiver can at most induces six posterior means. Notice that there are more possibilities regarding mixing regions. Firstly, if we only consider the structures with one mixing region, it can be placed to induce (x_2, x_3), (x_3, x_4), or (x_4, x_5). Secondly, a larger $N_{\max}$ also allows for more mixing regions. We can now place two mixing regions, one inducing (x_2, x_3), and the other inducing (x_4, x_5).

Given the uniform prior, all states are ex-ante equally important. This implies that there is no reason to consider asymmetric structures. If we consider the case with one mixing region, this suggests that we only need to consider mixing happens in the middle with x_3 and x_4. Also,

- x_1, x_2, x_5 and x_6 are generated by partition cells $[0, a_1]$, $[a_1, a_2]$, $[a_4, a_5]$ and $[a_5, 1]$, respectively.
- The IC condition is binding between x_2 and x_3, and between x_4 and x_5.

However, an easy argument shows that this is impossible. Note that the mixing region must be at least $4b$ in length. The two partition cells that induces x_1 and x_4 must be at least $4b$ length as well since $x_2 - x_1 \geqslant 2b$. The same holds for the two cells induces x_5 and x_6. This suggests the total length of the structure must be at least $12b$, which exceeds 1. A general version of this argument is formalized in Lemma 19 in Appendix B. We should also note that imposing symmetry is not necessary and just for the ease of exposition. In Appendix B, Theorem 21 shows that the optimal structure here cannot contain exactly one mixing region no matter where it is placed.

Therefore, we just need to consider the case that there are two mixing regions:

- x_1 is generated by the partition cell $[0, a_1]$, and x_6 is generated by the partition cell $[a_5, 1]$.
- x_2 and x_3 are induced by a mixing region $[a_1, a_3]$, and x_4 and x_5 are induced by a mixing region $[a_3, a_5]$.
- $x_2 - x_1 = 2b$, $x_4 - x_3 = 2b$, $x_6 - x_5 = 2b$.

An example of such structures is shown in Figure 9, and the states in the two shaded areas are in the two mixing regions.

We need x_1, x_3, x_6, and a_3 to pin down a structure.[9] It can be verified that all x_i's and w_i's can then be expressed using the four choice variables here. There are four constraints. The first two

[9]Of course, there are many other combinations of parameters that can pin down a structure within the family described here. But in general, one need four choice variables. Also, if one only considers symmetric structures, $a_3 = 1/2$ and $x_6 = x_1$, so only two choice variables are needed.

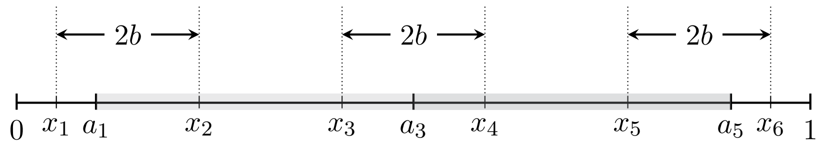


Figure 9 A Six-Signal Example

ensure that the mixing regions can be generated by an information structure:

$$2x_2 - a_1 \geqslant 2x_3 - a_3,$$
$$2x_4 - a_3 \geqslant 2x_5 - a_5.$$

The remaining two are the IC constraints, $x_3 - x_2 \geqslant 2b$ and $x_5 - x_4 \geqslant 2b$. Notice that by Theorem 6, the remaining IC constraints must bind. The optimal solution is

$$x_1 = \frac{1-10b}{2}, \quad x_3 = \frac{1-2b}{2}, \quad x_6 = \frac{1+10b}{2}, \quad a_3 = \frac{1}{2}.$$

This is a symmetric structure that converges to the six-cell even-partition structure as $b \to 1/12$. Again, the Receiver compares this structure with the five-cell even-parittion structure. The former is better when b is closer to $1/12$, while the latter is better when b is closer to $1/10$.

We plot the Receiver's optimal expected payoff under the optimal structures derived above for b not too small, which is the blue curve in Figure 10. As a comparison, the Receiver's optimal expected payoff in canonical cheap talk model without an information design layer is the orange curve. Not surprisingly, when $b < 0.25$ so that the informative equilibria are possible, the optimal structures with information design performs strictly better than the best structures in canonical cheap talk setting, which are always partition structures. In our setting, there are ranges of b where an even-partition structure is optimal, and the optimal expected payoffs for such a range are constant. These are the flat parts of the graph. For other b's, the optimal structure is either a structure with mixing regions or a structure with uneven partitions. In such cases, the way of mixing or the uneveness of the partitions depend on b, and the expected payoff is strictly decreasing in such b. Under the canonical model, Crawford and Sobel (1982) has shown that although the numbers of partition cells in the optimal structures may remain constant for a range of b, the specific partition points depend on b, leading to a curve that is strictly decreasing in b. Moreover, notice that the canonical setting usually allow much fewer cells (regions). For example, for b around $1/12$, the optimal structure with information design has 6 cells (regions), yet the optimal structure without information design has 3 cells when $b < 1/12$, and only 2 cells when $b > 1/12$.

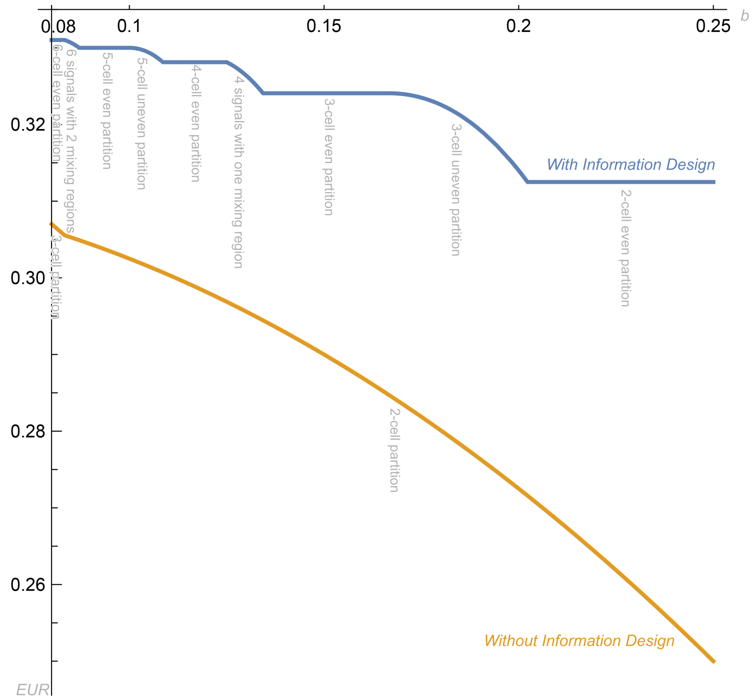


FIGURE 10 Receiver's Optimal Expected Payoff in Models with and without Information Design

7 Exogenous Capacity Constraints on Communication

We will next show that the graphical framework can be useful in more information design problems where the posterior mean approach is applicable. In the information control problem, the Receiver (the principal, or the information designer) faces endogenous constraints, namely, incentive compatibility constraints, that prevent the designer to use a perfectly informative information structure. In what follows, we will consider a principal restricted by exogenous or technical constraints. Such constraints are commonly observed in engineering problems (see Gray and Neuhoff (1998) for a survey). Below we consider the case when there is an upper bound on the number of signals available, and when there is an upper bound on the entropy of the information structure. In both problems we assume that the agent has the perfectly aligned preference as the principal, so the only relevant constraints stem from bounded capacity.

7.1 Upper Bounds on Signal Numbers

Suppose that the principal can only include at most K signals in any information structure, where K is finite. An easy argument establishes that the optimal information structure is partitional using the graphical framework.

Proposition 10. The optimal information structure is partitional.

Proof. Suppose not. Proposition 5 then suggests that the integrated CDF induced by the optimal structure must have a piece that is not tangent to $c_{\overline{\pi}}$, as shown in Figure 11. But then the principal can make a (slight) parallel shift of the non-tangent piece and achieve a strictly better payoff, which is a contradiction. ■

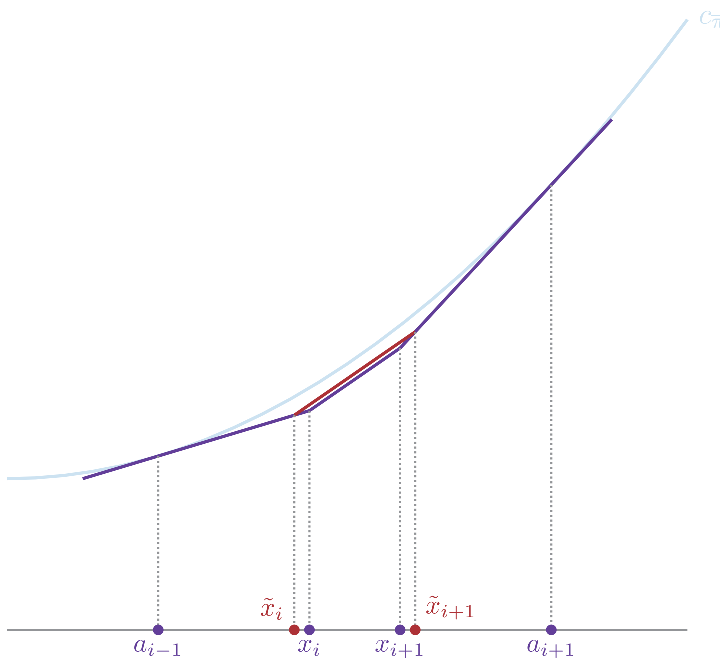


FIGURE 11 Parallel Shift

The Quadratic Utility Example. Once we determine that the optimal information structure is partitional, we can study the problem further by considering, for example, what specific features an optimal information structure have with a specific utility function. Here, we consider the quadratic utility. Since the optimal structure is partitional, we can just consider the K-cell partitional structures characterized by the partition points

$$0 = a_0 < a_1 < \cdots < a_{K-1} < a_K = 1.$$

Under quadratic utility, (2) gives the principal's objective function, which is

$$\sum_{k=1}^{K} \frac{\left[\int_{a_{k-1}}^{a_k} \omega f(\omega)\mathrm{d}\omega\right]^2}{\int_{a_{k-1}}^{a_k} f(\omega)\mathrm{d}\omega}. \tag{7}$$

For a generic $k = 1, \cdots, K-1$, there are only two terms in (7) containing a_k. That is,

$$\frac{\left[\int_{a_{k-1}}^{a_k} \omega f(\omega)\mathrm{d}\omega\right]^2}{\int_{a_{k-1}}^{a_k} f(\omega)\mathrm{d}\omega} + \frac{\left[\int_{a_k}^{a_{k+1}} \omega f(\omega)\mathrm{d}\omega\right]^2}{\int_{a_k}^{a_{k+1}} f(\omega)\mathrm{d}\omega}.$$

Since the density assumed to be positive everywhere, the first order condition can be simplified to

$$a_k = \frac{x_k + x_{k+1}}{2}. \tag{8}$$

This is summarized as follows.

Proposition 11. Suppose that utility is quadratic. In an optimal structure, the partition points must be the average of the two neighboring posterior means.

In the case of quadratic utility and uniform prior, (8) simply gives the even partition structures as expected. However, we should point out that under general prior distributions, the optimal structure does not assign equal weights to each cell, nor does it keep the conditional variances equalized among all the cells.[10]

7.2 Upper Bounds on Information Entropy

A different type of constraint on the information structure that can be used by the Sender is a cap on its entropy, a commonly used measure of the informativeness of the structure. This can be the case if there is a fixed budget for establishing the information structure and more informative structures are more expensive.

For an information structure that induces (discrete) distribution of posterior means $(w_k, x_k)_{k=1}^K$, the information entropy is defined by[11]

$$E(\pi) = -\sum_{k=1}^{K} w_k \ln w_k.$$

Consider the problem where the principal chooses an information structure π to maximizes the expected utility under the induced posterior distribution subject to the constraint

$$E(\pi) \leqslant E. \tag{9}$$

[10] As an example, consider the prior density $f(\theta) = 2 - 2\theta$, and let $K = 2$. The optimal partition point is approximately 0.382, and the weights of the two cells are approximately 0.618 and 0.382, respectively. The conditional variances are approximately 0.012 and 0.021, respectively.

[11] For the ease of notation, we adopt the natural logarithm. Changing the base of the logarithm will not alter the qualitative results below. Also, since the definition of differential entropy (or continuous entropy) is known to be not fully compatible with the discrete version, we will rule out any information structure that induce truth-telling in a non-trivial subset of the state space by setting the entropy of such information structure to be $+\infty$, which is the limit of the entropy of finer and finer partition structures.

Proposition 12. The optimal information structure is partitional. Also, the entropy constraint (9) must bind in an optimal structure.

To see that the optimal structure is partitional, recall that Segment i of the integrated CDF has the expression

$$\left(\sum_{j=1}^{i} w_j\right) x - \sum_{j=1}^{i} w_j x_j.$$

Therefore, as long as the slope of the integrated CDF are not altered, the entropy will not change. Then, the parallel shift argument in the proof of Proposition 10 applies again. For the proof that the entropy constraint must bind, see Appendix A. The idea is also simple – Partitioning an existing cell always provides more information and increases the entropy. The entropy capacity E in (9) then acts as a budget constraint: if the constraint does not bind, the principal can always partition one existing cell further, leading to a higher payoff.

The Quadratic Utility Example. As we now know that the optimal structure is partitional, we again consider the special case where the utility function is quadratic.

The Lagrangian is

$$\mathcal{L} = \sum_{k=1}^{K} \frac{\left[\int_{a_{k-1}}^{a_k} \omega f(\omega)\mathrm{d}\omega\right]^2}{\int_{a_{k-1}}^{a_k} f(\omega)\mathrm{d}\omega} + \lambda \left[E + \sum_{k=1}^{K} \int_{a_{k-1}}^{a_k} f(\omega)\mathrm{d}\omega \ln\left(\int_{a_{k-1}}^{a_k} f(\omega)\mathrm{d}\omega\right)\right].$$

The first order condition with respect to a_k simplifies to (again using $f(\cdot) > 0$ everywhere)

$$a_k = \frac{x_k + x_{k+1}}{2} + \lambda \frac{\ln w_k - \ln w_{k+1}}{x_{k+1} - x_k}. \tag{10}$$

Comparing this with (8), i.e., the optimal a_k when the only constraint is that the number of cells, the first term of (10) is exactly (8), while there is a second term adjusting the partition points to satisfies the entropy constraint. Due to the fact that (9) always binds in optimum, $\lambda > 0$. Since $x_{k+1} > x_k$ by construction, the sign of the second term is determined by the sizes of w_k and w_{k+1}.

However, the optimal structure can be more easily solved by reformulating the objective function. Importantly, since the values of posterior means do not enter the entropy, the problem can be further simplified if the conditional variances depend on the weights of the cells only. This is satisfied, for example, by uniform prior. Assume that the prior is uniform for the remainder of this section, so that the ex-ante expected payoff is

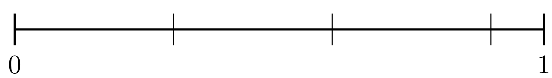
$$-\frac{1}{12}\sum_{i=1}^{K} w_i^3.$$

We can also without loss to consider the upper bound of the entropy takes the form of $\ln(K + \varepsilon)$ for some $\varepsilon \in [0, 1)$. Not only $\{[\ln K, \ln(K+1))\}_K$ covers the $\mathbb{R}_+$ space, $\ln K$ is also the entropy of the K-cell even partition structure. That is, for each K, we know that the optimal structure should perform at least as good as the K-cell even-partition structure.

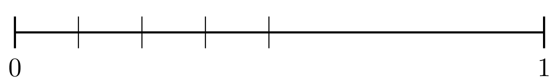
Proposition 13. Suppose that the prior distribution is uniform on $[0, 1]$, and the principal's payoff function is quadratic. If the upper bound of the entropy is $\ln K$ for some $K = 1, 2, 3, \cdots$, the optimal structure is a K-cell even partition structure. If the upper bound of the entropy is $\ln(K + \varepsilon)$ for some $\varepsilon \in (0, 1)$, the optimal structure is a $(K + 1)$-cell partition structure such that:

- All but one cells share the same cell size.
- The remaining one cell has a smaller cell size compared to all other cells.

The proof is in Appendix A. To show this result, we start by establishing the standard Lagrangian for a generic k-cell structure (where k is at least K). The first order condition indicates that the optimal structure can have at most two cell sizes, and all but one cell must share the same size. We can then focus on the structures with this feature. It is easy to check that with $\varepsilon \in (0, 1)$, there exists a $(K+1)$-cell structure such that it contains K equa-length, larger cells and one additional smaller cell. However, for any structure with more cells (and the feature aforementioned), the equal-length cells must be smaller than the one remaining cell (which holds for $(K + 1)$-cell structures when $\varepsilon = 0$). One example with $K = 3$ and $\varepsilon \in (0, 1)$ is shown in Figure 12. By further checking the second order condition, if the structure is at least locally optimal, the cell size shared by all but one cells must be at least as large as the size of the remaining one cell. This rules out the possibility that the optimal structure has $(K + 2)$ or more cells when $\varepsilon \in (0, 1)$, or the optimal structure has $(K + 1)$ or more cells when $\varepsilon = 0$.



(a) A Four-Cell Structure



(b) A Five-Cell Structure

FIGURE 12 Fitting One More Cell

References

Attila Ambrus, Eduardo M Azevedo, and Yuichiro Kamada. Hierarchical cheap talk. Theoretical Economics, 8(1):233–261, 2013.

David Austen-Smith. Strategic transmission of costly information. Econometrica: Journal of the Econometric Society, pages 955–963, 1994.

David Blackwell. Equivalent comparisons of experiments. The annals of mathematical statistics, pages 265–272, 1953.

Vincent P Crawford and Joel Sobel. Strategic information transmission. Econometrica: Journal of the Econometric Society, pages 1431–1451, 1982.

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

Paul E Fischer and Phillip C Stocken. Imperfect information and credible communication. Journal of Accounting Research, 39(1):119–134, 2001.

Matthew Gentzkow and Emir Kamenica. A rothschild-stiglitz approach to bayesian persuasion. American Economic Review, 106(5):597–601, 2016.

Robert M. Gray and David L. Neuhoff. Quantization. IEEE transactions on information theory, 44 (6):2325–2383, 1998.

Jerry R Green and Nancy L Stokey. A two-person game of information transmission. Journal of economic theory, 135(1):90–104, 2007.

Maxim Ivanov. Communication via a strategic mediator. Journal of Economic Theory, 145(2): 869–884, 2010a.

Maxim Ivanov. Informational control and organizational design. Journal of Economic Theory, 145 (2):721–751, 2010b.

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

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

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

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

Daniel Krähmer. Information design and strategic communication. American Economic Review: Insights, 3(1):51–66, 2021.

Sophie Kreutzkamp. Endogenous information acquisition in cheap-talk games. 2023.

Sophie Kreutzkamp and Yichuan Lou. Persuasion without ex-post commitment. Available at SSRN 4786596, 2024.

Yichuan Lou. Optimal delegation: Coarsification through information design. 2023.

Mark Machina and John Pratt. Increasing risk: some direct constructions. Journal of Risk and Uncertainty, 14:103–127, 1997.

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

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

Michael Rothschild and Joseph Stiglitz. Increasing risk: I. a definition. Journal of Economic Theory, 2(3):225–243, 1970.

Denis Shishkin. Evidence acquisition and voluntary disclosure. 2024.

Joel Watson. Information transmission when the informed party is confused. Games and Economic Behavior, 12(1):143–161, 1996.

A Proofs

Proof of Proposition 1

We are interested in the area below c_π, which is characterized by the intersection points of segments of c_π. It is clear that Segment i and Segment $(i+1)$ intersect at x_{i+1}, the $(i+1)$-th posterior mean. Let y_{i+1} be the corresponding vertical coordinate of the intersection point. Substitute x_{i+1} back to either Segment i or Segment $(i+1)$, we get that

$$y_{i+1} = \sum_{j=1}^{i} w_j (x_{i+1} - x_j). \tag{11}$$

In general, the area below c_π can be decomposed naturally into a triangle and several trapezoids, by vertically slicing at each x_i. An example with $N = 3$ is showing in Figure 5. Working towards a general conclusion, let's first consider the area excluding the last trapezoid (i.e., the trapezoid with the edge $(x_N, 0)$ – $(1, 0)$), denoting it by S_N.

When $N = 2$, the area excluding the last trapezoid is just the triangle, with the area

$$S_2 \equiv \frac{1}{2}(x_2 - x_1) y_2 = \frac{1}{2}(x_2 - x_1) w_1 (x_2 - x_1) = \frac{1}{2} w_1 (x_2 - x_1)^2.$$

For $N = 3$,

$$\begin{aligned}
S_3 &\equiv \frac{1}{2}(x_2 - x_1)y_2 + \frac{1}{2}(x_3 - x_2)(y_2 + y_3) \\
&= \frac{1}{2}w_1(x_2 - x_1)^2 + \frac{1}{2}(x_3 - x_2)\left[w_1(x_2 - x_1) + w_1(x_3 - x_1) + w_2(x_3 - x_2)\right] \\
&= \frac{1}{2}\left[w_1(x_3 - x_1)^2 + w_2(x_3 - x_2)^2\right].
\end{aligned}$$

This naturally leads to the following: For any finite N,

$$S_N = \frac{1}{2}\sum_{j=1}^{N-1} w_j(x_N - x_j)^2.$$

We can show this by induction. This has been verified for $N = 2$. Suppose that this holds for $N = k$,

$$S_k = \frac{1}{2}\sum_{j=1}^{k-1} w_j(x_k - x_j)^2.$$

Consider $N = k + 1$. Observe that

$$\begin{aligned}
S_{k+1} &= S_k + \frac{1}{2}(x_{k+1} - x_k)(y_k + y_{k+1}) \\
&= \frac{1}{2}\sum_{j=1}^{k-1} w_j(x_k - x_j)^2 + \frac{1}{2}(x_{k+1} - x_k)\left[\sum_{j=1}^{k-1} w_j(x_k - x_j) + \sum_{j=1}^{k} w_j(x_{k+1} - x_j)\right].
\end{aligned} \tag{12}$$

The first line follows from the fact that S_{k+1} can be decomposed into the last trapezoid it includes (i.e., the trapezoid with the edge $(x_k, 0) - (x_{k+1}, 0)$) and what left. Yet what left is the same "shape" as the shapes included in S_N. The second line follows from the induction assumption and (11).

For any $i = 1, \cdots, k-1$, let's collect the terms in (12) that contains the coefficient w_i. For now, let's also omit the coefficient $1/2$ for the ease of notations. These terms are

$$w_i(x_k - x_i)^2 + w_i(x_{k+1} - x_k)\left[(x_k - x_i) + (x_{k+1-x_i})\right] = w_i(x_{k+1} - x_i)^2.$$

What left is just the term contains the coefficient w_k, which is just $w_k(x_{k+1} - x_k)^2$. Substitute these back to (12) and put the coefficient $1/2$ back,

$$S_{k+1} = \frac{1}{2}\sum_{j=1}^{k} w_j(x_{k+1} - x_j)^2,$$

which is exactly what we need.

Now let's consider the last trapezoid. The area, denoted by T_N, is

$$T_N \equiv \frac{1}{2}\left[y_N + c_\pi(1)\right](1 - x_N),$$

where

$$
\begin{aligned}
c_{\pi}(1) = c_{\overline{\pi}}(1) &= \int_0^1 F(t)\mathrm{d}t \\
&= F(t)t|_0^1 - \int_0^1 tf(t)\mathrm{d}t \\
&= 1 - \sum_{j=1}^{N} w_j x_j.
\end{aligned}
$$

Following the expressions above,

$$
2(S_N + T_N) = \sum_{j=1}^{N-1} w_j (x_N - x_j)^2 + (1 - x_N)\left[\sum_{j=1}^{N-1} w_j (x_N - x_j) + 1 - \sum_{j=1}^{N} w_j x_j\right]. \tag{13}
$$

Notice that we consider $2(S_N + T_N)$ to remove the coefficient $1/2$ from the right hand side. Collect terms that contains the coefficient w_i, for $i = 1, \cdots, N-1$,

$$
w_i (x_N - x_i)^2 + w_i \left[(x_N - x_i) - x_i\right](1 - x_N) = w_i x_N - 2w_i x_i + w_i x_i^2.
$$

This shows, by substituting this back to (13),

$$
\begin{aligned}
2(S_N + T_N) &= \sum_{j=1}^{N-1} \left(w_j x_N - 2w_j x_j + w_j x_j^2\right) + (1 - w_N x_N)(1 - x_N) \\
&= 1 + \left(\sum_{j=1}^{N-1} w_j\right) x_N - x_N - w_N x_N - 2\sum_{j=1}^{N-1} w_j x_j + w_N x_N^2 + \sum_{j=1}^{N-1} w_j x_j^2 \\
&= 1 - 2\sum_{j=1}^{N} w_j x_j + \sum_{j=1}^{N} w_j x_j^2.
\end{aligned}
$$

Therefore,

$$
S_N + T_N = \frac{1}{2} EU^R(\pi) + \frac{1}{2} - \mu,
$$

where μ is the prior mean. This concludes the proof. ∎

Proof of Proposition 5

We start by noting that the conclusion in Gentzkow and Kamenica (2016) is not restricted to the case that the prior distribution has the support $[0, 1]$. Instead, it works on all closed intervals. The details are as follows:

- The part showing that the integrated CDF of an information structure must be between $c_{\overline{\pi}}$ and $c_{\underline{\pi}}$ uses the property of *mean-preserving spread*, which is not a concept restricted on $[0, 1]$.

- The part showing that a convex function between $c_{\overline{\pi}}$ and $c_{\underline{\pi}}$ corresponds to an information structure uses *Proposition 1* from Kolotilin (2014), which is shown under a more general interval $[\underline{r}, \overline{r}]$.

In particular, we will later use Gentzkow and Kamenica (2016) on $[0, m]$ for some $m < 1$.

Let c_π be the integrated CDF for some finite π such that segment i of c_π is tangent to $c_{\overline{\pi}}$ at some horizontal coordinate $m \in (0, 1)$. This implies the tangency point is

$$\left(m, \int_0^m F(t)\mathrm{d}t\right).$$

Segment i is a straight line passes through this point, with the slope of $F(m)$, the slope of $c_{\overline{\pi}}$ at this point. Therefore, we can write Segment i as

$$c_\pi(x) = F(m)x + c$$

for some constant c. Substitute the coordinate of the tangency point,

$$\begin{aligned}
c &= \int_0^m F(t)\mathrm{d}t - mF(m) \\
&= tF(t)|_0^m - \int_0^m t\mathrm{d}F(t) - mF(m) \\
&= -\int_0^m t\mathrm{d}F(t) \\
&= -F(m)\mathbb{E}(\theta \mid \theta \in [0, m]).
\end{aligned}$$

The second line uses integration by parts, the third line uses the fact that $F(0) = 0$, and the last line uses the definition of conditional expectation of θ on $[0, m]$. Therefore, the expression of Segment i is

$$c_\pi(x) = F(m)x - F(m)\mathbb{E}(\theta \mid \theta \in [0, m]). \tag{14}$$

Let's now take a detour and consider an information structure on $[0, m]$. Let the prior distribution on $[0, m]$ be

$$\tilde{F}(x) = \frac{F(x)}{F(m)}.$$

That is, we consider the original prior distribution F conditioned on $[0, m]$. An information structure based on $\tilde{F}$ has the upper bound

$$\tilde{c}_{\overline{\pi}}(x) = \int_0^x \tilde{F}(t)\mathrm{d}t = \frac{1}{F(m)} c_{\overline{\pi}}(x).$$

The lower bound $\tilde{c}_{\underline{\pi}}$ is a piecewise linear function which is 0 before $\mathbb{E}(\theta \mid \theta \in [0, m])$, and

$$x - \mathbb{E}(\theta \mid \theta \in [0, m])$$

for $x \in [\mathbb{E}(\theta \mid \theta \in [0,m]), m]$. Notice that $\tilde{c}_{\overline{\pi}}$ and $\tilde{c}_{\underline{\pi}}$ coincide at m.

Now consider c_π again, but scaled by $1/F(m)$. After this constant scaling, the new Segment 0 is still on the horizontal axis, and the new Segment i is, by using (14), exactly $\tilde{c}_{\underline{\pi}}$. Considering c_π only on $[0,m]$, the scaled version

$$\frac{c_\pi}{F(m)}$$

is a $(i+1)$-segment piecewise linear convex function between $\tilde{c}_{\underline{\pi}}$ and $\tilde{c}_{\overline{\pi}}$. This implies the scaled version, $c_\pi/F(m)$, defines an information structure on $[0,m]$. Correspondingly, c_π then also defines an information structure such that no signal realization sent at the states below m is also sent above m. (To further confirm this, one can check that the weighted average of x_i before m, using the scaled $c_\pi/F(m)$, is $\mathbb{E}(\theta \mid \theta \in [0,m])$.)

Conversely, consider an information structure such that, for some $m \in (0,1)$, no signal realization sent at the states below m is also sent above m. One can imagine that there is a "straight line" that divide $[0,1]$ into two independent parts, $[0,m]$ and $(m,1]$. Suppose that there are i posterior means induced before m; that is,

$$0 < x_1 < \cdots < x_i < m.$$

(Immediately, it is impossible that $x_i = m$ for some x_i. Otherwise, given that we have ruled out posterior means with zero weights, the signal realization that induces x_i must be sent above and below x_i.)

What is the expression of Segment i? Recall that in general, Segment i has the expression

$$c_\pi(x) = \left(\sum_{j=1}^{i} w_j\right) x - \sum_{j=1}^{i} w_j x_j.$$

Because of the straight line formulation,

$$\sum_{j=1}^{i} w_j = F(m),$$

since the signal realizations that induce $x_1, \cdots, x_i$ are sent on and only on $[0,m]$, and no other signal realizations are sent on $[0,m]$. For the same reason, $\sum_{j=1}^{i} w_j x_j$ must be the conditional mean on $[0,m]$ scaled by the weight, namely,

$$\sum_{j=1}^{i} w_j x_j = F(m)\mathbb{E}(\theta \mid \theta \in [0,m]).$$

But this means the expression of Segment i is exactly (14), which then shows that Segment i is tangent to $c_{\overline{\pi}}$. ■

Proof of Corollary 8

It suffices to show the following lemma.

Lemma 14. Suppose x_i and x_{i+1} are induced and only induced on states $[\ell, h]$. Let the weight of s_i (s_{i+1}) being sent be w_i (w_{i+1}). Regardless of the original information structure, the following modification leads to a payoff equivalent structure:

- In states $[\tilde{\ell}, \tilde{h}]$, signal s_{i+1} is sent, such that $\ell < \tilde{\ell} < \tilde{h} \leqslant h$, and

$$\int_{\tilde{\ell}}^{\tilde{h}} f(\theta)\mathrm{d}\theta = w_{i+1}, \quad \text{and} \quad \frac{1}{w_{i+1}} \int_{\tilde{\ell}}^{\tilde{h}} \theta f(\theta)\mathrm{d}\theta = x_{i+1}.$$

- In states $[\ell, \tilde{\ell}) \cup (\tilde{h}, h]$, signal s_i is sent.

We first consider the existences of $\tilde{\ell}$ and $\tilde{h}$. Let $\tilde{h}(t)$ be an implicit function defined by

$$\int_{t}^{\tilde{h}(t)} f(\theta)\mathrm{d}\theta = w_{i+1}.$$

Since f is assumed to have full support, $\tilde{h}(t)$ is strictly increasing. It is also clear that $\tilde{h}(\cdot)$ is continuous. To show this, first recall that f is continuous on $[0, 1]$, so that there exists m and M such that $m \leqslant f \leqslant M$ for all $\theta \in [0, 1]$. Let $\varepsilon > 0$ be given. We show that for any t_1, for t_2 satisfying $|t_2 - t_1| < \delta$, $|\tilde{h}(t_2) - \tilde{h}(t_1)| < \varepsilon$, where $\delta = \varepsilon m/M$. For simplicity, we show only when $t_2 > t_1$, and the other side is similar.

By definition, $\int_{t_1}^{\tilde{h}(t_1)} f(\theta)\mathrm{d}\theta = \int_{t_2}^{\tilde{h}(t_2)} f(\theta)\mathrm{d}\theta$, or equivalently,

$$\int_{t_1}^{t_2} f(\theta)\mathrm{d}\theta = \int_{\tilde{h}(t_1)}^{\tilde{h}(t_2)} f(\theta)\mathrm{d}\theta.$$

The left hand side is less than δM. Substitute $\delta = \varepsilon m/M$, this means

$$\int_{\tilde{h}(t_1)}^{\tilde{h}(t_2)} f(\theta)\mathrm{d}\theta < \varepsilon m.$$

Since $f \geqslant m$, $\tilde{h}(t_2) - \tilde{h}(t_1) < \varepsilon$.

Define function

$$\tilde{x}(t) = \frac{1}{w_{i+1}} \int_{t}^{\tilde{h}(t)} \theta f(\theta)\mathrm{d}\theta,$$

which is the posterior mean induced by the cell $[t, \tilde{h}(t)]$. Notice that this is also continuous (which can be shown similarly and the details are omitted). Let $\bar{t}$ be the point in $[\ell, h]$ such that $\tilde{h}(\bar{t}) = h$. Since $\tilde{x}(\ell) < x_{i+1} < \tilde{x}(\bar{t})$, by the intermediate value theorem, there exists $\tilde{\ell}$ such that $x_{\tilde{\ell}} = x_{i+1}$. In turn, this is exactly the ℓ we are looking for. Also, $\tilde{h} = \tilde{h}(\ell)$.

The rest of the proof is clear: since the structure after modifications keeps the same x_{i+1}, w_{i+1}, it must keep the same w_i, which then shows x_i must be unchanged as well. Given (2), this is all we need. ■

Proof of Theorem 9

The Receiver's ex-ante expected payoff under a four-signal structure is

$$\sum_{i=1}^{4} w_i x_i^2.$$

Substitute $x_2 = x_1 + 2b$, $x_3 = x_4 - 2b$, $w_1 = 2x_1$, $w_4 = 2(1 - 2x_4)$, and w_2 and w_3 from (4), the payoff is half of the following (That is, for simplicity, we drop the coefficient $1/2$ in the payoff in the remaining part of the proof.)

$$x_1 + x_4 - 2x_1x_4 - 8b^2(1 + 2x_1 - 2x_4) - 4b(x_1 + 2x_1^2 - 4x_1x_4 + x_4(2x_4 - 1)). \tag{15}$$

The Receiver chooses x_1 and x_4 to maximize this subject to $0 < x_1 < x_4 < 1$ and

$$x_4 - x_1 \geqslant 6b.$$

This is because the Receiver must also fit x_2 and x_3 in between, so that the IC condition implies there must be at least $6b$ distance between x_1 and x_4. The Hessian matrix of the objective (i.e., payoff) function in (15),

$$H = \begin{bmatrix} -16b & 16b - 2 \\ 16b - 2 & -16b \end{bmatrix}$$

Given the range of b, H is negative definite, showing that (15) is strictly concave in both variables. This implies that we can use the Kuhn-Tucker conditions to solve the optimzer. The Lagrange function is

$$\begin{aligned} \mathcal{L}(x_1, x_4, \lambda) = {} & x_1 + x_4 - 2x_1x_4 - 8b^2(1 + 2x_1 - 2x_4) - 4b(x_1 + 2x_1^2 - 4x_1x_4 + x_4(2x_4 - 1)) \\ & + \lambda(x_4 - x_1 - 6b). \end{aligned}$$

The Kuhn-Tucker conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= 1 - 16b^2 - 4b(1 + 4x_1 - 4x_4) - 2x_4 - \lambda = 0, \\ \frac{\partial \mathcal{L}}{\partial x_4} &= 1 + 16b^2 + 4b(1 + 4x_1 - 4x_4) - 2x_1 + \lambda = 0, \\ \lambda \cdot \frac{\partial \mathcal{L}}{\partial \lambda} &= \lambda(x_4 - x_1 - 6b) = 0, \qquad \lambda \geqslant 0. \end{aligned}$$

Notice that the conditions are slightly simplified since we know that x_1 and x_4 must be strictly between 0 and 1. We also know that the solution must exist since the objective function is strictly concave and the feasible region is convex. The Kuhn-Tucker conditions give

$$x_1 = \frac{1}{2} - 3b, \quad x_4 = \frac{1}{2} + 3b, \quad \text{and} \quad \lambda = 10b(8b-1).$$

Notice that since $\lambda > 0$, the constraint must bind, which implies that the IC condition binds everywhere. Therefore,

$$\begin{aligned} x_2 &= \frac{1}{2} - b, \quad x_3 = \frac{1}{2} + b, \\ w_2 &= w_3 = 6b - \frac{1}{2}. \end{aligned}$$

Notice that when $b = 1/8$, the distribution of posterior mean here is exactly the one induced by the four-cell even-partition structure.

We then need to verify that such distribution of posterior means can be induced by an information structure. Of course, one can check that the corresponding integrated CDF lies below $c_{\overline{\pi}}(x) = x^2/2$. Alternatively, since all mixing regions only induces two posterior means, a mixing region $[\ell, h]$ that induces x_i and x_{i+1} can be generated by an information structure if and only if

$$x_i + (x_i - \ell) \geqslant x_{i+1} - (h - x_{i+1}),$$

and the equality holds if the mixing is degenerate; that is, x_i and x_{i+1} are induced by two partition cells. The idea is as follows. Imagine x_i and x_{i+1} were induced by partition cells. If the right boundary of the cell induces x_i was smaller than the left boundary of the cell induces x_{i+1}, the two posterior would be two spreaded from each other such that even-partition structures could not generate such posterior means. In turn, the two cells must overlap with each other in a mixing structure. It turns out that this is also all one need for feasibility. It can be verified that the structure above is always feasible.

Given the Receiver's payoff is $\sum_{i=1}^{N} w_i x_i$, the payoffs of the structure propsed in Theorem 9 is

$$-96b^3 + 17b^2 + \frac{1}{4}. \tag{16}$$

The payoff of the three-cell even-partition strucuture is $35/108$. Notice that in $(1/8, 1/6)$, (16) is strictly monotone decreasing. When $b = 1/8$, (16) is $21/64$, greater than $35/108$. When $b = 1/6$, the structure proposed in Theorem 9 reduces to a two-signal structure with payoff $5/18$, lower than $35/108$. The existence of $\bar{b}$ is then guaranteed by the intermediate value theorem, and it must be unique. To solve for $\bar{b}$, let

$$-96b^3 + 17b^2 + \frac{1}{4} = \frac{35}{108},$$

which gives $\bar{b} \approx 0.134309$. ■

Proof of Proposition 12

To see that (9) must bind, suppose the contrary. That is, if $\pi^\star$ is the optimal structure, $E - E(\pi^\star) = \varepsilon > 0$. Pick an arbitrary signal realization s_k (that induces x_k in $\pi^\star$). Since $g(\cdot)$ is continuous, there exists $\bar{z} > 0$ such that

$$0 < g(\bar{z}) - (-w_k \ln w_k) < \frac{\varepsilon}{2}.$$

We now proposed the following modified information structure π'. Pick some $m \in (0,1)$ (defined later), and create a new signal s'_k. Then,

$$\pi'(s \mid \omega) = \begin{cases} \pi^\star(s \mid \omega) & \text{if } s \neq s_k \text{ or } s'_k, \\ \pi(s_k \mid \omega) & \text{if } s = s'_k, \\ 0 & \text{if } s = s_k \end{cases} \qquad \text{if } \omega < m,$$

and

$$\pi'(s \mid \omega) = \begin{cases} \pi^\star(s \mid \omega) & \text{if } s \neq s'_k, \\ 0 & \text{if } s = s'_k \end{cases} \qquad \text{if } \omega \geqslant m.$$

That is, we split the signal realization s_k in two by replace s_k by s'_k for all states below m. m is defined such that the overall weight of sending s'_k is less than $\bar{z}$, i.e.,

$$\int_0^m \pi^\star(s_k \mid \omega) f(\omega) \mathrm{d}\omega < \bar{z},$$

which is always feasible since f is continuous and every signal realization is assumed to be generated with positive weights. This modification only increase the entropy less than $\varepsilon/2$ so that (9) is still satisfied. Yet it is clear that the designer is better off. ∎

Proof of Proposition 13

Consider a k-cell structure for some arbitrary $k \geqslant 3$. (The problem is more or less trivial if there are at most two cells.) For a given upper bound E on the entropy, the information designer's problem is

$$\begin{aligned} \min_{w_1, \cdots, w_{k-1}} \quad & \sum_{i=1}^{k-1} w_i^3 + \left(1 - \sum_{i=1}^{k-1} w_i\right)^3 \\ \text{subject to} \quad & -\sum_{i=1}^{k-1} w_i \ln w_i - \left(1 - \sum_{i=1}^{k-1} w_i\right) \ln\left(1 - \sum_{i=1}^{k-1} w_i\right) \leqslant E, \end{aligned}$$

and $w_i \geqslant 0$ for all $i = 1, \cdots, k-1$. The Lagrangian is

$$\mathcal{L} = \sum_{i=1}^{k-1} w_i^3 + \left(1 - \sum_{i=1}^{k-1} w_i\right)^3 + \lambda \left[-E - \sum_{i=1}^{k-1} w_i \ln w_i - \left(1 - \sum_{i=1}^{k-1} w_i\right) \ln\left(1 - \sum_{i=1}^{k-1} w_i\right)\right].$$

For a generic i, the FOC is

$$\frac{\partial \mathcal{L}}{\partial w_i} = 3w_i^2 - 3\left(1 - \sum_{i=1}^{k-1} w_i\right)^2 + \lambda\left[\ln\left(1 - \sum_{i=1}^{k-1} w_i\right) - \ln w_i\right] = 0. \tag{17}$$

Let $1 - w_1 - \cdots - w_{k-1} = w_k$. The FOC indicates that, if $w_i \neq w_k$ for all $i = 1, \cdots, k-1$,

$$\lambda = \frac{3(w_i^2 - w_k^2)}{\ln w_i - \ln w_k}.$$

This further implies that for $i \neq j$,

$$\frac{(w_i/w_k)^2 - 1}{(w_j/w_k)^2 - 1} = \frac{\ln(w_i/w_k)}{\ln(w_j/w_k)}.$$

Since the function

$$g(x) = \frac{x^2 - 1}{\ln x} \tag{18}$$

is strictly increasing, $w_i/w_k = w_j/w_k$, or simply, $w_i = w_j$. This shows that in an optimal structure, there are at most two cell sizes, and all but one cell share the same size.

Let's further check the second order conditions. Using (17),

$$\frac{\partial^2 \mathcal{L}}{\partial w_i^2} = 6w_i + 6w_k - \frac{\lambda}{w_i} - \frac{\lambda}{w_k},$$
$$\frac{\partial^2 \mathcal{L}}{\partial w_i \partial w_j} = 6w_k - \frac{\lambda}{w_k}.$$

We can then write the bordered Hessian matrix H_k^B. Since $w_1 = \cdots = w_{k-1}$ in the solution of the FOCs, we will write all w_i as w_1 for $i = 1, \cdots, k-1$. Then,

$$H_k^B = \begin{bmatrix} 0 & \ln\frac{w_k}{w_1} & \ln\frac{w_k}{w_1} & \cdots & \ln\frac{w_k}{w_1} \\ \ln\frac{w_k}{w_1} & 6w_1 + 6w_k - \frac{\lambda}{w_1} - \frac{\lambda}{w_k} & 6w_k - \frac{\lambda}{w_k} & \cdots & 6w_k - \frac{\lambda}{w_k} \\ \ln\frac{w_k}{w_1} & 6w_k - \frac{\lambda}{w_k} & 6w_1 + 6w_k - \frac{\lambda}{w_1} - \frac{\lambda}{w_k} & \cdots & 6w_k - \frac{\lambda}{w_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ln\frac{w_k}{w_1} & 6w_k - \frac{\lambda}{w_k} & 6w_k - \frac{\lambda}{w_k} & \cdots & 6w_1 + 6w_k - \frac{\lambda}{w_1} - \frac{\lambda}{w_k} \end{bmatrix}$$

To calculate the determinant of this matrix, we first subtract Row 2 from Rows $3, 4 \cdots, k$. We then subtract Column k from Columns $2, 3 \cdots k-1$. All these operations do not change the determinant.

The matrix has been transformed into the following matrix,

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \ln\frac{w_k}{w_1} \\ \ln\frac{w_k}{w_1} & 6w_1-\frac{\lambda}{w_1} & 0 & 0 & \cdots & 0 & 6w_k-\frac{\lambda}{w_k} \\ 0 & -6w_1+\frac{\lambda}{w_1} & 6w_1-\frac{\lambda}{w_1} & 0 & \cdots & 0 & 0 \\ 0 & -6w_1+\frac{\lambda}{w_1} & 0 & 6w_1-\frac{\lambda}{w_1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & -6w_1+\frac{\lambda}{w_1} & 0 & 0 & \cdots & 6w_1-\frac{\lambda}{w_1} & 0 \\ 0 & -12w_1+\frac{2\lambda}{w_1} & -6w_1+\frac{\lambda}{w_1} & -6w_1+\frac{\lambda}{w_1} & \cdots & -6w_1+\frac{\lambda}{w_1} & 6w_1-\frac{\lambda}{w_1} \end{bmatrix}$$

We can then expand this matrix from the first column. It suffices to consider the determinant of the following matrix

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & \ln\frac{w_k}{w_1} \\ -6w_1+\frac{\lambda}{w_1} & 6w_1-\frac{\lambda}{w_1} & 0 & \cdots & 0 & 0 \\ -6w_1+\frac{\lambda}{w_1} & 0 & 6w_1-\frac{\lambda}{w_1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -6w_1+\frac{\lambda}{w_1} & 0 & 0 & \cdots & 6w_1-\frac{\lambda}{w_1} & 0 \\ -12w_1+\frac{2\lambda}{w_1} & -6w_1+\frac{\lambda}{w_1} & -6w_1+\frac{\lambda}{w_1} & \cdots & -6w_1+\frac{\lambda}{w_1} & 6w_1-\frac{\lambda}{w_1} \end{bmatrix} \tag{19}$$

Add Column $2, \cdots, k-2$ to the first column, and then switch Column 1 and Column $k-1$ to get

$$\begin{bmatrix} \ln\frac{w_k}{w_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 6w_1-\frac{\lambda}{w_1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & 6w_1-\frac{\lambda}{w_1} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 6w_1-\frac{\lambda}{w_1} & 0 \\ 6w_1-\frac{\lambda}{w_1} & -6w_1+\frac{\lambda}{w_1} & -6w_1+\frac{\lambda}{w_1} & \cdots & -6w_1+\frac{\lambda}{w_1} & (k-1)\left(-6w_1+\frac{\lambda}{w_1}\right) \end{bmatrix}$$

This is a triangular matrix, and the determinant is the product of the diagonal elements,

$$-(k-1)\ln\frac{w_k}{w_1}\left(6w_1-\frac{\lambda}{w_1}\right)^{k-2}.$$

Also, as we have switch columns once, the determinant of (19) needs to be multiplied by an additional (-1). The overall determinant of H_k^B is then

$$-(k-1)\left(\ln\frac{w_k}{w_1}\right)^2\left(6w_1-\frac{\lambda}{w_1}\right)^{k-2}. \tag{20}$$

We should note that all principal minor of H_k^B share the same structure, since they are just H_{k-1}^B, H_{k-2}^B, etc. Therefore, if the determinant H_k^B and the determinant of all principal minors of H_k^B are negative as required by the second order conditions, it must be the case that

$$6w_1 - \frac{\lambda}{w_1} > 0,$$

so that (20) is negative for all k. We already know that $\lambda = 3(w_1^2 - w_k^2)/\ln(w_1/w_k)$, so that we need to consider the sign of

$$6w_1 - \frac{3(w_1^2 - w_k^2)}{w_1 \ln \frac{w_1}{w_k}} = 3w_1 \left[2 + \frac{1 - \left(\frac{w_k}{w_1}\right)^2}{\ln \frac{w_k}{w_1}} \right].$$

Notice that the terms in the bracket is $2 - g(w_k/w_1)$. Since $\lim_{x\to 1} g(x) = 2$, and $g(\cdots)$ is strictly increasing, we know that $6w_1 - (\lambda/w_1) > 0$ if and only if $w_k/w_1 < 1$, or $w_k < w_1$. Therefore, any solution with $w_k > w_1$ will violate the second order conditions, implying that the solution does not attain a local minimum. That is, the optimal structure must have the following features:

- It is either a even-partition structure; or
- It is a partition structure with two types of cell sizes. All but one cell share the same, larger cell size, and the one cell left has the smaller cell size.

Let's now return to the case where the entropy upper bound E is $-\ln(K+\varepsilon)$ for some $\varepsilon \in [0, 1)$. Consider a $(K+2)$-cell structure with $w_{K+2} = \delta$, so that $w_1 = \cdots = w_{K+1} = (1-\delta)/(K+1)$. The entropy of the structure is

$$-\left[(K+1)\frac{1-\delta}{K+1} \ln \frac{1-\delta}{K+1} + \delta \ln \delta\right] = -(1-\delta)\ln\frac{1-\delta}{K+1} - \delta \ln \delta.$$

If $\delta < (1-\delta)/(K+1)$,

$$-(1-\delta)\ln\frac{1-\delta}{K+1} - \delta\ln\delta > -(1-\delta)\ln\frac{1-\delta}{K+1} - \delta\ln\frac{1-\delta}{K+1} = \ln\frac{K+1}{1-\delta} > \ln(K+1).$$

More generally, for any generic k, the structure such that $k-1$ cells share a larger cell size, and the remaining one cell has a smaller cell size must have an entropy greater than $\ln k$. In other words, for the upper bound $-\ln(K+\varepsilon)$ for some $\varepsilon \in [0, 1)$, a structure with $K+2$ or more cells that satisfies the first and the second order conditions above must exceed the upper bound of the entropy. Thus, the optimal structure is at most $K+1$ cell. When $\varepsilon = 0$, the $(K+1)$-cell structure is also impossible (since it must have an entropy strictly greater than $\ln K$), while we have known that among structures with K or fewer cells, the K-cell even-partition structure performs the best. When $\varepsilon \in (0, 1)$, a $(K+1)$-cell structure is possible, leading to a structure with two cell sizes described above. ■

B More Results under the Uniform Prior

In this section, we show that for the case that $N_{\max} = N$ is even, the optimal structure can never contain one and only one mixing region. To show this, we prove several lemmas that first show that all IC constraints must be binding.

We first note that we are considering the biases that are large enough so that fitting N posterior means via *partition* structures is impossible, yet the biases are not too large so that it is impossible to fit N posterior means via *any* structures. Formally, to fit N posterior means, the upper bound of the bias is determined by

$$(N-1) \times 2b < 1, \quad \Rightarrow \quad b < \frac{1}{2(N-1)}.$$

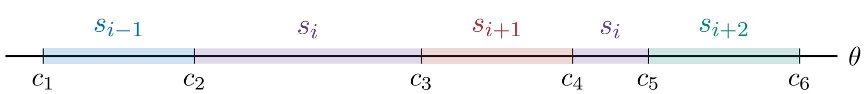
Meanwhile, if $b \leqslant 1/2N$, one can employ a partition structure. Therefore, we consider

$$\frac{1}{2N} < b < \frac{1}{2(N-1)}.$$

Consider the following structure:

- s_{i-1} and s_{i+2} are sent in partition cells. That is, posterior means x_{i-1} and x_{i+2} are induced by partition cells.
- s_i and s_{i+1}, which induces x_i and x_{i+1}, respectively, are mixed in some states.

That is, there are partition cells right before and right after the mixing region. Using Corollary 8, we know that we can consider the following structure instead:



That is, we convert the original structure into a structure that does not involve mixing when inducing x_i and x_{i+1}.

Lemma 15. In an optimal structure, if a mixing region that induces x_i and x_{i+1} is sandwiched by two partition cells that induces x_{i-1} and x_{i+2}, it must be the case that $x_{i+1} - x_i = 2b$.

Proof. We show this by contradiction. Suppose that $x_{i+1} - x_i > 2b$. Since this structure is optimal (which implies that the original structure is also optimal), so that

$$x_i - x_{i-1} = 2b, \quad \text{and} \quad x_{i+2} - x_{i+1} = 2b.$$

Also notice that this no-mixing structure allow us to write x_i and x_{i+1} easily in terms of the "partition" points:

$$x_i = \frac{c_3 - c_2}{c_3 - c_2 + c_5 - c_4} \frac{c_2 + c_3}{2} + \frac{c_5 - c_4}{c_3 - c_2 + c_5 - c_4} \frac{c_4 + c_5}{2},$$
$$x_{i+1} = \frac{c_3 + c_4}{2}.$$

Also, $x_{i-1} = (c_1 + c_2)/2$, and $x_{i+2} = (c_5 + c_6)/2$.

Consider the modification that moves c_2 to $c_2 + \varepsilon$ for some arbitrarily small ε. First observe that this does not violate IC conditions – the only one that needs to be examined is the distance between x_{i-1} and x_i. Using the partition points,

$$2(x_i - x_{i-1}) = \frac{c_3^2 - c_2^2 + c_5^2 - c_4^2}{c_3 - c_2 + c_5 - c_4} - (c_1 + c_2).$$

Differentiate this with respect to c_2 to get

$$\frac{-2(c_3 - c_2 + c_5 - c_4) + (c_3^2 - c_2^2 + c_5^2 - c_4^2)}{(c_3 - c_2 + c_5 - c_4)^2} - 1,$$

whose sign depends on

$$\left(-c_2^2 + c_3^2 - c_4^2 + c_5^2\right) - (-c_2 + c_3 - c_4 + c_5)^2 - 2c_2(-c_2 + c_3 - c_4 + c_5),$$

which simplifies to

$$2(c_5 - c_4)(c_3 - c_2) > 0.$$

Therefore, the IC conditions are still preserved.

Next, let's consider whether this modification also increases the payoffs of the Receiver. Since x_{i+1}, x_{i+2} and their weights are unchanged, it suffices to examine $w_{i-1}x_{i-1}^2 + w_i x_i^2$, which is

$$(c_2 - c_1)\frac{(c_1 + c_2)^2}{4} + \frac{(c_3^2 - c_2^2 + c_5^2 - c_4^2)^2}{4(c_3 - c_2 + c_5 - c_4)}.$$

Similarly, differentiate with c_2 to get

$$\frac{1}{4}\left((c_1 + c_2)^2 + 2(c_2 - c_1)(c_1 + c_2) + \frac{\left(c_2^2 - c_3^2 + c_4^2 - c_5^2\right)^2}{(-c_2 + c_3 - c_4 + c_5)^2} + \frac{4c_2\left(c_2^2 - c_3^2 + c_4^2 - c_5^2\right)}{-c_2 + c_3 - c_4 + c_5}\right),$$

whose sign depends on

$$\left[c_2^2 - c_3^2 + c_4^2 - c_5^2 + 2c_2(-c_2 + c_3 - c_4 + c_5)\right]^2 - (c_2 - c_1)^2(-c_2 + c_3 - c_4 + c_5)^2. \tag{21}$$

To consider the sign of (21), we further need $x_i - x_{i-1} = 2b$, which is

$$\frac{c_3 - c_2}{c_3 - c_2 + c_5 - c_4} \frac{c_2 + c_3}{2} + \frac{c_5 - c_4}{c_3 - c_2 + c_5 - c_4} \frac{c_4 + c_5}{2} - \frac{c_1 + c_2}{2} = 2b,$$

or equivalently,

$$c_3^2 - c_2^2 + c_5^2 - c_4^2 - (c_1 + c_2)(c_3 - c_2 + c_5 - c_4) = 4b(c_3 - c_2 + c_5 - c_4).$$

Now consider (21) again. It can be decomposed into two factors,

$$\begin{aligned}
&c_2^2 - c_3^2 + c_4^2 - c_5^2 + 2c_2(c_3 - c_2 + c_5 - c_4) - (c_2 - c_1)(c_3 - c_2 + c_5 - c_4) \\
= &c_2^2 - c_3^2 + c_4^2 - c_5^2 + (c_1 + c_2)(c_3 - c_2 + c_5 - c_4) \\
= &-4b(c_3 - c_2 + c_5 - c_4) < 0,
\end{aligned}$$

and

$$\begin{aligned}
&c_2^2 - c_3^2 + c_4^2 - c_5^2 + 2c_2(c_3 - c_2 + c_5 - c_4) + (c_2 - c_1)(c_3 - c_2 + c_5 - c_4) \\
= &c_2^2 - c_3^2 + c_4^2 - c_5^2 + (3c_2 - c_1)(c_3 - c_2 + c_5 - c_4) \\
= &c_2^2 - c_3^2 + c_4^2 - c_5^2 + (c_1 + c_2)(c_3 - c_2 + c_5 - c_4) + 2(c_2 - c_1)(c_3 - c_2 + c_5 - c_4) \\
= &2(c_3 - c_2 + c_5 - c_4)(c_2 - c_1 - 2b).
\end{aligned}$$

Therefore, if $c_2 - c_1 \geqslant 2b$, the sign of (21) (hence the sign of the derivative of the payoff of the Receiver) is negative. Given the assumption that the original structure is optimal, this must be true. In other words, the size of the cell that induces x_{i-1} is at least $2b$. A symmetric argument also shows that $c_6 - c_5 \geqslant 2b$.

Since the length of the cell that induces x_{i-1} is at least $2b$, $c_2 - x_{i-1} \geqslant b$. Similarly, $x_{i+2} - c_5 \geqslant b$. Because of the binding IC constraints,

$$x_i - c_2 \leqslant b, \quad \text{and} \quad c_5 - x_{i+1} \leqslant b.$$

Recall that by assumption, $x_{i+1} - x_i > 2b$. Then, it is easy to see that this structure (in particular, the mixing region in the middle) is not feasible. ■

We now turn to the partition cells before the first mixing region and after the last. As a preparation, we have the following.

Lemma 16. The first and the last IC constraints must bind in an optimal structure.

Proof. Let's consider the first IC, $x_2 - x_1 \geqslant 2b$ in this proof. The part regarding the last IC can be shown symmetrically.

First notice that if x_2 and x_3 are generated by a mixing, then the conclusion holds using Lemma 4. Suppose from now on that x_2 is generated by a partition cell $[a_1, a_2]$.

Suppose the contrary; that is, $x_2 - x_1 > 2b$. In turn,

$$\frac{a_1 + a_2}{2} - \frac{a_1}{2} > 2b, \quad \Rightarrow \quad a_2 > 4b.$$

It follows that either $[0, a_1]$ has length larger than $2b$, or $[a_1, a_2]$ has length larger than $2b$ (or both).

<u>*Case 1: Suppose $a_2 - a_1 > 2b$, but $a_1 \leqslant 2b$.*</u> There are two subcases:

- <u>*Case 1.1: $x_3 - x_2 > 2b$.*</u> In this case, consider the modification that slightly increases a_1. IC conditions are still preserved, since $x_2 - x_1 > 2b$ and $x_3 - x_2 > 2b$ before the modification.

  In terms of the payoffs, notice that only w_1, x_1, w_2, x_2 are affected by this modification, so that we consider

$$\begin{aligned} w_1 x_1^2 + w_2 x_2^2 &= \frac{1}{4}\left[a_1^3 + (a_2 - a_1)(a_2 + a_1)^2\right] \\ &= \frac{1}{4}(a_2^3 + a_2^2 a_1 - a_2 a_1^2) \end{aligned}$$

  Omit the coefficient $1/4$ and take the derivative with respect to a_1 to get

$$(a_2 - a_1)^2 - a_1^2.$$

  By assumption, this is positive, so that slightly increasing a_1 also increases the expected payoff.

- <u>*Case 1.2: $x_3 - x_2 = 2b$.*</u> First assume that x_3 is induced by a cell. Consider the modification that slightly decreases a_2. This modification decreases x_3 and x_2, but the distance

$$x_3 - x_2 = \frac{a_2 + a_3}{2} - \frac{a_1 + a_2}{2} = \frac{a_3 - a_1}{2}$$

  does not change. Also, since $x_2 - x_1 > 2b$, this modification does not break IC conditions.

  The case that x_3 is induced by a mixing region (together with x_4) is more complicated. Using Corollary 8, we can say that there exists d_2, d_3 such that $a_2 < d_2 < d_3 < a_4$, such that x_3 is induced by $[a_2, d_2) \cup [d_3, a_4]$, and x_4 is induced by $[d_2, d_3)$. The modification needs to take care the case that $x_3 - x_2 = 2b$, $x_4 - x_3 = 2b$, and $x_5 - x_4 = 2b$ – which gives the most strict constraints on the locations of posterior means. The modification involves decreasing a_2 and s_3 in a proper way. The idea is that given x_2 is induced in a large region, and x_3 is close to a_2, we need to redistribute weight from x_2 to x_3 by reducing a_2. Yet we also need to preserve the IC constraints, so that we also adjust s_3, so that x_3 does not decrease too fast. We omit the math details.

<u>*Case 2: Suppose $a_1 > 2b$.*</u> Immediately, due to the size of b, we must have $1 - a_{N-1} < 2b$. That is, the first cell is larger than the last cell. Consider the modification that reduces the first cell by slightly decreasing a_1. Also, we make a parallel shift for all the structure between a_1 and a_{N-1}. That is, the first cell is shrinked, the last cell is enlarged, and nothing is changed in the middle in terms of the structure. Notice that only the first and the last IC conditions change under this modification. The last IC condition is still easily preserved, since

$$x_N - x_{N-1} = (x_N - a_{N-1}) + (a_{N-1} - x_{N-1}).$$

$a_{N-1} - x_{N-1}$ is not changed under this modification since the parts in the middle is shifted in a parallel way. $x_N - a_{N-1}$ is larger after the modification since

$$x_N - a_{N-1} = \frac{1 + a_{N-1}}{2} - a_{N-1} = \frac{1 - a_{N-1}}{2},$$

and a_{N-1} is smaller after the modification. The first IC condition is also preserved (even though $x_2 - x_1$ is now smaller), since it is not binding originally.

To see that the Receiver's expected payoff is increased after the modification, one can use the argument above and calculate the derivative again. Intuitively, the first cell and the last cell become more even after the modification. ■

Lemma 17. In an optimal structure, all IC conditions are binding until the first mixing, and after the last mixing.

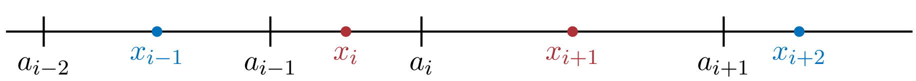


FIGURE 13 Part of the Structure with A Non-Binding IC condition

Proof. Suppose on the contrary that some IC conditions before the first mixing or after the last are not binding. Then, at least the part before first mixing or the part after the last mixing contains at least one non-binding IC. Consider the part regarding before first mixing. The part regarding after last mixing can be shown symmetrically.

Suppose that x_i and x_{i+1} is the first pair where the IC condition is not binding. By Lemma 16, this cannot be the first IC condition. That is, there must exist x_{i-1} that is also generated by a partition cell.

First notice that since the structure is optimal, it cannot be the case that x_{i+1} is generated by a mixing. It is easy to see that if $a_i - a_{i-1} > a_{i-1} - a_{i-2}$, then slightly increasing a_{i-1} always increases the expected payoff of the Receiver without violating any IC conditions. Therefore, it suffices to consider that $a_i - a_{i-1} \leqslant a_{i-1} - a_{i-2}$, as shown in Figure 13.

Since $x_i - x_{i-1} = 2b$, $a_i - a_{i-2} = 4b$, so that $a_i - x_i \leqslant b$. This means $x_{i+1} - a_i = a_{i+1} - x_{i+1} > b$, which in turn suggests that $a_{i+1} - a_i > 2b$. In words, the cell $[a_i, a_{i+1}]$ must be strictly larger than the cell $[a_{i-1}, a_i]$.

Suppose that $x_{i+2} - x_{i+1} > 2b$. Then a similar argument implies that one can slightly increase a_i, maintain all IC constraints, and make the two cells $[a_i, a_{i+1}]$ and $[a_{i-1}, a_i]$ more even, so that the Receiver's payoff is higher. This indicates that $x_{i+2} - x_{i+1} = 2b$.

We now need to consider two cases:

- *Case 1. x_{i+2} is generated by partition cell $[a_{i+1}, a_{i+2}]$.* The binding IC condition between x_{i+2} and x_{i+1} then suggests $[a_{i+1}, a_{i+2}]$ must be strictly smaller than $[a_i, a_{i+1}]$. In turn, it is payoff-increasing and feasible (i.e., not violating any IC) to decrease a_{i+1}, using the logic that the two cells $[a_i, a_{i+1}]$ and $[a_{i+1}, a_{i+2}]$ are more even.
- *Case 2. x_{i+2} is generated by a mixing.* In this case, decreasing a_{i+1} still works. Again, the IC conditions are not violated anywhere by slightly decreasing a_{i+1}. Moreover, since $a_{i+1} - a_i > 2b$, the proof of Lemma 15 has already shown that slightly decreasing a_{i+1} is beneficial.

Therefore, one can always find a payoff-improving operation if $x_{i+1} - x_i > 2b$. In turn, if there exists at least one IC condition that is not binding before the first mixing, the structure cannot be optimal. ■

Immediately we have the following conclusion.

Corollary 18. If an optimal structure only contains one pair of mixing, all IC conditions bind.

We can use Corollary 18 to learn the optimal structure under the restriction that there is only one mixing region. In particular, we consider fitting $N = 2k$ signal realizations when $b \in (\frac{1}{2N}, \frac{1}{2(N-1)})$. Under this range of b, it is impossible to fit N signal realizations using a partition structure, so that mixing is necessary.

We start with a feasibility result.

Lemma 19. It is impossible to have even number of cells on the left of the mixing region.

Proof. Suppose not. That is, there are $2M$ partition cells on the left of the mixing region (where M can be 0). In turn, there are $N-2M-2$ partition cells on the right of the mixing region. By Corollary 18, the combined length of two neighboring partition cells must be $4b$. The length of the mixing region must be larger than $4b$, so that the total length of the structure must be greater than

$$\left(\frac{2M}{2}+\frac{N-2M-2}{2}\right)\times 4b+4b=2Nb>1,$$

which is impossible. ■

Notice that Lemma 19 suggests that it is also impossible to have even number of cells on the right of the mixing region. This gives us the structure in Figure 14.

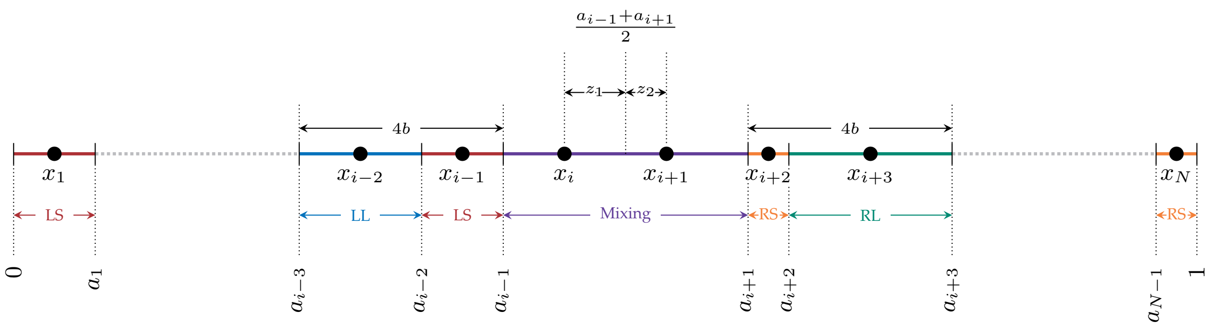


FIGURE 14 Structure with One Mixing

The way of mixing not only determines the lengths of the first and the last cells; it also affects the lengths of the cells in each group. Although the total length of each group is fixed at $4b$, how this $4b$ length is assigned to either cells in the group is determined by the way of mixing. To be precise, since $x_i - x_{i-1} = 2b$, the last cell before mixing (i.e., the cell corresponding to signal realization s_{i-1}) must have length

$$2\times(2b-(x_i-a_{i-1}))\,,$$

where $x_i - a_{i-1}$ is determined by the way of mixing. Then, since $x_{i-1}-x_{i-2}=2b$, the length of the cell corresponding to signal realization s_{i-2} is also determined by the way of mixing.

Therefore, as shown in Figure 14, we only have four types of cells: LL, LS, RL, RS. "L/R" means left/right (toward the mixing region), while "L/S" means long/short. We should note that "long/short" does not always imply, for example, RL type is longer than RS. In general, it should be the case: the total length of the mixing region is greater than $4b$, so that

$$(x_i-a_{i-1})+(a_{i+1}-x_{i+1})>2b.$$

The intuition above – squeezing the first and the last cell implies the first distance $x_i - a_{i-1}$ and the second distance $(a_{i+1} - x_{i+1})$ should both be greater than b.[12] But it is also possible that only one distance is greater than b, while the other is smaller.

If we want to write down the payoff of the structure shown in Figure 14, there are two things we need to consider:

- The way of mixing (i.e., the length of the mixing region and the location of the posterior means inside the mixing region) – this determines the length of each type of partition cells. Therefore, the payoff from the mixing region and the payoff of each cell can be calculated by knowing the way of mixing.
- The number of each type of the cells. This is determined by the location of the mixing region.

We introduce the following result that indicates we only need the way of mixing if we only care about the optimal structure.

Lemma 20. For a given way of mixing and an optimal structure, it is without loss to consider the case that the signal realizations involved in the mixing are either s_2 and s_3, or s_{N-2} and s_{N-1}.

Proof. It is much easier to see this result if we use the true quadratic preferences (so that the payoff of each cell is the conditional variance), which is equivalent to what we have been used so far. The way of mixing determines the first and the last cell (as well as the mixing region itself). Therefore, we can consider the cell groups in the middle, x_{i-2} and x_{i-1}, x_{i+2} and x_{i+3}, and so on. For each cell group, the total payoff is higher if the two cells contained in the group is more balanced. Therefore, if

$$|b - (x_i - a_{i-1})| < |(a_{i+1} - x_{i+1}) - b|,$$

the cell groups on the left of the mixing region is more balanced, so that the optimal structure should maximize the number of the left groups, and the mixing region is placed at s_{N-2} and s_{N-1}. Conversely, if

$$|b - (x_i - a_{i-1})| > |(a_{i+1} - x_{i+1}) - b|,$$

the cell groups on the right of the mixing region is more balanced, so that the optimal structure should maximize the number of the right groups, and the mixing region is placed at s_2 and s_3. When the two absolute values are equal (so that the mixing is symmetric), then it does not matter where to place the mixing region, so that without loss we can consider placing the

[12] In fact, one can easily show that if we do not consider all posterior means but the two in the mixing, the optimal mixing region should be symmetric.

region at either s_2 and s_3, or s_{N-2} and s_{N-1}. ■

We can further without loss consider the case that the mixing happens at s_{N-2} and s_{N-1} by restricting the mixing to satisfy

$$|b - (x_{N-2} - a_{N-3})| < |(a_{N-1} - x_{N-1}) - b| \,.$$

If not, we can just relabel every a_i and x_i by $1 - a_i$ and $1 - x_i$ (since the prior density is uniform – completely flat). Notice that this kills RL type. Given $N = 2k$, we now have 1 RS type cell, $(k-1)$ LS type cells, and $(k-2)$ LL type cells. Now we are ready to write the payoffs. Still, it is easier to just use the true quadratic preferences

$$\sum_{i=1}^{N} \int_0^1 -(\theta - x_i)^2 f(\theta)\pi(s_i \mid \theta)\mathrm{d}\theta,$$

since the payoff in this case does not depend on the location of the cells. By abstracting away from the specific locations, let c be the length of the mixing region, and z_1, z_2 defined as in Figure 14. In particular, $z_2 = 2b - z_1$. Using the system of equations

$$\begin{cases} w_{N-2} + w_{N-1} = c \\ w_{N-1}z_2 - w_{N-2}z_1 = 0, \end{cases}$$

we get

$$w_{N-2} = \frac{c(2b - z_1)}{2b} \quad \text{and} \quad w_{N-1} = \frac{cz_1}{2b}. \tag{22}$$

The second equation of in the system is acquired from the following

$$\begin{aligned} \frac{a_{N-1} + a_{N-3}}{2} &= \frac{w_{N-2}}{w_{N-2} + w_{N-1}}x_{N-2} + \frac{w_{N-1}}{w_{N-2} + w_{N-1}}x_{N-1} \\ &= \frac{w_{N-2}}{w_{N-2} + w_{N-1}}\left(\frac{a_{N-1} + a_{N-3}}{2} - z_1\right) + \frac{w_{N-1}}{w_{N-2} + w_{N-1}}\left(\frac{a_{N-1} + a_{N-3}}{2} + z_2\right) \\ &= \frac{a_{N-1} + a_{N-3}}{2} + \frac{w_{N-1}}{w_{N-2} + w_{N-1}}z_2 - \frac{w_{N-2}}{w_{N-2} + w_{N-1}}z_1. \end{aligned}$$

Since

$$\int_{a_{N-3}}^{a_{N-1}} -(\theta - x_{N-2})^2 f(\theta)\pi(s_{N-2} \mid \theta)\mathrm{d}\theta = w_{N-2}x_{N-2}^2 - \int_{a_{N-3}}^{a_{N-1}} \theta^2 \pi(s_{N-2} \mid \theta)\mathrm{d}\theta,$$

we can write the payoff from the mixing region as

$$\begin{aligned}
& w_{N-2}x_{N-2}^2 + w_{N-1}x_{N-1}^2 - \int_{a_{N-3}}^{a_{N-1}} \theta^2 \mathrm{d}\theta \\
= & \; w_{N-2}\left(\frac{a_{N-1}+a_{N-3}}{2} - z_1\right)^2 + w_{N-1}\left(\frac{a_{N-1}+a_{N-3}}{2} + z_2\right)^2 - \frac{1}{3}\left(a_{N-1}^3 - a_{N-3}^3\right) \\
= & \; \underbrace{(w_{N-2}+w_{N-1})}_{=(a_{N-1}-a_{N-3})}\left(\frac{a_{N-1}+a_{N-3}}{2}\right)^2 + w_{N-2}z_1^2 + w_{N-1}z_2^2 - \frac{1}{3}\left(a_{N-1}^3 - a_{N-3}^3\right) \\
= & \; w_{N-2}z_1^2 + w_{N-1}z_2^2 - \frac{1}{12}c^3 \\
= & \; \frac{c(2b-z_1)}{2b}z_1^2 + \frac{cz_1}{2b}(2b-z_1)^2 - \frac{1}{12}c^3 \\
= & \; c(2b-z_1)z_1 - \frac{1}{12}c^3
\end{aligned} \tag{23}$$

In the third line, we again use the fact that $w_{N-1}z_2 - w_{N-2}z_1 = 0$. The fourth line uses the definition of c, $a_{N-1} - a_{N-3} = c$. The fifth line uses the weights in (22).

For a generic cell,

$$\int_{a_{j-1}}^{a_j} -(\theta - x_j)\mathrm{d}\theta = -\frac{1}{12}(a_j - a_{j-1})^3, \tag{24}$$

since $x_j = (a_{j-1} + a_j)/2$. Therefore, to find out the payoff of a cell, it suffices to find out the length of the cell. Type RS, LS, LL have lengths

$$8b - c - 2z_1, \quad 4b - c + 2z_1, \quad \text{and} \quad c - 2z_1,$$

respectively. Therefore, the expected payoff of the Receiver, using these length, (24) and (23) is

$$(k-1)\left[-\frac{1}{12}(4b-c+2z_1)^3\right] + (k-2)\left[-\frac{1}{12}(c-2z_1)^3\right] + \left[c(2b-z_1)z_1 - \frac{1}{12}c^3\right] + \left[-\frac{1}{12}(8b-c-2z_1)^3\right].$$

The Receiver needs to choose c and z_1 to maximize this. Equivalently, by multiplying -12, the Receiver minimizes the following

$$(k-1)(4b-c+2z_1)^3 + (k-2)(c-2z_1)^3 + c^3 - 12c(2b-z_1)z_1 + (8b-c-2z_1)^3. \tag{25}$$

Until now, we treat c as a choice variable. But this is in fact not true. Notice that

- There are $(k-2)$ $4b$-long groups, containing $2k-4$ signal realizations; namely, the signal realizations $s_2, \cdots, s_{N-3}$.
- The remaining parts are the first cell (Type LS with length $4b - c + 2z_1$), last cell (Type RS with length $8b - c - 2z_1$), and the mixing region (with length c).

Since the total length sums up to 1,

$$(k-2)\times 4b+4b-c+2z_1+8b-c-2z_1+c=1 \quad\Rightarrow\quad c=4(k+1)b-1.$$

Substitute this back to (25), the Receiver in fact solves the following problem

$$\min_{z_1} 4\left[4b^2(k-1)(k(4b(2k+5)-3)-6)+6bz_1(-8b(k-2)k+2k-5)+3z_1^2\right]+1, \tag{26}$$

subject to the condition that the length of Type-LS cell is closer to $2b$ than the length of Type-RS cell (so that the cell group on the left of the mixing region is more evenly distributed).

Removing all terms that do not contain z_1, we can easily see that (26) is a quadratic function, and the global minimum is attained at

$$b\left[5-8k(k-2)b+2k\right].$$

We need to evaluate this value to determine the monotonicity of (26). Rewrite the terms in the bracket as

$$5+2k-2k(2k-4)2b=5+N-N(N-4)2b.$$

Since b is at most $1/2(N-1)$, this value is larger than

$$5+N-N\times\frac{N-4}{N-1}>5.$$

Therefore, for all feasible z_1 (since $0<z_1<2b$ by definition), (26) is monotone decreasing. The problem now is to find the largest possible z_1 given the constraint

$$|2b-\text{the length of Type-LS cell}|\leqslant|2b-\text{the length of Type-RS cell}|. \tag{27}$$

Here, using $c=4(k+1)b-1$, the length of Type-LS cell is $1+2z_1-4kb$, and the length of Type-RS cell is $1-2z_1-4(k-1)b$. To remove the absolute values, we consider the following three cases:

- <u>*Case 1.*</u> Both lengths are less than $2b$. In this case,

$$\begin{cases} 1+2z_1-4kb<2b \\ 1-2z_1-4(k-1)b<2b, \end{cases} \quad\Rightarrow\quad \frac{1}{2}-(N-1)b<z_1<(N+1)b-\frac{1}{2}.$$

  It can be verified that if $b\in(\frac{1}{2N},\frac{1}{2(N-1)})$, we have

$$\frac{1}{2}-(N-1)b<b<(N+1)b-\frac{1}{2},$$

  so that the range above is valid. The constraint (27) now simplifies to

$$2b-(1+2z_1-4kb)\leqslant 2b-[1-2z_1-4(k-1)b]$$

which gives $z_1 \geqslant b$. Therefore, it is ideal to pick $z_1 = (N+1)b - 1/2$, although this is not feasible since we need the length of Type-LS cell to be strictly less than $2b$ in this case. (In fact, when $z_1 = (N+1)b - 1/2$, the length of Type-LS cell is exactly $2b$. We will go back to this solution in the last case.)

Still, there is another restriction we have not considered: we need the length of Type-RS cell to be positive. When $z_1 = (N+1)b - 1/2$, the length of Type-RS cell is

$$1 - 2\left[(N+1)b - \frac{1}{2}\right] - 4(k-1)b = 2 - (2N-1)2b.$$

When $b \to 1/2(N-1)$, this length is negative. In turn, when

$$2 - (2N-1)2b < 0 \quad \Rightarrow \quad b > \frac{1}{2N-1},$$

one has to pick a further smaller z_1 that keeps the length of Type-RS cell positive. This implies

$$\frac{c}{2} - z_2 < 2b.$$

The LHS is the distance between a_{N-1} and x_{N-1}. This further gives

$$z_1 < z_1^\star = \frac{1}{2} - (N-2)b.$$

Notice that this value is indeed between $(N+1)b - 1/2$ and $\frac{1}{2} - (N-1)b$ when $b > 1/(2N-1)$.

- _Case 2._ The length of Type-LS cell is less than $2b$, but the length of Type-RS cell $\geqslant 2b$. This means that

  $$z_1 \leqslant \frac{1}{2} - (N-1)b$$

  in this case. The constraint (27) now simplifies to

  $$2b - (1 + 2z_1 - 4kb) \leqslant 1 - 2z_1 - 4(k-1)b - 2b \quad \Rightarrow \quad 2Nb - 1 \leqslant 1 - 2Nb.$$

  Given the range of b, the LHS is positive, yet the RHS is negative. Therefore, this case is impossible to satisfy (27).
- _Case 3._ The length of Type-LS cell $\geqslant 2b$, but the length of Type-RS cell is less than $2b$. This gives $z_1 \geqslant (N+1)b - 1/2$. In terms of the constraint in (27),

  $$1 + 2z_1 - 4kb - 2b \leqslant 2b - [1 - 2z_1 - 4(k-1)b] \quad \Rightarrow \quad 1 - 2Nb \leqslant 2Nb - 1,$$

  which always hold. Again, we need the length of Type-RS cell positive, meaning $z_1 < z_1^\star$. When $b \leqslant 1/(2N-1)$, $z_1^\star > (N+1)b - 1/2$, so that picking $z_1^\star$ leads to the length of Type-RS cell being 0, while the length of Type-LS cell larger than $2b$ (so that we are in Case 3). If $b > 1/(2N-1)$, keeping length of Type-RS cell positive means that we are back to Case 1.

We should further note that there is no "Case 4" in which both Type-LS and Type-RS cells have lengths at least $2b$, since the overall length of the structure exceeds 1.

What we learned in the discussion so far is as follows. For any $b \in (\frac{1}{2N}, \frac{1}{2(N-1)})$, due to the quadratic form of (26), the Receiver should pick z_1 as large as possible. For b relatively small ($b < 1/(2N-1)$), so that the length of the mixing region $c = 4(k+1)b - 1$ is relatively small, pushing z_1 to $z_1^\star$ makes the length of Type-LS cell greater than $2b$. When b relatively large, pushing z_1 to $z_1^\star$ makes the length of Type-LS cell smaller than $2b$. But in any case, this always leads to the limit $z_1^\star$, under which the last cell is shrinked to the degenerate (length-0) cell.

We should also note that while we consider that cases that N is even, the argument clearly does not apply to $N = 2$ or $N = 4$. For $N = 2$, it is already clear that the optimal structure must be a partition structure. For $N = 4$, there is no "$4b$-long group" used in the argument of Lemma 20, and the payoff is different from (25) since $k = 2$, and the second term disappears.

Since an optimal structure must exist, we have the following.

Theorem 21. Let $N \geqslant 6$ and N is even. For $b \in \left(\frac{1}{2N}, \frac{1}{2(N-1)}\right)$, one of the following must be true:

- The optimal structure is an even-partition structure with $N-1$ cells.
- The optimal structure sends N signal realizations with at least two pairs of mixing.

Proof. Suppose the optimal structure sends N signal realizations. If it contains only one pair of mixing, it must look like the category characterized in Lemma 20. But when we consider the optimal structure within this category, we found that the optimal structure does not exist unless one signal realization is send with probability zero. This means any structures with all non-trivial signal realizations and exactly one-pair of mixing cannot be optimal. ∎