Information Design in Cheap Talk*

Qianjun Lyu
University of Bonn

Wing Suen
University of Hong Kong

September 15, 2022

Abstract

An uninformed sender publicly commits to an informative experiment about an uncertain state, privately observes its outcome, and sends a cheap-talk message to a receiver. We provide an algorithm valid for arbitrary state-dependent preferences that will determine the sender's optimal experiment, and give sufficient conditions for information design to be valuable or not under different payoff structures. These conditions depend more on marginal incentives—how payoffs vary with the state—than on the alignment of sender's and receiver's rankings over actions within a state.

Keywords: marginal incentives, common interest, concave envelope, quasiconcave envelope, double randomization

JEL Classification: D82, D83

*This paper supersedes the solo project by the first author under the same title. Lyu acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1– 390838866. We would also like to thank Ricardo Alonso, Ettore Damiano, Francesc Dilme, Stephan Lauermann, Gilat Levy, Jin Li, Elliot Lipnowski, Francesco Nava, Matthew Mitchell, Xiaosheng Mu, Ronny Razin, Christopher Sandmann, Xianwen Shi, Dezsö Szalay, Balazs Szentes, Jianrong Tian and Yimeng Zhang for helpful discussions at different stages of this research project.

1 Introduction

Starting from Crawford and Sobel (1982), there is a large economics literature that studies how a biased sender can gain from strategic communication with an uninformed receiver. Much of this literature assumes that the sender is endowed with superior expertise. In many scenarios, however, the sender needs to learn about the payoff-relevant state before communicating with the receiver. For example, news media and think tanks that are biased for or against a political candidate or a government policy often collect information and conduct research in order to influence public opinion. Since the public may not have direct access to the data sources, nor the incentive to use time and effort to assess whether the conclusions drawn indeed follows from the original data, these conclusions effectively become cheap-talk messages. Similarly, financial institutions often have research departments whose work provide the basis for their portfolio recommendations to clients, but whether their investment advice is consistent with the findings of their research is often unverifiable. This paper studies optimal information acquisition when the sender cannot commit to communicating the outcome of his investigations in a verifiable way.

Specifically, we consider a strategic communication game where an imperfectly informed sender can acquire costless information and privately observes the information outcome before sending a cheap-talk message to a receiver, who then takes an action. The sender can commit to an arbitrary information structure. One can interpret this game as a bridge between strategic communication (Crawford and Sobel, 1982) and Bayesian persuasion (Kamenica and Gentzkow, 2011), in the sense that the sender can commit to the information structure but not to truthful reporting.[1]

We study this game under a binary state space and finite action space. Other than this, we allow the sender to have arbitrary state-dependent preferences, which generalizes Lipnowski and Ravid's (2020) analysis of the case where sender has transparent motives (i.e., state-independent preferences). State-dependent preferences create a tension between acquiring more information and alleviating the conflicts of interests. The first incentive is straightforward: since the sender's preferences depend on the true state, acquiring more information allows him to make better use of it. However, more

[1] A standard cheap talk game with fully informed sender can be considered as a game where sender has no commitment to both information structure and truthful reporting, because it is without loss of generality to assume that sender will acquire perfect information in this case.

information may intensify the conflict of interests between sender and receiver, and so affect the sender's incentive to misreport. When designing the information structure, the sender needs to consider the credibility issues in the interim stage after different information outcomes are realized.

A generic feature of a model with discrete action space is that the receiver is indifferent between several actions at certain beliefs, even though the sender may not be indifferent over those actions. Lipnowski and Ravid (2020) and Lipnowski et al. (2022) leverage this observation to show that the sender can benefit from greater credibility if the receiver randomizes over the sender's most preferred action and his less preferred ones at such beliefs. To characterize optimal information design for the sender, it is therefore important not to ignore the actions that are suboptimal for the sender at the interim stage, because inducing such actions may help relax his incentive compatibility constraints. Although the set of receiver actions is infinite when mixing is allowed, in Section 3 we identify a finite set of mixed actions that could yield the optimal equilibrium outcomes for the sender. An algorithm searching over this finite set of candidate actions then yields the optimal information design. The optimal experiment generally induces two possible posterior beliefs, and the receiver may take pure or mixed actions at each of these two beliefs. "Double randomization" (the receiver taking mixed actions at both posterior beliefs) is never optimal if the sender has transparent motives, but it can be part of optimal information design when the sender has state-dependent preferences.

We call $m_i(a) = u_i(a,1) - u_i(a,0)$ player i's *marginal incentive* for action a. It is the difference in utility of action a between state 1 and state 0. Given receiver's best responses, sender's marginal incentives are crucial for determining his truth-telling constraints at any interim beliefs. Moreover, receiver's randomization can smooth sender's marginal incentives, thereby dampening the sender's incentive to misreport. Therefore, mixture over actions can enlarge the set of equilibria and increase the sender's highest achievable payoff. To put it differently, letting the receiver to always choose the sender-preferred action whenever receiver is indifferent can backfire for the sender.

The algorithm we develop in Section 3 allows one to find the highest achievable payoff for the sender at a given prior belief. Information design is said to be *valuable* if there exist some prior beliefs such that the highest achievable payoff is strictly greater than sender's maximum payoff under no information. In Section 4 we provide some sufficient conditions on payoff structures which can guarantee whether information design is valuable or not. Intuitively one would expect that information design is valuable if sender and receiver have "similar" preferences. It turns out that "similarity" is not captured by whether sender and receiver have the same ranking over actions given some particular state; instead the key is the alignment of *marginal incentives* between them. With opposite aligned marginal incentives, information design is not valuable even if the sender's value function is not concave (and even if sender and receiver have identical ranking over actions in one state). With perfectly aligned marginal incentives, information design is valuable if, from sender's perspective, (i) no action *blocks* all other actions; or (ii) no action is *worst* (i.e., worse than all other actions in both states). We also consider the case where sender's preferences are *ordinally state-independent* (i.e., his ranking over actions is the same in the two states). In this case, if sender and receiver have aligned marginal incentives, information design is valuable if and only if the sender's ranking over actions is not identical to receiver's ranking in either of the two states. Such information design requires the receiver to randomize between sender's more-preferred and less-preferred actions.

Section 4 also discusses an application of our framework. Suppose an action a is receiver's best response in one of the states, e.g., state 0, and is also the sender's *best* action in that state. If this action is the sender's *best* action in both states, then standard Bayesian persuasion (with full commitment) would suggest that sender designs an information structure to raise the chances of the receiver taking pure action a by pooling state 0 and state 1 with specific probabilities. If this action is not the sender's best action in state 1, Bayesian persuasion with full commitment would have no generic prediction to the optimal information structure. In our model (with no commitment to truth-telling), if a is the best action in both states, then there is no information structure that can raise the chances of the receiver taking pure action a. If a is not the best action in both states, and specifically a is worse than some action that can be induced at a belief higher than the prior, then the optimal information structure generates a conclusive signal about state 0. This result is quite surprising: despite the fact that sender and receiver have common interest in state 0, the optimal information structure does not necessarily reveal the true state with probability one when the state is 0.

The optimal experiment in our model can be more or less informative than that under Bayesian persuasion. Section 5 provides a sufficient condition under which more information can be transmitted under communication. Specifically, we consider the scenarios where there is a best action that the sender prefers the most across states and is chosen by the receiver at moderate beliefs. In such settings, with mild restrictions, the optimal information structure in our model is strictly more informative than the optimal experiment under full commitment for some prior belief.

In Section 6 we extend our analysis to a model with binary state and a continuum of actions. For simplicity, we assume the sender's indirect value function $\overline{v}(\cdot)$ is continuous on the induced posterior belief held by receiver. Kamenica and Gentzkow (2011) and Lipnowski and Ravid (2020) show that sender's highest achievable payoff is the concave envelope or the quasiconcave envelope of $\overline{v}(\cdot)$ in their respective models. Our model does not permit a similar characterization, because the indirect value function $\bar{v}(\cdot)$ alone is not sufficient to pin down the optimal information design with cheap talk. One will also need to specify the sender's marginal incentives given the induced posterior belief held by receiver. Although there is no simple characterization of the sender's highest achievable payoff as a function of the prior belief, we show that given any prior the highest payoff (and the corresponding optimal information structure) can be determined via an optimization procedure. An immediate implication of this procedure is that in the special case where sender's marginal incentives are always equal to zero, which corresponds to the case of state-independent preferences, then the quasiconcave envelope of $\overline{v}(\cdot)$ gives the highest achievable payoff to the sender—a result first derived by Lipnowski and Ravid (2020).

Related literature. This paper describes a model of Bayesian persuasion with limited commitment, and is especially close to those papers in this literature that relax the commitment assumption at the communication stage. In Guo and Shmaya (2021) and Nguyen and Tan (2021), the sender cannot commit to reporting the true information outcomes but he incurs a cost of making incorrect claims. Alonso and Camara (2021) allow the receiver to endogenously design an audit scheme, which in turn affects the sender's cost of misreporting. Lipnowski et al. (2022) discuss the situation where the sender can misreport the information outcomes at an exogenously given probability. Shishkin (2022) studies test design where the information outcome is hard evidence, but the sender can claim not to have obtained any evidence. In Krähmer (2021), the receiver can cross-check the sender's reports by privately randomizing over information structures. In contrast to these papers, we fully relax the commitment assumption at the communication stage, so that the sender's messages about information outcomes become pure cheap talk.

Ivanov (2010) investigates endogenous information design followed by cheap talk in a uniform-quadratic environment. He characterizes the optimal interval information structures. Deimen and Szalay (2019) consider a two-dimensional state space and the sender can commit to a normally distributed signal structure. Instead, we consider arbitrary payoff functions and general information structures.

Our model of information design with cheap talk follows Lipnowski and Ravid (2020). They focus on situations where the sender has transparent motives, and find that the highest equilibrium payoff the sender can achieve is the quasiconcave envelope of the sender's value function. In contrast, we characterize the solution to a model where sender has arbitrary state-dependent preferences under binary state space. The relevance of the quasiconcave envelope comes from the fact that sender's marginal incentives are identically zero under transparent motives. Two other closely related papers are Lipnowski (2020) and Barros (2022). Instead of characterizing the optimal information design, they provide conditions such that the optimal equilibrium outcome under cheap talk is equivalent with Bayesian persuasion.[2]

Lin and Liu (2022) study the credibility of persuasion assuming that the sender's deviation in messages is not detectable if the marginal distribution of messages remains the same. Their sender's incentive constraints arrive at the ex-ante stage, in the sense that the gain from swapping messages in one state cannot outweigh the loss from that in another state. However, our sender's incentive constraints arrive at the interim stage after the outcome of the experiment is privately revealed to the sender. The incentive constraints in these two papers are not nested. Moreover, Lin and Liu (2022) focus on pure strategy equilibrium where the receiver cannot randomize. Salamanca (2021) studies a mediated communication game in which an informed sender sends an unverifiable message to a mediator, who can commit to a reporting rule based on sender's message. The receiver then takes an action based on the mediator's report. This model reverses the order of information acquisition and communication in our paper in the sense that sender's communication with the mediator can be interpreted as mediator acquiring information. Interestingly, our solution provides a lower bound to sender's

[2]Same as our model, Barros (2022) allows information design prior to communication. He shows that if the optimal equilibrium outcome coincides with Bayesian persuasion, then cheap talk (with a perfectly informed sender) equilibrium outcome coincides with Bayesian persuasion.

highest achievable payoff in Salamanca (2021) under binary state space. We provide a more thorough discussion of the relationship between these two papers and ours in Section 7.

2 The Model

A sender (S) and a receiver (R) initially share a common prior belief about some state θ. The state space $\Theta = \{0, 1\}$ is binary. We use $\mu \in \Delta\Theta$ to represent a probability distribution over the state, where $\mu(\theta)$ stands for the probability of state θ. The prior belief about the state is μ_0.

There is a finite set A of actions, with $|A| \geq 2$. We use a to represent a typical element of A, and use $\alpha \in \Delta A$ to represent a mixed action (i.e., a probability distribution over A). Each player $i \in \{S, R\}$ is an expected utility maximizer, whose utility $u_i(a, \theta)$ generally depends on both the action and the state. We assume no action is strictly dominated for the receiver.

The game consists of two stages. In the first stage, the sender commits to choosing a Blackwell experiment (a mapping from the state space to probability distributions over signals) and conducts the experiment at zero cost. As is standard in the Bayesian persuasion literature, this is equivalent to choosing a distribution of posterior beliefs induced by the experiment. In other words, the sender commits to a simple random posterior $P \in \Delta(\Delta\Theta)$ such that $\mathbb{E}_P[\mu] = \mu_0$, and P has finite support.[3] After the sender conducts the experiment, he privately observes the realization of the random posterior $\mu \in \operatorname{supp} P$. We use $P(\mu)$ to denote the ex-ante probability that the experiment induces posterior μ for the sender (given the prior belief μ_0).

In the second stage, the sender interacts with the receiver in a game of strategic information transmission. Denote M as a rich finite message space. The sender's reporting strategy, $\sigma_S : \Delta(\Delta\Theta) \times \Delta\Theta \to \Delta M$, maps the random posterior P chosen in the first stage and its privately observed realization μ to a distribution of messages. The receiver's decision rule, $\sigma_R : \Delta(\Delta\Theta) \times M \to \Delta A$, maps the random posterior P and the sender's message to a distribution of actions. Each player i's expected utility can be

[3] See Denti et al. (2022). Because we are directly working with the random posterior induced by a Blackwell experiment, we implicitly assume that distinct signals induce different posterior beliefs. It is without loss of generality.

written as:

$$U_i(\sigma_S, \sigma_R, P) = \sum_{\mu \in \operatorname{supp} P, \theta \in \Theta, m \in M, a \in A} P(\mu)\mu(\theta)\sigma_S(m|\mu)\sigma_R(a|m)u_i(a, \theta),$$

where we omit the random posterior P in $\sigma_i(\cdot)$ to simplify the notation.

In this framework the sender's posterior belief formation is trivial, and the receiver's posterior belief is obtained from P and σ_S using Bayes' rule. We focus on Perfect Bayesian Equilibrium, and call (σ_S, σ_R, P) an equilibrium strategy profile if σ_S and σ_R are mutual best responses given P and the belief system. The sender chooses the random posterior $P \in \Delta(\Delta\Theta)$ to maximize his expected utility subject to an equilibrium. If there are multiple equilibria for a given P, we let the sender choose the one that gives him the highest expected utility.

Notice that each player's equilibrium payoff only depends on the joint distribution of receiver's posterior belief and the action induced. Therefore, for every equilibrium such that the sender conceals information through mixed reporting strategy, we can find another truth-telling equilibrium where the sender directly coarsens the experiment in the first place and the equilibrium outcome remains the same. The following result is standard, and its proof is provided in the Appendix.

Lemma 1. *It is without loss of generality to focus on truth-telling equilibria and a binary random posterior, with* $|\operatorname{supp} P| = |\Theta| = 2$.

Because there are only two states, it is often simpler to represent a probability distribution over the state by the probability of state 1. Henceforth, we use μ to stand for the probability of state 1. With slightly abuse of notation, let

$$u_i(a, \mu) := \mu u_i(a, 1) + (1 - \mu)u_i(a, 0)$$

be player i's expected utility from action a when player i has posterior belief μ. Let

$$A_R(\mu) := \operatorname*{argmax}_{a \in A} u_R(a, \mu)$$

be the receiver's best response correspondence, mapping from belief into a non-empty set of actions. We use $v(\mu) := \operatorname{co} u_S(A_R(\mu), \mu)$ to denote the sender's value correspondence given that both the sender and the receiver hold the same posterior belief μ and the receiver responds optimally to this belief. Finally, let

$$\overline{v}(\mu) := \max_{a \in A_R(\mu)} u_S(a, \mu)$$

be sender's value function when both sender and receiver hold the same belief μ and the receiver takes the sender-preferred action in his best response correspondence.

Given Lemma 1, the sender's information design problem can be written as:

$$\max_{P \in \Delta(\Delta\Theta),\, \sigma_R(a|\cdot) \in \Delta A_R(\cdot)} \quad \sum_{\mu \in \operatorname{supp} P} P(\mu) \sum_{a \in A_R(\mu)} \sigma_R(a|\mu) u_S(a, \mu),$$

subject to sender's incentive constraints: for every $\mu', \mu'' \in \operatorname{supp} P$,

$$\begin{aligned} \sum_{a \in A_R(\mu')} \sigma_R(a|\mu')\, u_S(a, \mu') \quad &\geq \quad \sum_{a \in A_R(\mu'')} \sigma_R(a|\mu'')\, u_S(a, \mu'), \\ \sum_{a \in A_R(\mu'')} \sigma_R(a|\mu'')\, u_S(a, \mu'') \quad &\geq \quad \sum_{a \in A_R(\mu')} \sigma_R(a|\mu')\, u_S(a, \mu''); \end{aligned} \tag{1}$$

and subject to the requirement that $|\operatorname{supp} P| = 2$ and P is a mean-preserving spread of μ_0. We denote $W^*(\mu_0)$ as the solution value to this program at prior μ_0.

Figure 1 give two examples of the sender's value function $\overline{v}$. The left panel refers to the case where the sender has state-dependent preferences (the piecewise slope of $\overline{v}$ is arbitrary). The right panel refers to the case where the sender has state-independent preferences ($\overline{v}$ is piecewise constant). The red dashed curves W^* represent the highest payoff the sender can achieve for each prior belief (we will elaborate the algorithm to determine W^* in the next section). The function W^* is piecewise affine.

If the sender with arbitrary preferences has full commitment power to truthfully report the outcome of the experiment, then the concave envelope of $\overline{v}$ determines the highest equilibrium payoff the sender can achieve (Kamenica and Gentzkow, 2011). If the sender with state-independent preferences has no commitment power, the quasi-concave envelope of $\overline{v}$ determines the highest equilibrium payoff the sender can achieve (Lipnowski and Ravid, 2020). In our model, the sender has arbitrary preferences and no commitment power. Therefore, $W^*(\cdot)$ is bounded above by the concave envelope of $\overline{v}(\cdot)$. The relationship between $W^*(\cdot)$ and the quasiconcave envelope of $\overline{v}(\cdot)$ is in general ambiguous (see the red curve in the left panel). We will elaborate more on this

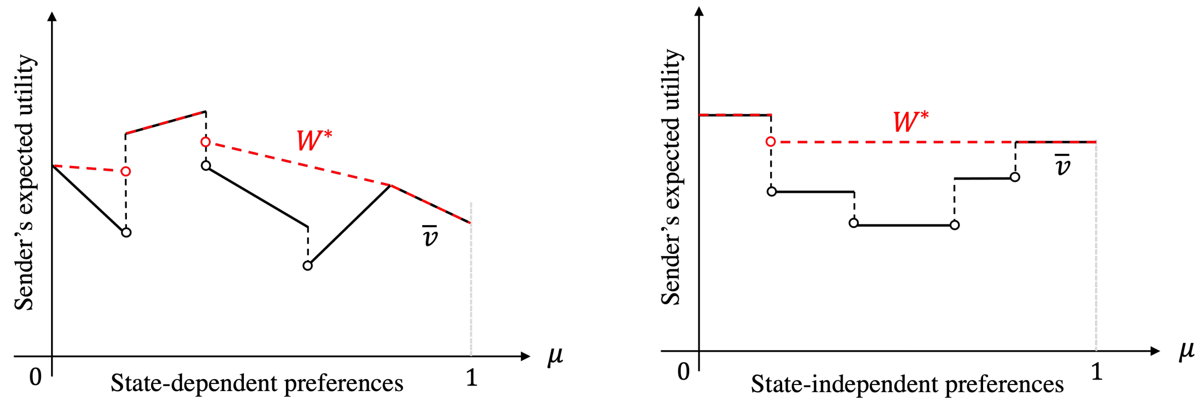

Figure 1: The sender's value function and his highest achievable payoff.

point later.

Lemma 1 suggests that we can focus on random posteriors with a binary support. For a given prior belief μ_0, a binary random posterior is completely pinned down by its support. For example, if $\operatorname{supp} P = \{\mu', \mu''\}$, then the requirement that P is a mean-preserving spread of the prior belief μ_0 implies that μ' and μ'' are induced with probabilities $P(\mu')$ and $1-P(\mu')$, where $P(\mu') = (\mu''-\mu_0)/(\mu''-\mu')$. Therefore, we sometimes refer to a binary random posterior simply by its support.

3 Optimal Information Design

In this section and the next, we make an assumption about A in order to clarify the exposition while avoiding burdensome notation. We assume that every element in A is uniquely optimal for the receiver at some belief. This rules out the possibility that an action $a \in A$ is an exact duplicate of another action $a' \in A$ according to the receiver's preferences (i.e., $u_R(a,\theta) = u_R(a',\theta)$ for all θ). It also rules out the possibility that $a \in A$ is weakly optimal (together with $a', a'' \in A$) for the receiver at exactly one belief, but is strictly worse than a' or a'' at any other belief. The analysis in this paper can be suitably extended to handle situations when this assumption does not hold, but at the cost of more clumsy notation.

Given the assumption that every element of A is a unique best response for the receiver at some belief, we have $|A_R(\mu)| \leq 2$ for all $\mu \in [0,1]$. Moreover, we can

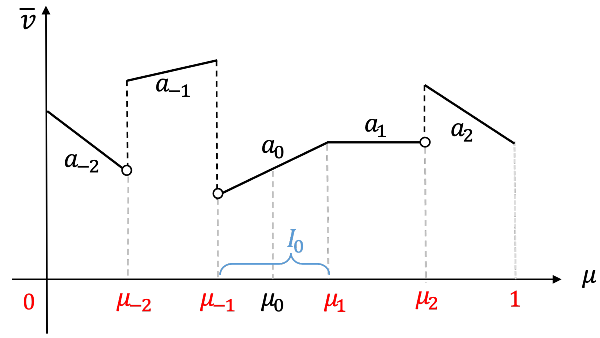

Figure 2: The set of boundary beliefs.

order the actions in A in an increasing sequence, $\{a_{-J}, \ldots, a_{-1}, a_0, a_1, \ldots, a_K\}$, such that action a_n is receiver's best response on a closed interval of beliefs I_n, where the lowest belief in I_n is equal to the highest belief in I_{n-1}.[4] Here, we let $a_0 = A_R(\mu_0)$ be the default action of the receiver when she has no information. For actions higher than a_0, we use μ_k to denote the highest belief that a_k is a best response for the receiver. For actions lower than a_0, we use μ_{-j} to denote the lowest belief that a_{-j+1} is a best response for the receiver. For completeness, we let $\mu_{K+1} = 1$ and $\mu_{-J-1} = 0$. We call $B := \{\mu_{-J-1}, \ldots, \mu_{-1}, \mu_1, \ldots, \mu_{K+1}\}$ the set of *boundary beliefs*. The notation adopted under this convention is illustrated by Figure 2. Elements of B are highlighted in red. We assume the prior μ_0 is in the interior of I_0 in the figure, but this is not important for our analysis.

Proposition 1. *For any prior belief, there exists an optimal binary random posterior whose support is a subset of B.*

Proof. If $W^*(\mu_0) = \overline{v}(\mu_0)$, the random posterior with support $\{\mu_{-1}, \mu_1\}$ (which induces the default action a_0) is optimal. Suppose $W^*(\mu_0) > \overline{v}(\mu_0)$. Then there is an incentive compatible (non-degenerate) random posterior P with $\operatorname{supp} P = \{\mu', \mu''\}$ which induces the receiver to take different responses after different messages. Suppose that at least one element of $\operatorname{supp} P$ does not belong to B, say $\mu'' \in (\mu_k, \mu_{k+1})$. Then, the receiver takes pure action a_k after sender's message μ''. Notice that another random posterior P' with a more spreading support $\{\mu', \mu_{k+1}\}$ is strictly more informative than P. Furthermore, since $u_S(a, \cdot)$ is linear, sender's incentive compatibility constraints (1) still hold under the new random posterior P' if the receiver remain to choose optimally

[4] Specifically, $I_n := \{\mu \in [0,1] : a_n \in A_R(\mu)\}$.

between $\sigma_R(a|\mu')$ and a_k. In other words, under both P and P', if the sender were the decision maker, he would take the same action as the receiver. Therefore, sender's payoff is weakly higher under P' because P' is more informative than P and the sender has the same preference over the relevant two actions with the decision maker (implied by Blackwell's theorem). A similar reasoning applies when μ' does not belong to B. $\square$

Proposition 1 is driven by the observation that, for a given pair of actions, if a less informative information structure is incentive compatible, then the two parties' interests are aligned for each information outcome, which further implies that a more informative information structure is also incentive compatible and provides the sender with a higher expected utility conditional on that the more informative information structure induces the same pair of actions on path. Therefore, it is without loss of generality to consider the most informative information structure that can induce a given pair of actions. Every posterior belief induced by this information structure belongs to the set B. Henceforth, we can focus on binary random posterior P such that $\operatorname{supp} P = \{\mu_{-j}, \mu_k\}$ for some j and k.

For a binary random posterior $\{\mu_{-j}, \mu_k\}$, use $\alpha_{-j} \in \Delta A_R(\mu_{-j})$ and $\alpha_k \in \Delta A_R(\mu_k)$ to represent the mixed strategy taken after message μ_{-j} and μ_k, respectively. Let

$$\mathbb{E}_{\alpha_k}[u_S(a,\mu_k)] = \sum_{a \in A_R(\mu_k)} \alpha_k(a)\, u_S(a,\mu_k)$$

be the sender's expected utility if he has a posterior belief μ_k and the receiver takes mixed strategy α_k, where $\alpha_k(a)$ stands for the probability of taking action a under mixed strategy α_k. Define $\mathbb{E}_{\alpha_{-j}}[u_S(a,\mu_{-j})]$ similarly.

Starting with initial belief $\mu \in (\mu_{-j}, \mu_k)$ (i.e., expectation of the random posterior), the payoff from an experiment that generates posteriors μ_{-j} and μ_k and induces α_{-j} and α_k is:

$$W_{-j,k}(\mu; \alpha_{-j}, \alpha_k) := \frac{\mu_k - \mu}{\mu_k - \mu_{-j}} \mathbb{E}_{\alpha_{-j}}[u_S(a,\mu_{-j})] + \frac{\mu - \mu_{-j}}{\mu_k - \mu_{-j}} \mathbb{E}_{\alpha_k}[u_S(a,\mu_k)].$$

This payoff is linear in μ with a constant derivative,

$$W'_{-j,k}(\cdot\,; \alpha_{-j}, \alpha_k) = \frac{\mathbb{E}_{\alpha_k}[u_S(a,\mu_k)] - \mathbb{E}_{\alpha_{-j}}[u_S(a,\mu_{-j})]}{\mu_k - \mu_{-j}}.$$

If α puts probability one on an action $a \in A_R(\mu)$, then it represents a pure strategy. We sometimes replace α by a to emphasize the difference between pure strategy and mixed strategy.

To analyze incentive compatibility issues, both the level of $\mathbb{E}_\alpha[u_S(a,\mu)]$ and its slope with respect to μ matter because we need to consider the sender's payoff when he deviates from truth-telling to induce α at a different belief. We define the *marginal incentive* corresponding to a mixed strategy α as:

$$m_S(\alpha) := \mathbb{E}_\alpha[u'_S(a,\cdot)].$$

We sometimes use $m_S(a) = u_S(a,1) - u_S(a,0)$ to represent the marginal incentive for a pure action a.

Lemma 2. *An information structure that generates posterior beliefs in $\{\mu_{-j}, \mu_k\}$ and induces α_{-j} and α_k at these two beliefs satisfies sender's incentive compatibility constraints (1) if and only if*

$$m_S(\alpha_{-j}) \leq W'_{-j,k}(\cdot\,; \alpha_{-j}, \alpha_k) \leq m_S(\alpha_k). \tag{IC}$$

Proof. Sender's payoff from inducing α_{-j} at belief μ_k is $\mathbb{E}_{\alpha_{-j}}[u_S(a,\mu_{-j})] + m_S(\alpha_{-j})(\mu_k - \mu_{-j})$. Incentive compatibility requires that this payoff be lower than $\mathbb{E}_{\alpha_k}[u_S(a,\mu_k)]$, which is sender's payoff from inducing α_k at belief μ_k. This is equivalent to $m_S(\alpha_{-j}) \leq W'_{-j,k}(\cdot\,; \alpha_{-j}, \alpha_k)$. The second inequality in (IC) follows similarly from the requirement that sender has no incentive to induce α_k when his private belief is μ_{-j}. □

Lemma 2 suggests a way to find the optimal information structure. For each binary random posterior $\{\mu_{-j}, \mu_k\}$, we check condition (IC) for all pairs $(\alpha_{-j}, \alpha_k) \in \Delta A_R(\mu_{-j}) \times \Delta A_R(\mu_k)$, and select the pair with the highest value of $W_{-j,k}(\mu_0; \alpha_{-j}, \alpha_k)$. Optimizing over j and k would then give the highest achievable payoff $W^*(\mu_0)$ for the sender. The difficulty is that there are infinitely many pairs (α_{-j}, α_k). We now identify the most relevant pairs that will guarantee a solution by searching over such pairs.

For a random posterior P with support $\{\mu_{-j}, \mu_k\}$, there are three types of receiver's best response we need to consider.

Pure strategy (PP). Suppose the receiver takes a pure action after each message. Because receiver's best response at each boundary belief typically contains two elements, there are four possible PP pairs. We only consider one particular pair. Let $\overline{a}_{-j}$ be

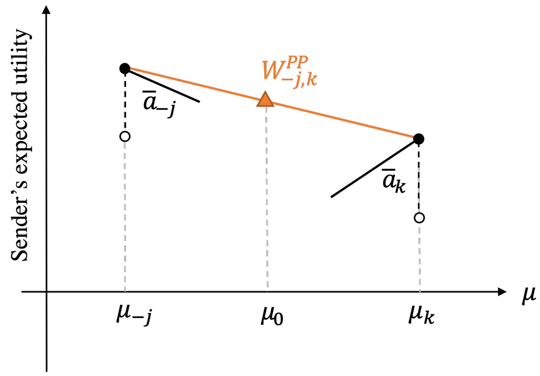

Figure 3: Incentive compatibility for pure strategy.

the sender-preferred action in $A_R(\mu_{-j})$ at belief μ_{-j}; if the sender is indifferent between $A_R(\mu_{-j})$ at belief μ_{-j}, choose $\overline{a}_{-j} = a_{-j+1}$. Let $\underline{a}_{-j}$ be the remaining action (less preferred by the sender) in $A_R(\mu_{-j})$. Similarly, let $\overline{a}_k$ be the sender-preferred action in $A_R(\mu_k)$ at belief μ_k; if the sender is indifferent, choose $\overline{a}_k = a_{k-1}$. Let $\underline{a}_k$ be the remaining action in $A_R(\mu_k)$. We break the indifference in this way because then the random posterior with support $\{\mu_{-j}, \mu_k\}$ is the most informative information structure that can induce $\overline{a}_{-j}$ and $\overline{a}_k$ if the sender reports truthfully.

If inequality (IC) holds for $(\alpha_{-j}, \alpha_k) = (\overline{a}_{-j}, \overline{a}_k)$, we say that the random posterior P is "IC-PP," and we define $W^{PP}_{-j,k} := W_{-j,k}(\mu_0; \overline{a}_{-j}, \overline{a}_k)$.

Figure 3 illustrates an incentive compatible pair $(\overline{a}_{-j}, \overline{a}_k)$. When $u_S(\overline{a}_{-j}, \cdot)$ (the black line on the left) is extended to μ_k, its value is below $u_S(\overline{a}_k, \mu_k)$ (the black dot on the right). This indicates that the sender would not misreport μ_{-j} when his true belief is μ_k. Similarly he has no incentive to misreport μ_k when his true belief is μ_{-j}.

One-sided randomization (PM or MP). Suppose the receiver takes mixed strategy after one of the messages. Consider the case of PM (the MP case is symmetric), and consider the pair $(\alpha_{-j}, \alpha_k) = (\overline{a}_{-j}, \alpha_k^{PM})$, where α_k^{PM} puts weight γ_k on $\overline{a}_k$ and weight $1-\gamma_k$ on $\underline{a}_k$. The value of γ_k is determined by the requirement that the sender is indifferent between $\overline{a}_{-j}$ and α_k^{PM} at belief μ_{-j}:[5]

$$u_S(\overline{a}_{-j}, \mu_{-j}) = \gamma_k u_S(\overline{a}_k, \mu_{-j}) + (1-\gamma_k) u_S(\underline{a}_k, \mu_{-j}). \tag{2}$$

[5]It is possible that γ_k is not uniquely pinned down by the indifference condition. However such situation will not arise in the algorithm we describe below.

(a) One side randomization (PM) (b) Double randomization (MM)

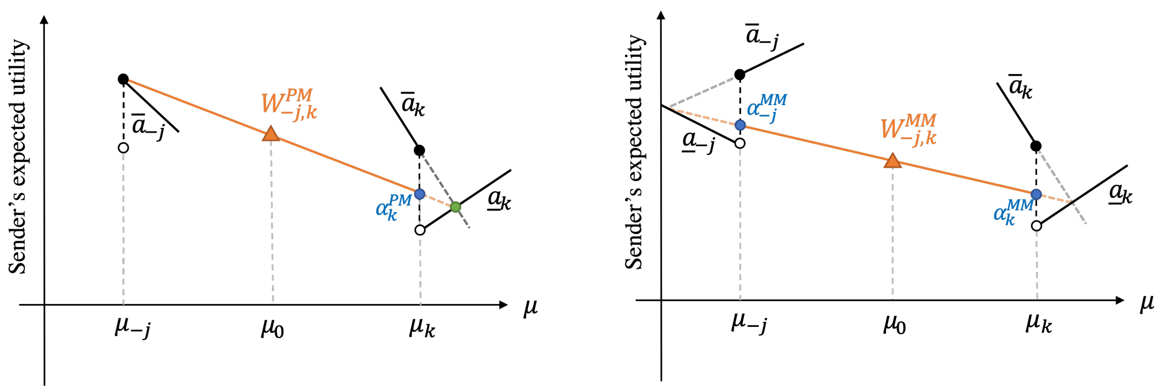

Figure 4: Relaxing incentive constraints by randomization.

Note that the value of γ_k that satisfies this equation may be outside $[0,1]$, in which case α_k^{PM} is not a probability distribution. By construction, the pair $(\overline{a}_{-j}, \alpha_k^{PM})$ satisfies the second inequality in (IC) with equality. If it also satisfies and first inequality in (IC), and if α_k^{PM} is a probability distribution and therefore a valid mixed action, we say that the information structure P is "IC-PM," and we define $W^{PM}_{-j,k} := W_{-j,k}(\mu_0; \overline{a}_{-j}, \alpha_k^{PM})$.[6]

The left panel of Figure 4 illustrates this construction. Since the sender is indifferent between $\overline{a}_j$ and α_k^{PM} at belief μ_{-j}, his expected utility from such one-sided randomization equals the expected utility from α_k^{PM} itself. Therefore, it is easy to pin down $W_{-j,k}(\cdot\,; \overline{a}_j, \alpha_k^{PM})$ graphically. First, draw an affine curve connecting $u_S(\overline{a}_{-j}, \mu_{-j})$ and the green dot—the intersection point between the extended curves of $u_S(\overline{a}_k, \cdot)$ and $u_S(\underline{a}_k, \cdot)$. If the sender's expected utility from α_k^{PM}—the blue dot—lies in the sender's value correspondence $v(\mu_k)$, which is the range between $u_S(\underline{a}_k, \mu_k)$ and $u_S(\overline{a}_k, \mu_k)$, then α_k^{PM} is a valid mixed action. The value of the affine curve $W_{-j,k}(\cdot\,; \overline{a}_j, \alpha_k^{PM})$ at μ_0 is the sender's expected utility from the one side randomization that we identify.

Double randomization (MM). This involves the receiver taking mixed strategy after each message. Let α_{-j}^{MM} be a mixed action that puts weight γ_{-j} on $\overline{a}_{-j}$ and weight $1-\gamma_{-j}$ on $\underline{a}_{-j}$. Let α_k^{MM} be a mixed action that puts weight γ_k on $\overline{a}_k$ and weight $1-\gamma_k$ on $\underline{a}_k$. The weights γ_{-j} and γ_k are chosen in such way that the sender is indifferent between

[6] If α_k^{PM} is not a valid probability distribution, we let $\mathbb{E}_{\alpha_k}[u_S(a, \mu_k)] := \gamma_k u_S(\overline{a}_k, \mu_{-j}) + (1 - \gamma_k) u_S(\underline{a}_k, \mu_{-j})$ given that γ_k is the solution to equation (2). The corresponding value of $W^{MP}_{-j,k}$ is defined accordingly. We adopt a similar convention for the cases of PM and MM.

α_{-j}^{MM} and α_k^{MM} both at belief μ_{-j} and at belief μ_k:

$$\mathbb{E}_{\alpha_{-j}^{MM}}[u_S(a,\mu_{-j})] = \mathbb{E}_{\alpha_k^{MM}}[u_S(a,\mu_{-j})], \qquad \mathbb{E}_{\alpha_k^{MM}}[u_S(a,\mu_k)] = \mathbb{E}_{\alpha_{-j}^{MM}}[u_S(a,\mu_k)].$$

The value of (γ_{-j},γ_k) that solves these two equations may be outside $[0,1]^2$, in which case one of α_{-j}^{MM} and α_k^{MM} is not a valid mixed action.[7] By construction, $(\alpha_{-j}^{MM},\alpha_k^{MM})$ satisfies $m_S(\alpha_{-j}^{MM}) = W'_{-j,k}(\cdot\,;\alpha_{-j}^{MM},\alpha_k^{MM}) = m_S(\alpha_k^{MM})$, and so the incentive constraints (IC) hold. If both α_{-j}^{MM} and α_k^{MM} are valid mixed actions, we say that the random posterior P is "IC-MM," and we define $W_{-j,k}^{MM} = W_{-j,k}(\mu_0;\alpha_{-j}^{MM},\alpha_k^{MM})$. This construction is illustrated graphically in the right panel of Figure 4. To verify that α_{-j}^{MM} and α_{-j}^{MM} are valid mixed actions, we just need to make sure that the blue dots in that figure lie on the sender's value correspondence $v(\cdot)$ at the respective beliefs.

Now we introduce an algorithm that yields the highest achievable payoff $W^*(\mu_0)$, together with an implied optimal random posterior P^*.

Algorithm 1:

1. For every pair $(-j,k)\in\{1,\ldots,J+1\}\times\{1,\ldots,K+1\}$, compute $W_{-j,k}(\mu_0;\overline{a}_{-j},\overline{a}_k)$ and rank these values from highest to lowest.[8] Starting from the pair with the highest value, verify whether it is IC-PP or not. Stop the first time an IC-PP pair is found. Assign $W^1 = W_{-j,k}^{PP}$ for such pair and let the set of $(-j,k)$ pairs with $W_{-j,k}^{PP}$ strictly higher than W^1 be S_1. If there does not exist an IC-PP pair, assign $W^1 = \overline{v}(\mu_0)$ and let $S_1 = \{1,\ldots,J+1\}\times\{1,\ldots,K+1\}$,
2. For every pair $(-j,k)$ in S_1:
   (a) Compute $W_{-j,k}(\mu_0;\overline{a}_{-j},\alpha_k^{PM})$ and re-rank these values from highest to lowest. Starting with the pair with the highest value, verify whether it is IC-PM or not. Stop the first time when an IC-PM pair is found. Assign $W^{(a)} = W_{-j,k}^{PM}$ for such pair and let the set of $(-j,k)$ pairs with $W_{-j,k}^{PM}$ strictly higher than $W^{(a)}$ be $S^{(a)}$. If none of them is IC-PM, assign $W^{(a)} = \overline{v}(\mu_0)$ and $S^{(a)} = S_1$
   (b) Go through a symmetric procedure in the case for MP. Assign $W^{(b)} = W_{-j,k}^{MP}$ the first time an IC-MP pair is found and let the set of $(-j,k)$ pairs with $W_{-j,k}^{MP}$ strictly higher than $W^{(b)}$ be $S^{(b)}$. If none of them is IC-PM, assign $W^{(b)} = \overline{v}(\mu_0)$ and $S^{(b)} = S_1$.

[7] It is possible that γ_j and γ_k are not uniquely pinned down by the indifference conditions. However such situation will not arise in the algorithm we describe below.

[8] It is not important how we break ties.

(c) Let $W^2 = \max\{W^{(a)}, W^{(b)}\}$. Let $S_2 = S^{(a)} \cup S^{(b)}$.

3. For every pair $(-j,k)$ in S_2, compute $W_{-j,k}(\mu_0; \alpha_{-j}^{MM}, \alpha_k^{MM})$ and re-rank these values from the highest to lowest. Starting with the pair with the highest value, verify whether it is IC-MM or not. Stop the first time an IC-MM pair is found and assign $W^3 = W_{-j,k}^{MM}$ for such pair. If none of them is IC-MM, assign $W^3 = \overline{v}(\mu_0)$.
4. Assign $W^*(\mu_0) = \max\{W^1, W^2, W^3\}$. The random posterior with support $\{\mu_{-j}, \mu_k\}$ corresponding to the $(-j,k)$ pair that yields $W^*(\mu_0)$ is optimal.

Theorem 1. *Algorithm 1 determines the highest achievable payoff for the sender.*

In the algorithm, although there are infinitely many possible mixed actions that the receiver would take for each pair of $(-j,k)$, we only check four possibilities, namely IC-PP, IC-PM, IC-MP and IC-MM. The procedure we describe guarantees a solution without searching over all possibilities across all $(-j,k)$. We prove the sufficiency of such simplification in the appendix.

The construction behind this algorithm generalizes Lipnowski and Ravid (2020) to the case of binary state with arbitrary preferences. When the sender has state-independent preferences (transparent motives), the marginal incentive $m_S(\alpha)$ is equal to 0 for every mixed action α (including pure action). The incentive compatibility requirement (IC) in Lemma 1 would then require $W'_{-j,k}(\cdot\,; \alpha_{-j}, \alpha_k) = 0$ for any action pair, which implies that sender's highest achievable payoff will be given by the quasiconcave envelope of $\overline{v}(\cdot)$. In our setup, the fact that $m_S(\alpha_{-j})$ is in general different from $m_S(\alpha_k)$ means that $W'_{-j,k}(\cdot\,; \alpha_{-j}, \alpha_k)$ is not restricted to be equal to 0. The sender in our setup can achieve a payoff greater than or less than the quasiconcave envelope of $\overline{v}(\cdot)$.

The use of randomization to relax incentive compatibility constraints also follows Lipnowski and Ravid (2020), who point out that sender can gain credibility by degrading self-serving information. In our model, randomizing over recommending the sender-preferred action $\overline{a}_k$ in $A_R(\mu_k)$ with the less preferred one $\underline{a}_k$ may make the sender less inclined to misreport μ_k when his true belief is μ_{-j}. Thus the sender in our model can benefit from inducing the receiver to choose actions which are ex post suboptimal from sender's point of view. The algorithm provides a general method of determining the mixture actions that will maximize the sender's payoff for general state-dependent preferences.

If the sender is recommending mixed actions α_{-j} and α_k at beliefs μ_{-j} and μ_k, he could strictly raise his payoff by putting more weight on $\overline{a}_{-j}$ and $\overline{a}_k$ in these mixed actions, provided that the new pair of mixed actions are still incentive compatible. Such deviation is always feasible as long as marginal incentives $m_S(\cdot)$ are equal for all actions, which explains why double randomization is never optimal under transparent motives ($m_S(\cdot)$ identically equal to 0). In our model with general preferences, unlike Lipnowski and Ravid (2020), such deviation may not be feasible, and therefore double randomization can remain a candidate as part of optimal information design.

4 When is Information Design Valuable?

The algorithm in Section 3 provides a systematic way to check whether sender's maximum payoff $W^*(\mu_0)$ under an optimal information structure strictly exceeds his default payoff $\overline{v}(\mu_0)$ for a given prior belief μ_0. We say that information design is *valuable* if $W^*(\mu_0) > \overline{v}(\mu_0)$ for some prior belief $\mu_0 \in [0,1]$. Thus information design is *not valuable* if $W^*(\mu_0) = \overline{v}(\mu_0)$ for all μ_0. To decide whether information design is valuable or not, however, is laborious because one would have to run the algorithm in Section 3 for every prior belief. Unlike Kamenica and Gentzkow (2011) or Lipnowski and Ravid (2020), there is no easy way to characterize the necessary and sufficient condition for information design to be valuable in our model based simply on the concavity or quasiconcavity of $\overline{v}(\cdot)$.[9] In our model, whether information design is valuable depends less on the concavity properties of $\overline{v}(\cdot)$ than on the structure of marginal incentives $m_S(\cdot)$. We provide some economically meaningful sufficient conditions in this section that will settle this question.

We introduce the following concepts that relate to the conflict of interest between sender and receiver.

Definition 1. Sender and receiver have *opposite marginal incentives* if, for any $a', a'' \in A$,

$$m_R(a') < m_R(a'') \iff m_S(a') > m_S(a'').$$

[9] In a model with discrete action space, sender's value function $\overline{v}(\cdot)$ is (generically) discontinuous at beliefs for which the receiver is indifferent between different actions. Since a discontinuous function is not concave, information design is always valuable according to our definition when there is full commitment.

They have *aligned marginal incentives* if, for any $a', a'' \in A$,

$$m_R(a') < m_R(a'') \iff m_S(a') < m_S(a'').$$

The notion of opposite or aligned marginal incentives has little to do with comparing the level (or the ranking) of utilities attached to different actions at a given belief by the receiver and by the sender. For example, sender and receiver may have identical preference ranking over actions in A if they know the true state is, say, state 0; yet they may still have opposite marginal incentives according to Definition 1.

Our definition is related to supermodularity or submodularity between action and state. With a binary state space, it is without loss of generality to assume that the receiver preferences are supermodular in (a, θ) (because we order actions in such a way that higher actions are chosen at higher beliefs). According to this convention, if $u_S(\cdot,\cdot)$ is strictly submodular, then sender and receiver have opposite marginal incentives. If $u_S(\cdot,\cdot)$ is strictly supermodular, they have aligned marginal incentives.

Proposition 2. *If sender and receiver have opposite marginal incentives, then information design is not valuable.*

Proof. Consider an arbitrary prior belief $\mu_0 \in (0,1)$. Take any pair of boundary beliefs such that $\mu_{-j} < \mu_0 < \mu_k$. Take any arbitrary receiver's best responses $\alpha_{-j} \in \Delta A_R(\mu_{-j})$ and $\alpha_k \in \Delta A_R(\mu_k)$, with $\alpha_{-j} \neq \alpha_k$. Our convention of ordering actions implies that $m_R(\alpha_{-j}) < m_R(\alpha_k)$, and hence $m_S(\alpha_{-j}) > m_S(\alpha_k)$. By Lemma 1, this pair of actions (α_{-j}, α_k) cannot be incentive compatible. This means that there is no incentive compatible binary information structure that can induce different actions at the boundary beliefs. Therefore an optimal information structure cannot outperform an uninformative experiment. □

Proposition 2 is valid regardless of how the sender's and receiver's preferences compare in any one of the two states. As long as their marginal incentives are opposite, information design has no value. Figure 5 shows one such example. The sender's value function $\overline{v}(\cdot)$ in this figure is obviously not concave. Nevertheless, because the slope in each separate segment of $\overline{v}(\cdot)$ is decreasing, Proposition 2 implies that, for any prior belief, information design cannot improve the sender's payoff when he cannot commit to truth telling.

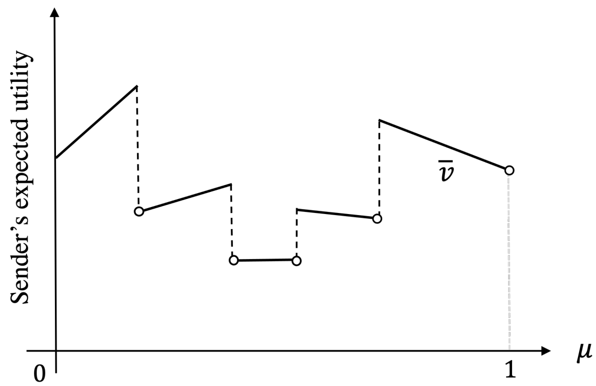

Figure 5: Information design has no value when sender and receiver have opposite marginal incentives.

Next, we turn to the case where sender and receiver have aligned marginal incentives.

Definition 2. An action $a' \in A$ *blocks* $a'' \in A$ if

$$u_S(a', \mu'') \geq u_S(a'', \mu'') \quad \text{for all } \mu'' \in \{\mu : a'' \in A_R(\mu)\}.$$

Action $a' \in A$ is an *all-blocker* if it blocks all actions in A.

According to Definition 2, action $a' \in A$ is an all-blocker if and only if

$$u_S(a', \mu) \geq \overline{v}(\mu) \quad \text{for all } \mu \in [0, 1].$$

If a' does not block a'' and a'' does not block a', then the incentive compatibility constraints (1) can be satisfied and there is an IC-PP information structure at some initial belief that will induce these two actions.

Definition 3. An action $a' \in A$ is *worst* if, for all $a'' \in A$,

$$u_S(a', \theta) \leq u_S(a'', \theta) \quad \text{for all } \theta \in \{0, 1\}.$$

An action $a' \in A$ is *best* if, for all $a'' \in A$,

$$u_S(a', \theta) \geq u_S(a'', \theta) \quad \text{for all } \theta \in \{0, 1\}.$$

If action a' is worst, the sender prefers any action in A to this action at any belief μ.

It implies that any other action in A blocks a', and a' does not block any other action. The converse is not true. Similarly, a best action is necessarily an all-blocker, but an all-blocker need not be best.

Proposition 3. *If the sender and the receiver have aligned marginal incentives, then information design is valuable if either of the following holds:*

(a) No action is an all-blocker for the sender.

(b) No action is worst for the sender.

Proof of part (a). For any pair of distinct actions $a', a'' \in A$, there are four mutually exclusive possibilities: (1) a' blocks a'' and a'' does not block a'; (2) a'' blocks a' and a' does not block a''; (3) neither action blocks the other; or (4) each action blocks the other. Case (4) is impossible under aligned marginal incentives. We claim that at least one pair of actions in A must fall under case (3). Suppose this claim is false, so that case (1) and case (2) mutually exhaust all possibilities on A. Then the binary relation "block" on A would be reflexive, complete, and antisymmetric. In the next paragraph, we show that it would also be transitive, and therefore "block" would be a total order on the finite set A, which would further imply that there is a maximal action on A, i.e., an all-blocker action exists in A. This is a contradiction, and therefore we conclude that at least one pair of actions, a' and a'', must fall under case (3). This pair of actions are strictly IC-PP because the complement of Definition 2 imposes strict inequality. Thus, an information structure that induces these two actions will improve sender's payoff when, for example, the prior belief is in the interior of $\{\mu : a' \in A_R(\mu)\}$.

To see why transitivity holds under the premise that cases (1) and (2) mutually exhaust all possibilities on A, consider $|A| \geq 3$. (If $|A| = 2$, it is immediate that "block" is a total order as the two actions are comparable.) Suppose a blocks b and b blocks c, and let μ_a, μ_b and μ_c be three distinct beliefs at which these three actions are respective best responses. (a) Suppose $\mu_a < \mu_b$. (a)(i) If $\mu_c < \mu_b$, then a blocks b implies $u_S(a, \mu_b) \geq u_S(b, \mu_b)$. Aligned marginal incentives (supermodularity of $u_S(\cdot, \cdot)$) then imply $u_S(a, \mu_c) \geq u_S(b, \mu_c) \geq u_S(c, \mu_c)$, where the last inequality follows because b blocks c. Since this argument holds for any $\mu_c < \mu_b$, we conclude that a blocks c. (a)(ii) If $\mu_c > \mu_b$, then b blocks c implies $u_S(b, \mu_c) \geq u_S(c, \mu_c)$. Aligned marginal incentives then imply that there exists $\mu_a \in \{\mu : a \in A_R(\mu)\}$ such that $u_S(a, \mu_a) > u_S(b, \mu_a) \geq u_S(c, \mu_a)$, where the first inequality follows because b does not block a. This shows that c does

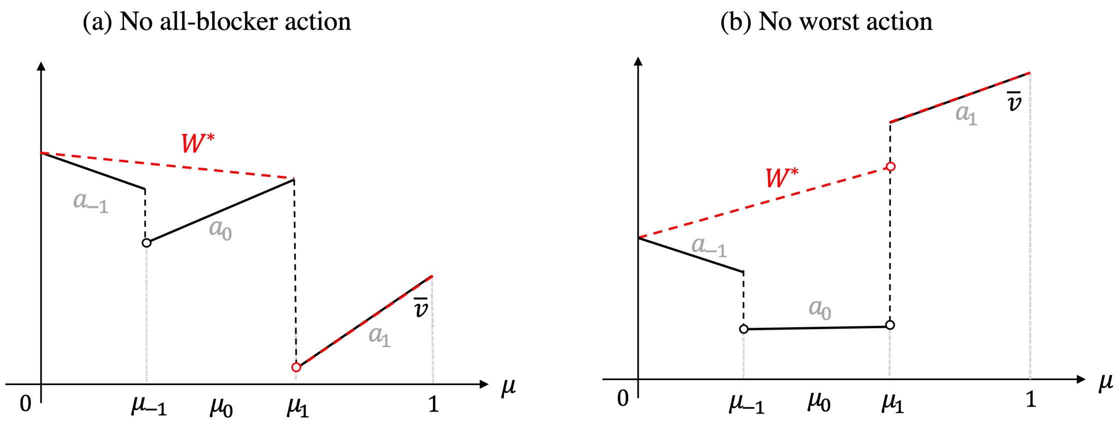

Figure 6: Aligned marginal incentives.

not block a. Since cases (1) and (2) are mutually exhaustive possibilities under the supposition that no pair of action falls under case (3), a blocks c whenever c does not block a. The analysis of (b), where $\mu_a > \mu_b$, is symmetric. In both cases, a blocks b and b blocks c implies a blocks c. □

The proof of part (b) of Proposition 3 involves finding IC-PM pairs and is more tedious; we leave it to the Appendix. Figure 6 provides two examples to illustrate this proposition. The left panel of Figure 6 shows a case where there is no all-blocker action. Proposition 3(a) implies that there must exist a pair of distinct actions such that neither action blocks the other action. In the figure, the information structure $\{0, \mu_1\}$ is IC-PP for a_{-1} and a_0 and it improves the sender's payoff at prior μ_0.

Next, consider the right panel, where a_{-1} is the least-preferred action in state 1 but is not the least-preferred action in state 0. There is no worst action. In this example, the sender prefers a_1 to a_{-1} to a_0 at belief 0. Therefore, we can find a randomization $\alpha_1^{PM} \in \Delta A_R(\mu_1)$ (shown by the red dot) such that the sender is indifferent between a_{-1} and α_1^{PM} at belief 0. Moreover, from aligned marginal incentives, we have $m_S(\alpha_1^{PM}) > m_S(a_{-1})$, implying that the sender must strictly prefer α_1^{PM} to a_{-1} at belief μ_1. Hence, $\{0, \mu_1\}$ is IC-PM and induces a_{-1} and α_1^{PM} at these two beliefs. This information structure improves the sender's payoff when, for example, the prior belief is μ_0.

Imagine that the payoff corresponding to action a_{-1} in the left panel of Figure 6 shifts down to such an extent that a_{-1} becomes the least-preferred action in both state 0 and state 1. Then any valid mixed action $\alpha_1 \in \Delta A_R(\mu_1)$ would be strictly preferred to a_{-1} at belief 0. Information design would not be valuable in this case because there are no incentive compatible pair of actions.

Note also that action a_1 in the left panel of Figure 6 is a worst action, and action a_1 in the right panel is an all-blocker action. This shows that information design can still be valuable when an all-blocker action or a worst action exists for the sender. In other words, conditions (a) and (b) in Proposition 3 are each sufficient for information design to be valuable, but neither of them is necessary.

Proposition 3 suggests that the alignment of marginal incentives between sender and receiver is important for determining whether information design is valuable or not. Given aligned marginal incentives, the alignment of preference ranking over actions matters a lot less. To see this point more clearly, consider a special, yet economically relevant, class of sender preferences.

Definition 4. Sender's preferences are *ordinally state-independent* if, for every $a', a'' \in A$,

$$u_S(a',1) > u_S(a'',1) \iff u_S(a',0) > u_S(a'',0).$$

This definition implies that sender's ranking over actions is the same at any $\mu \in [0,1]$. It is a generalization of transparent motives, because this class of preferences does not require $m_S(a)$ to be equal to 0 for all a.

Given the labeling we adopt on the action space, the receiver's ranking over action in state 0 is decreasing in the index of actions, and is increasing in the index of actions when the state is 1. A sender with ordinally state-independent preferences can have arbitrary ranking over actions even though his marginal incentives from each action is a monotone function in receiver's (when they have aligned marginal incentives).

Proposition 4. *Suppose sender and receiver have aligned marginal incentives, and the sender's preferences are ordinally state-independent. Information design is valuable if and only if and the sender's ranking of actions is non-monotone in the index of actions.*

Proof. The "only if" part is simple. If the sender's ranking is monotone in the index of the actions, then there does not exist an informative equilibrium outcome in which the receiver chooses different actions (including mixed actions) after different messages. This implies that information design is not valuable.

To show the "if" part, suppose the sender's ranking is non-monotone in the index of the actions. This implies that there must be at least three actions in A. Moreover there exists an index n such that either (1) the sender prefers a_{n-1} to a_n, but a_{n+1} is ranked above a_n; or (2) sender prefers a_n to a_{n-1}, but a_{n+1} is ranked below a_n. Let $\underline{I}_n := \min\{\mu : a_n \in A_R(\mu)\}$ and $\overline{I}_n := \max\{\mu : a_n \in A_R(\mu)\}$. In case (1a), the sender prefers a_{n+1} to a_{n-1} to a_n at all beliefs, including at belief $\underline{I}_{n-1}$. Therefore, there exists a mixture $\alpha_n \in \Delta\{a_n, a_{n+1}\}$ that the receiver would optimally choose at belief $\overline{I}_n$ such that the sender is indifferent between a_{n-1} and α_n at belief $\underline{I}_{n-1}$. Moreover, because $m_S(\alpha_n) > m_S(a_{n-1})$, the random posterior with support $\{\underline{I}_{n-1}, \overline{I}_n\}$ and an expectation $\mu_0 \in (\underline{I}_{n-1}, \overline{I}_n)$ is IC-PM given the receiver optimally chooses between a_{n-1} and α_n. In case (1b), the sender prefers a_{n-1} to a_{n+1} to a_n at any belief. With a similar reasoning, the random posterior with support $\{\underline{I}_n, \overline{I}_{n+1}\}$ is IC-MP given the receiver optimally chooses between some $\alpha_{n-1} \in \Delta\{a_{n-1}, a_n\}$ and a_{n+1}. In case (2a), the sender prefers a_n to a_{n-1} to a_{n+1}. Then the random posterior with support $\{\underline{I}_{n-1}, \overline{I}_n\}$ is IC-PM given the receiver optimally chooses between a_{n-1} some $\alpha'_n \in \Delta\{a_n, a_{n+1}\}$. In case (2b), the sender prefers a_n to a_{n+1} to a_{n-1}. Then the random posterior with support $\{\underline{I}_n, \overline{I}_{n+1}\}$ is IC-MP given the receiver optimally chooses between some $\alpha'_{n-1} \in \Delta\{a_{n-1}, a_n\}$ and a_{n+1}. □

Ordinal state-independence implies that there does not exist an incentive compatible information structure that induces pure actions by the receiver, because for any two distinct actions $a' \neq a''$, either $u_S(a', \mu) > u_S(a'', \mu)$ for all $\mu \in [0, 1]$, or the opposite (strict) inequality holds for all $\mu \in [0, 1]$. Nevertheless, provided marginal incentives are aligned, Proposition 4 shows that, information design is generally valuable to the sender except in the special case where his ranking over actions is identical to the receiver's ranking in one of the states. Such information design necessarily requires randomization to relax incentive constraints, and we rely on IC-PM or IC-MP information structures in the proof of Proposition 4.

Finally, in many situations, the sender and the receiver may have common interests in one state but conflicting interests in another state. By this, we mean that receiver's optimal action in one state is also sender's most-preferred action in that state (their rankings over other actions in that state can be different).

Definition 5. Sender and receiver have *common interest in one state* if, for $\theta = 0$ or $\theta = 1$,

$$u_S(a, \theta) \geq u_S(a', \theta) \quad \text{for all } a \in A_R(\theta) \text{ and all } a' \in A.$$

A recurring theme in the Bayesian persuasion literature is that, if preferences are state-independent and if there is full commitment, the sender would pool the other state with the common-interest state in order to raise the probability that the receiver will take the sender's most-preferred action. In our model, such pooling may not be feasible because of truth-telling constraints. Furthermore, because the sender has arbitrary preferences, he does not always want to induce the most-preferred action in the common-interest state even if he can.

Proposition 5. *Let the common-interest state be state 0, and let the optimal action corresponding to that state be a_{-J}.*

(a) If a_{-J} is not an all-blocker action, then information design is valuable.
(b) If there exists an action $a_k \in \{A_R(\mu) : \mu \in (\mu_0, 1]\}$ such that a_{-J} does not block a_k, then $0 \in \operatorname{supp} P^*$.

Proof. (a) Since a_{-J} is not an all-blocker action, there exists a different action a' such that $u_S(a_{-J}, \mu') < u_S(a', \mu')$ for some $\mu' \in \{\mu : a' \in A_R(\mu)\}$. Moreover, by the definition of common interest in state 0, $u_S(a_{-J}, 0) \geq u_S(a', 0)$. Thus, the random posterior with support $\{0, \mu'\}$ that induces a_{-J} and a' at these two beliefs is IC-PP. Furthermore, it strictly improves the sender's payoff, for example, when the prior belief is in the interior of $\{\mu : a_{-J} \in A_R(\mu)\}$.

(b) Since a_{-J} does not block a_k, we have $u_S(a_k, \mu_{k+1}) > u_S(a_{-J}, \mu_{k+1})$. Let $\overline{a}_{k+1}$ be the sender-preferred action in $A_R(\mu_{k+1})$. Then $u_S(\overline{a}_{k+1}, \mu_{k+1}) \geq u_S(a_k, \mu_{k+1}) > u_S(a_{-J}, \mu_{k+1})$. From the definition of common interest in state 0, $u_S(a_{-J}, 0) \geq u_S(\overline{a}_{k+1}, 0)$. Therefore, the random posterior with support $\{0, \mu_{k+1}\}$ is IC-PP if the receiver optimally chooses between a_{-J} and $\overline{a}_{k+1}$.

By Proposition 1, it is without loss of generality to only consider information structures that generate posteriors that are in the set of boundary beliefs B. Consider an incentive compatible random posterior P' with support $\{\mu_{-j'}, \mu_{k'}\}$ that induces (pure or mixed) actions $\alpha_{-j'} \in A_R(\mu_{-j'})$ and $\alpha_{k'} \in A_R(\mu_{k'})$. Consider another random posterior P with support $\{0, \mu_{k+1}\}$ (and with the same expectation μ_0 as P') that induces actions a_{-J} and $\overline{a}_{k+1}$ at these beliefs. There are two possibilities.

Case (1) $\mu_{k'} = \mu_{k+1}$. Since P' is incentive compatible, the payoff from this information structure is

$$\begin{aligned} W_{-j',k'}(\mu_0;\alpha_{-j'},\alpha_{k'}) &= \mathbb{E}_{P'}\left[\max\left\{\mathbb{E}_{\alpha_{-j'}}[u_S(a,\mu)],\mathbb{E}_{\alpha_{k'}}[u_S(a,\mu)]\right\}\right] \\ &\leq \mathbb{E}_{P}\left[\max\left\{\mathbb{E}_{\alpha_{-j'}}[u_S(a,\mu)],\mathbb{E}_{\alpha_{k'}}[u_S(a,\mu)]\right\}\right] \\ &\leq \mathbb{E}_{P}\left[\max\left\{u_S(a_{-J},\mu),\mathbb{E}_{\alpha_{k'}}[u_S(a,\mu)]\right\}\right] \\ &\leq \mathbb{E}_{P}\left[\max\left\{u_S(a_{-J},\mu),u_S(\overline{a}_{k+1},\mu)\right\}\right] \\ &= W_{-J,k+1}(\mu_0;a_{-J},\overline{a}_{k+1}), \end{aligned}$$

where the first inequality follows from the fact that P is a mean-preserving spread of P', and the second inequality follows from $\mathbb{E}_{\alpha_{-j'}}[u_S(a,0)] \leq u_S(a_{-J},0)$. The third inequality comes from $\mathbb{E}_{\alpha_{k'}}[u_S(a,\mu_{k+1})] \leq u_S(\overline{a}_{k+1},\mu_{k+1})$. The last equality comes from the fact that the random posterior with support $\{0,\mu_{k+1}\}$ is IC-PP for the pair of actions a_{-J} and $\overline{a}_{k+1}$.

Case (2) $\mu_{k'} \neq \mu_{k+1}$. If the information structure $\{0,\mu_{k'}\}$ that induces a_{-J} and $\alpha_{k'}$ is incentive compatible, then the same argument provided in case (1) shows that this information structure will give a higher payoff to the sender than does P'. So we only need to consider the case that $\{0,\mu_{k'}\}$ is not incentive compatible. In this case, because a_{-J} is sender's most-preferred action in state 0, incentive compatibility can fail only when $u_S(a_{-J},\mu_{k'}) > \mathbb{E}_{\alpha_{k'}}[u_S(a,\mu_{k'})]$ (i.e., the sender prefers a_{-J} to $\alpha_{k'}$ at belief $\mu_{k'}$). Moreover, since the sender prefers $\alpha_{k'}$ to $\alpha_{-j'}$ at belief $\mu_{k'}$ (incentive compatibility), by transitivity he prefers a_{-J} to $\alpha_{-j'}$ at belief $\mu_{k'}$. He also prefers a_{-J} to $\alpha_{-j'}$ at belief 0. Because preferences are linear in beliefs, this implies that he prefers a_{-J} to $\alpha_{-j'}$ at belief $\mu_{-j'}$. Therefore,

$$\begin{aligned} \mathbb{E}_{P'}\left[\max\left\{\mathbb{E}_{\alpha_{-j'}}[u_S(a,\mu)],\mathbb{E}_{\alpha_{k'}}[u_S(a,\mu)]\right\}\right] &< \mathbb{E}_{P'}\left[u_S(a_{-J},\mu)\right] \\ &= u_S(a_{-J},\mu_0) \\ &\leq \mathbb{E}_{P}\left[\max\left\{u_S(a_{-J},\mu),u_S(\overline{a}_{k+1},\mu)\right\}\right], \end{aligned}$$

where the first inequality follows from the fact that a_{-J} is strictly better than $\alpha_{-j'}$ and $\alpha_{k'}$ at belief $\mu_{-j'}$ and belief $\mu_{k'}$, respectively. The last inequality follows from the fact that the information structure P is incentive compatible for the pair of action a_{-J} and $\overline{a}_{k+1}$. □

Proposition 5 implies that as long as $u_S(a_{-J},\mu) \leq \overline{v}(\mu)$ for some $\mu > \mu_0$, the support of the optimal random posterior contains 0. If it also contains 1, then the optimal experiment reveals perfect information. If it does not contain 1, the optimal experiment will generate a conclusive signal of the common-interest state. In other words, the underlying Blackwell experiment corresponding to this optimal random posterior will produce a signal that reveals the common-interest state 0 with probability strictly less than 1 when the true state is 0, and never produces a signal that would suggest the state is 0 when the true state is 1. This means that the ex ante probability that the receiver takes action a_{-J} under the optimal information structure cannot exceed $1-\mu_0$.

This result is driven by the tension between acquiring more information and alleviating the conflict of interests. The sender may want to induce the receiver to take the common-interest action a_{-J} with probability larger than $1-\mu_0$. However, this may not be incentive compatible if a_{-J} blocks all other actions. And when there is an incentive compatible information structure that can induce the receiver to take action a_{-J} at some belief $\mu' > 0$ (so that the ex ante probability of her taking action a_{-J} is larger than $1-\mu_0$), incentive compatibility implies that the value of information is positive. With general state-dependent preferences, the sender is better off inducing the receiver to take action a_{-J} at belief 0 than inducing her to take the action a_{-J} at belief μ'.

5 Informativeness Compared to Bayesian Persuasion

In general, the optimal experiment when the sender has no commitment power can be more or less informative than (or not Blackwell-comparable to) the optimal experiment chosen when the sender can commit to truthfully revealing the outcome of the experiment. For example, when sender and receiver have opposite marginal incentives, Proposition 2 shows that the optimal experiment in our setup is a totally uninformative experiment, while the optimal experiment with full commitment is typically nondegenerate, as the concave envelope of $\overline{v}(\cdot)$ does not coincide $\overline{v}(\cdot)$ itself.

For an example in which the optimal experiment in our setup is more informative than that in a model with full commitment, consider the case where there is a best action a_n that the sender prefers the most in both states. Let a_n be the receiver's best response when the belief is in the interval $[\underline{I}_n, \overline{I}_n]$. Let the prior belief μ_0 be lower than $\underline{I}_n$. The lesson we learn from Kamenica and Gentzkow (2011) is that if the optimal experi-

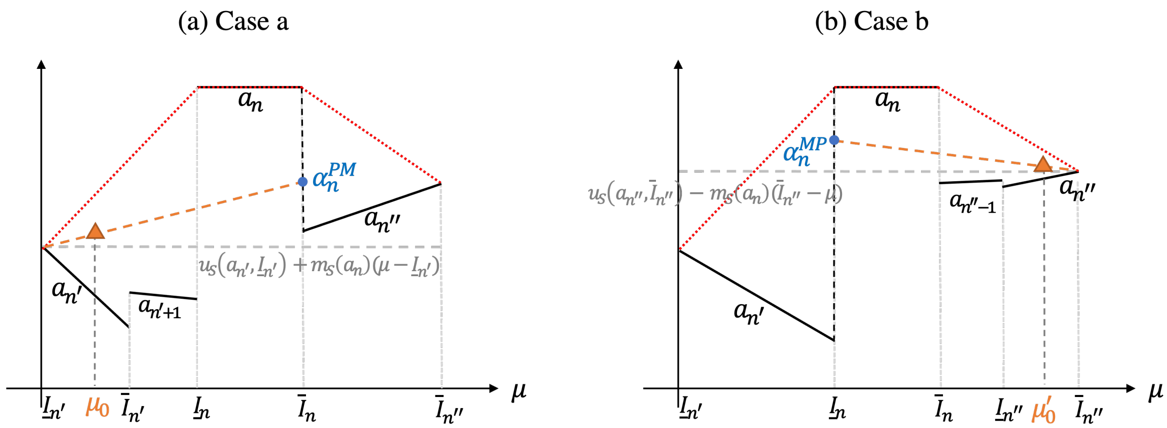

Figure 7: Optimal experiment with and without commitment.

ment with full commitment induces a_n and some other action, it maximizes the ex ante probability that a_n will be taken by inducing the smallest posterior belief $\underline{I}_n$ that is just enough to induce the receiver to choose a_n. When the sender lacks commitment power in communication, inducing the pure action a_n is not incentive compatible. However, it may be incentive compatible to induce a mixture action between a_n and a_{n+1} at belief $\overline{I}_n$. Because $\overline{I}_n$ is farther from μ_0 than $\underline{I}_n$ is from μ_0, the resulting experiment is more informative than the optimal experiment under full commitment. The next proposition specifies the precise conditions for an analogous argument to be valid.

Proposition 6. *Assume that sender and receiver have aligned marginal incentives and $|A| \geq 3$. If an action a_n ($n \neq -J-1, K+1$) is (strictly) best, then the optimal information structure in our model is (strictly) more informative than the optimal experiment under full commitment for some prior belief.*

The proof of this proposition is in the Appendix. The goal is to find a prior belief under which the optimal information design is more informative than the optimal experiment under full commitment. Consider Figure 7, a_n is the best action for the sender and the dotted red envelope is the concave envelope of the sender's value function. The optimal experiment under full commitment at prior μ_0 has support $\{\underline{I}_{n'}, \underline{I}_n\}$. The the optimal experiment under full commitment at belief μ_0' has support $\{\overline{I}_n, \overline{I}_{n''}\}$.

In the left panel (case a), the sender with belief $\underline{I}_{n'}$ strictly prefers a_n over $a_{n'}$ over $a_{n''}$. It implies that there exists a randomization α_n^{PM} between a_n and $a_{n''}$ such that the experiment with support $\{\underline{I}_{n'}, \overline{I}_n\}$ is IC-PM. Recall that with aligned marginal incentives,

$m_S(\alpha_n^{PM}) > m_S(a_n)$. Therefore the expected payoff from such IC-PM experiment (the orange triangle) is strictly higher than $u_S(a_{n'}, \underline{I}_{n'}) + m_S(a_n)(\mu_0 - \underline{I}_{n'})$ (lying on the gray dashed line). Moreover, with aligned marginal incentives, any incentive compatible experiment that inducing $a_{n'}$ and some action smaller than a_n can only lead to an expected payoff strictly below $u_S(a_{n'}, \underline{I}_{n'}) + m_S(a_n)(\mu_0 - \underline{I}_{n'})$. For example, in Figure 7(a), the experiment with support $\{\underline{I}_{n'}, \underline{I}_n\}$ is IC-PP for $a_{n'}$ and $a_{n'+1}$. However, sender's expected payoff from it is bounded by the gray dashed line because the slope of sender's expected payoff is smaller than the marginal incentives of $a_{n'+1}$ (implied by Lemma 2) which is smaller than $m_S(a_n)$. Thus, in this case, under the prior belief μ_0, the optimal experiment in our model is more informative than that under full commitment.

It is possible that the sender with belief $\underline{I}_{n'}$ strictly prefers all actions higher than a_n over $a_{n'}$, so that we cannot find an incentive compatible experiment that is more informative than $\{\underline{I}_{n'}, \underline{I}_n\}$. This happens in the right panel (case b). However, given the assumption of aligned marginal incentives, there must exist in this case two actions (weakly) smaller than a_n such that the sender with belief $\overline{I}_{n''}$ prefers one over $a_{n''}$ over the other one. In Figure 7(b), type-$\overline{I}_{n''}$ sender prefers a_n over $a_{n''}$ over $a_{n'}$. With a similar reasoning as in case (a), under the prior belief μ_0', the optimal experiment in our model is more informative than that under full commitment.

6 Continuous Action Space

In this section, we maintain the assumption that the state space is binary but allow the action space A to contain a continuum of possible actions. We assume in this section that the receiver has a unique best response $A_R(\mu)$ at every posterior belief μ, so that we only need to consider pure-strategy truth-telling equilibrium. Moreover, we assume A is compact and $u_i(a, \theta)$ is continuous. Note that a binary random posterior is still sufficient as the state space is binary.

One difference between continuous and finite action space is that the slope of sender's value function no longer represents his marginal incentives from an action. As an example, suppose the receiver's utility is $u_R(a, \theta) = -(a - \theta)^2$ for $\theta \in \{0, 1\}$ and $a \in [0, 1]$. Let the sender's utility be $u_S(a, \theta) = -(a - \theta - b_\theta)^2 + h(a)$, for some constants b_0 and

b_1 and for some function $h(\cdot)$. Then we have $A_R(\mu) = \mu$, and

$$\overline{v}(\mu) = -\mu(\mu - 1 - b_1)^2 - (1-\mu)(\mu - b_0)^2 + h(\mu).$$

The sender's value function $\overline{v}(\cdot)$ in this example depends on $h(\cdot)$, b_0 and b_1. More importantly, just knowing $\overline{v}(\cdot)$ without knowing b_0 and b_1 is not sufficient to determine the optimal information structure in our model, because we cannot derive the seller's marginal incentives from $\overline{v}(\cdot)$ alone. In particular, we have

$$m_S(A_R(\mu)) = (\mu - b_0)^2 - (\mu - 1 - b_1)^2 \neq \overline{v}'(\mu).$$

A full specification of our model requires information about both $\overline{v}(\cdot)$ and $m_S(A_R(\cdot))$. Therefore, we introduce a different approach to find the optimal information design under continuous action space.

For given $\mu' < \mu_0$, define a function $F_{\mu'} : [\mu_0, 1] \to \mathbb{R}$ by:

$$F_{\mu'}(\mu'') := \overline{v}(\mu') + m_S(A_R(\mu''))(\mu'' - \mu').$$

If $\overline{v}(\mu'') > F_{\mu'}(\mu'')$, then the sender would strictly prefer action $A_R(\mu'')$ to action $A_R(\mu')$ when his private belief is μ'. The following result is immediate by verifying incentive compatibility constraints (1).

Lemma 3. *A random posterior with support $\{\mu', \mu''\}$ is IC-PP if and only if*

$$u_S(A_R(\mu'), \mu'') \leq \bar{v}(\mu'') \leq F_{\mu'}(\mu'').$$

Denote $D_{\mu'} := \{\mu'' \in [\mu_0, 1] : u_S(A_R(\mu'), \mu'') \leq \bar{v}(\mu'') \leq F_{\mu'}(\mu'')\}$ as the set of μ'' such that the random posterior $\{\mu', \mu''\}$ is IC-PP. Define $c_{\mu'} : [\mu', 1] \to \mathbb{R}$ as the smallest concave function on $[\mu', 1]$ such that $c_{\mu'}(\mu'') \geq \overline{v}(\mu'')$ at all $\mu'' \in D_{\mu'} \cup \{\mu'\}$. If $D_{\mu'}$ is empty, define $c_{\mu'} = \overline{v}(\mu_0)$ on the domain $[\mu', 1]$.

Proposition 7. *The highest equilibrium payoff the sender can achieve is*

$$W^*(\mu_0) = \max_{\mu' \in [0, \mu_0]} c_{\mu'}(\mu_0).$$

Proof. Suppose a random posterior P has $\mu' \in [0, \mu_0)$ in its support. If $c_{\mu'}(\mu_0) > \overline{v}(\mu_0)$,

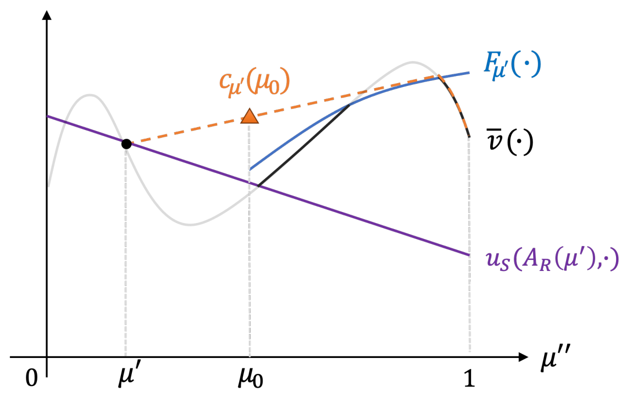

Figure 8: Continuous action space.

then there exists $\mu'' \in D_{\mu'} \setminus \{\mu_0\}$ such that the random posterior with support $\{\mu', \mu''\}$ and expectation μ_0 is IC-PP. Since $c_{\mu'}(\cdot)$ is the concave envelope, $c_{\mu'}(\mu_0)$ is the highest achievable payoff if $\mu' \in \operatorname{supp} P$. If $c_{\mu'}(\mu_0) = \overline{v}(\mu_0)$, then the random posterior that is degenerate at belief μ_0 can achieve this payoff. Next, suppose the random posterior has $\mu' = \mu_0$ in its support. Then, by our construction, $F_{\mu'}(\mu_0) = \overline{v}(\mu_0)$ and the domain of $c_{\mu'}(\cdot)$ is $[\mu_0, 1]$. Therefore, $c_{\mu'}(\mu_0) = \overline{v}(\mu_0)$, and the random posterior that is degenerate at belief μ_0 can achieve this payoff. Since it is without loss of generality to focus on a binary random posterior, $\max_{\mu' \in [0,\mu_0]} c_{\mu'}(\mu_0)$ determines the highest equilibrium payoff for the sender. □

In Figure 8, the gray curve is an arbitrary continuous $\overline{v}(\cdot)$. The black curve represent the segments of $\overline{v}(\cdot)$ that is smaller than $F_{\mu'}(\cdot)$ and greater than $u_S(A_R(\mu'), \cdot)$. The orange dashed curve $c_{\mu'}(\cdot)$ is the concave envelope of the black curve and $\overline{v}(\mu')$ (the black dot). Its value at μ_0 is the highest equilibrium payoff the sender can achieve if μ' belongs to the support of the optimal random posterior.

Proposition 7 immediately implies that the quasiconcave envelope determines the highest equilibrium payoff for the sender under transparent motives—a result first derived by Lipnowski and Ravid (2020).

Corollary 1. *If the sender's preferences are state-independent, the highest equilibrium payoff he can achieve is the quasiconcave envelope of* $\overline{v}(\cdot)$.

If the sender's preferences are state-independent, the marginal incentive $m_S(A_R(\mu''))$ equals 0 regardless of μ''. Thus, $F_{\mu'}(\cdot)$ is flat and equals to $\overline{v}(\mu')$. Moreover, $u_S(A_R(\mu'), \cdot)$

is also flat and equals to $\overline{v}(\mu')$ because $m_S(A_R(\mu')) = 0$. Therefore, $F_{\mu'}(\cdot)$ and $u_S(A_R(\mu'), \cdot)$ coincide on $[\mu_0, 1]$. Thus, if $D_{\mu'}$ is not empty, $c_{\mu'}(\mu_0) = \overline{v}(\mu')$ is the unique equilibrium payoff for the sender if μ' is in the support of the random posterior. The solution to $\max_{\mu' \in [0,\mu_0]} c_{\mu'}(\mu_0)$ is then reduced to be the quasiconcave envelope of $\overline{v}(\cdot)$ evaluated at μ_0.

7 Discussion

The model in this paper has a close relation with a particular scheme of mediated communication, in which a mediator maximizes the ex-ante welfare of an informed sender (Salamanca, 2021). Specifically, a perfectly informed sender sends a message about his private information to a mediator. The mediator then communicates a message to the receiver according to a noisy reporting rule that the mediator commits to at the beginning of the game. After receiving the message from the mediator, the receiver takes an action. If we consider the mediator's reporting rule as a mapping from the sender's private information to a distribution of action recommendations, this rule can be interpreted as an information structure. The incentive constraints for this mediated communication game are imposed at the *ex ante* stage, which require every type of sender who perfectly knows the state to report his private information truthfully before observing the message sent by the mediator.

In contrast, the sender in our model is uninformed when he commits to an information structure, and then reports his private information to the receiver after observing the outcome of the experiment. Therefore, our model requires *interim-stage* incentive constraints, such that the sender with an interim belief derived from the observed outcome prefers to report his private information truthfully. In spite of this difference, if our sender and the mediator in Salamanca (2021) commit to the same information structure in equilibrium, then both models will yield the same equilibrium outcomes.

Interestingly, under binary state space, the highest equilibrium payoff $W^*(\mu_0)$ that the sender can achieve in our model is always weakly lower than the maximum ex-ante welfare of the sender (i.e., evaluated at μ_0 before the sender becomes perfectly informed) in Salamanca (2021) for any μ_0.

To see this, suppose the optimal random posterior in our sender-receiver game has support $\{\mu', \mu''\}$ and the receiver optimally chooses $a' \in A_R(\mu')$ and $a'' \in A_R(\mu'')$ at the respective beliefs (the same argument will go through if the receiver takes mixed strategy). Without loss of generality, let $\mu' < \mu''$. Then incentive compatibility constraints (1) in our model implies that the following also holds:

$$u_S(a', 0) \geq u_S(a'', 0), \qquad u_S(a'', 1) \geq u_S(a', 1).$$

This means that it is incentive compatible for an informed sender to truthfully report his private information (belief 0 or 1) to the mediator, whenever the mediator commits to a reporting rule that recommends a' more often if the sender reports 0 and recommends a'' more often if the sender reports 1. Therefore, the sender's incentive constraints in the mediated communication game are satisfied if the mediator commits to the same information structure as the underlying experiment that induces our optimal random posterior. In other words, interim incentive compatibility in our model is more stringent than the incentive compatibility restrictions required by the mediator model, and therefore our model delivers a (weakly) lower expected payoff for the sender than that achievable in Salamanca (2021).

The connection between our paper and Lin and Liu (2022) is more subtle. Lin and Liu (2022) focus on a scenario where the sender's deviation in reporting strategy is not detectable if the marginal distribution of messages is unchanged. In other words, their sender maximizes the ex ante expected payoff subject to the constraint that the marginal distribution of messages remains the same. Therefore, sender's credibility requires either (1) the sender's gain from swapping messages in one state is smaller than the loss from swapping messages in another state, or (2) the sender cannot swap messages without affecting the marginal distributions. That is, the credibility constraint arrives at the ex ante stage. However, in our model, sender's incentive compatibility arrives at the interim stage.

Figure 9 provides two examples to show these two setups are not nested. In the left panel, the sender prefers a' to a in both states, therefore there is no incentive compatible experiment in our paper, i.e., the highest payoff for the sender under μ_0 is the default payoff (plotted as the orange circle). However, in Lin and Liu (2022), the blue square can be supported as equilibrium payoff, because the sender's gain from recommending a' instead of a at state 0 is smaller than his loss from recommending a instead of a' at state 1. In the right panel, on the contrary, our model can support a higher equilibrium

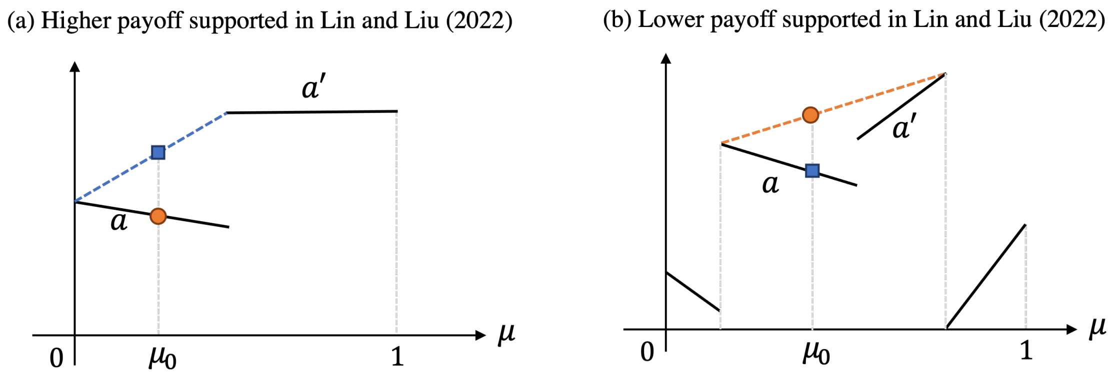

Figure 9: Comparing with Lin and Liu (2022).

payoff for the sender than Lin and Liu (2022). The random posterior that induce the orange circle is IC-PP in our model; however, such random posterior is not credible in Lin and Liu (2022) as its support contains two non-degenerate beliefs. In order to produce the payoff indicated by the orange circle, the sender recommends both actions a and a' with positive probabilities in both states. However, the sender strictly prefers a' over a at state 1 and strictly prefers a over a' at state 0, which implies the gains from swapping messages—replace the recommendation of a' by a in state 0 and replace the recommendation of a by a' in state 1—are positive in both states. Moreover, the sender can swap messages in such a way without affecting the marginal distributions. Therefore, the orange circle is not an equilibrium payoff in Lin and Liu (2022).

Appendix

Proof of Lemma 1. Given P, σ_S, and a message $m \in M$, the receiver forms a posterior belief $\hat{\mu}^m$, where

$$\hat{\mu}^m(\theta) = \sum_{\mu \in \operatorname{supp} P} \frac{P(\mu)\sigma_S(m|\mu)}{\sum_{\mu \in \operatorname{supp} P} P(\mu)\sigma_S(m|\mu)} \mu(\theta),$$

for $\theta \in \Theta$. We use $\hat{P} \in \Delta(\Delta\Theta)$ to denote the distribution of the receiver's posterior beliefs, with $\hat{P}(\hat{\mu}^m) = \sum_{\mu \in \operatorname{supp} P} P(\mu)\sigma_S(m|\mu)$. Then player i's expected utility can be simplified to:

$$U_i(\sigma_S, \sigma_R, P) = \sum_{m \in M,\, \theta \in \Theta,\, a \in A} \hat{P}(\hat{\mu}^m)\hat{\mu}^m(\theta)\sigma_R(a|m)u_i(a, \theta).$$

Thus each player's expected utility only depends on the joint distribution of receiver's posterior belief and the action. If we let the sender directly commits to the random posterior $\hat{P}$, and construct a (truth-telling) reporting strategy $\hat{\sigma}_S$ such that for all $\mu \in \operatorname{supp} \hat{P}$, $\hat{\sigma}_S(m|\hat{\mu}^m) = 1$, then player i's expected utility further simplifies to:

$$U_i(\sigma_S, \sigma_R, P) = U_i(\hat{\sigma}_S, \sigma_R, \hat{P}).$$

Moreover, (σ_S, σ_R, P) being an equilibrium strategy profile implies $(\hat{\sigma}_S, \sigma_R, \hat{P})$ is an equilibrium strategy profile. Since reporting $m \in M$ is a best response to σ_R for every sender type of $\{\mu \in \operatorname{supp} P : \sigma_S(m|\mu) > 0\}$, reporting m is also a best response for sender type $\hat{\mu}^m$, as $\hat{\mu}^m$ is a convex combination of $\{\mu \in \operatorname{supp} P : \sigma_S(m|\mu) > 0\}$.

To prove the second part, suppose a random posterior P with $|\operatorname{supp} P| > |\Theta|$ that can lead to a truth-telling equilibrium is optimal. By Carathéodory's Theorem and Krein-Milman Theorem, $\mu_0 = \mathbb{E}_P[\mu]$ can be written as a convex combination of $|\Theta|$ elements of $\operatorname{supp} P$, denoted as $P' \in \Delta(\Delta\Theta)$ with $|\operatorname{supp} P'| = |\Theta|$ and $\operatorname{supp} P' \subset \operatorname{supp} P$. Let the receiver preserve σ_R, then P' can lead to a truth-telling equilibrium. Let $c : \operatorname{co}(\operatorname{supp} P) \to \mathbb{R}$ be the smallest concave function such that $c(\mu) \geq \sum_{\theta, a} \mu(\theta)\sigma_R(a|\mu)u_S(a, \theta)$ at all $\mu \in \operatorname{supp} P$. The random posterior P' can perform equally well as P because c must be affine on $\operatorname{co}(\operatorname{supp} P)$. □

Proof of Theorem 1.

Claim 1. *If the optimal random posterior has support* $\{\mu_{-j}, \mu_k\}$*, and the receiver uses mixed strategies* $\alpha_{-j} \in \Delta A_R(\mu_{-j})$ *and* $\alpha_k \in \Delta A_R(\mu_k)$ *(with full support) in the sender-preferred equilibrium, then at least one of the following is true:*

(i) $\alpha_{-j} = \alpha_{-j}^{MM}$ *and* $\alpha_k = \alpha_k^{MM}$*;*

(ii) $W^*(\mu_0) \in \{W_{-j,k}^{PM}, W_{-j,k}^{MP}, W_{-j,k}^{PP}\}$.

Proof of Claim 1. Since the receiver uses mixed strategies at each belief, $\mu_{-j} \neq 0$ and $\mu_k \neq 1$. Let $A_R(\mu_{-j}) = \{a_{-j}, a_{-j+1}\}$ and $A_R(\mu_k) = \{a_{k-1}, a_k\}$. Since α_{-j} and α_k are incentive compatible for the sender, from (1),

$$\mathbb{E}_{\alpha_{-j}}[u_S(a, \mu_{-j})] \geq \mathbb{E}_{\alpha_k}[u_S(a, \mu_{-j})], \qquad \mathbb{E}_{\alpha_k}[u_S(a, \mu_k)] \geq \mathbb{E}_{\alpha_{-j}}[u_S(a, \mu_k)]. \tag{3}$$

Suppose $u_S(a_{-j}, \mu_{-j}) \neq u_S(a_{-j+1}, \mu_{-j})$ and $u_S(a_{k-1}, \mu_k) \neq u_S(a_k, \mu_k)$. Suppose further that both inequalities hold strictly and α_{-j} and α_k have full support. Then the sender can achieve a strictly higher expected payoff if the receiver deviates from α_{-j} by assigning a slightly larger probability on the sender-preferred action $\overline{a}_{-j} \in A_R(\mu_{-j})$. As long as the increase in probability of choosing $\overline{a}_{-j}$ is small enough, such deviation would raise the payoff from truth-telling at belief μ_{-j} without violating the truth-telling constraint at belief μ_k. Suppose the first inequality in (3) holds as an equality and the second inequality in (3) holds strictly. Then the sender can achieve a strictly higher expected payoff under the same argument. If the first inequality in (3) holds strictly and the second inequality in (3) holds as an equality, a symmetric argument will apply. Therefore, the optimality of α_{-j} and α_k implies that both inequalities in (3) hold as an equality.

If $m_S(a_{-j})$, $m_S(a_{-j+1})$, $m_S(a_{k-1})$ and $m_S(a_k)$ are not all equal, then there is a unique $(\alpha_{-j}^{MM}, \alpha_k^{MM}) \in \Delta A_R(\mu_{-j}) \times \Delta A_R(\mu_k)$ such that the constraints (3) hold as equalities. Next, if $m_S(a_{-j}) = m_S(a_{-j+1}) = m_S(a_{k-1}) = m_S(a_k)$, then there exist infinitely many solutions. The optimality of α_{-j} and α_k will then imply that the receiver takes either $\overline{a}_{-j}$ at belief μ_{-j} and α_k^{MM} at belief μ_k, or $\overline{a}_k$ at μ_k and α_{-j}^{MM} at belief μ_{-j}. This contradicts the premise that both α_{-j} and α_k have full support. Moreover, in this case $W^*(\mu_0) \in \{W_{-j,k}^{PM}, W_{-j,k}^{MP}\}$.

If at least one of $u_S(a_{-j}, \mu_{-j}) \neq u_S(a_{-j+1}, \mu_{-j})$ and $u_S(a_{k-1}, \mu_k) \neq u_S(a_k, \mu_k)$ does not hold, we can slightly alter the above argument to show $W^*(\mu_0) \in \{W_{-j,k}^{PM}, W_{-j,k}^{MP}, W_{-j,k}^{PP}\}$. □

Claim 2. *If the optimal random posterior has support $\{\mu_{-j}, \mu_k\}$, and the receivers takes pure action $a' \in A_R(\mu_{-j})$ and mixed action $\alpha_k \in \Delta A_R(\mu_k)$ (with full support) in the sender-preferred equilibrium, then at least one of the following is true:*

(i) $a' = \overline{a}_{-j}$ *and* $\alpha_k = \alpha_k^{PM}$*;*
(ii) $W^*(\mu_0) \in \{W_{-j,k}^{MM}, W_{-j,k}^{MP}, W_{-j,k}^{PP}\}$.

Proof of Claim 2. Since the receiver uses mixed strategies at belief μ_k, $\mu_k \neq 1$ and $A_R(\mu_k) = \{a_{k-1}, a_k\}$. Since a' and α_k are incentive compatible for the sender, from (1),

$$u_S(a', \mu_{-j}) \geq \mathbb{E}_{\alpha_k}[u_S(a, \mu_{-j})], \qquad \mathbb{E}_{\alpha_k}[u_S(a, \mu_k)] \geq u_S(a', \mu_k). \tag{4}$$

Suppose that $u_S(a_{k-1}, \mu_k) \neq u_S(a_k, \mu_k)$. We first show that the first inequality in (4) must hold as an equation. Suppose to the contrary that this inequality holds strictly. Then the sender can achieve a strictly higher payoff if the receiver deviates from α_k by assigning a slightly larger probability on the sender-preferred action $\overline{a}_k \in A_R(\mu_k)$. As long as the increase in probability of choosing $\overline{a}_k$ is small enough, such deviation would raise the sender's payoff from truth-telling at belief μ_k without violating (4) at belief μ_{-j}, leading to a contradiction.

Now, suppose the second inequality in (4) also holds as an equation. Given the result established above, the sender is indifferent between a' and α_k both at belief μ_{-j} and at belief μ_k. This case then reduces to the double-randomization case. Therefore, $W^*(\mu_0) = W_{-j,k}^{MM}$, and part (ii) of this claim is satisfied.

Next, suppose the second inequality in (4) holds strictly. There are two possibilities: (1) the sender obtains different payoffs from a_{-j} and a_{-j+1} at belief μ_{-j}, (2) the sender obtains the same payoff from a_{-j} and a_{-j+1} at belief μ_{-j}.

Case (1). Suppose a' is not the sender-preferred action $\overline{a}_{-j}$ at belief μ_{-j}. Then the sender can achieve a strictly higher payoff by inducing the receiver to deviate from a' to a mixed strategy α_{-j} that assigns a small positive probability on $\overline{a}_{-j}$, without violating the incentive constraints. This shows that a' must be equal to $\overline{a}_{-j}$. Because $a' = \overline{a}_{-j}$, and the first condition of (4) as an equation implies that $\alpha_k = \alpha_k^{PM}$, and part (i) of this claim is satisfied.

Case (2). When the sender obtains the same payoff from a_{-j} and a_{-j+1} at belief μ_{-j}, the convention we adopt is $\overline{a}_{-j} = a_{-j+1}$. Suppose $a' = a_{-j} \neq \overline{a}_{-j}$. Then the optimality of $\{\mu_{-j}, \mu_k\}$ implies that the sender is indifferent between α_k^{PM} and a_{-j} at belief μ_k. Otherwise, the random posterior with support $\{\mu_{-j-1}, \mu_k\}$ can performs strictly better. Therefore, the second inequality in (4) cannot hold strictly under the convention we adopt. This contradiction implies that we must have $a' = a_{-j+1} = \overline{a}_{-j}$. The first condition of (4) as an equation then implies that $\alpha_k = \alpha_k^{PM}$, and part (i) of this claim is satisfied.

Suppose $u_S(a_{k-1}, \mu_k) = u_S(a_k, \mu_k)$. If the first inequality in (4) holds with equality, we can use the same argument as above. Otherwise, if the first inequality in (4) holds strictly, then we can slightly alter the above argument to show $W^*(\mu_0) \in \{W_{-j,k}^{MM}, W_{-j,k}^{MP}, W_{-j,k}^{PP}\}$. □

Claim 3. *If the optimal random posterior has support $\{\mu_{-j}, \mu_k\}$, and the receivers uses only pure strategies $a' \in A_R(\mu_{-j})$ and $a'' \in A_R(\mu_k)$ in the sender-preferred equilibrium, then either one of the following is true:*

(i) $a' = \overline{a}_{-j}$ *and* $a'' = \overline{a}_k$;
(ii) $W^*(\mu_0) \in \{W_{-j,k}^{PM}, W_{-j,k}^{MP}, W_{-j,k}^{MM}\}$.

Proof of Claim 3. If $\mu_{-j} = 0$ and $\mu_k = 1$, then $A_R(\mu_{-j}) = \overline{a}_{-j}$, $A_R(\mu_k) = \overline{a}_k$, and part (i) is satisfied.

If $\mu_{-j} \neq 0$ and $\mu_k \neq 1$, then $A_R(\mu_{-j}) = \{a_{-j}, a_{-j+1}\}$, $A_R(\mu_k) = \{a_{k-1}, a_k\}$, and there are four possibilities. (1) The sender obtains different payoffs from a_{-j} and a_{-j+1} at belief μ_{-j}, and different payoffs from a_{k-1} and a_k at belief μ_k. (2) The sender obtains same payoff from a_{-j} and a_{-j+1} at belief μ_{-j}, but different payoffs from a_{k-1} and a_k at belief μ_k. (3) The sender obtains different payoffs from a_{-j} and a_{-j+1} at belief μ_{-j}, and same payoff from a_{k-1} and a_k at belief μ_k. (4) The sender obtains same payoff from a_{-j} and a_{-j+1} at belief μ_{-j}, and same payoff from a_{k-1} and a_k at belief μ_k.

Case (1). By incentive compatibility,

$$u_S(a', \mu_{-j}) \geq u_S(a'', \mu_{-j}), \qquad u_S(a'', \mu_k) \geq u_S(a', \mu_k). \tag{5}$$

Suppose both inequalities hold strictly. The optimality of (a', a'') implies $a' = \overline{a}_{-j}$ and $a'' = \overline{a}_k$, with a reasoning similar to that in the proof of Claim 1. Thus, part (i) is satisfied. Suppose the first inequality hold as an equality and the second inequality holds strictly. Then the optimality of (a', a'') implies $a' = \overline{a}_{-j}$, with a reasoning similar to that in the proof of Claim 2. Moreover, when $a' = \overline{a}_j$ and the first inequality hold as an equality, we have $a'' = \alpha_k^{PM}$. Therefore, $W^*(\mu_0) = W_{-j,k}^{PM}$, and part (ii) is satisfied. Similarly, if the first inequality hold strictly and the second inequality holds as an equality, then $W^*(\mu_0) = W_{-j,k}^{MP}$ and part (ii) is also satisfied. Finally, suppose that both inequalities hold as an equality. Then if $m_S(a_{-j})$, $m_S(a_{-j+1})$, $m_S(a_{k-1})$ and $m_S(a_k)$ are not all equal, $a' = \alpha_{-j}^{MM}$ and $a'' = \alpha_k^{MM}$. Therefore, $W^*(\mu_0) = W_{-j,k}^{MM}$. If $m_S(a_{-j})$, $m_S(a_{-j+1})$, $m_S(a_{k-1})$ and $m_S(a_k)$ are equal, then $W^*(\mu_0) \in \{W_{-j,k}^{PM}, W_{-j,k}^{MP}\}$. Part (ii) is again satisfied.

Case (2a). If $a' = a_{-j} \neq \overline{a}_{-j}$, then the optimality of $\{\mu_{-j}, \mu_k\}$ implies that the sender is indifferent between a'' and a_{-j} at belief μ_k; otherwise the random posterior with support $\{\mu_{-j-1}, \mu_k\}$ is strictly better. Moreover, if the sender is indifferent between a_{-j} and a'' at belief μ_{-j}, then $W^*(\mu_0) = W_{-j,k}^{MM}$. On the other hand, if the sender strictly prefers a_{-j} over a'' at belief μ_{-j}, then the optimality of a'' implies $a'' = \overline{a}_k$. That is, $W^*(\mu_0) = W_{-j,k}^{MP}$. In both sub-cases, part (ii) is satisfied. Case (2b). If $a' \neq a_{-j}$, then $a' = \overline{a}_{-j}$. Then if the first inequality in (5) hold strictly, the optimality of $(\overline{a}_{-j}, a'')$ implies that $a'' = \overline{a}_k$, and part (i) is satisfied. On the other hand, if the first inequality in (5) hold as equality, then $W^*(\mu_0) = W_{-j,k}^{PM}$, and part (ii) is satisfied.

Case (3) is symmetric to case (2), and if we apply all arguments above, case (4) implies $W^*(\mu_0) \in \{W_{-j,k}^{PP}, W_{-j,k}^{PM}, W_{-j,k}^{MP}, W_{-j,k}^{MM}\}$.

Finally, if either $\mu_{-j} = 0$ or $\mu_k = 1$, then with a similar reasoning we can conclude $W^*(\mu_0) \in \{W_{-j,k}^{PP}, W_{-j,k}^{PM}, W_{-j,k}^{MP}, W_{-j,k}^{MM}\}$. □

Claims 1–3 (together with an analogous claim for the case of MP) imply that it is sufficient to only focus on $W_{-j,k}^{PP}$, $W_{-j,k}^{PM}$, $W_{-j,k}^{MP}$, and $W_{-j,k}^{MM}$.

Lastly, in step 2 of the algorithm, we only consider $(-j,k) \in S_1$. To see the sufficiency of it, suppose a pair of $(-j,k)$ outside S_1 is IC-PP, or IC-PM, or IC-MP, or IC-MM. Then,

$$\max\{W_{-j,k}^{PP}, W_{-j,k}^{PM}, W_{-j,k}^{MP}, W_{-j,k}^{MM}\} \leq W_{-j,k}(\mu_0; \overline{a}_{-j}, \overline{a}_k) \leq W^1.$$

The first inequality follows from the fact that the sender's value correspondence $v(\mu_{-j})$ is lower than $u_S(\overline{a}_{-j}, \mu_{-j})$, and $v(\mu_k)$ is lower than $u_S(\overline{a}_k, \mu_k)$. The second inequality comes from the construction of S_1. So even if the random posterior with support $\{\mu_{-j}, \mu_k\}$ outside S_1 is incentive compatible, it will never improve on the outcome from step 1. Moreover, in step 3 of the algorithm, we only consider $(-j,k) \in S_2$. To see the sufficiency of it, suppose a pair of $(-j,k)$ outside S_2 is IC-MM. Then,

$$W^{MM}_{-j,k} \leq \max\{W_{-j,k}(\mu_0;\overline{a}_{-j},\alpha^{PM}_k), W_{-j,k}(\mu_0;\alpha^{PM}_{-j},\overline{a}_k)\} \leq \max\{W^{(a)}, W^{(b)}\} = W^2.$$

The second inequality is from our construction of S_2. To see the first inequality, notice that under IC-MM, the sender is indifferent between α^{MM}_{-j} and $\alpha^{MM}_k := (\gamma_k, 1-\gamma_k)$ at belief μ_{-j},

$$\mathbb{E}_{\alpha_{-j}}[u_S(a,\mu_{-j})] = \gamma_k u_S(\overline{a}_k,\mu_{-j}) + (1-\gamma_k)u_S(\underline{a}_k,\mu_{-j}).$$

From our construction of $\alpha^{PM}_k := (\gamma'_k, 1-\gamma'_k)$,

$$u_S(\overline{a}_{-j},\mu_{-j}) = \gamma'_k u_S(\overline{a}_k,\mu_{-j}) + (1-\gamma'_k)u_S(\underline{a}_k,\mu_{-j}).$$

Therefore $\mathbb{E}_{\alpha_{-j}}[u_S(a,\mu_{-j})] \leq u_S(\overline{a}_{-j},\mu_{-j})$ implies $\gamma_k \leq \gamma'_k$, which further implies that $\mathbb{E}_{\alpha^{MM}_k}[u_S(a,\mu_k)] \leq \mathbb{E}_{\alpha^{PM}_k}[u_S(a,\mu_k)]$. We thus have $W^{MM}_{-j,k} \leq W_{-j,k}(\mu_0;\overline{a}_{-j},\alpha^{PM}_k)$. For a similar reason, $W^{MM}_{-j,k} \leq W_{-j,k}(\mu_0;\alpha^{PM}_{-j},\overline{a}_k)$. Notice that this argument does not require $\{\mu_{-j},\mu_k\}$ to be IC-PM or IC-MP. □

Proof of Proposition 3, part (b). Let a_n be the least-preferred action for the sender in state 0 and a_l be his least-preferred action in state 1. (If there are multiple least-preferred actions in one state, just pick any one of them.) We have $a_n \neq a_l$, otherwise a_n would be a worst action. Moreover, $u_S(a_l,0) > u_S(a_n,0)$ and $u_S(a_n,1) > u_S(a_l,1)$. This implies $m_S(a_l) < m_S(a_n)$. By aligned marginal incentives, $m_R(a_l) < m_R(a_n)$, and therefore the interval of beliefs for which a_l is receiver's best response, denoted I_l, is to the left of the interval I_n for a_n. Following the same convention adopted in the text, we let $\underline{I}_l$ represent the lowest belief in I_l and let $\overline{I}_n$ represent the highest belief in I_n.

There are three mutually exclusive cases. (1) a_l and a_n are strictly IC-PP for $\{\underline{I}_l, \overline{I}_n\}$; (2) a_l blocks a_n (but a_n does not block a_l); and (3) a_n blocks a_l (but a_l does not block a_n). In case (1), information design is valuable, for example when μ_0 is in the interior of I_l. Cases (2) and (3) are symmetric; thus we consider case (2) only.

Denote a_{n+1} as the next action higher than a_n; i.e., the receiver is indifferent between a_{n+1} and a_n at belief $\overline{I}_n$. There are several possibilities:

(2a) Suppose a_l is (weakly) worse than a_{n+1} at belief $\underline{I}_l$. Note that under case (2) a_l is better than a_n at both belief $\underline{I}_l$ and $\overline{I}_n$. Therefore, there is a mixed action $\alpha_n \in \Delta\{a_n, a_{n+1}\}$ such that the sender with belief $\underline{I}_l$ is indifferent between a_l and α_n.

Moreover, by aligned marginal incentives, $m_S(\alpha_n) > m_S(a_l)$. Thus, the random posterior with support $\{\underline{I}_l, \overline{I}_n\}$ is IC-PM for action a_l and some mixed action α_n and information design is valuable, for example when μ_0 is in the interior of I_l.

(2b) Suppose a_l is (strictly) better than a_{n+1} at belief $\underline{I}_l$.

(i) If a_{n+1} is better than a_l at belief $\overline{I}_{n+1}$, then $\{\underline{I}_l, \overline{I}_{n+1}\}$ is IC-PP.

(ii) If a_{n+1} is worse than a_l at belief $\overline{I}_{n+1}$, then let $a_{n'}$ be the highest action that a_l blocks. Notice that $a_{n'} < a_K$, where $a_K = A_R(1)$, because $m_S(a_K) > m_S(a_n)$ and $u_S(a_K, 0) \geq u_S(a_n, 0)$ imply that $u_S(a_K, 1) > u_S(a_n, 1)$. Since a_l is the least-preferred action in state 1, we have $u_S(a_n, 1) \geq u_S(a_l, 1)$, and thereby $u_S(a_K, 1) > u_S(a_l, 1)$. Therefore a_l does not block a_K. Then with a similar argument as in (2a) and (2b-i), either of the following is true: $\{\underline{I}_l, \overline{I}_{n'}\}$ is IC-PM, or $\{\underline{I}_l, \overline{I}_{n'+1}\}$ is IC-PP. The existence of $a_{n'+1}$ comes from $a_{n'} < a_K$. □

Proof of Proposition 6. With aligned marginal incentives, the concavification result in Kamenica and Gentzkow (2011) implies that there exists an $n' < n$ such that with prior belief $\mu_0 \in (\underline{I}_{n'}, \overline{I}_{n'})$, the optimal experiment under full commitment has support $\{\underline{I}_{n'}, \underline{I}_n\}$. Similarly, there exists an $n'' > n$ such that with a different prior belief $\mu_0' \in (\underline{I}_{n''}, \overline{I}_{n''})$, the optimal experiment under full commitment has support $\{\overline{I}_n, \overline{I}_{n''}\}$. We want to show that, in our model, the optimal experiment is strictly more informative than that under full commitment either when the prior is μ_0 or when the prior is μ_0'. There are only two cases. (a) There exists a $k \geq n$ such that $u_S(a_k, \underline{I}_{n'}) \geq u_S(a_{n'}, \underline{I}_{n'}) \geq u_S(a_{k+1}, \underline{I}_{n'})$. (b) For every $k \geq n$, $u_S(a_k, \underline{I}_{n'}) > u_S(a_{n'}, \underline{I}_{n'})$. Notice that if there exists an experiment with support $\{\underline{I}_{n'}, \mu\}$ where $\mu \in (\overline{I}_{n'}, \underline{I}_n]$ that is incentive compatible, then the sender's expected payoff from such experiment is smaller than $u_S(a_{n'}, \underline{I}_{n'}) + m_S(a_n)(\mu_0 - \underline{I}_{n'})$ (this is implied by Lemma 2). Similarly, if there exists an experiment with support $\{\mu, \overline{I}_{n''}\}$ where $\mu \in (\overline{I}_n, \overline{I}_{n''})$ that is incentive compatible, then the sender's expected payoff under μ_0' from such experiment is smaller than $u_S(a_{n''}, \overline{I}_{n''}) - m_S(a_n)(\overline{I}_{n''} - \mu_0')$.

In case (a), the experiment with support $\{\underline{I}_{n'}, \overline{I}_k\}$ is IC-PM given aligned marginal incentives. Moreover, sender's expected payoff from such experiment is greater than $u_S(a_{n'}, \underline{I}_{n'}) + m_S(a_n)(\mu_0 - \underline{I}_{n'})$. Because the receiver randomizes between a_k and a_{k+1} at belief $\overline{I}_k$ and thereby the marginal incentive from such randomization α_k is greater than $m_S(a_n)$. Also, from the construction of IC-PM, sender's expected payoff equals $u_S(a_{n'}, \underline{I}_{n'}) + m_S(\alpha_k)(\mu_0 - \underline{I}_{n'})$. Therefore, under μ_0, there exists an experiment with support $\{\underline{I}_{n'}, \overline{I}_k\}$ that is better than any IC experiment with support $\{\underline{I}_{n'}, \mu\}$ where $\mu \in$

$(\overline{I}_{n'}, \underline{I}_n]$.

Moreover, there cannot exist an IC experiment $\{\mu, \mu'\}$ with $\mu < \underline{I}_{n'}$ and $\overline{I}_{n'} < \mu' < \underline{I}_n$ that yields the sender an expected payoff higher than $u_S(a_{n'}, \underline{I}_{n'}) + m_S(a_n)(\mu_0 - \underline{I}_{n'})$. To see this point, note that Lemma 2 implies that the slope of sender's expected payoff is smaller than $m_S(a')$ where $a' \in A_R(\mu')$, which in turn is smaller than $m_S(a_n)$ from aligned marginal incentives. This implies the sender's expected payoff from such experiment, if the prior belief is $\underline{I}_{n'}$, would be higher than $u_S(a_{n'}, \underline{I}_{n'})$, which contradicts the fact that $u_S(a_{n'}, \underline{I}_{n'})$ lies on the concave envelope of $\overline{v}(\cdot)$.

Since $\overline{I}_k > \underline{I}_n$, the optimal experiment in our model under μ_0, which has support $\{\underline{I}_{n'}, \overline{I}_k\}$, is strictly more informative than the (full commitment) experiment with support $\{\underline{I}_{n'}, \underline{I}_n\}$.

In case (b), since $u_S(a_{n''}, \underline{I}_{n'}) > u_S(a_{n'}, \underline{I}_{n'})$, we have $u_S(a_{n''}, \overline{I}_{n''}) > u_S(a_{n'}, \overline{I}_{n''})$ under the assumption of aligned marginal incentive. Then there must exist an a_k with $n' < k \leq n$ such that $u_S(a_k, \overline{I}_{n''}) \geq u_S(a_{n''}, \overline{I}_{n''}) \geq u_S(a_{k-1}, \overline{I}_{n''})$. Therefore, there exists an IC-MP experiment with support $\{\underline{I}_k, \overline{I}_{n''}\}$ such that the receiver randomizes between a_k and a_{k-1} at belief $\underline{I}_k$ and such randomization α_k has a marginal incentive $m_S(\alpha_k)$ smaller than $m_S(a_n)$. Under μ_0', this experiment generates the sender an expected payoff higher than $u_S(a_{n''}, \overline{I}_{n''}) - m_S(a_n)(\overline{I}_{n''} - \mu_0')$. Moreover, such experiment is better than any IC experiment with support $\{\mu, \overline{I}_{n''}\}$ where $\mu \in [\overline{I}_n, \overline{I}_{n''})$. Because $\underline{I}_k < \underline{I}_n$, the optimal experiment in our model under μ_0', which has support $\{\underline{I}_k, \overline{I}_{n''}\}$, is strictly more informative than the (full commitment) experiment with support $\{\overline{I}_n, \overline{I}_{n''}\}$. □

References

Alonso, Ricardo and Odilon Camara, "Organizing Data Analytics," *Working paper*, 2021.

Barros, Lucas, "Information Acquisition Design," *Working Paper*, 2022.

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

Deimen, Inga and Dezső Szalay, "Delegated Expertise, Authority, and Communication," *American Economic Review*, April 2019, *109* (4), 1349–1374.

Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini, "Experimental Cost of Information," *Working paper*, 2022.

Guo, Yingni and Eran Shmaya, "Costly Miscalibration," *Theoretical Economics*, May 2021, *16* (2), 477–506.

Ivanov, Maxim, "Informational Control and Organizational Design," *Journal of Economic Theory*, March 2010, *145* (2), 721–751.

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

Krähmer, Daniel, "Information Design and Strategic Communication," *American Economic Review: Insights*, March 2021, *3* (1), 51–66.

Lin, Xiao and Ce Liu, "Credible Persuasion," *Working paper*, 2022.

Lipnowski, Elliot, "Equivalence of Cheap Talk and Bayesian Persuasion in a Finite Continuous Model," *Working Paper*, 2020.

_ and Doron Ravid, "Cheap Talk with Transparent Motives," *Econometrica*, July 2020, *88* (4), 1631–1660.

_ , _ , and Denis Shishkin, "Persuasion via Weak Institutions," *Journal of Political Economy*, 2022, (forthcoming).

Nguyen, Anh and Teck Yong Tan, "Bayesian Persuasion with Costly Messages," *Journal of Economic Theory*, April 2021, *193*, 105212.

Salamanca, Andrés, “The Value of Mediated Communication,” *Journal of Economic Theory*, March 2021, *192*, 105191.

Shishkin, Denis, “Evidence Acquisition and Voluntary Disclosure,” *Working paper*, 2022.