Persuasion with Coarse Communication

Yunus C. Aybas†  Eray Turkel‡

May 9, 2024

Abstract

In many real-world scenarios, experts must convey complex information with limited message capacity. This paper explores how the availability of messages influences an expert’s persuasive ability. We develop a geometric representation of the expert’s payoff with limited message capacity and identify bounds on the value of an additional signal for the sender. In a special class of games, the marginal value of a signal increases as the receiver becomes more difficult to persuade. Moreover, we show that access to an additional signal does not necessarily translate into more information transmitted in equilibrium, and the receiver might prefer coarser communication. This suggests that regulations on communication capacity have the potential to shift the balance of power from the expert to the decision-maker, ultimately improving welfare. Finally, we study the geometric properties of optimal information structures and show that the complexity of the sender’s problem can be simplified to a finite algorithm.

Keywords: Bayesian Persuasion; Information Design; Coarse Communication

JEL Classification: D82, D83
1 Introduction

The information an expert holds is often much more complex than the language they can use to convey it. For example, credit rating agencies use simple ratings to describe complex financial risks to clients. Similarly, medical professionals often rely on basic health charts to communicate detailed patient information, and government agencies use simple grades to rate restaurant hygiene practices. This common approach of simplifying complex information is also seen in education grading systems, online product ratings, and electrical efficiency ratings. In all of these examples, experts communicate about a complex subject by using limited number of messages.

In this paper, we take coarseness as an exogenous constraint and study the optimal design of information. We explore several key questions: How does the persuasiveness of an expert change when faced with these constraints? Is there any advantage to the receiver in restricting the sender’s ability to communicate? How does coarse communication compare to richer communication?

We focus on settings where the sender has commitment power and utilize the Bayesian persuasion framework of Kamenica and Gentzkow (2011) to model these interactions. Without constraints on the available messages, the only obstacle to persuasive communication is the Bayes rationality of the receiver. The sender can implement any posterior distribution, as long as the expected posterior equals the prior.

When the sender has access to a limited set of messages, she is restricted to making a limited number of recommendations. Consequently, the prior must be represented with a convex combination of limited number of posteriors. We focus on these constrained convex combinations, and show how the concavification technique—the canonical characterization of the sender’s payoff—can be naturally adapted to coarse communication.

With a constraint on available messages, the sender performs worse and she values access to additional messages. We study how much the sender values relaxing this constraint by acquiring an extra message. By linking the sender’s optimal messages with finer and coarser communication, we establish bounds on the value of an additional message. We show that with a large number of states and actions, the marginal value becomes a relatively small fraction of the payoff achievable through richer communication. As we approach unconstrained communication, access to more messages does not change the sender’s payoff significantly. However, this does not imply the marginal value of a message is necessarily decreasing, and it can be non-monotonic.

To offer more precise insights into the value of an additional message, we analyze a class of games with a threshold preference structure. In these games, the receiver has a unique preferred action for each state, taking the corresponding action only if their posterior belief for that state is sufficiently high. Additionally, there is a default action chosen when the poste-
rior belief is highly uncertain, which is the sender’s least preferred action. These preferences
capture various economic settings that have been the focus of prior research, especially in
the context of state-independent cheap talk, such as buyer-seller interactions involving differ-
ent goods (Chakraborty and Harbaugh, 2010) and advice-seeking settings involving multiple
possible actions (Lipnowski and Ravid, 2020).

In this class of games, the utility of a message changes based on the nature of the prior
beliefs. Specifically, the marginal value of a message increases when the prior is skepti-
cal—meaning it is far from the action-belief thresholds. Conversely, this value decreases for
biased priors, which are already near one of these thresholds. Consider a scenario where the
message space is limited and the prior is skeptical. Under these conditions, the sender is con-
strained to satisfy Bayes’ plausibility by inducing their least preferred action with a positive
probability. As the level of skepticism in the prior increases, this probability must increase.
Consequently, the value derived from an additional message becomes greater.

Although we take coarseness as an exogenous constraint, it can also be a deliberate deci-
sion. Consider a patient’s interaction with a doctor. The patient might intentionally limit the
doctor’s options to binary decisions, such as whether to proceed with a specific treatment.
We show that a rational advice-seeker might strategically opt to restrict the advice-giver.
This insight stems from the surprising observation that limited access to messages does not
consistently result in reduced information transmission in equilibrium. In fact, the expert
may communicate in a way that provides more information about the states valuable to the
decision-maker’s choice when the number of available messages is reduced. Consequently, the
decision-maker may benefit from limiting the expert’s communication capacity. This suggests
that regulations on communication capacity can potentially shift the balance of power from
the expert to the decision-maker and enhance overall welfare. We study this feature with an
example in the context of targeted advertising.

Rest of our contributions are technical in nature. We identify geometric properties of the
sender’s optimal strategy using techniques from affine geometry, and expand the geometric
insights offered in the literature (Lipnowski and Mathevet, 2017). Our results demonstrate
how the techniques employed in the literature can be naturally extended, without the need to
assume a rich message space. We discuss how this approach can be extended to other settings
using belief-based approach, such as cheap talk with state-independent sender preferences
(Lipnowski and Ravid, 2020).

**Relationship to the Literature**

Questions relating to limitations of language and implications of coarse communication
have been studied in common-interest coordination games (Blume, 2000; Blume and Board,
2013; De Jaegher, 2003), cheap talk games (Jager, Metzger and Riedel, 2011; Hagenbach and
Koessler, 2020), and information design (Lyu, Suen and Zhang, 2023). Rubinstein (2000)
interprets the cardinality of vocabulary as a facet of bounded rationality. Arrow (1974) demonstrates how coarse communication effectively reflects information transmission within organizational constraints. The primary distinction that sets our work apart from this line of research is the presence of misaligned preferences between the sender and the receiver, and the sender’s ability to commit to a messaging strategy. When players share common interests, coarseness limits the amount of information that can be transmitted and makes both players worse-off. However, we show that when players have misaligned preferences, coarseness does not necessarily lead to reduced information transmission in equilibrium and it can increase the payoff of the uninformed party at the expense of the informed party.

Limitations on the sender’s ability to convey precise information can be represented through various models. One possible disruption to communication quality is exogenous noise. In the models that entertain this possibility, messages chosen by the sender can be misinterpreted due to the imperfections in the channel (Akyol, Langbort and Basar, 2016; Le Treust and Tomala, 2019; Tsakas and Tsakas, 2018). Alternatively, another class of models introduce a cost function based on the amount of information transmitted in a message—termed entropy costs—and apply these costs to the utility of either the sender (Gentzkow and Kamenica, 2014) or the receiver (Wei, 2018; Bloedel and Segal, 2018).

In this literature, Le Treust and Tomala (2019) provide a close comparison to our work. They study (infinitely) repeated Bayesian persuasion, where the sender has limited opportunities to intervene and send information through a noisy and cardinality-constrained channel. Their main result connects to Shannon’s coding theorem, demonstrating that the channel’s capacity, influenced by both noise and the cardinality of the signal space, determines the upper limit on sender utility when the persuasion game is repeated.

However, noisy and coarse signals have substantively different implications on the optimal information structure that will be chosen and the payoff sender gets, when a game is played once. In models of noisy or costly communication, the posterior distribution’s informativeness is restricted. Specifically, either the induced posteriors cannot be too close to the extremes of the belief space due to exogenous noise (Tsakas and Tsakas, 2018), or inducing beliefs closer to the extreme points of the belief space becomes increasingly costly (Bloedel and Segal, 2018; Gentzkow and Kamenica, 2014; Wei, 2018). While these models effectively capture the impact of information distortions and attention constraints, they do not address the effects of communication being finely or coarsely structured.

In contrast, our model posits that players face no restrictions on information processing or attention allocation, but they are limited to communicating with a fixed number of words. These constraints arise naturally due to bounded rationality (Rubinstein, 2000), or as an

---

1Other notable examples include: Organizational economics (Cremer, Garicano and Prat, 2007), matching (McAfee, 2002; Hoppe, Moldovanu and Ozdenoren, 2011) certification (Ostrovsky and Schwarz, 2010; Harbaugh and Rasmusen, 2018), limited memory (Wilson, 1989), and vague language (Lipman, 2009).
exogenous constraint within organizations (Cremer, Garicano and Prat, 2007). With limited access to messages, the sender is not limited in choosing the precision of posterior beliefs. However, she must carefully choose a limited number of actions to communicate with greater accuracy. Unlike costly or noisy communication, the sender in our setting faces a discrete prioritization question reminiscent of knapsack-style problems: choosing the best subset of actions they want to induce while also maintaining Bayes plausibility.

We believe that our approach complements the existing line of work by focusing on cases where communication constraints stem from external restrictions on the number of messages rather than direct limitations on information processing, such as rating systems, coarse grading scales, or organizational communications. Our findings specifically isolate and quantify the direct impact of coarse communication, even in environments where factors like limited attention or external noise may impede information processing. Therefore, our framework proves particularly useful for understanding regulatory measures that restrict communication capacity. Examples include limitations on the targeting abilities of online advertisements or the adoption of simplified hygiene rating systems.

2 Leading Example: Targeted Advertising

We start by analyzing a simple setting with three states to visualize our key insights. This example corresponds to the three-dimensional extension of the examples presented in Rayo and Segal (2010) and Kamenica and Gentzkow (2011), where the state is an underlying random ‘prospect’ capturing the quality of the match between the product characteristics and the customer.

Customers with varying characteristics such as demographics, location, and browsing history arrive at an online platform according to a known distribution. An advertiser, observing these diverse attributes, must decide which type of advertisement to display to optimize engagement.

Customer types are categorized into distinct segments of the population, represented in the state space $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Here, $\omega_1$ corresponds to customers whose preferences do not align with the product offered by the advertiser, $\omega_2$ includes those with weak alignment, and $\omega_3$ denotes those with strong alignment with the product. The prior belief about the distribution of these segments is given by $\mu_0 = (0.65, 0.1, 0.25)$. So, majority of customers initially observed are likely not well-aligned with the product, with a smaller proportion showing some or strong alignment.

---

Note that our analysis of coarse communication becomes interesting only if the state space (or the action space, depending on the binding constraint) has at least three elements. If the state space has two elements, constraining the message space to be smaller leads to no information transmission since the sender will have access to only one message.

---
The state (prospect) of the product is unknown to the customer, who is not aware of the product features before interacting with the advertisement. The advertiser, however, learns the state of the world by observing the characteristics of the incoming customers. Based on this information, the advertiser designs a targeted advertising campaign, deciding which specific advertisement to show to different types of customers. Formally, this is represented with an information structure—a map from states to distributions over the set of available messages.

The actions available to the receiver are represented by the set $A = \{a_3, a_2, a_1, a_0\}$. They correspond to different levels of engagement with the advertisement. Action $a_3$ represents a purchase, which is optimal if the customer’s preferences match the product sold by the advertiser ($\omega_3$). Action $a_2$ represents a click without a purchase, which is optimal when there is a weak match ($\omega_2$). Action $a_1$ represents ignoring or hiding the ad, which is optimal when the customer’s preferences are not aligned to the product ($\omega_1$). Default action $a_0$ represents an impression with no interaction, which is the optimal action when the receiver is sufficiently uncertain about the state.

The sender’s payoff depends solely on the action taken by the receiver, not on the state. She prioritizes engagement ($a_3$) or a click ($a_2$) over no engagement ($a_0$) or hiding the ad ($a_1$). For simplicity, we assume that receiver actions $a_3$ and $a_2$ yield the same high utility to the sender, while $a_1$ and $a_0$ yield the same low utility.\(^3\) We plot the sender and receiver utility in Figure 1.

Given access to three messages (the possibility of showing three different ads depending on customer characteristics), the advertiser induces actions $a_1$, $a_2$, and $a_3$. The optimal advertisement strategy induces the posteriors $\{(1,0,0), (\frac{1}{3}, \frac{2}{3}, 0), (\frac{1}{3}, 0, \frac{2}{3})\}$ with respective probabilities $(0.475, 0.15, 0.375)$. This strategy reveals the state $\omega_1$ with message $m_1$, but sends less precise

---

\(^3\)The assumption of equal utilities for $a_3$, $a_2$ and $a_1,a_0$ is for visual simplicity. The results presented generalize to the case with unequal utilities for different actions.
messages $m_2$ and $m_3$ that mix states $\omega_2$ and $\omega_3$ with $\omega_1$. The advertiser strategically induces the least convincing belief that renders the receiver indifferent between actions $a_0$ and $a_2$ (or $a_3$). This maximizes the total probability that the receiver will take actions $a_2$ or $a_3$.

The optimal solution with three messages can be easily identified by examining the concavification of the sender’s value function, as outlined by Kamenica and Gentzkow (2011). However, if the sender is restricted to using only two messages, the optimal strategy cannot be determined through the standard concavification method. In this scenario, the sender can only utilize convex combinations of at most two posterior beliefs to span the prior.

When picking which two actions to induce, the sender aims to maximize the probability that the receiver selects the action more preferable to her while minimizing the probability of the receiver choosing a less preferable action. Geometrically, this implies that the search can be confined to line segments that pass through the prior (Bayes plausibility) and supported on the ‘corners’ and ‘edges’ of the set of beliefs that lead to a fixed action (optimality). We illustrate some examples in Figure 2. As a preview of our results, observe that the posteriors demonstrated include at least one ‘corner’ (outer point) posterior from the set of beliefs that induce a fixed action. For any alternative combination of posteriors can be rotated in a way that increases the probability of the preferred action and decreases the probability of the less preferred action.

Sender’s optimal strategy with two messages implements actions $a_3$ and $a_1$, by inducing posteriors $(1, 0, 0)$ and $(0.07, 0.27, 0.66)$ with respective probabilities $0.63$ and $0.37$. This information structure maximizes the probability of the action 3 (a purchase, which is the most preferred action) while minimizing the probability of action 1 (hiding the ad, which is the least preferred action). Geometrically, this information structure minimizes the ratio of the distance between the prior and the posterior that leads to the desired action ($a_3$), and the distance between the prior and the posterior that leads to the undesired action ($a_1$). Sender payoff and the optimal information structures with three and two messages are shown in Figure 3.
Figure 3: Optimal information structures with 3 messages (blue, left) and 2 messages (red, right) shown over the sender utility function. The expected sender utility is the point at which the information structures intersect with the black line representing the prior.

Unsurprisingly, the sender performs worse with two messages. But what about the receiver? Could there be benefits to limiting the sender’s communication capabilities? We plot the receiver’s utility under the equilibrium with three and two messages in Figure 4. Note that the receiver has a higher payoff in the equilibrium with two messages. At first glance, this may seem counter-intuitive. The receiver’s payoff is convex and he benefits from increased information transmission. Yet, limiting the number of messages does not necessarily result in less precise posterior distributions at equilibrium. In fact, the optimal information structures for three and two messages are not comparable in the Blackwell sense.\(^4\)

Figure 4: Receiver’s utility over the belief space (yellow function). The beliefs induced by the optimal 3-message solution (blue) and the 2-message solution (red) to the sender’s problem.

To grasp the intuition behind this result, note that the sender is solely concerned with the implemented action. However, the receiver prefers more precise posteriors in specific directions of the belief space. Confining the sender’s targeting ability leads to an optimal information structure that produces more precise posteriors in the direction favored by the receiver (in the direction of \(\omega_2\) and \(\omega_3\)). This suggests that customers would be better off if the

\(^4\)It should be noted that increasing the number of allowed message realizations can never result in the optimal information structure being less informative in the Blackwell sense.
targeting capabilities of the advertiser are constrained. In the appendix, we characterize the conditions on receiver utility under which coarse communication enhances their well-being.\textsuperscript{5} Essentially, if the customers gain high enough utility from reducing the uncertainty about certain states, limiting the targeting capability of the advertiser would make them better off.

Can we describe the utility achievable by the sender for any prior belief? We can do this by employing a modified concavification method, where we define the set of points that are convex combinations of at most two points from the graph of the sender’s value function. These points represent the utilities achievable by the sender as a function of the prior. This is depicted in Figure 5.

![Figure 5: Maximum achievable sender utility with 3 messages (Left) and 2 messages (Right). The black line correspond to the prior belief.](image)

The sender’s utility is lower with two messages, and she values access to additional messages, or increased targeting ability. What can we infer about the marginal value of a signal? The marginal value of a message for any prior belief can be calculated through the difference of the two functions in Figure 5. Notably, for certain priors, the message space constraint is not restrictive, resulting in the marginal value of an additional message being zero. These correspond to priors where the probability of state $\omega_2$ and $\omega_3$ are high, so the sender can satisfy the Bayes plausibility constraint by inducing desirable actions $a_2$ (click) and $a_3$ (purchase) without inducing the undesirable $a_1$ (hide ad). On the other hand, having access to a third message is especially valuable for priors where the sender must induce their least favorite action $a_1$ frequently in a 2-message information structure. The value of an additional message increases as priors become more skeptical about taking actions 2 or 3, and it approaches the center of the simplex.

\textsuperscript{5}If $\beta_0^2, \beta_0^3$ for the set of beliefs inducing $a_0$ are high enough, the receiver will prefer the 2-message outcome over the 3-message outcome. Generally for the parametric preferences we defined, this condition can be written as $\beta_0^2 + \delta \beta_0^3 > 0$ with $\delta$ depending on the prior belief. For our example, $\delta \approx 0.85$. 

9
3 The Model

Let $\Omega$ be a finite state space and $A$ be a compact action space. There are two players, we call them the sender (she) and the receiver (he). They share a prior belief about the state of the world, $\mu_0 \in \Delta(\Omega)$.\(^6\) Both players have utility functions that depend on the state and the receiver’s action $a \in A$, respectively denoted by: $u^S, u^R : \Omega \times A \to \mathbb{R}$, respectively for the sender and the receiver.

The sender uses a finite set of messages available to her, $M$, to communicate the state. Our only deviation from the canonical Bayesian persuasion is to assume that the language is coarse, $|M| = k$ with $2 \leq k < \min\{|\Omega|, |A|\}.\(^7\)

The game starts with the sender committing to an information structure $\pi : \Omega \to \Delta(M)$. For convenience, we often denote $\pi$ as a collection of conditional probability mass functions $\{\pi(\cdot | \omega)\}_{\omega \in \Omega}$. We denote the set of all information structure with $\Pi$.

Once $\pi$ is chosen and announced to the receiver, a state $\omega \in \Omega$ is drawn according to $\mu_0$. Sender sends a message according to the committed information structure $\pi(\cdot | \omega)$ and communicates the realized message $m \in M$ to the receiver. Observing $m$, the receiver forms a posterior $\mu_m$ and choose an action $\hat{a}(\mu_m) \in \operatorname{arg\ max}_{a \in A} \mathbb{E}_{\omega \sim \mu_m} u^R(a, \omega)$. To have a unique selection of $\hat{a}(\mu_m)$, we focus on sender-preferred equilibria.\(^8\)

Given the receiver’s best response, $\hat{a}(\cdot)$, the sender’s expected utility from the information structure $\pi$ is given by:

$$U^S(\pi) := \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{m \in M} \pi(m | \omega) u^S(\hat{a}(\mu_m), \omega).$$

As in Kamenica and Gentzkow (2011), we can transform the problem of choosing an information structure $\pi : \Omega \to \Delta(M)$ to choosing a posterior distribution $\tau \in \Delta(\Delta(\Omega))$. Every information structure $\pi$ induces a posterior distribution $\tau$ with support $\{\mu_m\}_{m \in M}$:

$$\tau(\mu_m) = \sum_{\omega' \in \Omega} \pi(m | \omega') \mu_0(\omega').$$

However, any arbitrary posterior distribution $\tau \in \Delta(\Delta(\Omega))$ cannot be induced by an information structure $\pi$. A plausible posterior distribution $\tau$ induced by an information structure $\pi$ must satisfy two restrictions. First, Bayes updating necessitates that the expected posterior belief of the receiver must equal to her prior belief. This is commonly referred a the Bayes

\(^6\)We extend our results to the case where players have different priors, following Alonso and Camara (2016).

\(^7\)The setting where $k = 1$ is trivial since there will be no information transmission.

\(^8\)The existence of $\hat{a}(\mu_s)$ follows from $A$ being compact and $u(a, \omega)$ being continuous. If the receiver is indifferent between multiple actions, we assume that the indifferenc is resolved by picking the action that is preferred by the sender. If there are multiple such elements, we fix an arbitrary element from the set of maximizers as the choice $\hat{a}(\mu_m)$.
plausibility constraint and formally stated as \( \sum_{m \in M} \mu_m \tau_m(\mu_m) = \mu_0 \). Second, if \(|M| = k\) the sender can induce at most \( k \) different posteriors. We denote the set of all plausible posterior distributions \( \tau \) that can be induced by an information structure \( \pi \) by:

\[
I(k, \mu_0) = \left\{ \tau \in \Delta(\Delta(\Omega)) : \sum_{m \in M} \mu_m \tau_m(\mu_m) = \mu_0 \text{ and } \text{supp}(\tau) \leq k \right\}.
\]

We define the sender’s value function as \( \hat{u}_S(\mu_m) = \mathbb{E}_{\omega \sim \mu_m} u_S(\hat{a}(\mu_m), \omega) \) and similarly define \( \hat{u}_R(\mu_m) \) for the receiver. We use this to write the sender’s information design problem as the following constrained optimization problem:

\[
\max_{\tau \in I(k, \mu_0)} \mathbb{E}_{\mu_m \sim \tau} [\hat{u}_S(\mu_m)].
\]  (1)

There exists a plausible posterior distribution \( \tau \) that solves (1). This follows from an extension of the existence proof of Kamenica and Gentzkow (2011). \( \hat{u}_S \) is upper semi-continuous and attains a maximum over all Bayes plausible posterior distributions. We additionally show that \( I(k, \mu_0) \) is a closed subset of all Bayes plausible posterior distributions in the relevant topological space. This provides compactness of the domain in which the objective is considered. The result immediately follows from the extreme value theorem.

4 Set of Achievable Utilities

We start by providing the geometric characterization of the highest achievable sender payoffs. Let \( \text{CH}(\hat{u}_S) \) denote the convex hull of the graph of \( \hat{u}_S \). The seminal result of Kamenica and Gentzkow (2011) shows that if \((\mu_0, z) \in \text{CH}(\hat{u}_S)\) then the sender payoff \( z \) is achievable by a plausible posterior distribution \( \tau \) when the receiver prior is \( \mu_0 \). However, a prior-payoff pair \((\mu_0, z) \in \text{CH}(\hat{u}_S)\) might not be feasible under coarse communication if the corresponding posterior distribution \( \tau \) has a support with cardinality larger than \( k \).

Which prior-payoff pairs \((\mu_0, z)\) that are admissible under coarse communication? Given a set \( \Lambda \) and a positive integer \( k \), the \( k \)-convex hull of \( \Lambda \) is the set of all points that can be represented as the convex combination of at most \( k \) points in \( \Lambda \) and denote this as \( \text{co}_k(\Lambda) \).

We start with some elementary observations for this object. It follows from the Fenchel and Bunt’s strengthening of Caratheodory’s theorem that if \( n \leq k \) and \( \Lambda \) is a connected set then

\footnote{Throughout the paper, we assume that there are some gains to sending information i.e. there exists a \( \tau \) such that \( \mathbb{E}_{\tau}(\hat{u}_S) \geq \hat{u}_S(\mu_0) \). The other case is trivial and the sender always prefers sending no information.}

\footnote{Formally \( \text{CH}(\cdot) : \Delta(\Delta(\Omega)) \times \mathbb{R} \to \Delta(\Delta(\Omega)) \times \mathbb{R} \) is an operator taking a function and returning the convex hull of the graph of the function, that is \( \hat{u}_S \mapsto \text{co} \left( \text{graph}(\hat{u}_S) \right) \).

\footnote{We provide a formal definition of \( k \)-convex hull (Definition 2) in the appendix.}
co_k(Λ) coincides with co(Λ).\footnote{For a non-connected Λ this holds for n + 1 by Carathedory’s Theorem. See the discussion of Kamenica and Gentzkow (2011) on the rich message space assumption.} It is immediate that co_k is monotone in k, i.e. co_k(Λ) ⊆ co_{k'}(Λ) if k ≤ k'. We similarly define the k-convex hull of the graph of sender utility \( \hat{u}^S \) as \( \mathcal{CH}_k(\hat{u}^S) \). Let \( V(k, \mu_0) := \sup\{z : (\mu_0, z) \in \mathcal{CH}_k(\hat{u}^S)\} \). We show that if \( (\mu_0, z) \in \mathcal{CH}_k(\hat{u}^S) \) then there exists a posterior distribution \( \tau \in \mathcal{I}(k, \mu_0) \) giving payoff \( z \) to the sender.

**Proposition 1.** Let \( \tau \) be the optimal posterior distribution. Then \( V(k, \mu_0) = \mathbb{E}_{\tau} \hat{u}^S \).

This gives us the natural generalization of the concavification result to arbitrary message spaces, which we call \( k\)-concavification. Similar to the concavification approach, \( k\)-concavification can be used to identify the optimal information structure when plotted. An example is provided in Figure 5.

## 5 Marginal Value of a Message

An immediate corollary to the observed monotonicity of \( co_k \) is that \( V(k, \mu_0) \) weakly increasing in \( k \). How much does the sender value acquiring an extra message? We can describe the marginal value of a signal by \( V(k + 1, \mu_0) - V(k, \mu_0) \). From the discussion above, it directly follows that \( \mathcal{CH}_k(\hat{u}^S) = \mathcal{CH}(\hat{u}^S) \) if \( k \geq \min\{|\Omega|, |A|\} \). Thus, additional messages are valuable to the sender only if the communication is coarse.

To see how the marginal value of a message changes, consider a Bayesian persuasion game where the sender cannot induce all combinations of \( k \)-actions. If maintaining Bayes plausibility necessitates inducing a posterior inducing a lower-payoff yielding action, then the sender would be willing to pay more for increased precision in communication.

Unfortunately, the analysis of how optimal information structures change with respect to the number of messages is highly intractable. However, we can still partially characterize the expert’s value for acquiring an extra message by establishing bounds on the marginal value of a message.\footnote{The statement of Proposition 2 is valid for \( u^S \geq 0 \). This assumption is ‘without loss of generality’, as in the case where \( u^S \) can be negative, the utility function can be translated to achieve a minimum of zero, or we can simply change the statement by adding a constant proportional to the minimum of sender utility. We provide the general statement in the proof.}

**Proposition 2.** \( V(k + 1, \mu_0) - V(k, \mu_0) \leq \frac{2}{k+1} V(k + 1, \mu_0) \). This can be equivalently stated as: \( \frac{k-1}{k+1} V(k + 1, \mu_0) \leq V(k, \mu_0) \leq V(k + 1, \mu_0) \).

The \( \frac{2}{k+1} \) factor on the upper bound implies that in Bayesian persuasion games with large state and action spaces, the marginal value of a message cannot be too high as we approach rich communication. However, the result does not necessarily imply monotonicity, as we will
see in the next section. Moreover, this inequality can be recursively applied to get bounds on the value of attainable payoffs with any \( k \) number of messages.

The proof involves developing alternative \( k \)-dimensional plausible posterior distributions derived from the \((k+1)\)-optimal posterior distribution, \( \tau^*_{k+1} \). To do this, we generate \( k+1 \) distinct \( k \)-dimensional distributions. Each is formed by combining pairs of posteriors from the support of \( \tau^*_{k+1} \), while retaining the remaining posteriors unchanged. Specifically, we take two arbitrary posteriors, \( \mu_m \) and \( \mu_{m'} \) within the support of \( \tau^*_{k+1} \), and merge them into a single posterior, \( \mu_{m,m'} \). This new posterior is calculated as follows:

\[
\mu_{m,m'} := \frac{\tau(\mu_m)}{\tau(\mu_m) + \tau(\mu_{m'})} \mu_m + \frac{\tau(\mu_{m'})}{\tau(\mu_m) + \tau(\mu_{m'})} \mu_{m'}.
\]

Consequently, the constructed \( k \)-dimensional plausible posterior distribution, \( \tau_k \), assigns the probability \( \tau_k(\mu_{m,m'}) = \tau^*_{k+1}(\mu_m) + \tau^*_{k+1}(\mu_{m'}) \) and \( \tau_k(\mu_{m''}) = \tau^*_{k+1}(\mu_{m''}) \) for every \( m'' \in M \setminus \{m, m'\} \). The utilities provided by these new information structures are linked to \( V^*(k+1, \mu_0) \), because they contain \( k-1 \) posteriors which are also in the support of \( \tau^*_{k+1} \). Moreover, by the optimality of \( \tau^*_k \), new posterior distributions \( \tau_k \) must provide weakly less utility compared to \( \tau^*_k \). Combining these observations, we establish the upper bound.

In the next section, we study marginal value of a signal in games with a specific threshold preference structure. Proposition 3 in that section shows that the bound stated in Proposition 3 is tight when \( |\Omega| = 3 \). It can be similarly shown that the bound is tight for any \( k \) and \( \Omega \), by extending the the same threshold preference to higher dimensional state spaces.

Finally, by iteratively applying bounds in Proposition 2, we can provide an upper and lower bound on the payoffs attainable using \( k \) messages as a function of the payoff attainable with full communication and binary communication. We state these bounds in Corollary 1, and establish a relationship between achievable payoffs with coarse communication with the payoffs with unlimited communication and binary communication.

**Corollary 1.** \[
\frac{k(k-1)}{2} V(2, \mu_0) \geq V(k, \mu_0) \geq \frac{(k-1)k}{(k+1)(k+2)} V(|\Omega|, \mu_0) \] for every \( k > 2 \).

### 5.1 Belief Threshold Games

In this section we focus on a special class of preferences, to gain sharper insights on the marginal value of a message. We assume that the sender’s utility only depends on the action and not on the state, and the receiver’s default action under the prior is the least preferred action for the sender.

Examples involving these kinds of preferences have received interest in previous work. For instance, they have been used to capture buyer-seller interactions where the seller is trying to convince the buyer to purchase any one of multiple different products, and the buyer’s default action is buying nothing (Chakraborty and Harbaugh, 2010), or a think tank designing a study to persuade a politician to enact one of many possible policy reforms, where the default
action is a continuation of status quo (Lipnowski and Ravid, 2020).\footnote{Similar preferences are also studied by Sobel (2020) to analyze the conditions under which deception in communication will lead to loss in welfare.}

We provide a parametric formulation that captures these settings. For every action, there is a belief threshold above which the receiver finds it optimal to take the action. The default action is optimal if none of these thresholds are met. For a simple demonstration, we study the case where $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $A = \{a_0, a_1, a_2, a_3\}$. Formally, the preferences can be described with the following preference structure:

$$u^R(a, \omega_i) = \begin{cases} 0 & \text{if } a = a_0 \\ \frac{1-T}{T} & \text{if } a = a_i \forall i \in \{1, 2, 3\} \\ -1 & \text{if } a \neq a_i \forall i \in \{1, 2, 3\} \end{cases}$$

For each state $\omega_i$, matching the state with action $a_i$ is optimal, and mismatching the state with action $a_j$ is costly (with $i \neq j > 0$). The receiver can also take a safe action $a_0$ and obtain a payoff of zero. Under this specification of preferences, action $a_i$ is taken by the receiver if and only if the posterior probability of state $\omega_i$ is at least $T$.

Sender’s preferences are such that $u^S(a_0, \omega) = 0$ and $u^S(a_i, \omega) = 1$ for every $i > 0$. So, the sender only cares about the action, and not the realized state. In this setup, the sender wants to persuade the receiver to take one of the non-default actions $a_i$ (with $i \neq 0$) and the parameter $T$ can alternatively be interpreted as the ‘difficulty’ of persuading the desirable action to be taken by the receiver.

It is immediate that the sender can attain a payoff of 1 by using information structures with three messages.\footnote{Any prior $\mu_0$ can be represented as a convex combination of most extreme beliefs i.e. the information structure induces $\mu_1 = (1, 0, 0)$ inducing $a_1$, $\mu_2 = (0, 1, 0)$ inducing $a_2$ and $\mu_3 = (0, 0, 1)$ inducing $a_3$; and corresponding probabilities for those posteriors are $\mu_0(\omega_1)$, $\mu_0(\omega_2)$, and $\mu_0(\omega_3)$.} On the other extreme, with only a single message the sender’s payoff is immediately determined by the action under the prior belief.

We proceed by analyzing the non-trivial intermediate case with two messages and focus on priors $\mu_0$ for which the default action for the receiver is the safe action. Let $\Delta_T$ be the set of beliefs where two-message information structures attain a lower payoff compared to three-message information structures. Formally, we define $\Delta_T := \{\mu_0 \in \Delta(\Omega) : V(2, \mu_0) < V(3, \mu_0)\}$. We characterize the threshold $T$ such that this set is non-empty.

**Lemma 1.** $\Delta_T \neq \emptyset$ if and only if $T > \frac{2}{3}$.

For thresholds $T \leq \frac{2}{3}$, two-dimensional information structures suffice for achieving maximal utility, and value of an additional signal is zero. We restrict attention to cases where this value is positive. We characterize the range of the utilities that can be attained by two-message information structures in Proposition 3.
Figure 6: On the left, we have the action threshold $T = \frac{2}{3}$ so it is possible to maintain Bayes plausibility without inducing action 0 for any prior. On the right, $T > \frac{2}{3}$, so for the prior beliefs in the blue shaded region, the sender has to mix $\mu_0$ and another action when constrained to 2 messages. The blue shaded region in the right figure corresponds to $\Delta_T$.

**Proposition 3.** Let $T > \frac{2}{3}$ and $\mu_0 \in \Delta_T$. Then, $\frac{1}{2 < \frac{V(2, \mu_0)}{3T} < \frac{2T-1}{T} < V(3, \mu_0) = 1}$.

Note that as $T \to 1$, the lower bound identified in Proposition 2 is attained. This conclusion extends to higher dimensions, when the preference structure is appropriately extended. Using this result, we also show that the marginal value of a message can be a function with increasing or decreasing differences, depending on the location of the prior.

**Corollary 2.** Let $T > \frac{2}{3}$ and $\mu_0 \in \Delta_T$. There exists $\mu_0, \mu'_0 \in \Delta_T$ such that:

\[
V(3, \mu_0) - V(2, \mu_0) > V(2, \mu_0) - V(1, \mu_0) \quad \text{and} \quad V(3, \mu'_0) - V(2, \mu'_0) < V(2, \mu'_0) - V(1, \mu'_0)
\]

Since $V(3, \mu_0) = 1$ and $V(1, \mu_0) = 0$, the statement can be equivalently stated in terms of comparing $V(2, \mu_0)$ with $\frac{1}{2}$. The priors for which the marginal value of a message is increasing are the ones that are the furthest away from the desirable action regions. The only way to induce favorable actions with these priors is by also inducing the default action with high probability, getting an expected utility below $\frac{1}{2}$. Therefore, the value of the second message is also below $\frac{1}{2}$. Getting access to the third message allows the sender to maintain Bayes plausibility by not inducing the default action, guaranteeing a payoff of 1. Hence, the value of the third message is higher than $\frac{1}{2}$.

On the other hand, for priors that are already close to one of the action regions, the marginal value of an additional message is decreasing. Intuitively, if the receiver is already leaning towards taking an action, it is easy to induce that action with a high probability. This ensures that the sender can get an expected payoff above $\frac{1}{2}$, making the value of a second message higher than the value of a third message.
6 Properties of Optimal Information Structures.

In practical applications, the implementation of optimal information structures rely on the ability to compute the concavification of the sender’s utility function. This computational task is known to be challenging (Tardella, 2008), especially when dealing with limited messages (Dughmi, Kempe and Qiang, 2016). In this section, we determine the qualitative properties of optimal information structures with coarse communication, and use these properties to construct a finite algorithm to calculate the optimal information structure.

We begin with a result that simplifies the search for optimal information structures by focusing on those that induce affinely independent posteriors.16

Lemma 2. There exists an optimal information structure \( \tau \) such that \( \text{supp}(\tau) \) is affinely independent.

If an information structure \( \tau \) induces posteriors that are affinely dependent, certain posteriors are redundant and can be expressed as affine combinations of others. As a result, the sender can eliminate one of these redundant posteriors while still satisfying Bayes plausibility and weakly improving her payoff at the same time. Notably, we provide a constructive proof that identifies which posterior to drop from a given set of affinely dependent posteriors.

In our second result, we demonstrate that optimal information structures induce the most extreme beliefs possible. To formalize this result, we define the set of all posteriors for which the receiver finds it optimal to take action \( a \) as \( R_a = \{ \mu_m \in \Delta(\Omega) : a \in \arg\max_{a' \in A} \hat{u}^R(\mu_m) \} \).

We refer to these sets as action regions. Each action region can be characterized as the intersection of finitely many closed half-spaces. Thus, they are convex.

Extremeness of a belief that induces an action \( a \) can be defined using a term borrowed from convex analysis:

Definition 1. Let \( q \in \mathbb{N} \). \( \mu_m \) is \( q \)-extreme of \( R_a \), if it is in the interior of a \( q \)-dimensional convex subset of \( R_a \) but not in the interior of any \( (q+1) \)-dimensional convex subset of \( R_a \).

Intuitively, the extremeness of a belief comes from the following observation. A \( q \)-extreme belief that induces action \( a \) can be expressed as a convex combination of \( (q-1) \)-extreme beliefs that also induce action \( a \), but not the other way around. We say a \( q \)-extreme belief is more extreme than a \( q' \)-extreme belief if \( q < q' \). Consequently, a 0-extreme belief is the most extreme belief.

Lemma 3. There exists an optimal information structure \( \tau \) that induces \((k-1)\) posteriors that are 0-extreme points (of some \( R_a \)), and the remaining posterior is a \( q' \leq (n-k) \)-extreme-point.

16Results in this section relies on an assumption that rules out certain preference structures with ‘redundant’ states of the world which are irrelevant for the agents’ utilities. We discuss the formal details in the appendix.
Given an information structure with at least two posteriors that are not 0-extreme, it is always possible to move these posteriors in opposite directions while fixing every other posterior and maintaining Bayes plausibility. Essentially, this corresponds to rotating the information structure within the affine subspace spanned by the other posteriors.

Sender’s utility is convex in each action region, and the probabilities change linearly with this ‘rotation.’ This implies that the sender’s payoff is weakly increasing in either the direction of this rotation or the orthogonal direction. This rotation can be continued until one of the beliefs becomes 0-extreme, at which point any further rotation changes the action induced by the resulting posterior.

Using these result, we can reduce the size of our search space considerably from an infinite set (the set of all Bayes plausible information structures) to a search over a finite set.

**Corollary 3.** The sender’s information design problem described in (1) can be solved by checking finitely many candidate information structures.

The proof of the statement gives the explicit finite procedure to find an optimal information structure. It is straightforward to see that there are only finitely many ways to choose \((k-1)\) posteriors on 0-extreme beliefs of action regions \(\{R_a\}_{a \in A}\). Fixing \((k-1)\) posteriors, the \(k^{th}\) posterior must lie on an affine subspace characterized by \(\mu_0\) and the first \((k-1)\) posteriors, in order to ensure Bayes plausibility.

Searching for the \(k^{th}\) posterior in this affine subspace would still be a search over an infinite set over which the sender utility function is not guaranteed to be continuous and well-behaved. We show that it is without any loss to restrict the search for the optimal \(k^{th}\) posterior to the intersection of this affine subspace and the extreme points of \(\{R_a\}_{a \in A}\). The posteriors in this affine subspace correspond to \(q\)-extreme points of \(\{R_a\}_{a \in A}\) for \(q \leq (n-k)\). Thus, there are only finitely many candidates for the \(k^{th}\) posterior.

Our results in Lemmas 2 and 3 generalize the conclusions of Lipnowski and Mathevet (2017) to arbitrary message spaces. We note that an optimal \(k\)-message information structure can be found by solving \(\binom{|A|}{k}\) linear programs and picking the one with the largest value, based on the work of Bergemann and Morris (2016) and Taneva (2019). Our procedure, instead, compares the payoff under posterior distributions inducing \(k-1\) posteriors on 0-extreme beliefs of action regions \(\{R_a\}_{a \in A}\) and \(k^{th}\) posterior on the \(q\)-extreme points.

7 Discussion

**Optimal Compression.** We have previously shown that optimal information structures are affinely independent. Another way to interpret this result is that the optimal strategy for the sender compresses an \(|\Omega|\)-dimensional state space into a \(|M|\)-dimensional state space. Thus, instead of solving for the optimal \(k\)-message information structure, we can equivalently think
of the sender’s problem as picking a $k$-dimensional subspace and solving a full-dimensional Bayesian persuasion problem in this subspace. We interpret the optimal $k$-dimensional subspace as the optimal way for the sender to compress the higher-dimensional state space into $k$ new states that are affinely independent combinations of the original $n$ states. We provide the formal details of optimal compressions and the results in the appendix.

This approach suggests that our findings can be applied to other settings using the belief based approach in a natural way. In an application where the sender’s strategy is constrained to generate posteriors that span a lower-dimensional subspace, finding the optimal information structure can be described as a problem of optimal compression. In the appendix, we apply this approach to models of cheap talk with state-independent beliefs (Lipnowski and Ravid, 2020). Particularly, we show that when $M$ is constrained, the highest payoff sender can achieve in a state-independent cheap talk game corresponds to the quasi-concave envelope of the sender’s value function over the optimally compressed state space.

In concurrent work, Malamud and Schrimpf (2021) show that the sender can enhance her effectiveness by projecting multi-dimensional data onto an optimal information manifold. Our analysis and constructive methods for finding optimal solutions to constrained persuasion games also informs a recent line of research in the quantization and signal processing literature which studies optimal encoding and decoding schemes with misaligned preferences (Anand and Akyol, 2022, 2023; Akyol and Anand, 2023).

**Experiment Design.** A recent line of literature interpret information design as the strategic design of an experiment which reveals information about the state of the world to all parties (Kolotilin, 2015; Alonso and Camara, 2016). From this perspective, our model can be seen as imposing restrictions on the set of possible experimental procedures. Limited or constrained experiment design has been recently studied by Ball and Espín-Sánchez (2021) and Ichihashi (2019).

Ball and Espín-Sánchez (2021) study a setting where sender has access to a feasible set of experiments and can commit to garbling the outcomes. They analyze welfare implications of garbling the experiments. Through this lens, our model can be thought of as a setting where the sender has access to only a limited set of experiment designs, which naturally arises in settings where a social planner with welfare considerations limits the set of possible experiments. For example, FDA regulates the standards of a clinical trial, prosecutors are limited about what constitutes an evidence and who qualifies as a witness, and experiments on humans can only stratify and control certain variables due to ethical constraints.

**Linear persuasion.** In a concurrent paper, Lyu, Suen and Zhang (2023) extend the study of persuasion games with constrained signal spaces. They specifically focus on settings with continuous states and impose additional assumptions on the preference structures of the agents. They characterize the properties of optimal signaling schemes and analyze comparative statics
as the preference structures or the prior beliefs of the agents change. Similar to our results, their analysis reveals the interesting dynamic of allocating scarce signal resources and the tradeoff faced by the sender when deciding which regions of the state space to focus on.

We showed that limited access to signal spaces may not lead to less informative information structures. In a recent paper, Curello and Sinander (2022) study linear persuasion problems and identify the conditions under which a sender with more ‘convex’ value function will design a more informative signal structure. They extend their comparative statics result to our framework of coarse communication.

8 Conclusion

We set out to analyze the impact of limited access to messages on strategic communication. Specifically, we aimed to assess how the effectiveness of coarse communication compares to richer communication in influencing outcomes and the well-being of the involved parties.

Our findings reveal that the expert consistently performs worse and values gaining access to additional messages. We studied the marginal value of a message, and identified bounds for it. Our study uncovered that rational advice-seekers might find it beneficial to restrict the advice-giver. These findings suggest that regulations on communication capacity have the potential to rebalance power dynamics from the expert to the decision-maker, enhancing overall welfare.

Finally, we analyzed the properties of optimal information structures, using them to simplify the optimal information design problem into a finite procedure. Our results introduce new tools that can seamlessly extend existing findings in the Bayesian persuasion literature to coarse communication.

We believe our approach is useful for analyzing the interaction between the value of commitment and the value of richer communication. Coarseness can also be studied in richer settings, such as competition between senders with access to message spaces with different degrees of coarseness or the challenge of persuading a heterogeneous set of agents using public or private messages with different degrees of coarseness. These questions remain open for future work.
References

Akyol, Emrah and Anju Anand. 2023. Strategic Quantization. In 2023 IEEE International Symposium on Information Theory (ISIT). IEEE pp. 543–548.

Akyol, Emrah, Cedric Langbort and Tamer Basar. 2016. “Information-theoretic approach to strategic communication as a hierarchical game.” Proceedings of the IEEE 105(2):205–218.

Alonso, Ricardo and Odilon Camara. 2016. “Bayesian persuasion with heterogeneous priors.” Journal of Economic Theory 165:672–706.

Anand, Anju and Emrah Akyol. 2022. Optimal Strategic Quantizer Design via Dynamic Programming. In 2022 Data Compression Conference (DCC). IEEE pp. 173–181.

Anand, Anju and Emrah Akyol. 2023. Channel-Optimized Strategic Quantizer Design via Dynamic Programming. In 2023 IEEE Statistical Signal Processing Workshop (SSP). IEEE pp. 621–625.

Arrow, Kenneth J. 1974. “The limits of organization.” Fels Lectures on Public Policy Analysis.

Aumann, Robert J and Sergiu Hart. 2003. “Long cheap talk.” Econometrica 71(6):1619–1660.

Ball, Ian and José-Antonio Espín-Sánchez. 2021. “Experimental Persuasion.”.

Bergemann, Dirk and Stephen Morris. 2016. “Information design, Bayesian persuasion, and Bayes correlated equilibrium.” American Economic Review 106(5):586–91.

Bloedel, Alexander W and Ilya R Segal. 2018. “Persuasion with Rational Inattention.” Available at SSRN 3164033.

Blume, Andreas. 2000. “Coordination and learning with a partial language.” Journal of Economic Theory 95(1):1–36.

Blume, Andreas and Oliver Board. 2013. “Language barriers.” Econometrica 81(2):781–812.

Chakraborty, Archishman and Rick Harbaugh. 2010. “Persuasion by cheap talk.” American Economic Review 100(5):2361–82.

Cremer, Jacques, Luis Garicano and Andrea Prat. 2007. “Language and the Theory of the Firm.” The Quarterly Journal of Economics 122(1):373–407.

Curello, Gregorio and Ludvig Sinander. 2022. “The comparative statics of persuasion.” arXiv preprint arXiv:2204.07474.

De Jaegher, Kris. 2003. “A game-theoretic rationale for vagueness.” Linguistics and Philosophy 26(5):637–659.
Dughmi, Shaddin, David Kempe and Ruixin Qiang. 2016. Persuasion with limited communication. In Proceedings of the 2016 ACM Conference on Economics and Computation. ACM pp. 663–680.

Gentzkow, Matthew and Emir Kamenica. 2014. “Costly persuasion.” American Economic Review 104(5):457–62.

Hagenbach, Jeanne and Frédéric Koessler. 2020. “Cheap talk with coarse understanding.” Games and Economic Behavior 124:105–121.

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

Hoppe, Heidrun C, Benny Moldovanu and Emre Ozdenoren. 2011. “Coarse matching with incomplete information.” Economic Theory 47(1):75–104.

Ichihashi, Shota. 2019. “Limiting Sender’s information in Bayesian persuasion.” Games and Economic Behavior 117:276–288.

Jager, Gerhard, Lars P Metzger and Frank Riedel. 2011. “Voronoï languages: Equilibria in cheap-talk games with high-dimensional types and few signals.” Games and economic behavior 73(2):517–537.

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

Kolotilin, Anton. 2015. “Experimental design to persuade.” Games and Economic Behavior 90:215–226.

Le Treust, Mael and Tristan Tomala. 2019. “Persuasion with limited communication capacity.” Journal of Economic Theory p. 104940.

Lipman, Barton L. 2009. “Why is language vague.” Boston University.

Lipnowski, Elliot and Doron Ravid. 2020. “Cheap talk with transparent motives.” Econometrica 88(4):1631–1660.

Lipnowski, Elliot and Laurent Mathevet. 2017. Simplifying Bayesian Persuasion. Technical report mimeo.

Lyu, Qianjun, Wing Suen and Yimeng Zhang. 2023. “Coarse information design.” arXiv preprint arXiv:2305.18020.

Malamud, Semyon and Andreas Schrimpf. 2021. “Persuasion by Dimension Reduction.” arXiv preprint arXiv:2110.08884.

McAfee, R Preston. 2002. “Coarse matching.” Econometrica 70(5):2025–2034.
Ostrovsky, Michael and Michael Schwarz. 2010. “Information disclosure and unraveling in matching markets.” *American Economic Journal: Microeconomics* 2(2):34–63.

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

Rubinstein, Ariel. 2000. “Economics of language.” *Journal of Economic Literature* 38(3):557–574.

Sobel, Joel. 2020. “Lying and deception in games.” *Journal of Political Economy* 128(3):907–947.

Taneva, Ina. 2019. “Information design.” *American Economic Journal: Microeconomics* 11(4):151–85.

Tardella, Fabio. 2008. “Existence and sum decomposition of vertex polyhedral convex envelopes.” *Optimization Letters* 2(3):363–375.

Tsakas, Elias and Nikolas Tsakas. 2018. “Noisy persuasion.” *Available at SSRN 2940681*.

Volund, Rune T. 2018. Bayesian Persuasion on Compact Subsets. Technical report Aarhus University.

Wei, Dong. 2018. “Persuasion Under Costly Learning.”

Wilson, Robert. 1989. “Efficient and competitive rationing.” *Econometrica: Journal of the Econometric Society* pp. 1–40.
Appendix A  Proofs

A.1  k-convex hull and Preliminary Results

Definition 2. Let $\Lambda \in \mathbb{R}^n$ and $n, k \in \mathbb{N}$. We have that, $x \in \text{co}_k(\Lambda)$ of and only if there exists a set of at most $k$ points $\{\lambda_1, \ldots, \lambda_k\} \subseteq \Lambda$ and a set of corresponding convex weights $\{\gamma_1, \ldots, \gamma_k\}$ such that $\sum_{i \leq k} \gamma_i = 1$ and $\forall i, 1 \geq \gamma_i \geq 0$ such that $\lambda = \sum_{i \leq k} \gamma_i \lambda_i$. Equivalently, we can write:

$$\text{co}_k(\Lambda) = \{a \in \mathbb{R}^n : \exists B \subseteq \Lambda, s.t. \lambda \in \text{co}(B) \text{ with } |B| \leq k\}.$$ 

Lemma 4. For every action $a \in A$, the set $R_a$ is closed and convex.

Proof of Lemma 4. Given $a \in A$, $R_a$ is the intersection of $\Delta(\Omega)$, which is closed and convex, and finitely many closed half spaces defined by $\{\mu \in \mathbb{R}^{|\Omega|} : \sum_{\omega \in \Omega} \mu(\omega)(u(a, \omega) - u(a', \omega)) \geq 0\}$. It is therefore closed and convex. \qed

Lemma 5. The sender’s utility $\hat{u}^S$ is convex when restricted to each set $R_a$.

Proof of Lemma 5. Follows directly from Volund (2018), Theorem 1 or Lipnowski and Mathevet (2017), Theorem 1. \qed

A.2  Sender Payoff: k-concavification

Proof of Proposition 1. Let $\tau$ be the optimal information structure solving the sender’s maximization problem.

By definition $\tau \in I(k, \mu_0)$, so $\sum_{m \in M} \tau(\mu_m)\mu_m = \mu_0$ and $\sum_{m \in M} \tau(\mu_m) = 1$ and $1 \geq \tau(\mu_i) \geq 0$. Hence, using $\tau(\mu_m)$ as the convex weights and $(\mu_m, \mathbb{E}\hat{u}^S(\mu_m))$ as the points, we can show that $(\mu_0, \mathbb{E}_{\tau}\hat{u}^S) \in \text{CH}_k(\hat{u}^S)$. We conclude that $\sup\{z | (\mu_0, z) \in \text{CH}_k(\hat{u}^S)\} \geq \mathbb{E}_{\tau}\hat{u}^S$.

Since $(\mu_0, z) \in \text{CH}_k(\hat{u}^S)$, there exists $\{\hat{u}^S(\mu_m)\}_{m \in M}$ and convex weights $\{\gamma_m\}_{m \in M}$ with $\sum_{m \in M} \gamma_m \mu_m = \mu_0$ and $\sum_{m \in M} \gamma_m \hat{u}^S(\mu_m) = z$. Then, $\{\mu_m\}_{m \in M} \in I(k, \mu_0)$. Therefore, $\tau'$ could have been picked instead of $\tau$ in the sender’s maximization problem, contradicting the optimality of $\tau$. We conclude that $\sup\{z | (\mu_0, z) \in \text{CH}_k(\hat{u}^S)\} \leq \mathbb{E}_{\tau}\hat{u}^S$. \qed

A.3  Marginal Value of Signal

Proof of Proposition 2. Suppose $\tau_k$ is the optimal information structure with $k$ messages, and $\tau_{k-1}$ is the optimal information structure with $k - 1$ messages. Denote the utilities obtained using these information structures with $V(k, \mu_0)$ and $V(k - 1, \mu_0)$.
Let \( \text{supp}(\tau_k) = \{\mu_1, \ldots, \mu_k\} \). We can create a \( k - 1 \) dimensional information structure that maintains Bayes plausibility by choosing two posteriors, without loss say \( \mu_1 \) and \( \mu_2 \), and defining a new posterior \( \mu_{12} \) as their mixture:

\[
\mu_{12} = \frac{\tau_k(\mu_1)}{\tau_k(\mu_1) + \tau_k(\mu_2)} \mu_1 + \frac{\tau_k(\mu_2)}{\tau_k(\mu_1) + \tau_k(\mu_2)} \mu_2
\]

The resulting new information structure has \( \text{supp}(\tau'_{12}) = \{\mu_{12}, \mu_3, \ldots, \mu_k\} \), and new weights \( \{(\tau_k(\mu_1) + \tau_k(\mu_2)), \tau(\mu_3), \ldots, \tau(\mu_k)\} \). Note that \( \tau'_{12} \) maintains Bayes plausibility.

Now, we define \( k \) different information structures, each constructed the same way and containing \( k - 1 \) posteriors, denoted \( \tau_{12}, \tau_{23}, \ldots, \tau_{k-1,k}, \tau_{k1} \). By the optimality of \( \tau_{k-1} \) among the information structures with \( k - 1 \) messages, we have the following \( k \) inequalities:

\[
V(k-1, \mu_0) \geq (\tau_k(\mu_1) + \tau_k(\mu_2)) u^S \left( \frac{\tau_k(\mu_1)}{\tau_k(\mu_1) + \tau_k(\mu_2)} \mu_1 + \frac{\tau_k(\mu_2)}{\tau_k(\mu_1) + \tau_k(\mu_2)} \mu_2 \right) + \cdots + \tau_k(\mu_k) u^S(\mu_k),
\]

\[
V(k-1, \mu_0) \geq \tau_k(\mu_1) u^S(\mu_1) + (\tau_k(\mu_2) + \tau_k(\mu_3)) u^S \left( \frac{\tau_k(\mu_2)}{\tau_k(\mu_2) + \tau_k(\mu_3)} \mu_2 + \frac{\tau_k(\mu_3)}{\tau_k(\mu_2) + \tau_k(\mu_3)} \mu_3 \right) + \cdots + \tau_k(\mu_k) u^S(\mu_k),
\]

\[
V(k-1, \mu_0) \geq (\tau_k(\mu_1) + \cdots + \tau_k(\mu_{k-1}) + \tau_k(\mu_k)) u^S \left( \frac{\tau_k(\mu_{k-1})}{\tau_k(\mu_{k-1}) + \tau_k(\mu_k)} \mu_{k-1} + \frac{\tau_k(\mu_k)}{\tau_k(\mu_{k-1}) + \tau_k(\mu_k)} \mu_k \right),
\]

\[
V(k-1, \mu_0) \geq \tau_k(\mu_2) u^S(\mu_2) + \tau_k(\mu_3) u^S(\mu_3) + \cdots + (\tau_k(\mu_1) + \tau_k(\mu_k)) u^S \left( \frac{\tau_k(\mu_1)}{\tau_k(\mu_1) + \tau_k(\mu_k)} \mu_1 + \frac{\tau_k(\mu_k)}{\tau_k(\mu_1) + \tau_k(\mu_k)} \mu_k \right).
\]

Dividing all inequalities by \( k \) and summing up, we can write:

\[
V(k-1, \mu_0) \geq \frac{k-2}{k} V(k, \mu_0) + \frac{2}{k} V'(k, \mu_0) \geq \frac{k-2}{k} V(k, \mu_0)
\]

Where \( V'(k, \mu_0) \) is the utility of a \( k \) dimensional information structure that consists of the posteriors \( \{\mu_{12}, \mu_{23}, \ldots, \mu_{k-1,k}, \mu_{k1}\} \). For the largest gap possible, we set this equal to the minimum \( u^S \) which is 0. Then, we can rearrange this to obtain the following upper bound on the value of an additional message at \( k - 1 \) messages:

\[
V^*(k) - V^*(k-1) \leq \frac{2}{k} V^*(k)
\]

Analogously, it can be shown that the following relationship must hold between the maximum utilities attainable between \( k \) and \( k - 1 \) messages:

\[
\frac{k-2}{k} V(k, \mu_0) \leq V(k-1, \mu_0) \leq V(k, \mu_0)
\]

This concludes the proof of the claim in the text.
Note that, if the sender utility $u^S$ is allowed to be negative and has the infimum $\underline{u}^S$, then the above inequalities can be equivalently stated as follows:

$$V(k, \mu_0) - V(k - 1, \mu_0) \leq \frac{2}{k} \left( V(k, \mu_0) - \underline{u}^S \right),$$

and

$$\frac{k-2}{k} V(k, \mu_0) + \frac{2}{k} u^S \leq V(k-1, \mu_0) \leq V(k, \mu_0).$$

$$\square$$

### A.4 Threshold Games

Let $(E, \bar{E})$ denote an Euclidean affine space with $E$ being an affine space over the set of reals such that the associated vector space is an Euclidian vector space. We will call $E$ the Euclidean Space and $\bar{E}$ the space of its translations. For this example we will focus on three dimensional Euclidian affine space i.e. $\bar{E}$ has dimension 3. We equip $\bar{E}$ with Euclidean dot product as its inner product, inducing the Euclidean norm as a metric denoted by $d(\cdot)$.

Given this structure, we can define the unitary simplex in the affine space $\mathbb{R}^3$ by the following set where $\omega_i$ corresponds to the point with 1 in its $i$th coordinate and 0 in all of its other coordinates. We define the state space $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The simplex then becomes:

$$\Delta(\Omega) = \left\{ \mu \in \mathbb{R}^3 \mid \mu = \lambda_1 \omega_1 + \lambda_2 \omega_2 + \lambda_3 \omega_3 \text{ such that } \sum_{i=1}^{3} \lambda_i = 1 \text{ and } 1 > \lambda_i > 0 \forall i \in \{1, 2, 3\} \right\}$$

Building on the problem definition in the main text, we focus on Bayesian persuasion games where the receiver preferences are described with thresholds, i.e. the receiver prefers action $a_i \in \{a_1, a_2, a_3\}$ if and only if the posterior belief $\mu_i \in \Delta(\Omega)$ such that $\mu_i(\omega_i) \geq T_i$, and prefers $a_0$ otherwise. Hence, for $i \in \{1, 2, 3\}$ and $j \in \{0, 1, 2, 3\}$ with $i \neq j$ we have $\mathbb{E}_{\mu_i}[u^R(a_i, \omega)] \geq \mathbb{E}_{\mu_i}[u^R(a_j, \omega)]$ if and only if $\mu_i(\omega_i) \geq T_i$.

Define $\delta_1 = (0, 1 - T, -(1 - T))$, $\delta_2 = (1 - T, 0, -(1 - T))$ and $\delta_3 = (1 - T, -(1 - T), 0)$ and $\Gamma_1 = (T, 0, 1 - T)$, $\Gamma_2 = (0, T, 1 - T)$ and $\Gamma_3 = (0, 1 - T, T)$. We have that:

$$R_i = \{ \mu_m \in \Delta(\omega) \mid \mu_m \geq T_i \} = \Delta(\omega) \cap \{ (\mu - \Gamma_i) \cdot \delta_i \geq 0 \mid \mu \in \mathbb{R}^3 \},$$

where $\cdot$ denotes the Euclidean dot product.

**Proof of Lemma 1.** Let $\Delta_T = \Delta(\Omega) \setminus \text{co}_2(R_1 \cup R_2 \cup R_3)$. Note that:

$$\text{co}(R_1 \cup R_2) = \text{co}(\{\omega_1, (T, 1 - T, 0), (T, 0, 1 - T), \omega_2, (1 - T, T, 0), (0, T, 1 - T)\})$$

$$= \text{co}(\{\omega_1, (T, 0, 1 - T), \omega_2, (0, T, 1 - T)\})$$

(2)
The second line follows from the first line since the \{\omega_1, (T, 0, 1 - T), \omega_2, (0, T, 1 - T)\} corresponds to the extreme points of co\{\omega_1, (T, 1 - T, 0), (T, 0, 1 - T), \omega_2, (1 - T, T, 0), (0, T, 1 - T)\}). Similarly for co(R_1 \cup R_3) and co(R_2 \cup R_3) we have that

\[
\begin{align*}
\text{co}(R_1 \cup R_3) &= \text{co}\{\omega_1, (T, 1 - T, 0), \omega_3, (0, 1 - T, T)\} \quad (3) \\
\text{co}(R_2 \cup R_3) &= \text{co}\{\omega_2, (1 - T, 0, T), \omega_3, (1 - T, 0, T)\} \quad (4)
\end{align*}
\]

Using equation (2), (3) and (4), co(R_i \cup R_j) can be identified as the intersection of a half space and the simplex:

\[
\begin{align*}
\text{co}(R_1 \cup R_2) &= \Delta(\Omega) \cap \{ (\mu - (T, 0, 1 - T)) \cdot (-T, T, 0) \geq 0 | \mu \in \mathbb{R}^3 \} \quad (5) \\
\text{co}(R_1 \cup R_3) &= \Delta(\Omega) \cap \{ (\mu - (T, 1 - T, 0)) \cdot (-T, 0, T) \geq 0 | \mu \in \mathbb{R}^3 \} \quad (6) \\
\text{co}(R_2 \cup R_3) &= \Delta(\Omega) \cap \{ (\mu - (1 - T, T, 0)) \cdot (0, -T, T) \geq 0 | \mu \in \mathbb{R}^3 \} \quad (7)
\end{align*}
\]

Thus, \(\Delta_T = \Delta(\Omega) \setminus \text{co}(R_1 \cup R_2 \cup R_3)\). By (5), (6) and (7) we conclude:

\[
\Delta_T = \{ \mu = (\mu_1, \mu_2, \mu_3) \in \Delta(\Omega) | \forall i \in \{1, 2, 3\}, \mu_i > 1 - T \}
\]

This set is non-empty if and only if \(T > \frac{2}{3}\).

**Proof of Proposition 3** We can identify the upper bound through the following problem:

\[
\bar{V}(2, \mu_0) = \max_{i \in \{1, 2, 3\}} \max_{\mu_0 \in \Delta_T, \mu_i \in R_i, \mu_4 \in R_4} \left( 1 - \frac{d(\mu_i, \mu_0)}{d(\mu_4, \mu_0)} \right) \text{ subject to } \mu_0 \in \text{co}(\mu_i, \mu_4).
\]

Without loss of generality we pick \(i = 1\). The maximizing triple is \((\mu_0^*, \mu_1^*, \mu_4^*)\) with \(\mu_0^* = (1 - T, 1 - T, 2T - 1)\), \(\mu_1^* = \left(\frac{1 - T}{2}, \frac{1 - T}{2}, T\right)\) \(\mu_4^* = (0, \frac{1}{2}, \frac{1}{2})\). The solution follows from two observations. One is that given two points \(\mu_0\) and \(\mu_i\) there is a unique line passing through these points hence \(\mu_4\) is identified to be the furthest point on that line such that \(\mu_4 \in R_4\). The line always intersects with \(R_4\) as otherwise \(\mu_0 \notin \Delta_T\) by construction. Then we choose \(\mu_0\) and \(\mu_i\) to minimize \(d(\mu_0, \mu_i)\) where \(d(\mu_0, \mu_i)\) is measured in the space of translations of \(\mathbb{R}^3\). Given this solution, we have that:

\[
\begin{align*}
\| \left( T, \frac{1 - T}{2}, \frac{1 - T}{2} \right) - (2T - 1, 1 - T, 1 - T) \| &= \frac{\sqrt{6}}{2} (1 - T) \\
\| \left( T, \frac{1 - T}{2}, \frac{1 - T}{2} \right) - \left( 0, \frac{1}{2}, \frac{1}{2} \right) \| &= \frac{\sqrt{6}}{2} T
\end{align*}
\]

Giving us that \(\bar{V}(2, \mu_0) = 1 - \frac{1 - T}{T} = \frac{2T - 1}{T} \).

26
Similarly, we can solve:

\[ V(2, \mu_0) = \min_{i \in \{1,2,3\}} \left( \max_{\mu_i \in R_i, \mu_4 \in R_4} \left( \min_{\mu_0 \in \Delta_T} 1 - \frac{d(\mu_i, \mu_0)}{d(\mu_4, \mu_0)} \right) \right) \]

subject to \( \mu_0 \in \text{co}(\mu_i, \mu_4) \).

We observe that the point \( \mu^*_0 = B = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \) is a solution. This follows from the fact that \( B \) is the barycenter of the simplex, and \( R_1, R_2 \) and \( R_3 \) are defined with the same threshold \( T \). Thus, any prior \( \mu_0 \neq B \) implies that the \( \mu_0 \) is closer to one of the action zones. Minimizing the objective, we pick \( \mu^*_0 = B \). Now given this choice, we choose \( \mu_4 \) to maximize leading to the choice of \( \mu^*_4 = (0, \frac{1}{2}, \frac{1}{2}) \) and \( \mu^*_1 = (\frac{1-T}{2}, \frac{1-T}{2}, T) \). Thus, \( V(2, \mu_0) = \frac{1}{3T} \).  

**Proof of Corollary 2.** Observe that with fixed \( T = \frac{2}{3} \), we have \( V(2, \mu_0) = \frac{1}{2} = V(2, \mu_0) \). Also, \( V(2, \mu_0) = \frac{2T-1}{T} \) is increasing in \( T \) and \( V(2, \mu_0) = \frac{1}{3T} \) is decreasing in \( T \). By continuity of distance, the objective function in the definition of \( V(2, \mu_0) \) and \( V(2, \mu_0) \) are continuous. So for any other \( \mu_0 \in \Delta_T \), \( V(2, \mu_0) \) takes every value between \( V(2, \mu_0) \) and \( V(2, \mu_0) \) by intermediate value theorem. By definition, \( V(2, \mu_0) > \frac{1}{2} \) implies decreasing marginal value of a message and \( V(2, \mu_0) < \frac{1}{2} \) implies increasing marginal value of a message. 

### A.5 Properties of Optimal Information Structures

**Definition 3.** The affine hull \( \text{aff}(X) \) of \( X \) is the set of all affine combinations of elements of \( \Lambda \), that is,

\[ \text{aff}(X) = \left\{ \sum_{i=1}^{k} \alpha_i x_i \mid k > 0, x_i \in X, \alpha_i \in \mathbb{R}, \sum_{i=1}^{k} \alpha_i = 1 \right\} \]

**Assumption 1.** Receiver preferences over the simplex are such that the intersection of the affine spans of any two action regions are non-empty: \( \text{aff}(R_p) \cap \text{aff}(R_q) \neq \emptyset, \forall p, q \in A \).

Assumption 1 states that the game is already in this simplest possible representation. This assumption does not lead to any loss in generality and is only about the representation of the preference structure. It is satisfied when the (non-relative) interiors of the action regions \( \{R_a\}_{a \in A} \subseteq \Delta(\Omega) \) are non-empty. It is violated in the case when there are multiple states which are payoff irrelevant for the receiver under different actions, so that the affine spans of some action regions do not intersect.

In settings where Assumption 1 is violated, the persuasion game can be reduced to a simpler representation that satisfies it. Similarly, when Assumption 1 is satisfied, preferences and the state space can be reformulated in a way that violates Assumption 1.

To see this, consider a persuasion game that satisfies Assumption 1 with the state space \( \Omega = \{\theta_1, \theta_2, \theta_3\} \). We can add artificial ‘copies’ of the states to \( \Omega \) and transform it to \( \Omega = \)
\{\theta_1, \theta'_1, \theta_2, \theta'_2, \theta_3, \theta'_3\}$, update the preferences so that the players are indifferent between \{\theta_i, \theta'_i\} and split their prior belief between the copies of the states. However, these extra states only increase the dimensionality of the state space without any substantive difference in preferences, and the game has a simpler representation in a lower dimensional space which combines each \{\theta_i, \theta'_i\} to a single state.

**Proof of Lemma 2.** Let $\text{supp}(\tau) = \{\mu_1, \ldots, \mu_k\}$ be affinely dependent. Then, there must exist \{\lambda_1, \ldots, \lambda_k\} such that $\sum_{i=1}^k \lambda_i = 0$ and $\sum_{i=1}^k \lambda_i \mu_i = 0$. Since $\tau$ is Bayes plausible, we have $\mu_0 = \sum_{i=1}^k \tau(\mu_i) \mu_i$ for some $\tau(\mu_1), \ldots, \tau(\mu_k)$, which satisfy $\sum_{i} \tau(\mu_i) = 1$, and $\forall i, 1 > \tau(\mu_i) > 0$.

Now, from the set \{\lambda_1, \ldots, \lambda_k\}, some elements must be positive and some negative. Among the subset with negative weights, pick \(j^*\) such that $\frac{\tau(\mu_j)}{\lambda_j^*}$ is maximized. Among the subset with positive weights, pick \(p^*\) such that $\frac{\tau(\mu_j)}{\lambda_p^*}$ is minimized. Now, we can write

$$\mu_{j^*} = \sum_{i \neq j^*} -\frac{\lambda_i}{\lambda_{j^*}} \mu_i, \text{ and } \mu_{p^*} = \sum_{i \neq p^*} -\frac{\lambda_i}{\lambda_{p^*}} \mu_i.$$

Now, rewriting the Bayes plausibility condition, we get:

$$\tau(\mu_1) \mu_1 + \cdots + \tau(\mu_{j^*}) \left(\sum_{i \neq j^*} -\frac{\lambda_i}{\lambda_{j^*}} \mu_i\right) + \cdots + \tau(\mu_k) \mu_k = \mu_0$$

$$\Leftrightarrow \sum_{i \neq j^*} \left(\tau(\mu_i) - \frac{\tau(\mu_{j^*}) \lambda_i}{\lambda_{j^*}}\right) \mu_i = \mu_0, \text{ and analagously, } \sum_{i \neq p^*} \left(\tau(\mu_i) - \frac{\tau(\mu_{p^*}) \lambda_i}{\lambda_{p^*}}\right) \mu_i = \mu_0.$$

Now, we will show that $\forall i \neq j^*$, $\left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_{j^*})}{\lambda_{j^*}}\right) \geq 0$ and $\forall i \neq p^*$, $\left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_{p^*})}{\lambda_{p^*}}\right) \geq 0$.

If $\lambda_i = 0$, the inequalities hold trivially.

If $\lambda_i > 0$, the inequalities are equivalent to $\frac{\tau(\mu_i)}{\lambda_i} \geq \frac{\tau(\mu_{j^*})}{\lambda_{j^*}}$ and $\frac{\tau(\mu_i)}{\lambda_i} \geq \frac{\tau(\mu_{p^*})}{\lambda_{p^*}}$. In both cases, the condition holds, because $\lambda_{j^*}$ is negative and $\lambda_{p^*}$ is chosen to minimize this ratio.

If $\lambda_i < 0$, the inequalities are equivalent to $\frac{\tau(\mu_i)}{\lambda_i} \leq \frac{\tau(\mu_{j^*})}{\lambda_{j^*}}$ and $\frac{\tau(\mu_i)}{\lambda_i} \leq \frac{\tau(\mu_{p^*})}{\lambda_{p^*}}$. In both cases, the condition holds, because $\lambda_{j^*}$ is chosen to maximize this ratio and $\lambda_{p^*}$ is positive.

Moreover, note that $\sum_{i \neq j^*} \left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_{j^*})}{\lambda_{j^*}}\right) = (1 - \tau(\mu_{j^*})) + \frac{\tau(\mu_{j^*})}{\lambda_{j^*}} \lambda_{j^*} = 1$, and analagously for $p^*$. Therefore, we can define $\tau'$ and $\tau''$ respectively from $\tau$ by dropping $\mu_{j^*}$ or $\mu_{p^*}$, and we maintain Bayes plausibility using convex weights $\left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_{j^*})}{\lambda_{j^*}}\right)$ and $\left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_{p^*})}{\lambda_{p^*}}\right)$.

Now, writing $E_{\tau} \hat{u}^S - E_{\tau} u^S$ and $E_{\tau'} \hat{u}^S - E_{\tau'} u^S$, we get:

$$E_{\tau'} \hat{u}^S - E_{\tau} u^S = \sum_{i \neq j^*} \left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_{j^*})}{\lambda_{j^*}}\right) \hat{u}^S(\mu_i) - \sum_{i \leq k} \tau(\mu_i) \hat{u}^S(\mu_i)$$
\[
\mathbb{E}_{r'}\hat{u}^S - \mathbb{E}_r\hat{u}^S = \sum_{i \neq p^*} \left( \tau(\mu_i) - \lambda_i \frac{\tau(\mu_{p^*})}{\lambda_{p^*}} \right) \hat{u}^S(\mu_i) - \sum_{i \leq k} \tau(\mu_i)\hat{u}^S(\mu_i)
\]

\[
\Leftrightarrow \mathbb{E}_{r'}\hat{u}^S - \mathbb{E}_r\hat{u}^S = -\frac{\tau(\mu_{j^*})}{\lambda_{j^*}} \left( \sum_{i \neq j^*} \lambda_i \hat{u}^S(\mu_i) \right) - \tau(\mu_{j^*})\hat{u}^S(\mu_{j^*})
\]

\[
\Leftrightarrow \mathbb{E}_{r''}\hat{u}^S - \mathbb{E}_r\hat{u}^S = -\frac{\tau(\mu_{p^*})}{\lambda_{p^*}} \left( \sum_{i \neq p^*} \lambda_i \hat{u}^S(\mu_i) \right) - \tau(\mu_{p^*})\hat{u}^S(\mu_{p^*}).
\]

Suppose \(\mathbb{E}_{r'}\hat{u}^S - \mathbb{E}_r\hat{u}^S < 0\) and \(\mathbb{E}_{r''}\hat{u}^S - \mathbb{E}_r\hat{u}^S < 0\). This implies:

\[
-\frac{1}{\lambda_{j^*}} \left( \sum_{i \neq j^*} \lambda_i \hat{u}^S(\mu_i) \right) - \hat{u}^S(\mu_{j^*}) < 0, \quad \text{and} \quad -\frac{1}{\lambda_{p^*}} \left( \sum_{i \neq p^*} \lambda_i \hat{u}^S(\mu_i) \right) - \hat{u}^S(\mu_{p^*}) < 0
\]

\[
\Leftrightarrow \frac{1}{\lambda_{j^*}} \left( \sum_{i \neq j^*} \lambda_i \hat{u}^S(\mu_i) \right) + \hat{u}^S(\mu_{j^*}) > 0, \quad \text{and} \quad \frac{1}{\lambda_{p^*}} \left( \sum_{i \neq p^*} \lambda_i \hat{u}^S(\mu_i) \right) + \hat{u}^S(\mu_{p^*}) > 0.
\]

However, note that by assumption, \(\lambda_{j^*}\) and \(\lambda_{p^*}\) have opposite signs. Multiplying the first inequality by \(\lambda_{j^*}\) and the second inequality by \(\lambda_{p^*}\), we must have:

\[
\left( \sum_{i \leq k} \lambda_i \hat{u}^S(\mu_i) \right) < 0, \quad \text{and} \quad \left( \sum_{i \leq k} \lambda_i \hat{u}^S(\mu_i) \right) > 0.
\]

Which is a contradiction. So \(\mathbb{E}_{r'}\hat{u}^S - \mathbb{E}_r\hat{u}^S < 0\) and \(\mathbb{E}_{r''}\hat{u}^S - \mathbb{E}_r\hat{u}^S < 0\) cannot hold at the same time, and either \(\tau'\) or \(\tau''\) must yield weakly higher expected utility for the sender.

Replace \(\tau\) with the information structure that yields weakly higher utility using the process defined above, which drops one belief that is affinely dependent. If the resulting information structure is affinely independent, we’re done. If not, we can repeat the process described above and we will either reach an affinely independent set of vectors before we get to two, or we reach two vectors, which must be affinely independent. This completes the proof. \(\square\)

**Proof of Lemma 3.** Suppose \(\mu = \{\mu_1, \ldots, \mu_k\}\) is an information structure, and without loss of generality, let \(\mu_1, \mu_2\) be posteriors that are not 0-extreme points of any action region \(R_\alpha\). Let \(\mu_1 \in R_1\) and \(\mu_2 \in R_2\). Since they are not 0-extreme points, they are at least 1-extreme points. The proof proceeds analogously if they are \(p\)–extreme points for any \(p > 0\).

By Bayes plausibility, we know that \(\sum_{i=1}^{k} \tau(\mu_i)\mu_i = \mu_0\), for the given prior \(\mu_0\). We can
rearrange the Bayes plausibility condition and write:

\[
(\tau(\mu_1) + \tau(\mu_2)) \left( \frac{\tau(\mu_1)\mu_1 + \tau(\mu_2)\mu_2}{\tau(\mu_1) + \tau(\mu_2)} \right) + (1 - \tau(\mu_1) - \tau(\mu_2)) \left( \frac{\sum_{i>2} \tau(\mu_i)\mu_i}{1 - \tau(\mu_1) - \tau(\mu_2)} \right) = \mu_0.
\]

Denoting \( \tau(\mu_1) + \tau(\mu_2) = \tilde{\tau}_{12}, \) \( \tau(\mu_1)\tilde{\tau}_{12} = \tilde{\tau}_1, \) \( \tau(\mu_2)\tilde{\tau}_{12} = \tilde{\tau}_2, \) and \( \frac{\tau(\mu_1)\mu_1 + \tau(\mu_2)\mu_2}{\tau(\mu_1) + \tau(\mu_2)} = \tilde{\mu}_{12}, \) we note that we can replace \( \mu_1, \mu_2 \) with \( \mu'_1, \mu'_2 \) and still maintain Bayes plausibility if the following condition is satisfied:

\[
\alpha \mu'_1 + (1 - \alpha) \mu'_2 = \tilde{\mu}_{12}, \text{ for some } \alpha \in (0, 1).
\]

The new information structure \( \mu' = \{\mu'_1, \mu'_2, \mu_3, \ldots, \mu_k\} \) will be Bayes plausible with the weights \( \tau'(\mu'_i) = \alpha \tilde{\tau}_{12}, \tau'(\mu'_2) = (1 - \alpha) \tilde{\tau}_{12}, \) and \( \tau'(\mu_i) = \tau(\mu_i) \) for \( i > 2. \) Since we know \( \mu_1, \mu_2 \) are (at least) 1-extreme points, there exists line segments \( A_1 \subset R_1, A_2 \subset R_2 \) and \( \mu_1, \mu_2 \) are in the relative interior of \( A_1, A_2 \) respectively.

Now, let us choose \( \mu''_1, \mu''_2 \) that satisfy the following condition:

\[
\frac{2\tilde{\tau}_1 - 1}{\tilde{\tau}_1 - \tilde{\tau}_2} \mu_1 + \frac{2\tilde{\tau}_2 - 1}{\tilde{\tau}_1 - \tilde{\tau}_2} \mu_2 = \mu''_1 - \mu''_2. \tag{8}
\]

With any \( \mu''_1, \mu''_2 \) that satisfies the above condition, we can calculate the corresponding \( \mu'_1, \mu'_2 \) such that:

\[
\tilde{\tau}_1 \mu'_1 + \tilde{\tau}_2 \mu'' = \mu_1,
\]

\[
\tilde{\tau}_1 \mu'_2 + \tilde{\tau}_2 \mu'' = \mu_2.
\]

Moreover, \( \mu'_1, \mu''_1, \mu'_2, \mu''_2 \) will satisfy:

\[
\hat{\mu}_{12} = \tilde{\tau}_1 \mu'_1 + \tilde{\tau}_2 \mu'_2,
\]

\[
\hat{\mu}_{12} = \tilde{\tau}_1 \mu'' + \tilde{\tau}_2 \mu''.
\]

There will be infinitely many possible pairs \( (\mu''_1, \mu''_2) \) that satisfy equation 8, but let us pick an arbitrary pair that are within a sufficiently close radius of \( \mu_1, \mu_2. \) Since \( \hat{u}^S \) is piecewise affine and convex within every action region, let us choose a small enough radius so that \( (\mu''_1, \mu'_1, \mu_1) \) are on the same affine piece in \( R_1, \) and \( (\mu''_2, \mu'_2, \mu_2) \) are on the same affine piece in \( R_2. \) Since \( \mu_1, \mu_2 \) are 1-extreme points, hence relative interior points of the line segments \( A_1, A_2, \) we can find such \( \epsilon, \delta. \) Denoting the directional derivative of \( \hat{u}^S \) with \( \nabla_v \hat{u}^S, \) the piecewise affine nature of the sender utility function will imply the following:

\[
\{\mu'_1, \mu''_1\} \subset (A_1 \cap B_\epsilon(\mu_1)) \subset R_1,
\]

30
respectively. Now, we define the two candidate information structures that will replace \( \mu = \{ \mu_1, \mu_2, \mu_3, \ldots, \mu_k \} \) as follows:

\[
\mu' = \{ \mu_1', \mu_2', \mu_3, \ldots, \mu_k \}, \\
\mu'' = \{ \mu_1'', \mu_2'', \mu_3, \ldots, \mu_k \}.
\]

Denote the part of the sender utility that is coming from the 0-extreme points \( \{ \mu_3, \ldots, \mu_k \} \) as \( \bar{u} = \sum_{i>2}^k \tau(\mu_i)\hat{u}^S(\mu_i) \). Now, by our initial assumption, \( \mu \) is an optimal information structure, so we must have:

\[
\hat{\tau}_1 \hat{\tau}_2 \hat{u}^S(\mu_1') + \hat{\tau}_1 \hat{u}^S(\mu_2') + \bar{u} \leq \tau(\mu_1)\hat{u}^S(\mu_1) + \tau(\mu_2)\hat{u}^S(\mu_2) + \bar{u}, \\
\hat{\tau}_1 \hat{\tau}_2 \hat{u}^S(\mu_1'') + \hat{\tau}_1 \hat{u}^S(\mu_2'') + \bar{u} \leq \tau(\mu_1)\hat{u}^S(\mu_1) + \tau(\mu_2)\hat{u}^S(\mu_2) + \bar{u}
\]

\[
\iff \\
\hat{\tau}_1 \hat{u}^S(\mu_1') + \hat{\tau}_2 \hat{u}^S(\mu_2') \leq \hat{\tau}_1 \hat{u}^S(\mu_1) + \hat{\tau}_2 \hat{u}^S(\mu_2), \\
\hat{\tau}_1 \hat{u}^S(\mu_1'') + \hat{\tau}_2 \hat{u}^S(\mu_2'') \leq \hat{\tau}_1 \hat{u}^S(\mu_1) + \hat{\tau}_2 \hat{u}^S(\mu_2).
\]

\[
\iff \\
\hat{\tau}_1 / \hat{\tau}_2 (\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)) \leq (\hat{u}^S(\mu_2') - \hat{u}^S(\mu_1)), \\
\hat{\tau}_1 / \hat{\tau}_2 (\hat{u}^S(\mu_1'') - \hat{u}^S(\mu_1)) \leq (\hat{u}^S(\mu_2'') - \hat{u}^S(\mu_1)).
\]

Now, by the convexity of \( \hat{u}^S \) within each action region, \( (\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)) \) and \( (\hat{u}^S(\mu_1'') - \hat{u}^S(\mu_1)) \) can't both be negative. Similarly, \( (\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)) \) and \( (\hat{u}^S(\mu_1'') - \hat{u}^S(\mu_1)) \) can't both be positive, since it would imply that \( (\hat{u}^S(\mu_2') - \hat{u}^S(\mu_1')) \) and \( (\hat{u}^S(\mu_2') - \hat{u}^S(\mu_1'')) \) are both positive, which is in contradiction with convexity. This leaves us with two possible cases. We will focus on one case, and the proof proceeds analogously in the symmetric case.

Suppose \( (\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)) \) is positive and \( (\hat{u}^S(\mu_1'') - \hat{u}^S(\mu_1)) \) is negative. This implies \( (\hat{u}^S(\mu_2') - \hat{u}^S(\mu_2')) \) must also be positive. Therefore, \( (\hat{u}^S(\mu_2') - \hat{u}^S(\mu_2'')) \) is negative. Since sender utility is piecewise affine within \( R_1, R_2 \), we rewrite the above inequalities using the directional derivatives and the definitions of \( \mu_1', \mu_1'', \mu_2', \mu_2'' \):

\[
\hat{\tau}_1 / \hat{\tau}_2 (\hat{\tau}_2 \cdot (\mu_1' - \mu_1'')) \leq \gamma \cdot (\hat{\tau}_2 (\mu_2'' - \mu_2')).
\]
\[ \hat{\tau}_1/\hat{\tau}_2 \left( \hat{\tau}_1 \theta \cdot (\mu''_1 - \mu'_1) \right) \leq \gamma \cdot (\hat{\tau}_1 (\mu''_2 - \mu'_2)). \]

\[ \iff \]

\[ \hat{\tau}_1 \left( \theta \cdot (\mu'_1 - \mu''_1) \right) \leq \hat{\tau}_2 \left( \gamma \cdot (\mu''_2 - \mu'_2) \right), \]

\[ \iff \]

\[ \hat{\tau}_1 \left( \theta \cdot (\mu'_1 - \mu''_1) \right) = \hat{\tau}_2 \left( \gamma \cdot (\mu''_2 - \mu'_2) \right). \]

Therefore the information structure \( \mu = \{\mu_1, \mu_2, \mu_3 \ldots, \mu_k\} \) will at best yield the same sender utility with \( \mu' = \{\mu'_1, \mu'_2, \mu_3 \ldots, \mu_k\} \) and \( \mu'' = \{\mu''_1, \mu''_2, \mu_3 \ldots, \mu_k\} \).

The remaining part for the proof follows from the following claim:

**Claim 1.** Let \( |\Omega| = n \) and \( |A| = k \). Suppose we have an information structure \( \tau \) with \( \text{supp}(\tau) = \mu = \{\mu_1, \ldots, \mu_k\} \) satisfying Bayes plausibility. If there exists a posterior in \( \text{supp}(\tau) \) where \( \mu_a \in R_a \) such that \( \mu_a \) is a \( q \)-extreme point of \( R_a \), with \( q > (n-k) \), then there must exist a Bayes plausible \( \tau' \neq \tau \) that weakly improves sender utility.

**Proof.** By our previous results in Lemma 2, we know that \( k \)-dimensional information structures can be improved unless they consist of affinely independent posteriors. So without loss, we can restrict attention to affinely independent \( k \)-dimensional information structures. Since \( |\Omega| = n \), the beliefs over \( \Omega \) are represented in the \((n-1)\) dimensional space. Let \( \mu_1 \) be a \( q \)-extreme point of \( R_1 \) with \( q \geq (n-k) \). In other words, \( \mu_1 \) is in the interior of a \( q \)-dimensional convex set within \( R_1 \), but there is no \( q+1 \) dimensional convex set within \( R_1 \) such that \( \mu_1 \) is an interior point. By definition, there is a unique \( q \)-dimensional affine surface \( S \) containing this \( q \)-dimensional set of \( R_1 \). Note that \( S \) is at least \( n-k+1 \) dimensional.

Moreover, \( \mu_1 \in Y := \text{aff}(\mu) \), which is \((k-1)\)-dimensional. The intersection \( S \cap Y \) is non-empty and includes \( \mu_1 \) by construction. Moreover, it is at least 1 dimensional since \( n-k+1 + k-1 = n > n-1 \).

We can find a radius \( \varepsilon \) small enough such that \( B_\varepsilon(\mu_1) \cap (S \cap Y \cap R_1) \neq \emptyset \). Within this intersection, there exists a line segment, since \( S \cap Y \) is at least 1 dimensional. We can find two points from this line segment \( \mu'_1, \mu''_1 \) such that \( \mu_1 \) is a convex combination of \( \mu'_1, \mu''_1 \) with \( \alpha \mu'_1 + (1-\alpha) \mu''_1 = \mu_1 \).

Therefore we can ‘split’ \( \mu_1 \) into \( \mu'_1, \mu''_1 \) to build the \( k+1 \) dimensional information structure \( \tilde{\mu} = \{\mu'_1, \mu''_1, \mu_2, \ldots, \mu_k\} \) which will satisfy Bayes plausibility with the new adjusted weights \( \{\alpha \tau(\mu_1), (1-\alpha) \tau(\mu_1), \tau(\mu_2), \ldots, \tau(\mu_k)\} \). This yields utility:
\[
\tau(\mu_1)((\alpha)\hat{u}^S(\mu_1') + (1 - \alpha)\hat{u}^S(\mu_k')) + \sum_{i=2}^{k} \tau(\mu_i)\hat{u}^S(\mu_i) \geq \\
\tau(\mu_1)\hat{u}^S(\mu_1) + \sum_{i=2}^{k} \tau(\mu_i)\hat{u}^S(\mu_i),
\]

by convexity of \(\hat{u}^S\) within \(R_1\).

\(\tilde{\mu}\) consists of \(k+1\) points belonging to a \((k-1)\) dimensional affine surface. Thus, it cannot be affinely independent. Using Lemma 2, we can find an improvement by dropping one posterior from \(\tilde{\mu}\), which weakly improves on the utility gained by inducing \(\mu = \{\mu_1, \ldots, \mu_k\}\).

**Proof of Corollary 3.** We have \(|A|\) many action zones with finitely many 0-extreme points. Let us denote the total number of 0-extreme points of all the sets \(\{R_a\}_{a \in A} \subset \Delta(\Omega)\) with \(E\).

An optimal information structure \(\mu = (\mu_1, \ldots, \mu_k)\) should have a support with at least \((k-1)\) 0-extreme points. There are \(\binom{E}{k-1}\) way of picking \((k-1)\) different 0-extreme points. Let us denote an arbitrary choice of \((k-1)\) unique 0-extreme points with \(\mu_{k-1} = (\mu_1, \ldots, \mu_{k-1})\).

If \(\mu_0 \in \text{co}(\mu_{k-1})\) then the information structure \(\mu_{k-1}\) itself is a candidate for the optimal and in fact the optimal sender utility can be achieved with only \((k-1)\) messages.

If \(\mu_0 \not\in \text{co}(\mu_{k-1})\), we can define the set of \(\mu_k\) such that for \(\mu = (\mu_{k-1}, \mu_k)\) we get that \(\mu_0 \in \text{co}(\mu)\).

This set corresponds to the intersection of the affine polyhedral convex cone generated by \(\mu_{k-1} + \mu_0 = (\mu_1 + \mu_0, \ldots, \mu_{k-1} + \mu_0)\) - which we denote \(Y = \{\mu_0 = \sum_{i=1}^{k-1} (\alpha_i \mu_i + \mu_0) \mid \alpha_i \geq 0 \forall i \in \{1, \ldots, k-1\}\}\) and the simplex \(\Delta(\Omega)\). Define the set \(S = Y \cap \Delta(\Omega)\).

By the definition of the set \(Y\), we have that for each \(\mu_k \in S \subset \Delta(\Omega)\) there exists \(\alpha = (\alpha_1, \ldots, \alpha_k)\) with \(\alpha_i > 0\) for all \(i = 1, \ldots, k\) such that \(\sum \alpha_i \mu_i = \mu_0\).

Now if \(\mu = (\mu_{k-1}, \mu_k)\) is not affinely independent, then we can drop some posteriors from \(\tilde{\mu}\) using the protocol described in Lemma 2 and obtain an affinely independent information structure. Moreover, we know \(\tilde{\mu} \neq \mu_k\) since \(\mu_0 \not\in \text{co}(\mu_{k-1})\) violating Bayes plausibility.

If it is the case that \(\mu = (\mu_{k-1}, \mu_k)\) is affinely independent, we have established that for each \(\mu\) - hence for each choice of \(\mu_k \in Y\) - the weights \(\alpha\) are uniquely determined. Hence, given \(\mu_{k-1}\) the choice of \(\mu_k\) determines the sender utility uniquely.

Now we turn to the question of choosing \(\mu_k\). First note that \(Y\) is a polyhedral cone, so it defines a convex polyhedra in \(\mathbb{R}^n\), Moreover, its intersection with \(\Delta(\Omega)\) - an n-dimensional polytope- is a convex polytope. Moreover, \(S = Y \cap \Delta(\Omega)\) has at most dimension \(k < n\). By these facts, it follows that for every action region \(R_a\), the restriction of \(R_a\) to the set \(S\), denoted \(R_a = R_a \cap S\) is a convex polytope of dimension at most \(k\).

We will now show that when we are choosing \(\mu_k\) which must lie in a set \(R_a\), the optimal
choice of \( \mu_k \in \mathcal{R}_a \) can be always restricted to lie on the 0-extreme points of the sets \( \{ \mathcal{R}_a \}_{a \in A} \). Suppose not, let \( \mu_k \) be a \( q \)-extreme point for \( q > 0 \). We can now proceed analogously to proof of Lemma 3 and find a \( \epsilon \)-ball around \( \mu_k \) that will stay inside \( S \) and \( \mathcal{R}_a \). Our assumption on \( \mu_k \) being a \( q \)-extreme point implies that it belongs to a \( q \)-face of \( \mathcal{R}_a \). Moreover, since \( S \) is \( n \)-dimensional and the \( q \)-face \( \mu_k \) belongs to is \( q > 0 \) dimensional, their intersection has dimension of at least 1.

Within this intersection, we can therefore find a line segment and points on this line segment \( \mu'_k, \mu''_k \) such that \( \mu_k \) is a convex combination of \( \mu'_k, \mu''_k \) with \( (\alpha)\mu'_k + (1 - \alpha)\mu''_k = \mu_k \). Again following the same line of argument with Lemma 3, we can show that either the information structure \( \{ \mu_{-k}, \mu'_k \} \) or \( \{ \mu_{-k}, \mu''_k \} \) weakly improves over \( \{ \mu_{-k}, \mu_k \} \). This shows that we can, without loss, pick \( \mu_k \) from the 0-extreme points of \( \mathcal{R}_a \).

Hence, given a choice of \( (\mu_1, \ldots, \mu_{-k}) \) - which are all 0-extreme points of \( \{ \mathcal{R}_a \}_{a \in A} \), the choice of the \( k \)-th point has finitely many candidates identified as the 0-extreme points of the sets \( \{ \mathcal{R}_a \}_{a \in A} = \{ \mathcal{R}_a \cap S \}_{a \in A} \). There are at most \(|A| = m \) sets in this collection with finitely many 0-extreme points. So the optimal information structure can be found in finitely many steps, specifically by choosing the first \((k - 1)\) posteriors in \( \binom{E}{k-1} \) different ways, and adding the final \( k \)-th posterior by checking the 0-extreme points of the sets \( \{ \mathcal{R}_a \}_{a \in A} = \{ \mathcal{R}_a \cap S \}_{a \in A} \).

## Appendix B  Extensions

### B.1 Optimal Compression

We will begin by establishing a series of Lemmas that illustrate the connection between choosing \( k \)-dimensional information structures in \( \Delta(\Omega) \) and optimally compressing \( n \) states to \( k \) states. Subsequently, the Bayesian persuasion problem within the new belief space in \( \mathbb{R}^k \) has a well-known solution given by \( k \)-concavification.

**Definition 4.** A flat \( T \) belonging to the set \( T_k \) can be defined by linearly independent vectors \( \{ \tilde{\mu}_1, \ldots, \tilde{\mu}_k \} \in \mathbb{R}^{n \times k} \) as \( T = \{ \mu \in \mathbb{R}^n | \mu = \mu_0 + \sum_{i=1}^{k} \alpha_i \tilde{\mu}_i \} \subset \mathbb{R}^n \).

A \( k \)-dimensional flat in \( \mathbb{R}^n \) is defined as a subset of a \( \mathbb{R}^n \) that is itself homeomorphic to \( \mathbb{R}^k \). Essentially, flats are affine subspaces of Euclidian spaces. Formally, the coarse strategic communication problem for the sender is equivalent to an alternative formulation in which the sender first selects an ‘optimal \( k \)-dimensional compression,’ denoted as \( T_k \), of the state space. Subsequently, the sender solves a full-dimensional problem in \( \mathbb{R}^k \) with \( k \) messages. This allows us to reinterpret this \( k \)-dimensional summary as the optimal method for the sender to compress the higher-dimensional state space into \( k \) new states, which are mixtures of the former \( n \) states.
Lemma 6.
\[
\max_{\tau} \mathbb{E}_{\mu \sim \tau} \hat{u}^S(\mu_i) \mid \text{subject to } \mathbb{E}_{\mu \sim \tau} \mu = \mu_0, \ |\text{supp}(\tau)| \leq k
\]
(9)
achieves the same optimal value with the problem:
\[
\max_{T \in T_k} \max_{\tau} \mathbb{E}_{\mu \sim \tau} \hat{u}^S(\mu_i) \mid T \text{ subject to } \mathbb{E}_{\mu \sim \tau} \mu = \mu_0, \ |\text{supp}(\tau)| \leq k \text{ and } \text{supp}(\tau) \subset T_k
\]
(10)

Proof. We will first show that a solution to the second maximization problem exists. In order to see this we first establish the compactness of \( T_k \).

Lemma 7. \( T_k \) is a compact smooth manifold. Moreover, \( T \in T_k \) can be represented with the projection matrix of its parallel subspace \( W = \text{span}(\hat{\mu}_1, \ldots, \hat{\mu}_k) \).

Proof. \( T_k \) is homomorphic to the space that parameterizes all \( k \)-dimensional linear subspaces of the \( n \)-dimensional vector space.

This is called the Grassmannian space, which we will denote \( G_k(\mathbb{R}^n) \). The Grassmannian \( G_k(\mathbb{R}^n) \) is the manifold of all \( k \)-planes in \( \mathbb{R}^n \), or in other words, the set of all \( k \)-dimensional subspaces of \( \mathbb{R}^n \). The homeomorphism is obtained by subtracting \( \mu_0 \) from each line equation.

Define the Steifel manifold \( V_k(\mathbb{R}^n) \) as the set of all orthonormal \( k \)-frames of \( \mathbb{R}^n \).\(^{17}\) Hence, elements of \( V_k(\mathbb{R}^n) \) are \( k \)-tuples of orthonormal vectors in \( \mathbb{R}^n \). \( V_k(\mathbb{R}^n) \) is identified with a subset of the cartesian product of \( k \) many \( (n - 1) \) spheres \( S^{n-1} = \{ x \in \mathbb{R}^n : \|x\| = 1 \} \). It is an immediate observation that \( (S^{n-1})^k \) a closed subspace of a compact space. So, we can easily conclude the Steifel manifold \( V_k(\mathbb{R}^n) \) is compact in the inherited topology from \( R^{n \times k} \).

Next, we define a map \( V_k(\mathbb{R}^n) \rightarrow G_k(\mathbb{R}^n) \) which takes each \( n \)-frame to the subspace it spans. Letting \( G_k(\mathbb{R}^n) \) be constructed via the quotient topology from \( V_k(\mathbb{R}^n) \), we establish that \( G_k(\mathbb{R}^n) \) is also compact. This also establishes that \( T_k \) is a compact smooth manifold, as it is just an affine translation of \( G_k(\mathbb{R}^n) \).

Now we will show that \( T \in T_k \) can be represented with the projection matrix of its parallel subspace \( W = \text{span}(\hat{\mu}_1, \ldots, \hat{\mu}_k) \). Consider the set of real \( n \times n \) matrices \( X_k(n) \) that are (i) idempotent, (ii) symmetric and (iii) have rank \( k \). The requirement that a matrix \( X \in X_k(n) \) has rank \( k \) is equivalent to requiring \( X \) has trace \( k \).\(^{18}\)

To prove the second claim, it suffices to define a homeomorphism between \( X_k(n) \) and \( G_k(\mathbb{R}^n) \). The homeomorphism \( \phi \) is \( \phi(X) = C(X), \phi : X_k(n) \rightarrow G_k(\mathbb{R}^n) \) where \( C(X) \) denotes the column space of \( X \). Moreover, letting \( X_W \) be the operator for projection to subspace \( W \) and \( X_{W'} \) be the operator for projection to subspace \( W' \) we can define the metric \( d_{G_k(\mathbb{R}^n)}(W, W') = \|X_W - X_{W'}\| \) where \( \|\cdot\| \) is the operator norm, that metrizes \( G_k(\mathbb{R}^n) \).

\(^{17}\)A \( k \)-frame is is an ordered set of \( k \) linearly independent vectors in a vector space. It is called an orthogonal frame if the set of vectors are orthonormal

\(^{18}\)This follows the fact that \( X \) is idempotent. An idempotent matrix is always diagonalizable and its eigenvalues are either 0 or 1 (Horn and Johnson, 1991). Trace of \( X \) is the sum of its eigenvalues, hence gives the rank of \( X \).
We call the projections from $\Delta(\Omega)$ onto the flat $T \in T_k$ a \textit{k-dimensional summary}, as it is a lower dimensional representation of the $n-$dimensional state space. When we talk about the flat $T$, we will be actually talking about its intersection with the simplex, $T \cap \Delta(\Omega)$, but we will be omitting the intersection for brevity. We will now show that the value of the interior maximization problem is upper-semi continuous in $T$. Formally we prove this with the following Lemma:

\textbf{Lemma 8. The optimal value of the maximization problem:}

$$W(T, \mu_0) = \max_{\tau} \left( \mathbb{E}_{\mu, \sim \tau} \hat{u}^S(\mu_i)|_T \right) \text{ subject to } \mathbb{E}_{\mu, \sim \tau}(\mu_i) = \mu_0, \supp(\tau) = \mu \subseteq T$$

is upper semi-continuous in $T$.

\textit{Proof.} We will start with discussing some preliminary facts. The value function $W(T, \mu_0)$ exists, as shown by Kamenica and Gentzkow (2011).

Let $\tau_T$ be the optimal information structure with support $\mu_T$ on the flat $T$ that is represented with the parallel subspace $W$ and projection matrix $X_T$. Let $\tau_{T'}$ with support $\mu_{T'}$ be the optimal information structure on the flat $T'$ represented with the parallel subspace $W'$ and projection matrix $X_{T'}$, formally $\forall \varepsilon > 0$, there exists a $\delta > 0$ such that whenever we have $|X_T - X_{T'}| < \delta$, we get $W(T', \mu_0) \leq W(T, \mu_0) + \varepsilon$.

We know that $(\mathbb{E}_{\mu, \sim \tau} \hat{u}^S(\mu_i))$ is upper semi-continuous in $\mu$. So for any $\varepsilon$, there exists a $\delta_\varepsilon$ such that whenever $||\mu - \mu'|| < \delta_\varepsilon$, we get $V(\mu') \leq V(\mu) + \varepsilon$. Observe that:

$$||X_T - X_{T'}|| = \sup_{\tilde{\mu}} \{||(X_T - X_{T'})\mu|\mu \in \mathbb{R}^n \text{ and } ||\mu|| \leq 1|| = \sup_{T'} \{||(X_T - X_{T'})\mu|\mu \in \Delta(\Omega)||}\}.

Define $P_T$ and $P_{T'}$ be the projection matrices corresponding to parallel subspace consisting of vectors $\{m_T \mid m_T \in T \cap \text{Bd}\Delta(\Omega)\}$ and $\{m_{T'} \mid m_{T'} \in T' \cap \text{Bd}\Delta(\Omega)\}$. We will show:

$$||X_T - X_{T'}|| = ||(X_T - X_{T'})\tilde{\mu}|| \geq \gamma ||P_T - P_{T'}|| \geq \gamma ||\mu_T - \mu_{T'}||$$

We start by showing that $||(X_T - X_{T'})\tilde{\mu}|| \geq \gamma ||P_T - P_{T'}||$. First, by definition of matrix norm $||(X_T - X_{T'})\tilde{\mu}|| \geq ||P_T - P_{T'}||_{\text{max}} = \max_{r \in R} ||m_T - m_{T'}||_2$. By equivalence of finite dimensional norms, there exists a constant $\gamma$ such that $||P_T - P_{T'}||_{\text{max}} \geq \gamma ||P_T - P_{T'}||$. Hence, we obtain that $||(X_T - X_{T'})\tilde{\mu}|| \geq \gamma ||P_T - P_{T'}||$.

Now let us turn to the last inequality $\gamma ||P_T - P_{T'}|| \geq \gamma ||\mu_T - \mu_{T'}||$. This follows by making $\mu_0$ the origin via subtracting $\mu_0$ i.e. $P_T - \mu_0$, $P_{T'} - \mu_0, \mu_T - \mu_0, \mu_{T'} - \mu_0$ in $\mathbb{R}^N$ and noticing that for $u$ and $v$ in $\mathbb{R}^N$, $||\alpha u - \beta v||$ is monotone in $\alpha$ and $\beta$.

Recall that, $(\mathbb{E}_{\mu, \sim \tau} \hat{u}^S(\mu_i))$ is upper semi-continuous in $\mu$. So for any $\varepsilon$, there exists a $\delta$, such that whenever $||\mu - \mu'|| < \delta$, we get $V(\mu') \leq V(\mu) + \varepsilon$. Then for each $\varepsilon > 0$ one can pick
\[ \delta = \frac{1}{\gamma} \delta_\varepsilon \] to ensure that
\[ \frac{1}{\gamma} \delta_\varepsilon > ||X_T - X_T'|| \geq ||\mu_T - \mu_T'||. \]
This ensures the upper semicontinuity of \( V(T) \) i.e. \( \forall \varepsilon > 0 \), there exists a \( \delta > 0 \) such that whenever we have \( |X_T - X_T'| < \delta \), we get \( V(\mu_T') \leq V(\mu_T) + \varepsilon \).

By above Lemmas, the existence of the optimal for the second maximization problem in Lemma 6 follows from topological extreme value theorem as it is shown to be an upper semi-continuous function maximized over a compact smooth manifold to reals. To complete the proof of Lemma 6, it is straightforward to show that the two maximization problems yield the same maximum.

Let \( \tau_1 \) be the maximizer of equation (9) and \( \tau_2 \in T \) be the maximizer of equation (10). We show that \( \mathbb{E}(\tau_1) \hat{u}^S(\mu) = \mathbb{E}(\tau_1) \hat{u}^S(\mu) \). Suppose not, let \( \mathbb{E}(\tau_1) \hat{u}^S(\mu) > \mathbb{E}(\tau_1) \hat{u}^S(\mu) \). But then in the second problem, we could have picked \( T_{\tau_1} = \text{aff}(\mu_1) \) where \( \text{aff} \) denotes affine hull, and \( \tau = \tau_1 \) to get a higher value, contradicting the optimality of \( \mu_2 \). Now suppose \( \mathbb{E}(\tau_1) \hat{u}^S(\mu) < \mathbb{E}(\tau_1) \hat{u}^S(\mu) \), but then directly picking \( \tau = \tau_2 \) in the first problem yields a better payoff, contradicting to the optimality \( \tau_1 \) in the first problem.

### B.2 Cheap Talk with Transparent Motives

Lipnowski and Ravid (2020) study an abstract cheap-talk model in a recent paper. Their model is identical to our setup, except for three major changes. First, the communication protocol is cheap talk, the expert cannot commit to an information structure \( \pi \). Second, the sender’s utility is independent of the state but only depends on the action taken i.e. \( u^S : A \rightarrow \mathbb{R} \), and hence their paper is titled ‘Cheap Talk with Transparent Motives.’ Finally, in they assume rich message spaces: \( |M| \geq |\Omega| \). We consider a variation of their model where only the last assumption is changed to \( |M| < |\Omega| \).

Throughout this appendix, we will focus on the Perfect Bayesian Equilibria - hereinafter referred as the \( \text{equilibrium-} \mathcal{E}(\pi, \rho, \beta) \) of this cheap talk game. Formally, the equilibrium is defined by three measurable maps: a messaging strategy for the sender \( \pi : \Omega \rightarrow \Delta(M) \); a receiver strategy \( \hat{a} : M \rightarrow \Delta A \); and a belief system for the receiver \( \mu_m : M \rightarrow \Delta \Omega \); such that:

1. \( \mu_m \) is obtained from \( \mu_0 \), given \( \pi \), using Bayes rule;
2. \( \hat{a}(m) \) is supported on \( \arg \max_{a \in A} \int_{\Omega} u_R(a, \cdot) d\beta(\cdot | m) \) for all \( m \in M \); and
3. \( \pi(\omega) \) is supported on \( \arg \max_{m \in M} \int_{A} u_S(\cdot) d\rho(\cdot | m) \) for all \( \omega \in \Omega \).
Lipnowski and Ravid (2020) approach this problem using the belief based approach, similar to the Bayesian persuasion framework we described in the main text, by focusing on information structures $\tau \in \Delta(\Delta(\Omega))$.

As discussed in the main text, every belief system and sender strategy leads to an ex-ante distribution over receiver’s posteriors. By Bayes Rule these posteriors should be equal to the prior on average. Hence, the set of Bayes plausible information structures can be identified by every equilibrium sender strategy which leads to a posterior belief that is an element of $\mathcal{I}(\mu_0) = \{\tau \in \Delta(\Delta(\Omega)) | \int \mu d\tau(\mu) = \mu_0\}$.

However, if the sender is constrained to sending only $k$ messages it can only induce an ex-ante distribution over receiver’s posterior with $k$ elements in the its support and this is the only restriction imposed by access to limited number of messages. Recall that, the set of possible ex-ante distributions was $\mathcal{I}(k, \mu_0) = \{\tau \in \Delta(\Delta(\Omega)) | \int \mu d\tau(\mu) = \mu_0$ and $|\text{supp} (\tau)| \leq k\}.$

Let the sender’s possible continuation values from the receiver having $\mu$ as his posterior be defined with the correspondence $V(\mu) := \text{co} u^S(\text{arg max}_{a \in A} \int u^R(a, \cdot) d\mu).$ Aumann and Hart (2003) and Lipnowski and Ravid (2020) show that an outcome $(\tau, z)$ is an equilibrium outcome if and only if (i) $\tau \in \mathcal{I}(\mu_0)$, and (ii) $z \in \bigcap_{\mu \in \text{supp}(\tau)} V(\mu)$.

Building on their insight, we can show that this result directly extends to the coarse communication environment. Let $(\tau, z)$ be an outcome pair describing a distribution over posterior beliefs $\tau$, and a utility level $z$. When the receiver is constrained to sending $k$-message i.e $|M| \leq k$ we can characterize equilibrium outcomes as follows. The proof of the following Lemma is identical to the proof of Lipnowski and Ravid (2020).

**Lemma 9.** $(\tau, z)$ is an equilibrium outcome if and only if: $\tau \in \mathcal{I}_k(\mu_0)$ and $z \in \bigcap_{\mu \in \text{supp}(\tau)} V(\mu)$.

Essentially, the first condition, $\tau \in \mathcal{I}_k(\mu_0)$, is identical to the condition imposed in the problem in main text. Limiting the available messages limit the set of inducable posteriors with a one-to-one relationship, hence replacing $\tau \in \mathcal{I}(\mu_0)$ with $\tau \in \mathcal{I}_k(\mu_0)$. The second condition - $z \in \bigcap_{\mu \in \text{supp}(\tau)} V(\mu)$ - is a combination of sender and receiver incentive compatibility constraints and shown to be the equilibrium IC condition by Lipnowski and Ravid (2020).

In their paper, Lipnowski and Ravid (2020) also provide a novel way of using non-equilibrium information structures to infer possible equilibrium payoffs of the sender. Formally, they say that an information structure $\tau \in \mathcal{I}(\mu_0)$ secures $z$ if and only if $\mathbb{P}_{\mu \sim \tau}(V(\mu) \geq z) = 1$. Using this definition they show that an equilibrium inducing sender payoff $z$ exists if and only if $z$ is securable.

The only difference in coarse communication is that the sender is restricted use an information structure $\tau$ from $\mathcal{I}_k(\mu_0)$. Hence, we say that an information structure $\tau \in \mathcal{I}_k(\mu_0)$ k-secures $z$ if and only if $\mathbb{P}_{\mu \sim \tau}(V(\mu) \geq z) = 1$. Following the exact arguments in Lipnowski

---

19This is a generalization of $\hat{u}^S(\cdot)$ in the main text. The key difference is that we are no longer focusing on sender-preferred equilibrium.
and Ravid (2020), when $|M| \leq k$ an equilibrium inducing sender payoff $z$ exists if and only if $z$ is $k$-securable.

Using this equilibrium characterization via $k$-securability, we can state that a sender-preferred equilibrium exists and the payoff of the sender in this equilibrium can be characterized by $V_*(\cdot) := \max_{\tau \in \mathcal{I}_k(\cdot)} \inf \{ V(\text{supp } \tau) \}$. In this setting, the sender is maximizing the highest payoff value it can secure across all $k$-dimensional information policies, as $\inf \{ V(\text{supp } \tau) \}$ corresponds to the highest value which the information structure $\tau$ $k$-secures. By comparison, with unlimited messages this value is characterized by $V^*(\cdot) := \max_{\tau \in \mathcal{I}(\cdot)} \inf \{ V(\text{supp } \tau) \}$. Lipnowski and Ravid (2020) show that $V^*(\cdot)$ corresponds to the quasiconcave envelope of sender’s value function $V(\mu)$. This means that it is the pointwise lowest quasi-concave and upper semi-continuous function that majorizes $v$.

**Proposition 4.** In the setting of Lipnowski and Ravid (2020) with a coarse message space $|M| = k$, a sender preferred equilibrium exists. Defining all Bayes plausible information structures within a new compressed space $T_k$ by $\mathcal{I}_{T_k}(\mu_0) = \{ \tau \in \Delta(\Delta(T_k)) \mid \int \mu d\tau(\mu) = \mu_0 \}$, the sender’s utility with the optimal information structure can be characterized by:

$$v_k^* = \max_{T_k \in \mathcal{I}(\mu_0)} \left( \max_{\tau \in \mathcal{I}_{T_k}(\mu_0)} \left( \min_{\mu \in \text{supp } \tau} \mathbb{E}_{\omega \sim \mu} u^S(\mu) \right) \right).$$

Proposition 4 shows quasi-concavification can be used on lower-dimensional linear compressions of the state space, which is equivalent to the solution of the cheap talk game with coarse communication. This is to say that, given sender’s optimal choice of optimal $k$-compression $T_k$, the solution to the sender’s problem is identical to solving an unconstrained problem over the compressed state space $T_k$. Lipnowski and Ravid (2020) point out that the difference between the quasi-concave envelope and the concave envelope at a fixed prior can be interpreted as the value of commitment power for the sender. The methods we develop in this paper can then be used to analyze the interaction between commitment power and communication complexity, to compare the achievable utilities with and without commitment, and with message spaces of different size.

**Proof.** This follows directly from Lipnowski and Ravid (2020) and the result in Lemma 6. To see the equivalence of the maximization problem in Lipnowski and Ravid (2020) with the $V_k^* = \max_{T_k \in T_k} \left( \max_{\tau \in T_k} \left( \min_{\mu \in \text{supp } \tau} \mathbb{E}_{\omega \sim \mu} u^S(\mu) \right) \right)$, it suffices to show that

$$\max_{\tau} \min_{\mu \in \text{supp } (\tau)} \mathcal{C}(k)(\hat{u}^S(\mu)) \text{ subject to } \mathbb{E}_{\mu \sim \tau} \mu = \mu_0$$

Given $x, y \in \mathbb{R}^n$, we say that $x$ weakly majorizes (or dominates) $y$ from below (or equivalently, we say that $y$ is weakly majorized (or dominated) by $x$ from below) denoted as $x \succ_{\omega} y$ if $\sum_{i=1}^k x_i \geq \sum_{i=1}^k y_i$ for all $k = 1, \ldots, n$. 

---

20Given $x, y \in \mathbb{R}^n$, we say that $x$ weakly majorizes (or dominates) $y$ from below (or equivalently, we say that $y$ is weakly majorized (or dominated) by $x$ from below) denoted as $x \succ_{\omega} y$ if $\sum_{i=1}^k x_i \geq \sum_{i=1}^k y_i$ for all $k = 1, \ldots, n$. 

---

39
is equivalent to
\[
\max_{T_k \in T_k} \max_{\tau \in T_k} \min_{\mu \in \text{supp} \tau} \mathbb{C}(\hat{u}^S)(\mu) \text{ subject to } \mathbb{E}_{\mu \sim \tau} \mu = \mu_0.
\]

Existence for the first maximum problem follows from existence results in Lipnowski and Ravid (2020) and the fact that \( \{ \tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_{\mu \sim \tau} \mu = \mu_0 \text{ and } |\text{supp} \tau| \leq k \} \) is a closed subset of \( \{ \tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_{\mu \sim \tau} \mu = \mu_0 \} \). The equivalence follows from Lemma 6 proven above. First it is already shown that \( T_k \) is compact, and secondly \( \max_{\tau \in T_k} \min_{\mu \in \text{supp} \tau} \mathbb{C}(\hat{u}^S)(\mu) \) is upper semicontinuous due to upper semi-continuity of \( \hat{u}^S \).

\[\square\]

**B.3 A Model with Heterogeneous Priors**

We can also easily use our framework to persuasion games in which the sender and the receiver have different priors about the state, originally studied by Alonso and Camara (2016).

Let \( \mu_0^S \) be the sender’s prior, and \( \mu_0^R \) be the receiver’s prior. We will adopt the perspective of the sender. For any posterior belief \( \mu_k \) of the sender, let \( t(\mu_k, \mu_0^S, \mu_0^R) \) denote the perspective transformation function giving us the receiver’s posterior belief, given the priors for the two agents. Alonso and Camara (2016) show that this is a bijective function and provide additional details. For every posterior belief of the sender induced by a signal, there is a unique corresponding posterior for the receiver which can be derived using this simple perspective transformation function. For brevity, we suppress the last two arguments of the function \( t \) and simply write \( t(\mu_k) \) to denote the corresponding receiver posterior given the sender posterior \( \mu_k \).

Re-defining the expected sender utility to reflect heterogeneity in priors, we can write \( \hat{u}^S(\mu_k) = \mathbb{E}_{\omega \sim \mu_k} u^S(\hat{a}(t(\mu_k), \omega)) \), mindful of the fact that when the sender’s posterior is \( \mu_k \), receiver’s will be \( t(\mu_k) \) and the receiver-optimal action \( \hat{a}(t(\mu_k)) \) will be potentially different from \( \hat{a}(\mu_k) \).

Under coarse communication, the sender will solve the following maximization problem

\[
\max_{\tau \in \Delta(\Delta(\Omega))} \mathbb{E}_{\mu_m \sim \tau} \hat{u}^S(\mu_m) \text{ subject to } |\text{supp} \tau| \leq k \text{ and } \mathbb{E}_\tau(\mu_m) = \mu_0^S.
\] (11)

Our framework can be used to analyze the achievable utilities, and the concavification result described in Proposition 2 in Alonso and Camara (2016) can be extended to the case of \( k \)-concavification. Simply, the \( k \)-dimensional optimal information structure given the sender prior \( \mu_0^S \) will be equal to the \( k \)-concavification of the perspective transformed-sender utility function \( V_k(\mu_0^S) = \sup \{ z | (\mu_0^S, \hat{a}) \in \text{CH}_k(\hat{u}^S) \} \).

Therefore, we can quite easily generalize our example in section 2, or our parametric
analysis of threshold games in Section 5.1 to settings where there are disagreements about the prior likelihoods of different states between agents. For example, the voter’s initial beliefs that they would get an ad from an ideologically aligned politician could be different from the politician’s prior belief that they would interact with a voter with aligned ideologies. Or, in the case of threshold games, the buyer’s initial belief on which one of the multiple possible products is a better fit for their preferences could be different from the seller’s beliefs. The k-concavification method can then be used to analyze how the value of increased precision in communication will depend on the level of disagreement (in terms of prior beliefs) between the sender and the receiver.