Persuasion with Coarse Communication*

Yunus C. Aybas[†]   Eray Turkel[‡]

January 31, 2024

Abstract

In many real-world scenarios, experts must convey complex information using a limited number of messages. This paper explores how an expert's ability to persuade changes with the availability of messages. We develop a geometric representation of the expert's payoff when using a limited number of messages. We 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 an additional signal does not directly translate into more information in equilibrium, and the receiver might prefer coarse 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 the sender's problem can be simplified to a finite algorithm.

Keywords: Bayesian Persuasion; Information Design; Coarse Communication

JEL Classification: D82, D83

*We are grateful to Avidit Acharya, Steven Callander, Mine Su Erturk, Francoise Forges, Matthew Gentzkow, Arda Gitmez, Edoardo Grillo, Matthew Jackson, Semih Kara, Tarik Kara, Emin Karagozoglu, Elliot Lipnowski, Cem Tutuncu, Robert Wilson, Kemal Yildiz, Weijie Zhong, seminar audiences in Stanford University, Bilkent University, Stony Brook Game Theory Festival 2022, and Econometric Society Meetings in 2020 and 2021 for helpful comments.

†Stanford University, Department of Economics. Corresponding author: aybas@stanford.edu.

‡Stanford University, Stanford Graduate School of Business.

1 Introduction

In many instances of communication, the information held by a party is often much more complex than the language they can use to convey it. For example, credit rating agencies employ coarse ratings to describe the riskiness of financial assets to their clients, medical professionals often use simplified health charts to communicate complex patient information, and governmental agencies utilize coarse grades to assess hygiene practices in restaurants. Additionally, grading systems in education, product ratings in online reviews, traffic signals, nutritional labels on food packaging, sports statistics, and project management status updates further exemplify instances where coarseness is inherent. In all of these examples, experts communicate about a complex subject by using limited number of messages.

How does the persuasiveness of an expert evolve in the face of these constraints? This is the question at the heart of our investigation. Our primary objective in this study is not to delve into why coarseness arises but, rather, to gain insights into its implications for communication. Specifically, we aim to determine how coarse communication compares to richer communication in influencing outcomes and the well-being of the parties involved. We focus on settings where the expert 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 effective communication is the credibility of the expert. In the canonical Bayesian persuasion framework, the expert can implement any combination of actions, as long as the convex combination of the posterior beliefs induced by the expert's recommendations equal the prior belief of the receiver.

When the expert has access to a limited set of messages, coarseness introduces an additional challenge. The expert is restricted to making a fixed number of recommendations while still adhering to the Bayesian rationality constraint. This limitation results in the prior belief being represented by a convex combination of only a restricted set of posterior beliefs.

We focus on these constrained convex combinations, and show how the concavification approach, the canonical characterization of attainable payoffs, can be seamlessly adapted to our context of coarse communication. This adaptation provides a visual representation of how the expert's utility changes concerning the number of available messages.

The expert performs worse when her communication capacity is constrained and she values access to additional messages. We study how much the expert values acquiring an extra message, and provide a partial answer by establishing bounds on the marginal value of a message that applies to all persuasion games. This result is derived by linking the sender's optimal messages with finer and coarser communication.

The upper-bound shows that in settings with large number of states and actions, the marginal value of a message becomes a relatively small fraction of the payoff achievable through richer communication. Approaching unconstrained communication, having access to more messages does not significantly change the sender's payoff. It's crucial to note that this does not imply the marginal value of a message is necessarily a decreasing function; and it can be non-monotonic.

To offer more precise insights into the value of an additional message, we analyze a class of games characterized by a specific preference structure termed *belief-threshold games*. 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's a default action chosen when the prior belief is 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 different goods (Chakraborty and Harbaugh, 2010) and advice-seeking settings involving multiple possible actions (Lipnowski and Ravid, 2020).

In belief-threshold games, the marginal value of a message increases for *skeptical* priors that are maximally distant from the action-belief thresholds and decreases for *biased* priors that are already close to one of the belief thresholds. When the message space is constrained and the priors are skeptical, the sender can only meet Bayes plausibility by inducing their least preferred action with positive probability. For highly skeptical priors, this probability must increase with restrictions on communication capacity, leading to a higher value from an additional message.

In certain cases coarseness is inevitable, whereas in others, it is a deliberate decision. 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 or not. 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.

We show that, for a class of games, the expert communicates in a way that provides more information about the states critical 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 with an example in the context of targeted advertising.

We provide a simple proof of the existence of an optimal messaging strategy for arbitrary message space. 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 our 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) and cheap talk games (Jager, Metzger and Riedel, 2011; Hagenbach and Koessler, 2020).[1] 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 provide precise information can be modeled in other ways. 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). Another approach is defining a cost function over the amount of information conveyed in the message (i.e. entropy costs), and imposing these costs on the utility of the sender (Gentzkow and Kamenica, 2014) or the receiver (Wei, 2018; Bloedel and Segal, 2018). We discuss how our model differs from these approaches in Section 7.

In terms of the mathematical techniques we develop, our work is also related to Lipnowski and Mathevet (2017) and Dughmi, Kempe and Qiang (2016). Lipnowski and Mathevet (2017) characterize the properties of optimal information structures in message-rich settings relying on extremal representation theorems from convex analysis. We extend their results to settings with general message spaces. Dughmi, Kempe and Qiang (2016) also analyze limited message spaces, but take a computational perspective and focus characterizing the algorithmic complexity of approximating optimal sender utility.

2 Leading Example: Targeted Advertising

We begin by analyzing a simple setting with three states, in order to visualize our key insights using a utility function defined over the space of posterior beliefs for the receiver.[2]

Consider a scenario where various types of customers arrive at an online platform based on

[1]Other notable examples of coarse information include Wilson (1989); McAfee (2002); Hoppe, Moldovanu and Ozdenoren (2011); Ostrovsky and Schwarz (2010); Cremer, Garicano and Prat (2007); Lipman (2009); Harbaugh and Rasmusen (2018).

[2]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.

a known distribution. An advertiser observes diverse characteristics (demographics, location, browsing history, etc.) of incoming customers and must decide which type of advertisement to display based on these observations.

We assume that different customer types correspond to three distinct segments of the population. We represent these three segments as the state space $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The state ω_1 represents preferences and tastes that are not aligned with the product sold by the advertiser, ω_2 represents weak alignment, and ω_3 represents strong alignment. We assume the prior $\mu_0 = (0.65, 0.1, 0.25)$, which is a vector representing the fraction of ω_1, ω_2 and ω_3 type customers in the population.

Our example is 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.

The advertiser learns the state of the world, or the quality of the prospect, by observing the characteristics of the customer. The state is unknown to the customer, who does not know the features of the product sold by the advertiser ex-ante. Formally, the sender's (advertiser's) messaging strategy is a map from Ω to distributions over the set of available of messages $\Delta(S)$. In practical terms, this means the advertiser selects a distribution over various types of advertisements based on the observed characteristics of a customer. The commitment assumption is consistent with advertisers setting up a targeted advertising campaign specifying which ad to show to each type of customer.

The actions available to the receiver are represented by the set $A = \{a_3, a_2, a_1, a_0\}$, and the optimal action depends on their beliefs. The actions 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 (ω_3). Action a_2 represents a click without a purchase, which is optimal when there is a weak match (ω_2). Action a_1 represents ignoring or hiding the ad, which is optimal when the customer's preferences are not aligned to the product (ω_1). Default action a_0 represents an impression with no interaction, which is the optimal when the receiver is sufficiently uncertain about the state.[3]

The sender only cares about the action taken by the receiver, and not the state. Hence, the sender utility function is constant when the receiver's action is fixed. The sender prioritizes engagement, meaning a purchase (a_3) or a click (a_2) is preferred over no engagement (a_0) or hiding the ad (a_1). For simplicity, we assume that receiver actions a_3 and a_2 yield equal utility to the sender, while a_1 and a_0 are the least preferred actions.[4] We plot the sender and

[3]In this example, we suppose that for every state there exists a unique optimal action and there is a unique safe action when there is large uncertainty. Similar preferences are studied in the literature for different contexts (Sobel, 2020; Chakraborty and Harbaugh, 2010; Lipnowski and Ravid, 2020).

[4]The assumption of equal utilities is for visual clarity and can be easily relaxed. The results generalize to the case with unequal utilities for different actions. We set receiver utility to be $u^R(a_i, \mu) = \langle \beta_i, \mu \rangle$ for

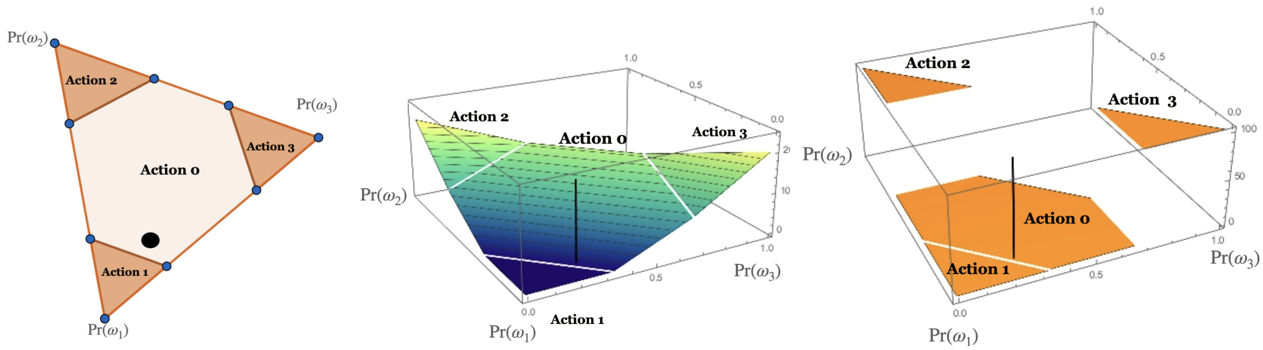

Figure 1: Action regions (Left), and the receiver (Middle) and sender utility over the belief space (Right). We plot sender and receiver utility over the simplex that represents receiver beliefs. The black dot and the black line represent the location of the prior.

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 $\{(1, 0, 0), (1/3, 2/3, 0), (1/3, 0, 2/3)\}$ with respective probabilities $(0.475, 0.15, 0.375)$. This strategy reveals the state ω_1 with message s_1, but sends less precise messages s_2 and s_3 that mix states ω_2 and ω_3 with ω_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 choice maximizes the ex-ante probability that the receiver will opt for actions a_2 or a_3. Moreover, this optimal solution 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 case, the sender can only utilize convex combinations of at most two posterior beliefs to represent the prior. This means that the sender can induce two actions, while adhering to the Bayesian rationality constraint of the receiver.

The sender aims to maximize the likelihood 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. We illustrate some examples in Figure 2.[5]

some coefficient vectors β_i, where $\langle \cdot, \cdot \rangle$ denotes scalar product. We specify the β coefficients so that when the belief $\mu = (\mu_1, \mu_2, \mu_3)$ has coordinate $\mu_i > T_i$, the action a_i is optimal. Namely, for a given $\beta_0 = [\beta_0^1, \beta_0^2, \beta_0^3]$ vector for the action a_0, and k_1, k_2, k_3, representing how much the receiver prefers actions a_1, a_2, a_3 compared to a_0, we define the remaining vectors β_i as : $\beta_i^j = \beta_0^j + k_j$ if $j = i$, and $\beta_0^j - \frac{T_j}{1-T_j} k_j$ if $j \neq i$. For this specific example, we draw and solve for the optimal sender strategy with the receiver preferences defined using $\beta_0 = [-250/3, 500/3, 500/3], \beta_1 = [0, 0, 0], \beta_2 = [-150, 200, 100], \beta_3 = [-150, 100, 200]$.

[5]For simplicity of illustration, we parameterize preferences such that the action region boundaries are parallel to the simplex boundaries, and the sender utility is state-independent. Our conclusions in this section

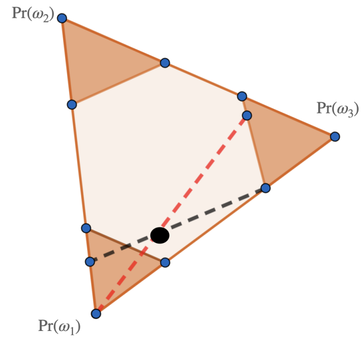

Figure 2: Two-message information structures drawn over the belief space. The black dot represents the prior, and the dashed red and black lines represent information structures.

As a preview of our results, observe that the posteriors demonstrated in Figure 2 include at least one 'corner' (outer point) posterior from the set of beliefs that induce a fixed action. Any alternative combination of posteriors can be rotated to either increase or decrease the probability of one of the induced actions. We generalize this geometric property to higher dimensions and different orders of extreme points.

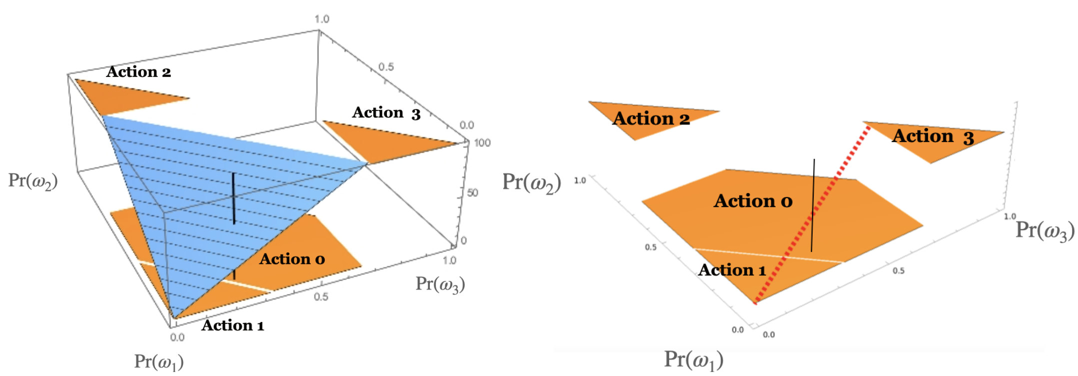

Figure 3: Optimal information structures with 3 messages (blue, left) and 2 messages (red, right) shown over the sender utility function. The expected sender utility is the point at which the information structures intersect with the black line representing the prior.

Sender's optimal strategy induces 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 (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 utility and the optimal information structures with three and two messages are shown in Figure 3.

are not restricted to this case.

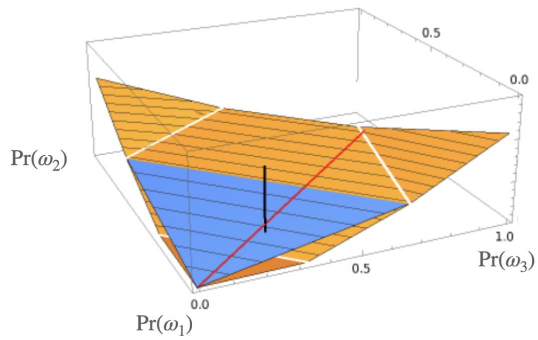

Figure 4: Receiver's Bernoulli utility over the belief space (yellow function). The beliefs induced by the optimal 3-message solution (blue) and the 2-message solution (red) to the sender's problem. Expected receiver utility is the point at which the information structures intersect with the black line representing the prior.

We plot the receiver's utility under the equilibrium with three messages and two messages in Figure 4. We observe that the receiver has a higher payoff in the equilibrium with two messages. This might appear counter-intuitive at first sight, given that the receiver benefits from more information. In fact, the receiver preferences always exhibit convexity over the belief space. However, as depicted in Figure 3 constraining the cardinality of the message space doesn't necessarily lead to less precise posteriors being induced at the equilibrium—the optimal information structures for three and two messages are not Blackwell-comparable.[6]

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. Confining the sender's targeting ability leads to an optimal messaging strategy that produces more precise posteriors in the direction favored by the receiver (in the direction of ω_2 and ω_3). This suggests that customers would be better off if the targeting capabilities of the advertiser are constrained. Naturally, in more general settings, if the preferences of the two agents are perfectly aligned, the receiver would never wish to limit the sender.

In the appendix, we characterize the conditions on receiver utility under which coarse communication enhances their well-being.[7] Essentially, if the customers gain high enough utility from reducing the uncertainty about certain specific states, limiting the targeting capability of the advertiser would make them better off.

The utilities achievable by the sender for any prior belief can be described using a modified concavification method. We do this by plotting the set of points that can be represented as

[6]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.

[7]As long as the slope parameters β_0^2, β_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 δ depending on the prior belief. For our example, $\delta \approx 0.85$.

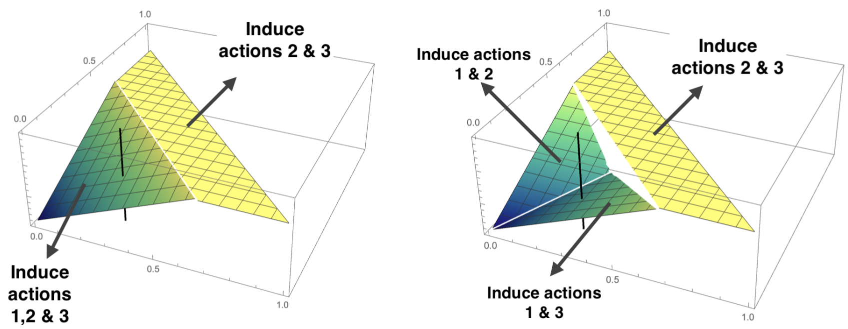

Figure 5: Maximum achievable sender utility with 3 messages (Left) and 2 messages (Right). The black line correspond to the prior belief.

the convex combination of at most 2 points from the graph of the sender's value function. This technique allows us to represent the achievable utilities for the sender as a function of the prior in Figure 5.

The sender's utility is lower with two messages, and she values access to additional messages, or increased targeting ability. The marginal value of a message for any prior belief can be calculated through the difference of the two functions in Figure 5. We see that there are priors where the message space constraint is not binding, and the value of an additional message is zero. These correspond to priors where the probability of state ω_2 and ω_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). This shows that having access to a third message is especially valuable for priors where the sender has to induce their least favorite action a_1 frequently in a 2-message information structure.

3 The Model

We study the canonical Bayesian persuasion game of Kamenica and Gentzkow (2011) extended to environments with limited access to messages.

Let Ω be a (finite) state space and A be a compact action space. There are two agents. We call them the sender (she) and the receiver (he). They share a prior belief about the state of the world, $\mu_0 \in \Delta(\Omega)$.[8] Both players have utility functions that depend on the state and the receiver's action $a \in A$, respectively denoted by: $u^S, u^R : \Omega \times A \to \mathbb{R}$ for the sender and the receiver.

The sender uses a language (message space) S, a finite set of messages available to her, to communicate the state. Critically, our only deviation from the canonical Bayesian persuasion

[8]We extend our results to when players do not share a common prior, following the approach of Alonso and Camara (2016).

is to assume that the language is coarse, i.e. $|S| = k$ with $2 \leq k < \min\{|\Omega|, |A|\}$.[9]

The game starts with the sender committing to a messaging strategy $\pi : \Omega \to \Delta(S)$. For convenience, we also denote a messaging strategy as a collection of conditional probability mass functions $\{\pi(\cdot \mid \omega)\}_{\omega \in \Omega}$. We denote the set of all messaging policies with Π.

Once π is chosen and announced to the receiver, a state ω is drawn from Ω according to μ_0. Sender sends a message according to the committed messaging strategy $\pi(s|\omega)$ and communicates the realized message s with the receiver. Observing s, the receiver forms a posterior μ_s and chose an action $\hat{a}(\mu_s) \in \arg\max_{a \in A} \mathbb{E}_{\omega \sim \mu_s} u^R(a, \omega)$.

The existence of $\hat{a}(\mu_s)$ follows from A being compact and $u(a, \omega)$ being continuous. To have a unique selection of $\hat{a}(\mu_s)$, we focus on sender-preferred equilibria. More precisely, if the receiver is indifferent between multiple actions, we assume that the indifference 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 $\hat{a}(\mu_s)$.[10]

Given the receiver's best response, $\hat{a}(\cdot)$, the sender's expected utility from committing to a messaging strategy π is given by:

$$U^S(\pi) := \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{s \in S} \pi(s \mid \omega) u^S(\hat{a}(\mu_s), \omega).$$

The optimal messaging strategy π maximizes $U^S(\pi)$ over Π. As in Kamenica and Gentzkow (2011), we can transform the problem of choosing a messaging strategy $\pi : \Omega \to \Delta(S)$ to choosing an *information structure* $\tau \in \Delta(\Delta(\Omega))$. Formally, every messaging strategy π induces a distribution τ with support $\{\mu_s\}_{s \in S}$:[11]

$$\tau(\mu_s) = \sum_{\omega' \in \Omega} \pi\left(s \mid \omega'\right) \mu_0\left(\omega'\right).$$

However, any arbitrary distribution in $\tau \in \Delta(\Delta(\Omega))$ cannot be induced by a messaging strategy π. An information structure τ induced by a messaging strategy π must satisfy two restrictions. First, Bayesian updating necessitates that the expected posterior belief of the receiver must equal to her prior belief. This is commonly referred a the *Bayes plausibility* constraint and formally stated as $\sum_{\mu_s \in \mathsf{supp}\,(\tau)} \mu_s \tau(\mu_s) = \mu_0$.

Second constraint is special to our focus on coarse communication. If $|S| = k$ the sender can induce at most k different posteriors, i.e. $\mathsf{supp}\,(\tau) \leq k$.

[9] The setting where $|S| = k = 1$ is trivial since there will be no information transmission.

[10] The literature on Bayesian persuasion generally focuses on sender-preferred equilibrium for existence. Lipnowski and Ravid (2020) studies the robustness of this assumption.

[11] Formally there might be multiple messages that induce a posterior $\tilde{\mu}$. To minimize the notational clutter, we do not entertain this possibility in the main text. The only difference is that the formula for $\tau(\tilde{\mu})$ has an additional sum $\sum_{s:\mu_s = \tilde{\mu}}$.

We denote the set of all distributions τ that can be induced by a messaging strategy π by:

$$\mathcal{I}(k,\mu_0) = \left\{ \tau \in \Delta(\Delta(\Omega)) \; : \sum_{\mu_s \in \mathsf{supp}\,(\tau)} \mu_s \tau(\mu_s) = \mu_0 \text{ and } \mathsf{supp}\,(\tau) \leq k \right\}.$$

Let $\hat{u}^S(\mu_s) = \mathbb{E}_{\omega\sim\mu_s} u^S(\hat{a}(\mu_s),\omega)$ and similarly define $\hat{u}^R(\mu_s)$ for the receiver. $\hat{u}^S$ is the sender's value function (as a function of the receiver's posterior μ_s). We use this to write the sender's information design problem as the following constrained optimization problem:

$$\max_{\tau\in\Delta(\Delta(\Omega))} \mathbb{E}_{\mu_s\sim\tau}[\hat{u}^S(\mu_s)] \text{ subject to } \tau \in \mathcal{I}(k,\mu_0). \tag{1}$$

We first show that the existence of a solution to the sender's problem in equation (1).[12]

Proposition 1. *There exists an optimal information structure τ solving the optimization problem in equation (1).*

The result follows from an extension of the existence proof of Kamenica and Gentzkow (2011). Their result established that $\hat{u}^S$ is upper semi-continuous and attains a maximum over all Bayes plausible information structures. We additionally show that $\mathcal{I}(k,\mu_0)$ is a closed subset of all Bayes plausible information structures 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 a geometric characterization of the highest achievable sender payoffs. Let $\mathbb{CH}(\hat{u}^S)$ denote the convex hull of the graph of $\hat{u}^S$.[13] The seminal result of Kamenica and Gentzkow (2011) shows that if $(\mu_0,z) \in \mathbb{CH}(\hat{u}^S)$ then the sender payoff z is achievable by an information structure τ when the receiver prior is μ_0. However, a prior-payoff pair $(\mu_0,z) \in \mathbb{CH}(\hat{u}^S)$ might not be feasible under coarse communication if the corresponding τ has a support with cardinality larger than k.

We focus on prior-payoff pairs (μ_0,z) that are admissible under coarse communication. Given a set Λ and a positive integer k, the *k-convex hull* of Λ is the the set of all points that can be represented as the convex combination of at most k points in the set Λ and denote this as $\mathrm{co}_k(\Lambda)$.[14]

[12] Throughout the paper, we assume that there are some gains to sending information i.e. there is some τ 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.

[13] Since $\hat{u}^S : \Delta(\Omega) \to R$, we can represent any belief μ with $|\Omega| - 1 = n - 1$ dimensions. Thus, formally $\mathbb{CH}(\cdot) : \mathbb{R}^{n-1} \times \mathbb{R} \to \mathbb{R}^{n-1} \times \mathbb{R}$ is an operator taking a function whose graph can be represented in $\mathbb{R}^n$, and returning the convex hull of the graph of the function in $\mathbb{R}^n$ that is $\hat{u}^S \mapsto \mathrm{co}(\mathrm{graph}(\hat{u}^S))$.

[14] We provide a formal definition of k-convex hull (Definition 2) in the appendix.

We start with some elementary observations for this object. Using the Fenchel and Bunt's strengthening of Caratheodory's theorem it follows that whenever $n \geq k$ and Λ is a connected set $\text{co}_k(\Lambda)$ coincides with $\text{co}(\Lambda)$.[15] It is immediate that co_k is monotone in k, i.e. $\text{co}_k(\Lambda) \subseteq \text{co}_{k'}(\Lambda)$ if $k \leq k'$.

We similarly define the k-convex hull of the graph of sender utility $\hat{u}^S$ as $\mathbb{CH}_k(\hat{u}^S)$. By following the classical arguments in the literature, it can be shown that if $(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)$, then there exists an information structure $\tau \in \mathcal{I}(k, \mu_0)$. Thus, the set of payoffs sender can achieve, when the prior is μ_0 and there are k messages, is given by:

$$V(k, \mu_0) := \sup\{z : (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\}.$$

Formally, we can state this result in Proposition 2.

Proposition 2. *Let τ be a solution to equtation (1) 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 Section 2 and Figure 5. We discuss the broader scope of this approach and explore how it can be applied to other problems where the concavification technique is employed, later in the paper.

5 Marginal Value of a Message

An immediate corollary to the observed monotonicity of co_k is that $V(k, \mu_0)$ weakly increases in k. Thus, the sender values access to more messages. This begs the question: how much does the expert value acquiring an extra message?

We can describe the marginal value of a signal by $V(k+1, \mu_0) - V(k, \mu_0)$. First note that the marginal value of a message is zero whenever the language is rich, i.e. $k \geq \min\{|\Omega|, |A|\}$. This follows from the previous observation that $co_k(\Lambda) = co(\Lambda)$ if $k \geq \min\{|\Omega|, |A|\}$. So, additional messages are valuable to the sender only if the communication is coarse.

To see how the marginal value of a message can change, consider a communication game where the sender cannot induce all combinations of k-actions. If maintaining Bayes plausibility with lower dimensional messages requires 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 number of messages is highly intractable. However, we can still partially characterize the

[15]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.

expert's value for acquiring an extra message by establishing bounds on the marginal value of a message.[16]

Proposition 3. *Let $|S| = k \geq 2$, then $V(k, \mu_0) - V(k-1, \mu_0) \leq \frac{2}{k} V(k, \mu_0)$, or equivalently $\frac{k-2}{k} V(k, \mu_0) \leq V(k-1, \mu_0) \leq V(k, \mu_0)$.*

The $\frac{2}{k}$ 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 through our analysis 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 relies on creating alternative $k-1$ message information structures from the k-optimal information structure τ_k^* and comparing them to the $k-1$-optimal information structure τ_{k-1}^*. We observe that τ_k^* can be 'collapsed' to get an information structure with $k-1$ messages. By optimality of τ_{k-1}^* new information structures must provide weakly less utility compared to τ_{k-1}^*. We can construct k different $k-1$ dimensional information structures using this method by combining the posteriors that are in the support of τ_k^* pairwise and leaving the rest of the posteriors the same as τ_k^*. The utilities provided by these new information structures are related to $V^*(k, \mu_0)$, because they contain $k-2$ posteriors which are also in the support of τ_k^*.

We can use Proposition 3 to provide an upper and lower bound on the payoffs attainable using k messages as a function of the payoff attainable with full communication ($k = |\Omega|$) and binary communication ($k = 2$).

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 $k > 2$.*

The bounds in Corollary 1 are obtained by iteratively applying bounds in Proposition 3. This establishes a relationship between achievable payoffs with intermediate communication with the payoffs with unlimited communication and binary communication.

5.1 Belief Threshold Games

We focus on a special class of preferences in this section 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.

[16] The statement of Proposition 3 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.

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). Similar preferences are also studied by Sobel (2020) to analyze the conditions under which deception in communication will lead to loss in welfare.

We study 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 ω_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 ω_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.

Given these preferences, it is immediate that the sender can attain a payoff of 1 by using information structures with three messages.[17] 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 μ_0 for which the the default action for the receiver is the safe action. Let Δ_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.

[17] Any prior μ_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 $\mu_0(\omega_3)$.

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

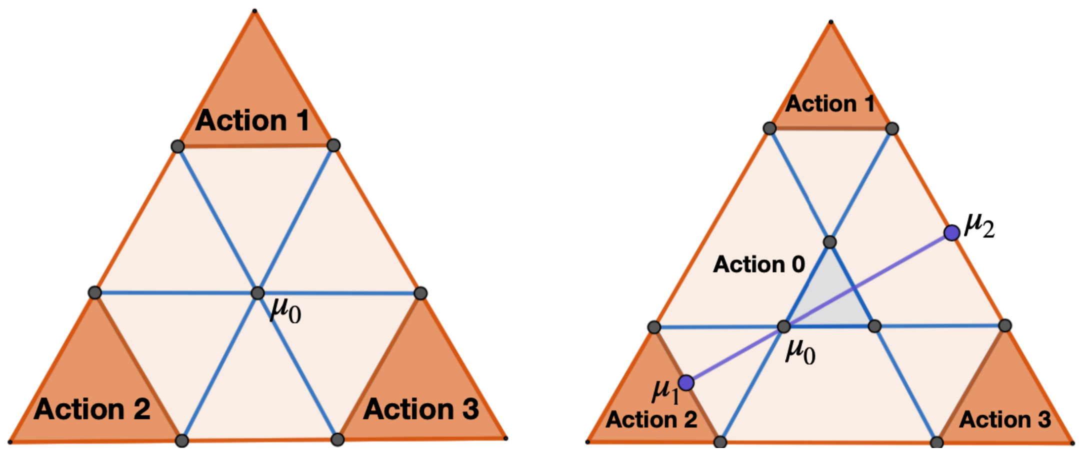

Figure 6: On the left, we have the action threshold $T = \frac{2}{3}$ so it is possible to maintain Bayes plausibility without inducing action 0 for any prior. On the right, $T > \frac{2}{3}$, so for the prior beliefs in the blue shaded region, the sender has to mix a_0 and another action when constrained to 2 messages. The blue shaded region in the right figure corresponds to Δ_T.

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 Lemma 2.

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

Using this result, we can show that depending on the location of the prior the marginal value of a message can be a function with increasing or decreasing differences.

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 a decreasing function. 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). When dealing with limited messages, the search for information structures becomes even more demanding (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.[18]

Lemma 3. *There exists an information structure τ such that $\mathsf{supp}(\tau)$ is affinely independent and solves the sender's information design problem in equation* (1).

If an information structure τ 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 $R_a = \{\mu_i \in \Delta(\Omega) : a \in \arg\max_{a' \in A} \hat{u}^R(\mu_i)\}$. 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. *$\mu_i \in R_a$ is q-extreme if it is in the interior of a q-dimensional convex set within R_a, but not in the interior any (q+1)-dimensional convex set within R_a.*

[18] Results 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.

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, and 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 4. *There exists an information structure τ 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 (of some R_a), which solves the sender's information design problem in equation* (1).

Given an information structure with at least two posteriors that are not 0-extreme there is always a way 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 that point any further rotation changes the action induced by the resulting posterior.

Using this 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 optimization 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 μ_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 $q \leq (n-k)$.

Our results in Lemmata 3 and 4 are directly related with the results of Lipnowski and Mathevet (2017). They show that the optimal information structures induce affinely independent beliefs supported at 0-extreme points of action regions R_a, under the assumption that the message space is rich. We generalize their conclusions to arbitrary message spaces.

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 $|S|$-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 $|S|$. This approach allows us to reinterpret the 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 exemplify this in models of cheap talk with state-independent beliefs (Lipnowski and Ravid, 2020) and Bayesian persuasion with heterogeneous priors (Alonso and Camara, 2016).

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).

Noise v. Coarseness. The difficulty in communication we analyze in our setting is substantively different from noisy or costly persuasion (Akyol, Langbort and Basar, 2016; Le Treust and Tomala, 2019; Tsakas and Tsakas, 2018; Bloedel and Segal, 2018; Wei, 2018; Gentzkow and Kamenica, 2016). While exogenous noise, entropy or Blackwell-informativeness costs also limit the sender's problem, these approaches still allow for arbitrarily many action recommendations, and existence results in these models still rely on having a rich message space. Our setting with a limited message space complements this line of work.

With noisy or costly communication, the sender's choice is restricted to information structures in which posteriors are not too close to the extreme points of the belief space, or inducing beliefs closer to the extreme points of the belief space gets increasingly costly.

On the other hand, with a cardinality constraint on the messages, there are no restrictions on the locations of the posteriors, but the sender has to strategically choose a limited set of directions in the belief space to convey more precise information. While the sender can induce precise posteriors in the sense that the receiver can be arbitrarily certain about the state of the world, it's never possible to perfectly inform the receiver about *all* states of the world at the same time. Unlike costly or noisy games, 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 with limited communication capacity while also maintaining Bayes plausibility. We believe that our approach captures many real-life situations where constraints on available messages are natural.

Experiment Design. Some recent papers interpret the communication procedure with commitment as the strategic design of an experiment which reveals information about the state of the world (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 project, 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. Their insights provide an interesting research avenue for extending our observations.

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.

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.".

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.

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

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. "Voronoi 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.

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

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 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*** *$\{\lambda_1, \dots, \lambda_k\} \subseteq \Lambda$ and a set of corresponding convex weights $\{\gamma_1, \dots, \gamma_k\}$ such that $\sum_{i \le k} \gamma_i = 1$ and $\forall i, 1 > \gamma_i > 0$ such that $a = \sum_{i \le k} \gamma_i a_i$. Equivalently, we can write:*

$$co_k(A) = \{a \in \mathbb{R}^n : \exists B \subseteq A,\ s.t.\ a \in co(B)\ with\ |B| \le k\}.$$

Lemma 5. *For every action $a \in A$, the set R_a is closed and convex.*

Proof of Lemma 5 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 $\{\mu \in \mathbb{R}^{|\Omega|} : \sum_{\omega \in \Omega} \mu(\omega)(u(a, \omega) - u(a', \omega)) \ge 0\}_{a' \in A}$. It is therefore closed and convex. □

Lemma 6. *The sender's utility $\hat{}\hat{u}^S$ is convex when restricted to each set R_a.*

Proof of Lemma 6 Follows directly from Volund (2018), Theorem 1 or Lipnowski and Mathevet (2017), Theorem 1. □

Appendix B Proofs of Statements in the Main Text

B.1 Proofs on Attainable Payoffs and Marginal Value

Proof of Proposition 2. Let τ be the optimal information structure solving the sender's maximization problem.

Let $\mathsf{supp}\,(\tau) = \{\mu_1, \dots, \mu_k\}$. By definition $\tau \in \mathcal{I}(k, \mu_0)$, so $\sum_{i \le k} \tau(\mu_i)\mu_i = \mu_0$ and $\sum_{i \le k} \tau(\mu_i) = 1$ and $1 \ge \tau(\mu_i) \ge 0$. Hence, using $\tau(\mu_i)$ as the convex weights and $(\mu_s, \mathbb{E}\hat{u}^S(\mu_s))$ as the points, we can how that $(\mu_0, \mathbb{E}_\tau \hat{u}^S) \in \mathbb{CH}_k(\hat{u}^S)$. We conclude that $\sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} \ge \mathbb{E}_\tau \hat{u}^S$.

Since $(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)$, there exists $\{\hat{u}^S(\mu_1), \dots, \hat{u}^S(\mu_k)\}$ and convex weights $\{\alpha_1, \dots, \alpha_k\}$ with $\sum_{i \le k} \alpha_i \mu_i = \mu_0$ and $\sum_{i \le k} \alpha_i \hat{u}^S(\mu_i) = z$. Then, $\{\mu_1, \dots, \mu_k\} \in \mathcal{I}(k, \mu_0)$. Therefore τ' could have been picked instead of τ in the sender's maximization problem, contradicting the optimality of τ. We conclude that $\sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} \le \mathbb{E}_\tau \hat{u}^S$. □

Proof of Proposition 3. Suppose τ_k is the optimal information structure with k messages, and τ_{k-1} is the optimal information structure with $k-1$ messages. Denote the utilities obtained using these information structures with $V(k, \mu_0), V(k-1, \mu_0)$

Let $\mathsf{supp}\,(\tau_k) = \{\mu_1, \dots, \mu_k\}$. We can create a $k-1$ dimensional information structure that maintains Bayes plausibility by choosing two posteriors, without loss say μ_1 and μ_2, and defining a new posterior μ_{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 $\mathsf{supp}\,(\tau'_{12}) = \{\mu_{12}, \mu_3, \dots, \mu_k\}$, and new weights $\{(\tau_k(\mu_1) + \tau_k(\mu_2)), \tau(\mu_3), \dots, \tau(\mu_k)\}$. Note that τ'_{12} maintains Bayes plausibility .

Now, we define k different information structures, each constructed the same way and containing $k-1$ posteriors, denoted $\tau_{12}, \tau_{23}, \dots, \tau_{k-1,k}, \tau_{k1}$.

By the optimally of τ_{k-1} among the information structures with $k-1$ messages, we have the following k inequalities:

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

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

$$\vdots$$

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

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

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 $\{\mu_{12}, \mu_{23}, \dots, \mu_{k-1,k}, \mu_{k1}\}$. For the largest gap possible, we set this equal to the minimum $\underline{u^S}$ 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

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

□

Proof of Lemma 2

Let $\mathsf{supp}\,(\tau) = \{\mu_1,\ldots,\mu_k\}$ be affinely dependent. Then, there must exist $\{\lambda_1,\ldots,\lambda_k\}$ such that $\sum_{i\leq k}\lambda_i = 0$ and $\sum_{i\leq k}\lambda_i\mu_i = 0$. Since τ 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 $\frac{\tau(\mu_p)}{\lambda_p}$ 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 $\forall i \neq p^*, \left(\tau(\mu_i) - \lambda_i\frac{\tau(\mu_k)}{\lambda_{p^*}}\right) \geq 0$.
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 λ_{j^*} is negative and λ_{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 λ_{j^*} is chosen to maximize this ratio and λ_{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 analogously for p^*. Therefore, we can define τ' and τ'' respectively from τ by dropping μ_{j^*} or μ_{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 $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S$ and $\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S$, we get:

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

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

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

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

Suppose $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_\tau \hat{u}^S < 0$ and $\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_\tau \hat{u}^S < 0$. This implies:

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

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

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

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

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

Replace τ 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. □

B.2 Proofs for Section 5.1

Let $(E, \vec{E})$ 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 $\vec{E}$ the space of its translations. For this example we will focus on three dimensional Euclidian affine space i.e. $\vec{E}$ has dimension 3. We equip $\vec{E}$ with Euclidean dot product as its inner product, inducing the Euclidian norm as a metric. To simplify notation, we will simply write $(\mathbb{R}^3, \vec{\mathbb{R}}^3)$. Given this structure, we can define the unitary simplex in the affine space $\mathbb{R}^3$ by the following set where ω_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:

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

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 $\mu_s \in \Delta(\Omega)$ such that $\mu_s(\omega_i) \geq T$, and prefers a_0 otherwise. Hence, we can say that for $i \in \{1, 2, 3\}$, $j \in \{0, 1, 2, 3\}$ and $j \neq i$ we have $\mathbb{E}_{\mu_s}[u^R(a_i, \omega)] \geq \mathbb{E}_{\mu_s}[u^R(a_j, \omega)]$ if and only if $\mu_s(\omega_i) > T$. 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 $\Gamma_3 = (0, 1-T, T)$. The action zones will become:

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

where $\cdot$ denotes the Euclidean dot product.

Proof of Lemma 1 Let us first characterize the set Δ_T. We have[19] $\Delta_T = \Delta(\Omega) \setminus \mathrm{co}_2(R_1 \cup R_2 \cup R_3))$. We note that:

$$\begin{aligned}\mathrm{co}(R_1 \cup R_2) =& \mathrm{co}(\{\omega_1, (T, 1-T, 0), (T, 0, 1-T), \omega_2, (1-T, T, 0), (0, T, 1-T)\}) \\ =& \mathrm{co}\{\omega_1, (T, 0, 1-T), \omega_2, (0, T, 1-T)\}\end{aligned} \tag{2}$$

[19] co denotes convex hull operator and co_k denotes k-convex hull i.e. $\mathrm{co}_k(A)$ are the points that can be represented as convex combination of k elements in A.

and similarly for $\text{co}(R_1 \cup R_3)$ and $\text{co}(R_2 \cup R_3)$ we have that

$$\text{co}(R_1 \cup R_3) = \text{co}\{\omega_1, (T, 1-T, 0), \omega_3, (0, 1-T, T)\} \tag{3}$$
$$\text{co}(R_2 \cup R_3) = \text{co}\{\omega_2, (1-T, 0, T), \omega_3, (1-T, 0, T)\} \tag{4}$$

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 $\text{co}(\{\omega_1, (T, 1-T, 0), (T, 0, 1-T), \omega_2, (1-T, T, 0), (0, T, 1-T)\})$. Similarly using equation (2), (3) and (4), $\text{co}(R_i \cup R_j)$ can be identified as the intersection of a half space and the simplex i.e.

$$\text{co}(R_1 \cup R_2) = \Delta(\Omega) \cap \{(\mu - (T, 0, 1-T)) \cdot (-T, T, 0) \geq 0 | \mu \in \mathbb{R}^3\} \tag{5}$$
$$\text{co}(R_1 \cup R_3) = \Delta(\Omega) \cap \{(\mu - (T, 1-T, 0)) \cdot (-T, 0, T) \geq 0 | \mu \in \mathbb{R}^3\} \tag{6}$$
$$\text{co}(R_2 \cup R_3) = \Delta(\Omega) \cap \{(\mu - (1-T, T, 0)) \cdot (0, -T, T) \geq 0 | \mu \in \mathbb{R}^3\} \tag{7}$$

So we can define $\Delta_T \subset \Delta(\Omega)$ as $\Delta_T = \Delta(\Omega) \setminus \text{co}_2(R_1 \cup R_2 \cup R_3)$. By (5), (6) and (7) we can see that Δ_T is defined as

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

By definition of Δ_T and $\Delta(\Omega)$ this set is non-empty if and only if $T > \frac{2}{3}$. □

Proof of Lemma 2 We can identify the upper bounds through the following problem:

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

First note that by the symmetry of the problem choice of i is not relevant. Without loss of generality we pick $i = 1$. Moreover, the constraint that $\mu_0 \in \text{co}(\mu_i, \mu_4)$ implies that we are searching for a point with the goal of minimizing the distance with μ_i and maximizing the distance with μ_4. The maximizing triple is therefore $(\mu_0^*, \mu_1^*, \mu_4^*)$ with $\mu_0^* = (1-T, 1-T, 2T-1)$, $\mu_1^* = (\frac{1-T}{2}, \frac{1-T}{2}, T)$ $\mu_4^* = (0, \frac{1}{2}, \frac{1}{2})$. The solution follows from two observations. One is that given two points μ_0 and μ_i there is a unique line passing through these points hence μ_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 μ_0 and μ_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:

$$||(T, \frac{1-T}{2}, \frac{1-T}{2}) - (2T-1), 1-T, 1-T|| = \frac{\sqrt{6}}{2}(1-T)$$
$$||(T, \frac{1-T}{2}, \frac{1-T}{2}) - (0, \frac{1}{2}, \frac{1}{2}))|| = \frac{\sqrt{6}}{2}T$$

Giving us that $\overline{V(2, \mu_0)} = \frac{2T-1}{T}$. Similarly, we can solve:

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

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 μ_0 is closer to one of the action zones. Minimizing the objective, we pick $\mu_0^* = B$. Now given this choice, we choose μ_4 to maximize leading to the choice of $\mu_4^* = (0, \frac{1}{2}, \frac{1}{2})$ and $\mu_1^* = (\frac{1-T}{2}, \frac{1-T}{2}, T)$.

Interestingly, the posteriors induced in the optimal information structure for the two problems are the same, but they are induced with different probabilities. This follows from the fact that the hyperplanes defining the action zones is parallel to one of the hyperplanes defining the simplex. So we can write $\underline{V(2, \mu_0)} = \frac{1}{3T}$. □

Proof of corollary 2 Observe that with fixed $T = 2/3$, we have $\overline{V(2, \mu_0)} = \frac{1}{2} = \underline{V(2, \mu_0)}$. Also, $\overline{V(2, \mu_0)} = \frac{2T-1}{T}$ is increasing in T and $\underline{V(2, \mu_0)} = \frac{1}{3T}$ is decreasing in T. By continuity of distance, the objective function in the definition of $\overline{V(2, \mu_0)}$ and $\underline{V(2, \mu_0)}$ are continuous. So for any other $\mu_0 \in \Delta_T$, $V(2, \mu_0)$ takes every value between $\underline{V(2, \mu_0)}$ and $\overline{V(2, \mu_0)}$ 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. □

B.3 Proofs on Properties of Optimal Information Structures

Definition 3. *The affine hull aff (S) of S is the set of all affine combinations of elements of S, that is,*

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

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

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 Ω 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 3. Let $\mathsf{supp}\,(\tau) = \{\mu_1, \ldots, \mu_k\}$ be affinely dependent. Then, there must exist $\{\lambda_1, \ldots, \lambda_k\}$ such that $\sum_{i\leq k} \lambda_i = 0$ and $\sum_{i\leq k} \lambda_i\mu_i = 0$. Since τ 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 $\frac{\tau(\mu_p)}{\lambda_p}$ 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 $\forall i \neq p^*, \left(\tau(\mu_i) - \lambda_i\frac{\tau(\mu_k)}{\lambda_{p^*}}\right) \geq 0$.

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 λ_{j^*} is negative and λ_{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 λ_{j^*} is chosen to maximize this ratio and λ_{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 analogously for p^*. Therefore, we can define τ' and τ'' respectively from τ by dropping μ_{j^*} or μ_{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 $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S$ and $\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S$, we get:

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

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

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

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

Suppose $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S < 0$ and $\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S < 0$. This implies:

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

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

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

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

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

Replace τ 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 4. Suppose $\mu = \{\mu_1, \ldots, \mu_k\}$ is an information structure, and without loss of generality, let μ_1, μ_2 be posteriors that are not 0-extreme points of any action region R_a. 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 $\sum_{i=1}^{k} \tau(\mu_1)\mu_i = \mu_0$, for the given prior μ_0. We can rearrange the Bayes plausibility condition and write:

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

Denoting $\tau(\mu_1) + \tau(\mu_2) = \hat{\tau}_{12}$, $\frac{\tau(\mu_1)}{\hat{\tau}_{12}} = \hat{\tau}_1$, $\frac{\tau(\mu_2)}{\hat{\tau}_{12}} = \hat{\tau}_2$, and $\frac{\tau(\mu_1)\mu_1 + \tau(\mu_2)\mu_2}{\tau(\mu_1) + \tau(\mu_2)} = \hat{\mu}_{12}$, we note that we can replace μ_1, μ_2 with μ_1', μ_2' and still maintain Bayes plausibility if the following condition is satisfied:

$$\alpha\mu_1' + (1 - \alpha)\mu_2' = \hat{\mu}_{12}, \ \textit{for some } \alpha \in (0, 1).$$

The new information structure $\mu' = \{\mu_1', \mu_2', \mu_3 \ldots, \mu_k\}$ will be Bayes plausible with the weights $\tau'(\mu_1') = \alpha\hat{\tau}_{12}, \tau'(\mu_2') = (1 - \alpha)\hat{\tau}_{12}$, and $\tau'(\mu_i) = \tau(\mu_i)$ for $i > 2$. Since we know μ_1, μ_2 are (at least) 1-extreme points, there exists line segments $A_1 \subset R_1, A_2 \subset R_2$ and μ_1, μ_2 are in the relative interior of A_1, A_2 respectively.

Now, let us choose μ_1'', μ_2' that satisfy the following condition:

$$\frac{2\hat{\tau}_1 - 1}{\hat{\tau}_1 - \hat{\tau}_2}\mu_1 + \frac{2\hat{\tau}_2 - 1}{\hat{\tau}_1 - \hat{\tau}_2}\mu_2 = \mu_1'' - \mu_2'. \tag{8}$$

With any μ_1'', μ_2' that satisfies the above condition, we can calculate the corresponding μ_1', μ_2'' such that:

$$\hat{\tau}_1\mu_1' + \hat{\tau}_2\mu_1'' = \mu_1,$$
$$\hat{\tau}_1\mu_2' + \hat{\tau}_2\mu_2'' = \mu_2.$$

Moreover, $\mu_1', \mu_1'', \mu_2', \mu_2''$ will satisfy:

$$\hat{\mu}_{12} = \hat{\tau}_1\mu_1' + \hat{\tau}_2\mu_2',$$
$$\hat{\mu}_{12} = \hat{\tau}_1\mu_1'' + \hat{\tau}_2\mu_2''.$$

There will be infinitely many possible pairs (μ_1'', μ_2') that satisfy equation 8, but let us pick an arbitrary pair that are within a sufficiently close radius of μ_1, μ_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 μ_1, μ_2 are 1-extreme points, hence relative interior points of the line segments A_1, A_2, we can find such ϵ, δ. 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:

$$\{\mu_1', \mu_1^{''}\} \subset (A_1 \cap B_\epsilon(\mu_1)) \subset R_1,$$

$$\{\mu_2', \mu_2^{''}\} \subset (A_2 \cap B_\delta(\mu_2)) \subset R_2,$$

$$\nabla_{(\mu_1^{''}-\mu_1')}\hat{u}^s(\mu_1) = \nabla_{(\mu_1^{''}-\mu_1')}\hat{u}^s(\mu_1') = \nabla_{(\mu_1^{''}-\mu_1')}\hat{u}^s(\mu_1^{''}) = \theta,$$

$$\nabla_{(\mu_2^{''}-\mu_2')}\hat{u}^s(\mu_2) = \nabla_{(\mu_2^{''}-\mu_2')}\hat{u}^s(\mu_2') = \nabla_{(\mu_2^{''}-\mu_2')}\hat{u}^s(\mu_2^{''}) = \gamma,$$

where γ and θ are the directional derivatives of $\hat{u}_s$ in the directions $(\mu_2^{''} - \mu_2')$, $(\mu_1^{''} - \mu_1')$ 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, μ is an optimal information structure, so we must have:

$$\hat{\tau}_1\hat{\tau}_{12}\hat{u}^S(\mu_1') + \hat{\tau}_2\hat{\tau}_{12}\hat{u}^S(\mu_2') + \bar{u} \leq \tau(\mu_1)\hat{u}^S(\mu_1) + \tau(\mu_2)\hat{u}^S(\mu_2) + \bar{u},$$

$$\hat{\tau}_1\hat{\tau}_{12}\hat{u}^S(\mu_1^{''}) + \hat{\tau}_2\hat{\tau}_{12}\hat{u}^S(\mu_2^{''}) + \bar{u} \leq \tau(\mu_1)\hat{u}^S(\mu_1) + \tau(\mu_2)\hat{u}^S(\mu_2) + \bar{u}$$

$$\iff$$

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

$$\hat{\tau}_1\hat{u}^S(\mu_1^{''}) + \hat{\tau}_2\hat{u}^S(\mu_2^{''}) \leq \hat{\tau}_1\hat{u}^S(\mu_1) + \hat{\tau}_2\hat{u}^S(\mu_2).$$

$$\iff$$

$$\hat{\tau}_1/\hat{\tau}_2 \left(\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)\right) \leq \left(\hat{u}^S(\mu_2) - \hat{u}^S(\mu_2')\right),$$

$$\hat{\tau}_1/\hat{\tau}_2 \left(\hat{u}^S(\mu_1^{''}) - \hat{u}^S(\mu_1)\right) \leq \left(\hat{u}^S(\mu_2) - \hat{u}^S(\mu_2^{''})\right).$$

Now, by the convexity of $\hat{u}^S$ within each action region, $\left(\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)\right)$ and $(\hat{u}^S(\mu_1^{''}) - \hat{u}^S(\mu_1))$ can't both be negative. Similarly, $\left(\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)\right)$ and $\left(\hat{u}^S(\mu_1^{''}) - \hat{u}^S(\mu_1)\right)$ can't both be positive, since it would imply that $\left(\hat{u}^S(\mu_2) - \hat{u}^S(\mu_2')\right)$ and $\left(\hat{u}^S(\mu_2) - \hat{u}^S(\mu_2^{''})\right)$ 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 $\left(\hat{u}^S(\mu_1') - \hat{u}^S(\mu_1)\right)$ is positive and $\left(\hat{u}^S(\mu_1^{''}) - \hat{u}^S(\mu_1)\right)$ is negative. This implies $\left(\hat{u}^S(\mu_2) - \hat{u}^S(\mu_2')\right)$ must also be positive. Therefore, $\left(\hat{u}^S(\mu_2) - \hat{u}^S(\mu_2^{''})\right)$ 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 $\mu_1', \mu_1^{''}, \mu_2', \mu_2^{''}$:

$$\hat{\tau}_1/\hat{\tau}_2 \left(\hat{\tau}_2\theta \cdot (\mu_1' - \mu_1^{''})\right) \leq \gamma \cdot (\hat{\tau}_2(\mu_2'' - \mu_2')),$$

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

$$\Longleftrightarrow$$

$$\hat{\tau}_1 \left(\theta \cdot (\mu_1' - \mu_1^{''})\right) \leq \hat{\tau}_2 \left(\gamma \cdot (\mu_2'' - \mu_2')\right),$$

$$\hat{\tau}_1 \left(\theta \cdot (\mu_1^{''} - \mu_1')\right) \leq \hat{\tau}_2 \left(\gamma \cdot (\mu_2' - \mu_2'')\right).$$

$$\Longleftrightarrow$$

$$\hat{\tau}_1 \left(\theta \cdot (\mu_1' - \mu_1^{''})\right) = \hat{\tau}_2 \left(\gamma \cdot (\mu_2'' - \mu_2')\right).$$

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 τ with supp$(\tau) = \mu = \{\mu_1, \ldots, \mu_k\}$ satisfying Bayes plausibility. If there exists a posterior in supp(τ) where $\mu_a \in R_a$ such that μ_a is a q-extreme points 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 3, 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 Ω are represented in the $(n-1)$ dimensional space. Let μ_1 be a $q-$extreme point of R_1 with $q \geq (n-k)$. In other words, μ_1 is in the interior of a q-dimensional convex set S within R_1, but there is no $q+1$ dimensional convex set within R_1 such that μ_1 is an interior point.

Since R_1 is a polyhedron, μ_1 belongs to the interior of a q-dimensional face of R_1. Moreover, μ_1 belongs to μ, which consists of k affinely independent points, so it belongs to the $(k-1)$-dimensional affine surface M which consists of the affine hull of μ. Since μ_1 belongs to a q-dimensional face of R_1, by definition, there is a unique q-dimensional affine surface S containing this face. Additionally, M is $(k-1)-$dimensional, and S is at least $n-k+1$ dimensional by definition, their intersection $S \cap M$ is non-empty and includes μ_1 by construction and it is at least 1 dimensional (since $\underbrace{n-k+1}_{\dim S} + \underbrace{k-1}_{\dim M} = n > n-1$).

We can find a radius ε small enough such that $B_\varepsilon(\mu_1) \cap (S \cap M \cap R_1) \neq \emptyset$, and within this intersection a line segment , since $S \cap M$ is at least 1 dimensional. We can find two points from this line segment μ_1', μ_1'' such that μ_1 is a convex combination of μ_1', μ_1'' with $(\alpha)\mu_1' + (1-\alpha)\mu_1'' = \mu_1$.

Therefore we can 'split' μ_1 into μ_1', μ_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:

$$\tau(\mu_1)((\alpha)\hat{u}^s(\mu_1') + (1-\alpha)\hat{u}^s(\mu_1'')) + \sum_{i=2}^{k} \tau(\mu_i)\hat{u}^s(\mu_i) \geq$$

$$\tau(\mu_1)\hat{u}^s(\mu_1) + \sum_{i=2}^{k} \tau(\mu_i)\hat{u}^s(\mu_i),$$

by convexity of $\hat{u}^s$ within R_1.

Since $\tilde{\mu}$ consists of $k+1$ points belonging to a $k-1$ dimensional affine surface, it cannot be affinely independent. Then, using Lemma 3, 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 $\mu_{-k} = (\mu_1, \ldots, \mu_{k-1})$.

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

If $\mu_0 \notin \mathrm{co}(\mu_{-k})$, we can define the set of μ_k such that for $\mu = (\mu_{-k}, \mu_k)$ we get that $\mu_0 \in \mathrm{co}(\mu)$.

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

By the definition of the set M, 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 $\mu = (\mu_{-k}, \mu_k)$ is not affinely independent, then we can drop some posteriors from $\tilde{\mu}$ using the protocol described in Lemma 3 and obtain an affinely independent information structure. Moreover, we know $\tilde{\mu} \neq \mu_k$ since $\mu_0 \notin \mathrm{co}(\mu_{-k})$ violating Bayes plausibility.

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

Now we turn to the question of choosing μ_k. First note that M 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, $S = M \cap \Delta(\Omega)$ 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 $\mathcal{R}_a = R_a \cap S$ is a convex polytope of dimension at most k.

We will now show that when we are choosing μ_k which must lie in a set $\mathcal{R}_a$, 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 μ_k be a q-extreme point for $q > 0$. We can now proceed analogously to proof of Lemma 4 and find a ϵ-ball around μ_k that will stay inside S and $\mathcal{R}_a$. Our assumption on μ_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 μ_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 μ_k', μ_k'' such that μ_k is a convex combination of μ_k', μ_k'' with $(\alpha)\mu_k' + (1-\alpha)\mu_k'' = \mu_k$. Again following the same line of argument with Lemma 4, 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 μ_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 $\{R_a\}_{a\in A}$, the choice of the k^{th} point has finitely many candidates identified as the 0-extreme points of the sets $\{\mathcal{R}_a\}_{a\in A} = \{R_a \cap S\}_{a\in A}$. 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 $\{\mathcal{R}_a\}_{a\in A} = \{R_a \cap S\}_{a\in A}$.

Appendix C Optimal Compressions

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 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.*

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 $T = \left\{\mu \in \mathbb{R}^n | \mu = \mu_0 + \sum_{i=1}^{k} \alpha_i \tilde{\mu}_i\right\} \subset \mathbb{R}^n$.

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 $\mathcal{T}_k$, 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 7.

$$\max_{\tau} \mathbb{E}_{\mu\sim\tau} \hat{u}^s(\mu_i)| \text{ subject to } \mathbb{E}_{\mu\sim\tau}\mu = \mu_0 \ , \ |\mathsf{supp}\,(\tau)| \leq k \tag{9}$$

achieves the same optimal value with the problem:

$$\max_{T\in\mathcal{T}_k} \max_{\tau} \mathbb{E}_{\mu\sim\tau} \hat{u}^s(\mu_i)|_T \text{ subject to } \mathbb{E}_{\mu\sim\tau}\mu = \mu_0 \ , \ |\mathsf{supp}\,(\tau)| \leq k \text{ and } \mathsf{supp}\,(\tau) \subset T_k \tag{10}$$

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 $\mathcal{T}_k$.

Lemma 8. *$\mathcal{T}_k$ is a compact smooth manifold. Moreover, $T \in \mathcal{T}_k$ can be represented with the projection matrix of its parallel subspace $W =$ **span**$(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$.*

Proof. $\mathcal{T}_k$ is homemorphic 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 $\mathcal{G}_k(\mathbb{R}^n)$. The Grassmannian $\mathcal{G}_k(\mathbb{R}^n)$ 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 μ_0 from each line equation.

Define the Steifel manifold $\mathcal{V}_k(\mathbb{R}^n)$ as the set of all orthonormal k-frames of $\mathbb{R}^n$.[20] Hence, elements of $\mathcal{V}_k(\mathbb{R}^n)$ are k-tuples of orthonormal vectors in $\mathbb{R}^n$. $\mathcal{V}_k(\mathbb{R}^n)$ is identified with a subset of the cartesian product of k many $(n-1)$ spheres $\mathbb{S}^{n-1} = \{x \in \mathbf{R}^n : \|x\| = 1\}$. It is an immediate observation that $\left(\mathbb{S}^{n-1}\right)^k$ a closed subspace of a compact space. So, we can easily conclude the Steifel manifold $\mathcal{V}_k(\mathbb{R}^n)$ is compact in the inherited topology from $R^{n\times k}$

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

Now we will show that $T \in \mathcal{T}_k$ can be represented with the projection matrix of its parallel subspace $W = \mathbf{span}(\tilde{\mu}_1, \ldots, \tilde{\mu}_k)$. Consider the set of real $n \times n$ matrices $\mathcal{X}_k(n)$ that are (i)

[20] 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

idempotent, (ii) symmetric and (iii) have rank k. The requirement that a matrix $X \in \mathcal{X}_k(n)$ has rank k is equivalent to requiring X has trace k.[21]

To prove the second claim, it suffices to define a homeomorophism between $\mathcal{X}_k(n)$ and $\mathcal{G}_k(\mathbb{R}^n)$. The homeomorphism ϕ is $\phi(X) = C(X)$, $\phi : \mathcal{X}_k(n) \to \mathcal{G}_k(\mathbb{R}^n)$ 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 $d_{\mathcal{G}_k(\mathbb{R}^n)}(W, W') = \|X_W - X_{W'}\|$ where$\|\cdot\|$ is the operator norm, that metrizes $\mathcal{G}_k(n)$. $\square$

We call the projections from $\Delta(\Omega)$ onto the flat $T \in \mathcal{T}_k$ a *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:

Lemma 9. *The optimal value of the maximization problem:*

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

is upper semi-continous in T.

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 τ_T be the optimal information structure with support μ_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 $\forall \epsilon > 0$, 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 $(\mathbb{E}_{\mu_i \sim \tau} \hat{u}^s(\mu_i))$ is upper semi-continuous in μ. So for any ε, there exists a δ_ϵ such that whenever $||\mu - \mu'|| < \delta_\epsilon$, we get $V(\mu') \leq V(\mu) + \epsilon$. 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 M_T and $M_{T'}$ be the projection matrices corresponding to parallel subspace consisting of vectors $\{m_T \mid m_T \in T \cap \mathbf{Bd}\Delta(\Omega)\}$ and $\{m_{T'} \mid m_{T'} \in T' \cap \mathbf{Bd}\Delta(\Omega)\}$. We will

[21] 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.

show:

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

We start by showing that $||(X_T - X_{T'})\tilde{\mu}|| \geq \gamma||M_T - M_{T'}||$. First, by definition of matrix norm $||(X_T - X_{T'})\tilde{\mu}|| \geq ||M_T - M_{T'}||_{\max} = \max_{r \in R} ||m_T^r - m_{T'}^r||_2$. By equivalence of finite dimensional norms, there exists a constant γ such that $||M_T - M_{T'}||_{\max} \geq \gamma||M_T - M_{T'}||$. Hence, we obtain that $||(X_T - X_{T'})\tilde{\mu}|| \geq \gamma||M_T - M_{T'}||$.
Now let us turn to the last inequality $\gamma||M_T - M_{T'}|| \geq \gamma||\mu_T - \mu_{T'}||$. This follows by making μ_0 the origin via subtracting μ_0 i.e. $M_T - \mu_0$, $M_{T'} - \mu_0, \mu_T - \mu_0, \mu_{T'} - \mu_0$ in $\mathbb{R}^N$ and noticing that for u and v in $\mathbb{R}^N$ $||\alpha u - \beta v||$ is monotone in α and β.
Recall that, $(\mathbb{E}_{\mu_i \sim \tau} \hat{u}^s(\mu_i))$ is upper semi-continuous in μ. So for any ε, there exists a δ_ϵ such that whenever $||\mu - \mu'|| < \delta_\epsilon$, we get $V(\mu') \leq V(\mu) + \epsilon$. 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 semicontinutity of $V(T)$ i.e. $\forall \epsilon > 0$, 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 7 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 7, it is straightforward to show that the two maximization problems yield the same maximum.

Let τ_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 aff denotes affine hull, and $\tau = \tau_1$ to get a higher value, contradicting the optimality of μ_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 τ_1 in the first problem. □

Appendix D Cheap Talk with Transperant 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 committ to a messaging strategy π. Second, the sender's utility is independent of the state but only depends on the action taken i.e. $u^S : A \to \mathbb{R}$, and hence their paper is titled 'Cheap Talk with Transparent Motives.' Finally, in they assume rich message spaces: $|S| \geq |\Omega|$. We consider a variation of their model where only the last assumption is changed to $|S| < |\Omega|$.

Throughout this appendix, we will focus on the Perfect Bayesian Equilibria - hereinafter referred as the *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 $\pi : \Omega \to \Delta(S)$; a receiver strategy $\hat{a} : S \to \Delta A$; and a belief system for the receiver $\mu_s : S \to \Delta\Omega$; such that:

1. μ_s is obtained from μ_0, given π, using Bayes's rule;
2. $\hat{a}(s)$ is supported on $\arg\max_{a \in A} \int_{\Omega} u_R(a, \cdot) \mathrm{d}\beta(\cdot \mid s)$ for all $s \in S$; and
3. $\pi(\omega)$ is supported on $\arg\max_{s \in S} \int_{A} u_S(\cdot) \mathrm{d}\rho(\cdot \mid s)$ 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 $\mathcal{I}(k, \mu_0) = \{\tau \in \Delta(\Delta(\Omega)) | \int \mu d\tau(\mu) = \mu_0 \text{ and } |\operatorname{supp}(\tau)| \leq k\}$.

Let the sender's possible continuation values from the receiver having μ as his posterior be defined with the correspondence $V(\mu) := \operatorname{co} u^S \left(\arg\max_{a \in A} \int u^R(a, \cdot) \mathrm{d}\mu\right)$.[22] Aumann and Hart (2003) and Lipnowski and Ravid (2020) show that an outcome (τ, z) is an equilibrium outcome if and only if it holds that (i) $\tau \in \mathcal{I}(\mu_0)$, and (ii) $z \in \bigcap_{\mu \in \operatorname{supp}(\tau)} V(\mu)$.

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

Lemma 10. *(τ, z) is an equilibrium outcome if and only if: $\tau \in \mathcal{I}_k(\mu_0)$ and $z \in \bigcap_{\mu \in \operatorname{supp}(\tau)} V(\mu)$.*

[22]This 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.

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 τ 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 and Ravid (2020), when $|S| \leq k$ an equilibrium inducing sender payoff z exists if and only if z is k-securable.

Using this equilibrium characterization via k-securablity, we can state that a sender-preferred equilibrium exists and the payoff of the sender in this equilibrium can be characterized by $V_k^*(\cdot) := \max_{\tau\in\mathcal{I}_k(\cdot)} \inf V(\text{supp}\,\tau)$. 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 τ 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 the quasiconcave envelope of sender's value function $V(\mu)$. This means that it is the the pointwise lowest quasi-concave and upper semi-continuous function that majorizes v.[23]

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

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

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

[23] Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we say that $\mathbf{x}$ weakly majorizes (or dominates) $\mathbf{y}$ from below (or equivalently, we say that $\mathbf{y}$ is weakly majorized (or dominated) by $\mathbf{x}$ from below) denoted as $\mathbf{x} \succ_w \mathbf{y}$ if $\sum^k x_{(i)} \geq \sum_{i=1}^k y_{(i)}$ for all $k = 1, \ldots, n$.

problem over the 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 7. To see the equivalence of the maximization problem in Lipnowski and Ravid (2020) with the $V_k^* = \max_{T_k \in \mathcal{T}_k} \left(\max_{\tau \in T_k} \left(\min_{\mu \in \mathsf{supp}\, \tau} \mathbb{E}_{\omega \sim \mu} u^S(\mu) \right) \right)$, it suffices to show that

$$\max_{\tau} \min_{\mu \in \mathsf{supp}\,(\tau)} \mathbb{CH}_k(\hat{u}^s)(\mu) \text{ subject to } \mathbb{E}_{\mu \sim \tau} \mu = \mu_0$$

is equivalent to

$$\max_{T_k \in \mathcal{T}_k} \max_{\tau \in T_k} \min_{\mu \in \mathsf{supp}\, \tau} \mathbb{CH}(\hat{u}^s)(\mu) \text{ subject to } \mathbb{E}_{\mu \sim \tau} \mu = \mu_0.$$

Existence for the first maximum problem follows from existence results in Lipnowski and Ravid(2020) and the fact that $\{\tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_{\mu \sim \tau} \mu = \mu_0 \text{ and } |\mathsf{supp}\, \tau| \leq k\}$ is a closed subset of $\{\tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_{\mu \sim \tau} \mu = \mu_0\}$. The equivalence follows from lemma 7 proven above. First it is already shown that $\mathcal{T}_k$ is compact, and secondly $\max_{\tau \in T_k} \min_{\mu \in \mathsf{supp}\, \tau} \mathbb{CH}(\hat{u}^s)(\mu)$ is upper semicontinuous due to upper semi-continuity of $\hat{u}^s$. □

Appendix E 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 μ_0^s be the sender's prior, and μ_0^r be the receiver's prior. We will adopt the perspective of the sender. For any posterior belief μ_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 μ_k.

Re-defining the expected sender utility to reflect heterogeneity in priors, we can write $\hat{u}_t^S(\mu_k) = \mathbb{E}_{\omega \sim \mu_k} u^s(\hat{a}(t(\mu_k), \omega)$, mindful of the fact that when the sender's posterior is μ_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

$$\max_{\tau\in\Delta(\Delta(\Omega))} \mathbb{E}_{\mu_s\sim\tau}\hat{u}_t^S(\mu_s) \text{ subject to } |\mathsf{supp}\,(\tau)| \leq k \text{ and } \mathbb{E}_\tau(\mu_s) = \mu_0^S. \tag{11}$$

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 μ_0^S will be equal to the k-concavification of the perspective transformed-sender utility function $V_k(\mu_0^s) = \sup\{z|(\mu_0, z) \in \mathbb{CH}_k(\hat{u}_t^S)\}$.

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.