Persuasion with Coarse Communication*

Yunus C. Aybas[†] ⓡ Eray Turkel[‡]

June 27, 2023

Abstract

How does an expert's ability persuade change with the availability of messages? We study games of Bayesian persuasion the sender is unable to fully describe every state of the world or recommend all possible actions. We characterize the set of attainable payoffs. Sender always does worse with coarse communication and values additional signals. We show that there exists an upper bound on the marginal value of a signal for the sender. In a special class of games, the marginal value of a signal is increasing when the receiver is difficult to persuade. We show that an additional signal does not directly translate into more information and the receiver might prefer coarse communication. Finally, we study the geometric properties of optimal information structures. Using these properties, we show that the sender's optimization problem can be solved by searching within a finite set.

Keywords: Bayesian Persuasion; Information Design; Coarse Communication; Information Structures

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, Cem Tutuncu, Robert Wilson, Kemal Yildiz, Weijie Zhong, seminar participants in Stanford University, Bilkent University and Econometric Society Meetings in 2020 and 2021 for helpful comments. Author names are randomized via AEA author randomization tool. Randomization can be verified at `https://www.aeaweb.org/journals/policies/random-author-order/search?RandomAuthorsSearch%5Bsearch%5D=XlTLPbzQO1gt`.

[†]Stanford University, Department of Economics. aybas@stanford.edu

[‡]Stanford University, Graduate School of Business. eturkel@stanford.edu

1 Introduction

Language is a coarse and imperfect tool, especially when the topic is complicated. Credit rating agencies use grades to describe the riskiness of a financial asset to their clients. Schools use letter grade schemes to summarize the performance of students to potential employers. Governmental agencies rate the hygiene practices of restaurants or the energy efficiency of electrical appliances using grades to provide information to consumers. In all of these examples, agents communicate about a complex subject by using a coarse set of signals.

In some settings coarseness is unavoidable, while in others it is a choice. A patient interacting with a doctor might ask for advice on whether they should go through with a specific treatment. There may be multiple actions that they do not wish to consider, such as a surgical procedure. In general, when an agent is asking for advice from an expert, they might find it beneficial to limit the expert's communication capacity, and what actions the expert can recommend. This is especially important when the preferences are misaligned. Limits on communication can also be imposed by a social planner or a regulator, when the parties involved lack such power. For example, in settings where firms use advertising to send product information, a regulator might find that limiting the targeting capability of advertisers is welfare improving for consumers.

In this paper, we study coarseness in communication in settings where the sender has commitment power. We use the Bayesian persuasion framework to model such interactions. Our purpose for this study is not to focus on why coarseness arises, but instead to understand how it affects communication, and how the resulting interaction compares with the rich communication setting. Coarseness naturally arises in many real life interactions due to technological, institutional or regulatory constraints. These settings cannot be fully captured using the standard Bayesian persuasion model, which assumes access to a large set of signals that can describe every state of the world. Our work also provides a natural extension of the current literature, by showing how existing techniques extend to these environments.

In the analysis of canonical games of Bayesian persuasion, existence proofs and solution techniques rely on having access to rich signal spaces.[1] The sender can induce as many actions as they want, and the only restriction is that a convex combination of the posterior beliefs generated is equal the prior. In our setting, the convex combinations must involve a limited number of posteriors. Focusing on these constrained convex combinations, we extend the concavification approach used in the literature to characterize the set of attainable sender

[1]These results generally leverage the extremal representation theorems of Caratheodory and Krein-Millman (Kamenica and Gentzkow 2011), which does not directly apply to our setting.

payoffs. The resulting function allows us to visualize how achievable sender utilities change as a function of the prior belief, and as a function of the cardinality of the signal space.

We then focus on the marginal value of a signal for the sender. In the applications of our model, this value is of particular economic interest. In the leading example we study, it corresponds to the value of increasing the targeting ability of an advertiser relaying product information to potential customers. We prove an upper bound on the marginal value of a signal that applies for any game of persuasion. The result is derived through a novel insight linking the sender's optimal signals with finer and coarser communication. Our result implies that in settings with large state and action spaces, the marginal value of a signal becomes a very small fraction of the rich communication payoff as we increase the size of the signal space. In other words, having access to more signals cannot change the sender's payoff by a large amount as we approach rich communication. However, this does not imply that the marginal value of a signal is necessarily a decreasing function in all persuasion games.

We provide a detailed analysis of the marginal value of signals under a general preference structure which we call belief-threshold games. In these settings, the receiver has a different preferred action for every state, taken only if their posterior belief for that state is above a certain threshold. There is a default action taken under the prior when none of the thresholds are met, which is the sender's least preferred action. The sender's payoff does not depend on the state and only depends on the receiver's action, and the threshold values represent the difficulty of convincing the receiver to take an action. These preferences capture many economic settings that have been the focus of previous work, 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 these games, we show that the marginal value of a signal is increasing for 'skeptical' priors that are maximally distant from the belief thresholds, and decreasing for priors that are already close to one of the belief thresholds. With a constrained signal space and maximally skeptical priors, the sender can satisfy Bayes plausibility only when they induce their least preferred action. If the prior is further away from the thresholds, the probability of inducing the least preferred action has to increase, which implies that the value of an additional signal will be higher.

The remainder of our contributions are theoretical in nature. We provide a simple proof showing the existence of an optimal strategy for the sender in coarse communication settings, and the tools to constructively find solutions of coarse persuasion games. We characterize the properties of the solutions of these games using techniques from affine geometry: one of our results shows that the search for optimal information structures can be restricted to affinely independent posteriors. The result is based on the observation that inducing affinely dependent posteriors proves to be 'inefficient' for the sender, even when there are no explicit costs associated with using more signals. We show that when inducing an action, the sender prefers to generate the most extreme beliefs possible through their signals. We capture extremeness of beliefs by studying different degrees of extreme points of a set. Using this property, we show that the search for an optimal solution can be restricted to a finite set. From this, we derive a finite procedure to find a sender-optimal information structure, which applies to both coarse and rich communication settings.

Finally, we show that the tools we develop can be extended and used in other settings involving strategic communication, such as cheap talk games with state-independent preferences. Our framework therefore opens many avenues for future research, and can be used to analyze how the sender's value for increased communication capacity depends on their commitment power and the level of disagreement in prior beliefs.

1.1 Related 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 et al. 2011; Hagenbach and Koessler 2020). The main difference which separates our work from this line of research is the potential for misaligned preferences between the sender and the receiver, and the sender's ability to commit to a signaling scheme (Kamenica and Gentzkow 2011).

Some recent papers interpret this 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. In this framework, sender publicly designs a Blackwell experiment. Then, the receiver sees the realization of the experiment and chooses an action. This line of work generally assumes that the sender can design any experiment. This assumption is recently challenged by Ball and Espín-Sánchez (2021) and Ichihashi (2019).

Ball and Espín-Sánchez (2021) studies a model of experimental persuasion, where sender has access to a feasible set of experiments and can commit to garble the outcomes. Our model can be thought as a scenario where the sender has access to experiments with only a limited set of outcomes. 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. Ichihashi (2019) studies a persuasion game in which the receiver can limit the Blackwell informativeness of the information structures. We show that optimal information structures under different cardinality constraints are not always Blackwell comparable. Hence, cardinality and Blackwell informativeness constraints lead to different outcomes in general.

Another stream of literature introduces a cost for generating more precise information structures e.g. Gentzkow and Kamenica (2014) assume that the costs are proportional to the reduction in the entropy of prior beliefs. While entropy or Blackwell informativeness measures also put constraints on the sender's problem, this approach still allows for arbitrarily many action recommendations (possibly subject to a cost), and existence results rely on having a high dimensional signal space.

Similiar papers analyze settings where agents have preferences that take into account both the outcome of the persuasion game and the complexity of information structures induced (Wei 2018; Bloedel and Segal 2018). Although we don't focus on behavioral limitations in this paper, the tools we develop can also be used to motivate settings where larger signal spaces and complicated signal structures create mental burdens on the receiver or the sender. Our paper complements this line of work, as it can be used to analyze the cost of sending or receiving one more message, instead of using information theoretic costs. In our setting with a smaller signal space, Bayesian updating involves fewer operations and there are fewer contingencies for the actors to take into account.

Another possible disruption to communication quality is noise. In the models that entertain this possibility, signals chosen by the sender can be misinterpreted or transformed due to the imperfections in the channel (Akyol et al. 2016; Le Treust and Tomala 2019; Tsakas and Tsakas 2018). Substantively, the difficulty in communication caused by noisy channels is different from our setting with coarse channels. With coarse communication, the sender strategically chooses which directions in the belief space they want to be more 'precise' about, instead of an exogenous noise structure making the communication imprecise.

In terms of the mathematical techniques we develop, our work is also related to Lipnowski and Mathevet (2017) and Dughmi et al. (2016). Lipnowski and Mathevet (2017) characterize the properties of optimal information structures in signal-rich settings relying on extremal representation theorems from convex analysis. We provide a similar result that applies to both rich and coarse communication settings. Dughmi et al. (2016) also analyze limited signal spaces, but take a computational perspective and focus characterizing on the algorithmic complexity of approximating optimal sender utility.

The rest of the paper is organized as follows. We start by analyzing an application of our model to targeted advertising in 2. Section 3 beings by giving the full mathematical description of the persuasion games we study, and provides the description of the adjusted concavification method in 3.1, and the properties of optimal information structures in 3.3. Section 3.2 proves a lower bound result on the marginal value of a signal, and 4.1 focuses on analyzing the marginal value of a signal in a specific class of persuasion games with assumptions on sender and receiver preferences. In Section 4, we show how to extend our methods to various other strategic communication games. Proofs and additional results are provided in the appendix.

2 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 a three dimensional simplex.[2]

Suppose that different types of customers arrive to an online platform, according to a known distribution. An advertiser observes the characteristics (demographics, location, browsing history etc.) of the arriving customers and must decide on what type of advertisement to show to the customer conditional on this observation. For ease of demonstration, we suppose different types of customers correspond to three different segments of the population. In this sense, our example is a three dimensional version of the examples in Rayo and Segal (2010) and Kamenica and Gentzkow (2011), where the state is an underlying random 'prospect' which captures the quality of the match between the product characteristics and the customer.

We represent the state space by $\Omega = \{\omega_1, \omega_2, \omega_3\}$, corresponding to customers with different characteristics arriving to the platform. 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 will assume a prior $\mu_0 = (0.65, 0.1, 0.25)$, which is a vector representing the prior probabilities of $\omega_1, \omega_2, \omega_3$.

The advertiser learns the state of the world by observing the characteristics of the cus-

[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 signal space to be smaller leads to no information transmission since the sender will have access to only one signal.

tomer and inferring the quality of the match. The state is initially unknown to the customer, who doesn't know the properties of the product sold by the advertiser. Formally, the sender's (advertiser's) signaling strategy is a mapping from the set Ω to the distributions over the set of available of signals $\Delta(S)$. In practice, the advertiser chooses a distribution over different kinds of ads given 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, that they will commit to for some fixed length of time. In this world, an additional signal can be interpreted as increased targeting ability: the sender can use another type of ad and show it to different types of customers. Limitations on the set of signals can represent institutional constraints such as regulation on targeting, or technological constraints on how fine targeting can be in this environment.

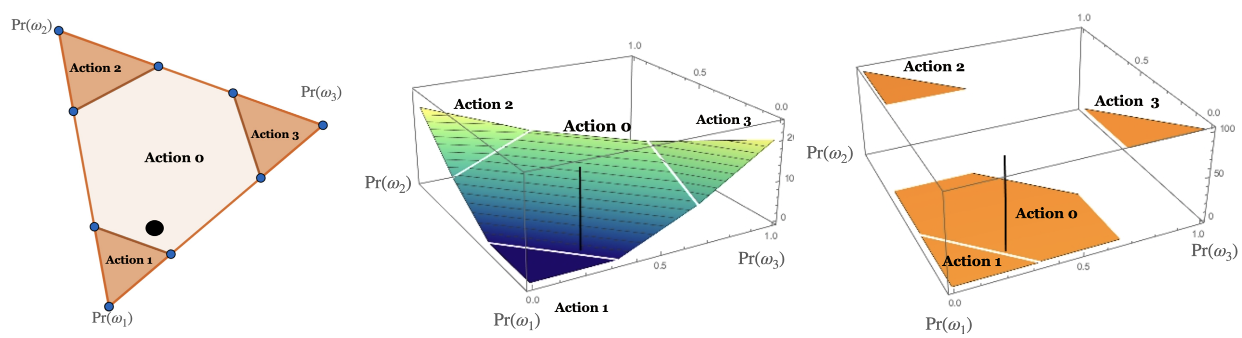

Figure 1: Action regions, and the receiver and sender utility over the belief space. The figure on the left gives a top-down view of the belief space, showing regions where each action is optimal. The three corners of the simplex correspond to the beliefs about each one of the states. The center figure plots the receiver utility as a function of induced beliefs. The receiver utility is piecewise maximum of expected utility of fixed actions that are optimally chosen depending on the posterior belief. The beliefs where a_i is the optimal action corresponds to the action zone i labeled on the top. Receiver prefers taking action a_i when the state is ω_i with high probability, and takes action a_0 when his beliefs are uncertain about all of the states. The right figure plots the sender utility as a function of induced beliefs. The sender prefers the receiver to take actions 2 and 3, over actions 0 and 1. The black dot and the black line represent the location of the prior.

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 ad.

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 uncertainity. Similiar preferences are studied in the literature for different contexts (Sobel 2020; Chakraborty and Harbaugh 2010; Lipnowski and Ravid 2020). To put this into the context of targeted advertising on a platform, action a_3 represents a purchase, which is optimal if the customer's preferences match the product sold by the advertiser (ω_3). The action a_2 represents a click without a purchase, which is optimal when there is a weak match (ω_2). The 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). The default action under the prior a_0 represents an impression with no interaction at all, and is the optimal action when the beliefs are not precise in any particular direction. Moreover, the receiver gains some utility from reducing the uncertainty in their prior. Their utility is convex and weakly increasing in all beliefs, and strictly increasing in their beliefs of the states ω_2 and ω_3. This represents the customers having preferences towards 'informative' advertising that reduces uncertainty in their beliefs, making them strictly better off if they learn about a product which at least partially suits their preferences. The receiver's utility function is shown in Figure 1.

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. They prefer engagement (either in the form of a donation a_3, or a click a_2) over no engagement or hiding the ad. Thus, we assume that receiver actions a_3 and a_2 give equal utility to the sender, and a_1 and a_0 are the least preferred actions. We plot the sender utility in Figure 1.[3]

Given access to three signals, or the possibility of showing three different ads depending on customer characteristics, the advertiser induces actions a_3, a_2, a_1 depending on the state. The optimal signaling 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 signal s_1, but sends less precise signals s_2 and s_3 that mix states ω_2 and ω_3 with ω_1. In other words, the sender induces the *least convincing* belief that will make the receiver indifferent between actions a_0 and a_2 (or a_3). This is the choice which maximizes the ex-ante probability that the receiver will take actions a_2 or a_3.

This solution can be found by inspecting the concavification of sender utility, as described by Kamenica and Gentzkow (2011). Given access to a rich signal space, the sender can induce as many posteriors as they want. If the sender is constrained to using two signals, the optimal strategy cannot be found using the standard concavification method, because we can only

[3]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 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 $\beta_0 = [-250/3, 500/3, 500/3], \beta_1 = [0,0,0], \beta_2 = [-150, 200, 100], \beta_3 = [-150, 100, 200]$.

use convex combinations consisting of at most two points from the graph of the sender utility function.

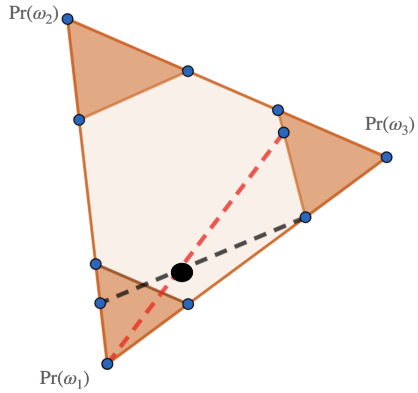

Figure 2: Two-signal information structures drawn over the belief space. The black dot represents the prior, and the dashed red and black lines represent information structures.

Since in this example sender's preferences are independent of the state, when they can induce two actions, they maximize the probability that the receiver takes the more preferable action, under the Bayes plausibility constraint. Given posteriors in a fixed action region, the sender will either want the posterior to be as close as possible, or as far as possible from the prior. Geometrically, this implies we can restrict our search to line segments supported on the corners and edges of the action regions, passing through the prior. There are finitely many such candidates, and we draw some examples in Figure 2.[4]

As a preview of our results, observe that the posteriors induced in figure 2 incorporate at least one extreme point of one of the three action regions. This geometric property generalizes to higher dimensions and different orders of extreme points, generalizing results from Lipnowski and Mathevet (2017). This property will help us provide a finite procedure to construct solutions to Bayesian persuasion games.

With the optimal signal, the sender will choose to induce actions a_3 and a_1, using the posteriors $\{(1,0,0),(0.07,0.27,0.66)\}$ with respective probabilities $(0.63,0.37)$. This information structure maximizes the probability of the action 3 while minimizing the probability of action 1. In other words, it minimizes the ratio of the distance between the prior and the posterior that leads to the desired action, and the distance between the prior and the posterior that leads to the undesired action. Sender utility and optimal information structures

[4]For simplicity of illustration, we parametrize 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 are not restricted to this case.

are shown in Figure 3.

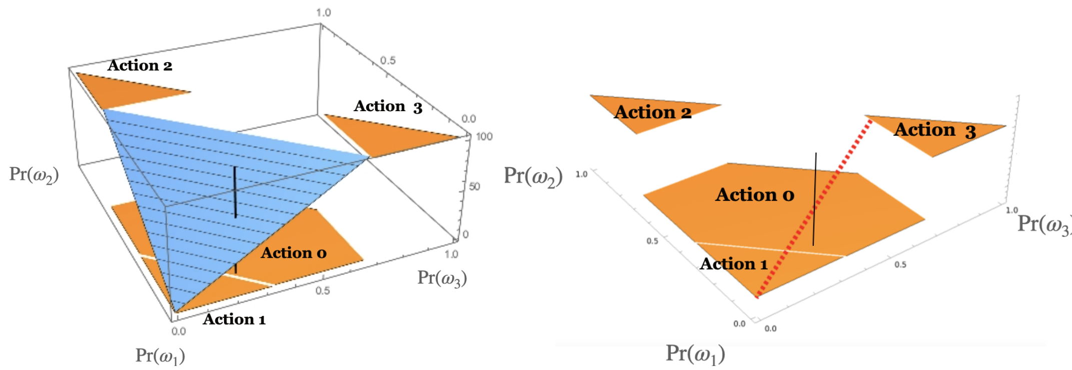

Figure 3: Optimal information structures with 3 signals (blue, left) and 2 signals (red, right) shown over the sender utility function. In this representation, a signaling scheme is a triangle (with 3 signals) or a line segment (with 2 signals), passing through the prior to satisfy the Bayes plausibility constraint. With 3 signals, the sender chooses to optimally induce actions 1,2, and 3. With 2 signals, their choice is limited, and they optimally induce actions 1 and 3. The expected sender utility is the point at which the information structures intersect with the black line representing the prior.

The receiver's utility in the equilibrium with three signals and two signals is drawn in Figure 4. We see that under certain conditions, the receiver will be better off in the equilibrium with two signals. This is only possible when there is a misalignment in preferences: in this setting, the sender only cares about actions, where the receiver wants to have more precise posteriors in certain directions. Limiting the sender's targeting ability results in an optimal signaling strategy which generates more precise posteriors in these directions. This implies that the customers will be better off if the targeting capabilities of the advertiser are limited. Note that if the preferences of the two agents are perfectly aligned, the receiver would never want to limit the Sender.

We can mathematically characterize conditions on receiver utility such that coarse communication makes them better off.[5] If the customers get high enough utility from reducing the uncertainty in their beliefs, limiting the targeting capability of the advertiser would make them better off.

It seems counter-intuitive that receivers who benefit from more precise posteriors would prefer to limit the communication capacity of the sender. Indeed, receiver preferences are convex over the belief space, so they (weakly) benefit from more precise information globally.

[5] As long as the slope parameters β_0^2, β_0^3 for the action region a_0 are high enough, the receiver will prefer the 2-signal outcome over the 3-signal 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, $\delta \approx 0.85$.

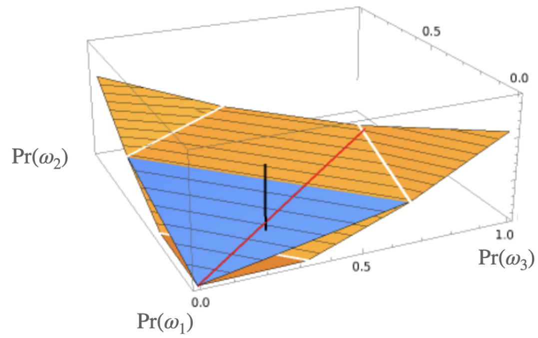

Figure 4: Receiver utility (yellow surface) over the belief space. The beliefs induced by the optimal 3-signal solution and the 2-signal solution to the sender's problem are drawn in blue and red, respectively. The expected receiver utility is the point at which the information structures intersect with the black line representing the prior. If the slope of the receiver utility is high enough in the middle region corresponding to a_0, the receiver will be better off if the sender has access to only 2 signals.

However, as one can see in Figure 3, limiting the cardinality of the signal space does not necessarily result in less precise posteriors being induced at the equilibrium. In fact, the three signal and the two signal optimal information structures are not Blackwell-comparable.

Intuitively, the sender has the ability to choose which directions in the belief space they want to be more precise about, using the limited set of signals they have access to. This stands in stark contrast to settings with noisy communication, where increasing the amount of noise would necessarily result in less precise posterior distributions at the equilibrium.

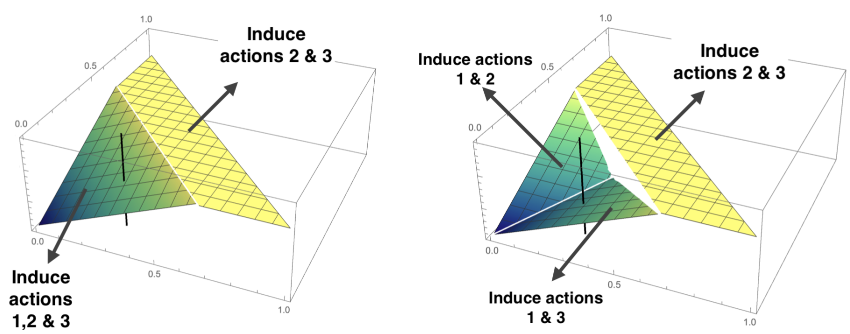

Figure 5: The left figure shows the maximum achievable sender utility with 3 signals, and the center figure shows the maximum achievable sender utility with 2 signals as a function of the prior beliefs. The black lines correspond to the prior belief given in the example. The right figure plots the value of having access to the third signal for the sender, as a function of the prior belief. This is the difference between the two surfaces plotted in the top row. The third signal is more valuable when the prior is in a region where the sender needs to induce their least favorite action (a_1) with high probability.

We can also characterize the utilities achievable by the sender for any prior belief, using a modified concavification method. We can plot the set of points that can be represented as the convex combination of at most 2 points from the graph of the sender utility 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 signals, hence they would be willing to pay to get access to additional signals, or increased targeting ability in their advertising strategy. The marginal value of a signal for any prior belief can be calculated through the difference of the two surfaces in the top row of Figure 5. We can see that there are priors where the signal space constraint is not binding, and the value of an additional signal 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 the actions a_2 (engagement) and a_3 (donation) without inducing a_1. Having access to a third signal is especially valuable for priors where the sender has to induce their least favorite action a_1 more frequently.

3 The Model

We consider the classic Bayesian persuasion game of Kamenica and Gentzkow (2011) extended to limited signal environments. As in their paper, we focus on a finite persuasion problem in this paper.

Players: There are two agents. We call them the Sender (she) and the Receiver (he), who are communicating about an uncertain state of the world.

Environment: The state of the world ω can take values from a finite set Ω, which has cardinality $|\Omega| = n$. The receiver's actions are denoted with $a \in A$ with $|A| = m$. Sender communicates with the receiver using signals $s \in S$ with $|S| = k$.[6] Critically, we focus on the case where $2 \leqslant k < \min\{m, n\}$. The setting where $|S| = k = 1$ is trivial since there will be no information transmission.

Payoffs: The sender and the receiver have utility functions which can depend on the state of the world and the receiver's action, respectively denoted by: $u^S, u^R : \Omega \times A \to \mathbb{R}$.

Information: The agents share a prior belief about the state of the world, μ_0, which is assumed to be in the interior of $\Delta(\Omega)$.

Timing: The sender chooses a signaling policy π which maps the realization of the states $\omega \in \Omega$ to a probability distribution over signals $\Delta(S)$. A more convenient way to denote signaling policy is to consider it as a collection of conditional probability mass functions

[6] As a general fact, signals don't carry any meaning ex-ante, and obtain a meaning via the signaling policy at the equilibrium.

$\{\pi(.|\omega)\}_{\omega\in\Omega}$ over the signal space S. We denote the set of all signaling policies $\pi : \Omega \to \Delta(S)$ with Π. After the choice of a signaling policy π, a state ω is realized. Sender follows the announced signaling policy and sends a signal according to probability distribution $\pi(s|\omega)$. Receiver observes the realized signal and takes an action.

Given the sender's signaling strategy $\pi \in \Pi$, realization of each signal s leads to a posterior μ_s for the receiver, calculated by Bayes' Rule. Before the realization of the state and hence the signal, the signaling strategy π induces a distribution over posterior beliefs $\tau \in \Delta(\Delta(\Omega))$ with $\mathsf{supp}\,(\tau) = \mu = \{\mu_s\}_{s\in S}$ defined by:[7]

$$\tau(\tilde{\mu}) = \sum_{s:\mu_s=\tilde{\mu}} \sum_{\omega'\in\Omega} \pi\left(s \mid \omega'\right) \mu_0\left(\omega'\right) \quad \forall \tilde{\mu} \in \Delta(\Omega).$$

We can interpret the support of this distribution as the set of posteriors induced by π. The induced set of posteriors are denoted as $\mu = (\mu_1, \ldots, \mu_k) \in \Delta(\Omega)^k$ with $\mu_i := \mu_{s_i}$. Note that, with only k signals, the sender can induce at most k different posteriors. The assumption that $k < \min\{m, n\}$ is the reason behind sender's inability to recommend every possible action or describe every possible state perfectly. Unlike the classical Bayesian persuasion setup, the sender faces an additional problem, and has to decide which collection of actions to induce. Formally, because sender has access to only k signals, they can induce at most k different beliefs. This is the only additional restriction imposed by coarse communication.

After forming the posterior μ_s, the receiver chooses an action from the set $\hat{A}(\mu_s) = \arg\max_{a\in A} \mathbb{E}_{\omega\sim\mu_s} u^R(a, \omega)$.[8] The existence of this maximum is guaranteed since A is a compact set and $u(a, \omega)$ is continuous. 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.[9] If there are multiple such elements that maximize the sender's utility, we pick an element from $\hat{A}(\mu_s)$ arbitrarily. We denote the sender-optimal action from the set of receiver-optimal actions at belief μ_s by $\hat{a}(\mu_s)$.Sender's expected utility from $\pi \in \Pi$ is:

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

[7] $\mathsf{supp}\,(\tau)$ denotes support of τ. μ_s denotes the posterior induced by s which is a generic element of S, and μ_i denotes the i$^{\text{th}}$ entry of $\mu = \mathsf{supp}\,(\tau)$. So we use μ_i to refer a specific entry of μ and μ_s to generic posteriors receiver forms upon observing a generic signal $s \in S$.

[8] The notation $\mathbb{E}_{\omega\sim\mu_s}$ is used to denote the expectation over the random variable ω taken with respect to the measure μ_s. When the random variable is clear, we will just use the measure that gives the probability distribution on the subscript.

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

An optimal signaling strategy π^* for sender is then defined by $\arg\max_{\pi\in\Pi} U^S(\pi)$ and has value $u^* = \max_{\pi\in\Pi} U^S(\pi)$.

Similar to Lemma 1 in Kamenica and Gentzkow (2011) we can transform the problem of choosing $\pi \in \Pi$ to choosing $\tau \in \Delta(\Delta(\Omega))$ such that $|\mathsf{supp}\,(\tau)| \leqslant k$. This is the belief based approach where sender's signaling strategy and receiver's equilibrium beliefs are replaced with the ex-ante distribution over the posterior beliefs.Throughout the paper, τ will be called an information structure (induced by the signaling strategy π).

Formulating the sender's problem as a search for an optimal information structure τ rather than a search for signaling strategy $\{\pi(.|\omega)\}_{\omega\in\Omega}$ makes the problem more tractable. In section 3.3 we will show that, in fact, using this approach reduces the candidate optimal information structures to a finite set.

The sender's utility when the posterior μ_s is induced will be denoted as $\hat{u}^S(\mu_s) = \mathbb{E}_{\omega\sim\mu_s} u^S(\hat{a}(\mu_s), \omega)$ and the receiver's utility with posterior belief μ_s is $\hat{u}^R(\mu_s)$ defined in the same way. Expected utility under the information structure τ is denoted by $\mathbb{E}_{\mu_s\sim\tau}\hat{u}^S(\mu_s)$ and $\mathbb{E}_{\mu_s\sim\tau}\hat{u}^R(\mu_s)$ for the sender and receiver respectively.

For a distribution of posteriors to be feasibly induced in the persuasion game with shared priors, we need the expected value of the posterior beliefs to be equal to the prior belief. This is also called the Bayes plausibility constraint (Kamenica and Gentzkow 2011), which we can state formally by $\mathbb{E}_{\mu_s\sim\tau}\mu_s = \sum_{\mu_s\in\mathsf{supp}\,(\tau)} \mu_s\tau(\mu_s) = \mu_0$ alongside with the cardinality constraint $\mathsf{supp}\,(\tau) \leqslant k$ due to coarse communication. Formally, we can state the following:

Lemma 1. *There exists a signal with value u^* if and only if there exists a Bayes plausible distribution of posteriors τ such that $E_\tau\hat{u}^S(\mu) = u^*$ and $|\mathsf{supp}\,(\tau)| \leqslant k$. If $k \geqslant \min\{m, n\}$, this is true for any Bayes plausible $\tau \in \Delta(\Delta(\Omega))$ such that $E_\tau\hat{u}^S(\mu) = u^*$.*

This statement is identical to Lemma 1 in Kamenica and Gentzkow (2011) when $k \geqslant \min\{m, n\}$. When $k \leqslant \min\{m, n\}$, given a signaling policy π, we can derive the equivalent distribution of posteriors $\tau(\mu_s)$ for any μ_s by Bayes' rule. This imposes $\sum_{s\in S}\tau(\mu_s)\mu_s = \mu_0$. From a given an information structure τ such that $E_\tau\hat{u}^S(\mu) = u^*$ and $|\mathsf{supp}\,(\tau)| \leqslant k$ we can always find the associated signals by writing $\pi(s|\omega) = \frac{\mu_s(\omega)\tau(\mu_s)}{\mu_0(\omega)}$ for each $\mu_s \in \mathsf{supp}\,(\tau)$.

Using 1 we can describe sender's problem as the following constrained optimization problem:

$$\max_{\tau\in\Delta(\Delta(\Omega))} \mathbb{E}_{\mu_s\sim\tau}\hat{u}^S(\mu_s) \text{ subject to } |\mathsf{supp}\,(\tau)| \leqslant k \text{ and } \mathbb{E}_\tau(\mu_s) = \mu_0. \tag{1}$$

Throughout the paper, our focus will be on games where there are some gains to sending information i.e. there is some τ such that $\mathbb{E}_\tau(\hat{u}^S) \geqslant \hat{u}^S(\mu_0)$.[10] We first show that sender's problem (1) has a solution.

Proposition 1. *An optimal information structure τ which solves the optimization problem described in (1) exists.*

Existence follows from extending the existence proof of Kamenica and Gentzkow (2011). They who show that $\hat{u}^S$ is upper semi-continuous and attains a maximum over all Bayes plausible information structures. We how that set of Bayes plausible information structures whose support has cardinality at most k is a closed subset of all Bayes plausible information structures in the relevant topological space. This provides compactness of the domain the objective is considered. We add on this result about existence, in section 2.2 by providing a finite algorithm that describes how to find a sender optimal information structure in any finite Bayesian persuasion game.

3.1 Achievable Utilities and k-Concavification

A key contribution of Kamenica and Gentzkow (2011) is the characterization of the set of attainable payoffs by the sender. This set is particularly useful to describe the best-case payoffs in a strategic communication setting. Existing literature studies similar descriptions of highest attainable payoffs for the sender in various models of strategic communication (Lipnowski and Ravid 2020; Aumann and Hart 2003; Aybas and Callander 2022).

In this paper, we add onto these results by providing a method to geometrically characterize the highest achievable sender payoffs for each prior μ_0 and cardinality of the signal space k. Our characterization directly parallels the geometric characterization provided Kamenica and Gentzkow (2011) with concavification. We call our characterization as the k-concavification of sender utility.

Let $\mathbb{CH}(\hat{u}^S)$ denote the convex hull of the graph of $\hat{u}^S$.[11] Without restrictions on the set of available signals, the point $(\mu_0, z) \in \mathbb{CH}(\hat{u}^S)$ represents a prior μ_0 and sender payoff z achievable by an information structure for prior μ_0.[12] This is the foundation of the concavification technique, first used in repeated games (Aumann and Maschler 1995) and then applied to Bayesian persuasion (Kamenica and Gentzkow 2011).

[10] Throughout the paper, our focus will be on games where there are some gains to sending information: the other case is trivial and the sender always prefers sending no information.

[11] Formally, $\mathbb{CH}(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ 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 $\mathbb{R}^n$ i.e. $f \mapsto \text{co}(\text{graph}(f))$.

[12] Since $\hat{u}^S : \Delta(\Omega) \to R$, we can represent any belief μ with $|\Omega| - 1 = n - 1$ dimensions, and $\hat{u}^S(\mu)$ with a real number, so $(\mu, z) \in \mathbb{R}^n$.

For any $(\mu_0, z) \in \mathbb{CH}(\hat{u}^S)$, Caratheodory's Theorem assures the existence of a τ such that $\mu_0 \in \text{co}(\mathsf{supp}\,(\tau))$ and $|\mathsf{supp}\,(\tau)| \leqslant n + 1$, where co denotes the convex hull operator. Note that the last condition possibly requires n posteriors in the support of τ to attain payoff z. With restricted communication, the point $(\mu, z) \in \mathbb{CH}(\hat{u}^S)$ might not be feasible. In this scenario, the construction of (μ, z) could require a convex combination of more than k points from the graph of $\hat{u}^S$. That is, any information structure giving payoff z could only be constructed using strictly more than k signals.

In the constrained communication case, a prior belief-utility pair (μ, z) is feasible if it can be contained in the convex hull of k or fewer points from the graph of $\hat{u}^S$. We formalize this by extending the definition of a convex hull.

The convex hull of a set Λ is the set of all points that can be represented by convex combination of points in Λ. We denote this by $\text{co}(\Lambda)$. As a generalization, we define the *k-convex* hull. Given a set $\Lambda \subseteq \mathbb{R}^n$ and an integer $0 < k$, the k-convex hull of Λ is the the set of points that can be represented as the convex combination of at most k points in Λ. We denote this by $\text{co}_k(\Lambda)$. Note that whenever $n \geqslant k$, $\text{co}_k(\Lambda)$ coincides with $\text{co}(\Lambda)$ and whenever $k < n$ it is a subset of $\text{co}(\Lambda)$. Below we provide a formal definition of k-convex hull.

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

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

Using this definition, we denote the k-convex hull of the graph of $\hat{u}^S$ as $\mathbb{CH}_k(\hat{u}^S)$. Note that if $(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)$, there exists an information structure τ with $\mathsf{supp}\,(\tau) \leqslant k$ and the $\mathbb{E}_\tau(\hat{u}^S) = z$.

In order to state the set of attainable payoffs, we define $V(\mu_0) = \sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\}$. $V(\mu_0)$ is the largest payoff the sender can achieve when the prior is μ_0. By construction, if $V(\mu_0) = z$, then we have k beliefs such that $\sum_{i \leqslant k} \tau(\mu_i)\mu_i = \mu_0$ for some set of weights $\{\tau(\mu_1), \ldots, \tau(\mu_k)\}$ and $\sum_{i \leqslant k} \tau(\mu_i)\hat{u}^S(\mu_i) = z$. This gives us the following equivalence between k-concavification and our previous result on the optimal information structures.

Proposition 2. *Let τ be an optimal information structure that solves the sender's maximization problem described in (1). Then $V(\mu_0) = \mathbb{E}_\tau \hat{u}^S$.*

Similar to concavification approach, k-concavification can be used to identify the optimal information structure when plotted. Moreover, the direct relationship between their characterization and ours suggests a wider scope for the concavification technique. Our results can be used to extend current concavification-based results relying on rich signals and generalize them to arbitrary signal spaces. This highlights the full power of the concavification technique. We show this by an example in the last section of the paper, by applying our results to extend Lipnowski and Ravid (2020).

Finally, this method is a useful tool to analyze and compare the possible gains from persuasive communication under different priors and different degrees of coarseness in communication. By comparing the $k-$concavification of the sender utility for different values of k one can characterize the value of additional signals for any prior belief.

3.2 The Marginal Value of a Signal

So far, our analysis focused on how limited availability of signals affects communication. In this section, we turn to the question on how it compares to unlimited communication.

Our particular focus is on analyzing how much the sender would be willing to pay for an additional signal. This value is of particular importance, especially when we consider cases where generating additional signals or having additional outcomes in an experiment is costly. In such examples, this value informs the problem of the sender and determines the number of signals or outcomes that are used in the equilibrium.

As expected, the sender attains the highest utility in a signal-rich setting. Additional signals always weakly improve the payoff for the Sender. We use the unlimited signal payoff for the sender to provide a lower bound on the optimal utility in the coarse communication setting. This result relies on structural relationship between higher and lower dimensional optimal information structures.

Let $V^*(k, \mu_0)$ be the value function of the sender with prior μ_0 when the signal space S is restricted to have k elements. Then, we can define the marginal value of a signal as $V^*(k+1, \mu_0) - V^*(k, \mu_0)$. This value represents what the sender would be willing to pay for access to an additional signal

By Caratheodory theorem, whenever $k \geqslant \min\{|\Omega|, |A|\}$, the marginal value of a signal will be zero. Hence, additional signals are only valuable in coarse environments.

The sender's willingness to pay for an additional signal is measured in terms of this gap $V^*(k+1, \mu_0) - V^*(k, \mu_0)$. This gap depends on the preferences, and the common prior μ_0.

Sender cannot induce all combinations of k-actions in coarse communication, as some might fail Bayes plausibility. This restriction is critical for the value of an additional signal.

If maintaining Bayes plausibility with lower dimensional signals requires inducing a posterior located in a low-payoff yielding portion of an action region, then the sender will be willing to pay more for more precise communication.

We establish an upper bound on the marginal value of a signal. This can be equivalently stated as a lower bound on the utility achievable with $k-1$ signals. Our result applies to any finite game of Bayesian persuasion. The inequality can be recursively applied to get bounds on the value of attainable payoffs with any number of signals.

Proposition 3. *Suppose $|S| = k \geqslant 2$, and the sender utility function u^S is positive everywhere.*[13] *Then it holds that*

$$V^*(k, \mu_0) - V^*(k-1, \mu_0) \leqslant \frac{2}{k} V^*(k, \mu_0).$$

Equivalently, we can state:

$$\frac{k-2}{k} V^*(k, \mu_0) \leqslant V^*(k-1, \mu_0) \leqslant V^*(k, \mu_0).$$

The $2/k$ factor on the upper bound implies that in persuasion games with large state and action spaces, the marginal value of a signal 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.

The proof relies on creating alternative $k-1$ signal 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 $k-1$ signals. 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 signals as a function of the payoff attainable with full communication and binary communication (the smallest non-trivial case where information transmission happens).

[13] This is without loss of generality. 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.

Corollary 1. *Define* $\underline{V}(q,\mu_0) = \frac{(q-1)q}{(q+1)(q+2)}V^*(k,\mu_0)$ *and* $\overline{V}(q,\mu_0) = \frac{q(q-1)}{2}V(2,\mu_0)$*. Then, it holds that* $\overline{V}(q,\mu_0) \geqslant V(q,\mu_0)$ *for every* $q > 2$ *and* $V(q,\mu_0) \geqslant \overline{V}(q,\mu_0)$ *for every* $q < k$*.*

3.3 Properties of Optimal Information Structures

An important behavioral critique for Bayesian persuasion is the difficulty (for the sender) of finding optimal information structures and the calculation of the concave envelope of sender utility. Lipnowski and Mathevet (2017) makes the first effort of simplifying this search to a finite problem in settings with rich signal spaces. The implementation of optimal information structures rely on the sender's ability to compute the concavification of their utility function, or derive qualitative properties of it, which is a difficult computational task (Tardella 2008).

Access to limited signals makes the problem for searching information structures more difficult (Dughmi et al. 2016). In this section, we determine the qualitative properties of optimal information structures with coarse communication. We will also show that the search for the optimal information structures can be finalized after searching over a finite set. We follow a the technique of using the underlying preference structure of the receiver to partition the space of posterior beliefs.

Formally, we can define subsets of $\Delta(\Omega)$ where the receiver's action is constant, and use the fact that sender utility is convex within these subsets.[14]

Definition 2. *The set* $R_a \subseteq \Delta(\Omega)$ *is the set of beliefs where the action* a *is receiver-optimal* $R_a = \{\mu_i \in \Delta(\Omega) : a \in \hat{A}(\mu_i)\}$*.* $R = \{R_a\}_{a\in A}$ *is the collection consisting of these sets for every action* $a \in A$*.*[15]

In order to characterize the conditions for optimality we use well known properties established in the literature. Lemmas 2 and 3 have been applied in (i) the context of persuasion games where the receiver has psychological preferences over different posterior beliefs (Volund 2018) and (ii) general games of persuasion to simplify the search for optimal information structures (Lipnowski and Mathevet 2018).

Lemma 2. *For every action* $a \in A$*, the set* R_a *is closed and convex.*

Lemma 3. *The sender's utility* $\hat{u}^S$ *is convex when restricted to each set* R_a*.*

[14] In the appendix, we also establish that expected sender utility is a continuous and piecewise affine function in the interior of these sets.

[15] R is a finite cover of $\Delta(\Omega)$.

Lemma 2 follows from the fact that each R_a can be written as the intersection of finitely many closed half spaces. The proof of Lemma 3 uses the definition of $\hat{u}^S$, which is a function of sender-optimal actions at every belief. For any two beliefs μ', μ'' in a given R_a, let the sender-optimal action be $\hat{a}(\mu)$ at their convex combination μ. This action must be among the set of receiver-optimal actions for the two original beliefs. Since the action $\hat{a}(\mu)$ is defined as the action that maximizes sender utility among the set of receiver-optimal actions $\hat{A}(\mu)$, and we have $\hat{a}(\mu) \in \hat{A}(\mu')$ and $\hat{a}(\mu) \in \hat{A}(\mu'')$, convexity of $\hat{u}^S$ follows.

Using Lemma 2 and 3, we can restrict our search to information structures inducing affinely independent posteriors. Lipnowski and Mathevet (2017) shows that in the canonical full communication Bayes persuasion model the optimal information structures are affinely independent. They use extreme point theorems of Krein-Millman and Caratheodory to argue the existence of an information structure with affinely independent support. Our proof generalizes their result to arbitrary signal spaces. Moreover, the proof provides a systematic way to modify any given information structure by dropping beliefs to reach an affinely independent set and improve payoffs simultaneously, while maintaining Bayes plausibility.

Lemma 4. *Let τ be a distribution of posteriors satisfying Bayes plausibility. Suppose that $\mathsf{supp}(\tau)$ is not affinely independent. Then, there must exist a Bayes plausible $\tau' \neq \tau$ such that $\mathsf{supp}(\tau')$ is affinely independent and $\mathbb{E}_{\tau'}\hat{u}^S \geqslant \mathbb{E}_{\tau}\hat{u}^S$.*

Intuitively, inducing affinely dependent beliefs is not a good use of the signals by the sender - even if there is no cost associated with using more signals. For an affinely dependent set of posteriors, some signals are 'redundant' and the resulting beliefs can be represented as affine combinations of each other. The sender can always drop one of them and still maintain Bayes plausibility and our proof outlines which posteriors to drop.

An immediate corollary of this result is that given an optimal information structure $\mu = \{\mu_1, \ldots, \mu_k\}$ that is affinely independent, the probability distribution τ that ensures Bayes plausibility is uniquely determined through Choquet Theorem.[16] The unique probability distribution can in fact be calculated through a series of matrix operations.

We show that an information structure can always be weakly improved by changing it in a way that maintains Bayes plausibility, and moving as many posteriors as possible to the most extreme beliefs inducing an action. We measure the extremeness of a belief using the following definition from convex analysis:

[16] See the appendix for a statement of this well known result in affine geometry.

Definition 3. *A point in an arbitrary convex set Λ is q-extreme if it lies in the interior of a q-dimensional convex set within the set, but not a q+1-dimensional convex set within Λ.*

In our setting, we will call q-extreme points of action regions q-extreme beliefs. Intuitively, a q-extreme belief can be represented as a convex combination of $(q-1)-$extreme beliefs. Whenever we compare a q-extreme belief and a q'-extreme belief with $q < q'$, we call q-extreme belief less extreme. Critically, less extreme beliefs can be described of as averages of more extreme beliefs, but not vice versa. A 0-extreme belief is therefore the most extreme belief and cannot be represented as the average of any other belief in a given action region.

For a belief in the interior of the simplex and in a given an action region R_a, and integers $q < r$, the receiver is indifferent between more of their actions at $q-$extreme beliefs of R_a, compared to $r-$extreme beliefs of R_a. And for a belief on the boundary of the simplex $\Delta(\Omega)$, the receiver's posterior gives 0 probability to more states at $q-$extreme beliefs, compared to $r-$extreme beliefs.

Thus intuitively, more extreme beliefs correspond to either more precise posteriors, or posteriors where the receiver is indifferent between more actions. The next lemma shows that these properties can be used to simplify the search for optimal information structures, by starting from 0-extreme beliefs of action regions.

Lemma 5. *Let τ be an information structure, and $\mathsf{supp}(\tau) = (\mu_1, \mu_2, \ldots, \mu_k)$ has fewer than $(k-1)$ posteriors that are 0-extreme beliefs of some action regions $\{R_a\}_{a \in A}$. Then, there must exist a Bayes plausible $\tau' \neq \tau$ that weakly improves sender utility, such that $\mathbb{E}_{\tau'}\hat{u}^S \geqslant \mathbb{E}_{\tau}\hat{u}^S$.*[17]

When $k = n$, we are back to the unrestricted communication, our result simplifies to sender inducing n many 0-extreme points (also called outer points). This special case is studied and proven by Lipnowski and Mathevet (2017).

In the proof of this Lemma, we show that given an information structure with at least two beliefs that are not 0-extreme, one can always find a direction to move these two posteriors towards more extreme beliefs, without affecting the probabilities of the other induced beliefs and maintaining Bayes plausibility. Sender either prefers moving posterior beliefs in this direction or the opposite direction by the linearity of expected utility in probabilities.

This result allows us to reduce the size of our search space considerably from an infinite set (the set of k-dimensional Bayes plausible information structures) to a search over a finite set, and explicitly characterize the sender-optimal information structure in any Bayesian persuasion game using a finite algorithm, under the following assumption that rules out

[17] We also show in the appendix that the remaining belief is at most $(n-k)$ extreme.

certain preference structures with 'redundant' states of the world which are irrelevant for the agents' utilities.

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

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. Assumption 1 states that the game is already in this simplest possible representation.

Corollary 2. *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. Using Lemma 3, we can 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 beliefs of $\{R_a\}_{a \in A}$. The posteriors in this affine subspace correspond to q-extreme points of $\{R_a\}_{a \in A}$ for $q \leqslant (n-k)$.

An alternative approach of finding an optimal information structure based on Bergemann and Morris (2019) could be solving multiple linear programs (one for each k-combination of the actions) and then comparing the results. Our procedure simply checks finitely many candidate information structures (that can be enumerated by listing combinations of 0-extreme points into sets of size (k-1), and the corresponding q-extreme point), and leverages the insights we have developed in the paper to reduce the size of the search space, compared to solving $\binom{M}{k}$ linear programs.

4 Applications and Extensions

4.1 Threshold Games

In this section, we focus on a class of games where the sender's utility only depends on the action and not on the state, and the receiver's default action under the prior is the least preferred action for the sender.Examples involving these kinds of preferences have received interest in previous work: e.g. 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).

Our parametric example captures settings where there are belief 'thresholds' above which the receiver finds it optimal to take a different action, and the default action is doing nothing. Similar preferences are studied by Sobel (2020) to analyze the conditions under which deception in communication will lead to loss in welfare.

We study the case with 3 states of the world to be able to visually demonstrate how the marginal value of a signal can depend on the location of the prior, and the threshold values- or equivalently, the difficulty of inducing desirable actions for the sender.

Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$. There are four actions available to the receiver $A = \{a_0, a_1, a_2, a_3\}$. We consider a Bayesian persuasion game where the sender has an optimal action for each state and a default safe action. This can be represented with receiver preferences of the form:

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

These preferences can be used to model situations in which for each state ω_i action a_i is optimal, and mismatching the state i.e. taking action a_j $j \neq 0$ and$j \neq i$ is costly, with cost normalized to unity. Finally, a_0 is the safe action. Such receiver preferences lead to action thresholds over the simplex of posterior beliefs.

Let us denote $\mu_s(\omega_i)$ by μ_s^i, where μ_s^i is the i^{th} coordinate of a given posterior belief μ_s. One can think of $\mu_s(\omega)$ as the probability distribution over Ω induced by μ_s. For each state, there is a corresponding preferred action a_i which is taken by the receiver if and only if the receiver believes the state of the world is ω_i with at least probability T. Specifically, 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 $\mu_s^i \geqslant T$, 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 that $\mathbb{E}_{\mu_s}[u^R(a_i, \omega)] \geqslant \mathbb{E}_{\mu_s}[u^R(a_j, \omega)]$ if and only if $\mu_s^i > T$. The action zones for these receiver preferences can be represented as:

$$R_i = \{\mu_s \in \Delta(\omega) | \mu_s^i \geqslant T\}$$

Sender preferences are such that $\forall \omega \in \Omega$, $u^s(a_0, \omega) = 0$ and $u^s(a_i, \omega) = 1$. Thus, the sender only cares about actions and not the states, and wants to induce one of the non-default actions. The parameter T can be interpreted as the difficulty of inducing the desirable actions for the sender - since the posteriors need to be above T to induce any desirable action for the sender.

Given this structure, it is immediately clear that sender can attain a payoff of 1 by using 3-signal information structures, as drawn in Figure 6. Every point inside simplex can be represented as the convex combination of the extreme points of the simplex, hence achieving the maximal utility with 3 signals is possible for every interior prior.

With a 1-signal information structure, we see that the sender's payoff is immediately determined by the prior belief. If no information is transmitted, the receiver will take whatever action is optimal under the prior belief, which will yield a 0 payoff to the sender unless the prior is located in one R_1, R_2 or R_3.

We proceed by analyzing the non-trivial case of 2 signals. We focus on priors μ_0 that are in R_0, as for priors in R_i for $i \in \{1, 2, 3\}$ the maximal payoff can be obtained with no information transmission at all. We use Δ_c to denote the set of beliefs where two-signal information structures attain lower payoff than three-signal information structures. The following Lemma characterizes the values of T such that this set is non-empty.

Lemma 6. *$\Delta_c \neq \emptyset$ if and only if $T \geqslant \frac{2}{3}$.*

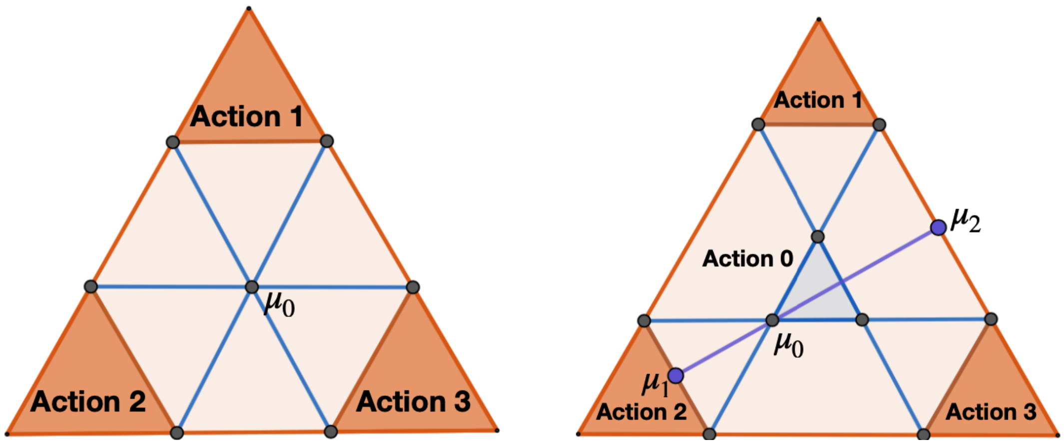

Figure 6: On the left, we have the action threshold $T = \frac{2}{3}$ so it is possible to maintain Bayes plausibility without inducing action 0 for any prior. On the right, $T > \frac{2}{3}$, so for the prior beliefs in the blue shaded region, the sender has to mix a_0 and another action when constrained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

For thresholds $T \leqslant \frac{2}{3}$, two-dimensional information structures suffice for achieving maximal utility. We restrict attention to cases where $T > \frac{2}{3}$. In this regime, we can state that for any prior in Δ_c, the utility attained by two-signal information structures is must lie between two values, determined by T.

Lemma 7. *If $T > \frac{2}{3}$, whenever $\mu_0 \in \Delta_c$, we have that $\frac{1}{3T} < V(2, \mu_0) < \frac{2T-1}{T} < V(3, \mu_0) = 1$. When $\mu_0 \notin \Delta_c$ we have that $V(2, \mu_0) = V(3, \mu_0) = 1$.*

Corollary 3. *Depending on the location of the prior inside Δ_c the marginal value of a signal can be a function with increasing or decreasing differences. That is $\frac{1}{3T} > \frac{1}{2}$ and $\frac{2T-1}{T} < \frac{1}{2}$.*

The priors for which the marginal value of a signal is increasing are the ones that are the furthest away from the desirable action regions. For the sender who only has access to two signals, 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 0.5. Therefore, the value of the second signal is also below 0.5. Getting access to the third signal allows the sender to maintain Bayes plausibility by not inducing the default action, guaranteeing a payoff of 1. Hence, the value of the third signal is higher than 0.5.

On the other hand, for some priors, the marginal value of an additional signal is a decreasing function. These are prior beliefs that are already close to one of the action regions. Intuitively, if the receiver is already leaning towards taking one action, it is easy to induce that action with a high probability, getting an expected payoff above 0.5. The value of the second signal is then higher than the value of the third signal. Note that additional signals always weakly increase the sender utility, because the feasible set in the optimization problem is expanding.

4.2 Cheap Talk with Transperant Motives

Lipnowski and Ravid (2020) study an abstract cheap-talk model in a recent paper. In this setting, there are two players: A sender and a receiver. The game proceeds identically to the persuasion game we describe, except for the fact that the signal $s \in S$ is chosen after the sender observes the state $\omega \in \Omega$.[18] Receiver, upon observing $s \in S$ decides which action a to take from set A. It is assumed that both players' utility functions are continuous, but only receiver's utility depends on the state i.e. $u^R : \Omega \times A \to \mathbb{R}$. Critically, 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 the games are called Cheap Talk with Transparent Motives. To contribute to existing results, we will impose $|S| \leqslant k$ to study implications of our theory in this environment.

Throughout this section we will focus on the Perfect Bayesian Equilibria - hereinafter succinctly referred as the *equilibrium*- $\mathcal{E}(\pi, \rho, \beta)$ of this cheap talk game. Formally, the equilibrium is defined by three measurable maps: a signaling strategy for the sender $\pi : \Omega \to \Delta(S)$; a receiver strategy $\rho : S \to \Delta A$; and a belief system for the receiver $\beta : S \to \Delta\Omega$; such that:

1. β is obtained from μ_0, given π, using Bayes's rule;
2. $\rho(s)$ 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 earlier sections. Hence, we can again focus on the ex-ante distributions over the receiver's posterior beliefs i.e. information structures $\tau \in \Delta(\Delta(\Omega))$.

As we have discussed in our model, every belief system and sender strategy leads to an ex-ante distribution over receiver's posteriors, and by Bayes Rule these posteriors should

[18]Each of Ω, A and S are assumed to be compact metrizable spaces containing at least two elements.

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 signals 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 signals . This set of possible ex-ante distributions is identified by $\mathcal{I}_k(\mu_0) = \{\tau \in \Delta(\Delta(\Omega)) | \int \mu d\tau(\mu) = \mu_0$ and $|\operatorname{supp}(\tau)| \leqslant k\}$. This set $\mathcal{I}_k(\mu_0)$ is identical to the set we maximized over in the problem of Bayes persuasion with coarse communication.

Using the sender's possible continuation values from the receiver having μ as his posterior - described by the correspondence $V(\mu) : \operatorname{co} u^S \left(\arg\max_{a \in A} \int u^R(a, \cdot) \mathrm{d}\mu\right)$ - 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. Formally, when the receiver is constrained to sending k-signal i.e $|S| \leqslant k$ we can characterize equilibrium outcomes as follows.

Lemma 8. *Let (τ, z) be an outcome pair describing a distribution over posterior beliefs τ, and a utility level z. (τ, z) is an equilibrium outcome if and only if: $\tau \in \mathcal{I}_k(\mu_0)$ and $z \in \bigcap_{\mu \in \operatorname{supp}(\tau)} V(\mu)$.*

Essentially, the first condition - $\tau \in \mathcal{I}_k(\mu_0)$ - follows from the equivalence between Bayesian updating and Bayes plausible information structures. Limiting the available signals limit the set of inducable posteriors with a one-to-one relationship, hence replacing $\tau \in \mathcal{I}(\mu_0)$ with $\tau \in \mathcal{I}_k(\mu_0)$ suffices. The second condition - $z \in \bigcap_{\mu \in \operatorname{supp}(\tau)} V(\mu)$ - is a combination of sender and receiver incentive compatibility constraints.

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 $\mathbb{P}_{\mu \sim \tau}(V(\mu) \geqslant z) = 1$. Using this definition they show that an equilibrium inducing sender payoff z exists if and only if z is securable.

With rich signal spaces, the sender can choose an information structure from $\tau \in \mathcal{I}(\mu_0)$ to secure a payoff. The only difference with 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 $\mathbb{P}_{\mu \sim \tau}(V(\mu) \geqslant z) = 1$. Following the exact arguments in Lipnowski and Ravid (2020), when $|S| \leqslant k$ 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 $v_k^*(\cdot) := \max_{\tau \in \mathcal{I}_k(\cdot)} \inf v(\operatorname{supp} \tau)$. In this setting, the sender is maximizing the highest payoff value it can secure across all k-dimensional information policies, as $\inf v(\operatorname{supp} \tau)$ corresponds to the highest value which the information structure τ k-secures. By comparison, with unlimited signals this value is characterized by $v^*(\cdot) := \max_{\tau \in \mathcal{I}(\cdot)} \inf v(\operatorname{supp} \tau)$. Lipnowski and Ravid (2020) show that $v^*(\cdot)$ corresponds to the the quasiconcave envelope of sender's value function $v(\mu) = \max V(\mu)$. This means that it is the the pointwise lowest quasi-concave and upper semi-continuous function that majorizes v.

In order to showcase how to apply our methods in their context, we provide an intuitive connection between constructing lower dimensional optimal information structures and optimal *linear compressions* of the state space. Formally, we first define $\mathcal{T}_k$ be the set of all k-dimensional *flats* that contain the prior μ_0.[19] Formally, we show the original coarse strategic communication problem for the sender is equivalent to an alternative formulation in which sender first picks an *optimal k-dimensional compression* T_k of the state space, and then solves a full-dimensional problem in $\mathbb{R}^k$ with k signals. One can then reinterpret this k-dimensional summary as the optimal way for sender to compress the higher dimensional state space into k new states that are mixture of the former n states. Using this observation we can restate the payoff the sender in the sender-preferred equilibrium with the following proposition:

Proposition 4. *In the setting of Lipnowski and Ravid (2020) with a coarse signal 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:*

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

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

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

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 signal spaces of different size.

5 Conclusion

We set out to analyze the effect of coarse signal spaces in Bayesian persuasion, which was left unexplored by previous literature.

We provide the tools to solve coarse persuasion games, and characterize the fundamental properties of optimal information structures, proving results that apply to both signal-rich and signal-poor environments. The tools we develop can be used to extend existing results in the Bayesian persuasion literature to coarse communication settings. This extends the range of interactions that Bayesian persuasion can be applied to, such as regulatory constraints on experiment designs in drug trials, technological or legislative constraints on communication complexity in settings such as targeted advertising, expert-layperson communication and advice seeking.

We show that the sender prefers to induce the most extreme beliefs that are feasible. We simplify the sender's optimization problem and show that it can be solved by a finite procedure, and describe achievable sender utilities using k-concavification. Through our example in targeted advertising, we demonstrate that receivers might prefer coarse communication, and fewer signals do not necessarily lead to less informative posteriors. Our model is therefore a novel and useful theoretical framework in analyzing settings in which the communication between parties can be limited by the receiver, or a regulator who cares about the aggregate welfare.

With this general model, we then analyze the properties of the marginal value of a signal for the sender. We prove a fundamental property of persuasion games with signal spaces of different sizes, and show that the loss in utility due to a limited signal space can be bounded below. Our lower bound result also implies that as a persuasion game approaches rich communication, additional signals become marginally less valuable for the sender. However in a finite setting the marginal value is not necessarily monotonically decreasing, as our example demonstrates.

We apply our framework to analyze how much a sender would be willing to pay for a larger signal space, focusing on specific preference structures. The belief-threshold games we study capture some of the most important questions studied in persuasion, such as firms sending product information, or lobbyists commissioning studies to convince politicians. In these threshold games, we show that precise communication is more valuable when desirable actions are more difficult to induce for the sender.

The framework we develop opens many avenues for future research. As we show through our extensions, our model is flexible enough to be applied to cheap talk settings and other models of Bayesian persuasion. Our framework is a useful to analyze the interaction between the value of commitment and the value of richer communication, and how these depend on the level of disagreement between agents. We can study competition between senders who have access to signal spaces with different degrees of complexity, or the problem of trying to persuade a heterogeneous set of agents using public or private signals with different degrees of coarseness. We leave these questions for future work.

Appendices

A Additional Results

We begin by providing additional results that will be referenced in the proofs of the statements in the main text.

Choquet Theorem for Simplices

Theorem (Choquet Theorem)**.** *Suppose that P is a metrizable compact convex subset of a locally convex Hausdorff topological vector space, and that μ_0 is an element of P. Then there is a probability measure τ on P which represents μ_0 i.e. $\sum_{p \in P} \tau(p)p = \mu_0$ s.t. $\mathsf{supp}(\tau) = \boldsymbol{Ext}(P)$, where $\boldsymbol{Ext}(P)$ denotes the extreme points of P. Furthermore, if $\boldsymbol{Ext}(P)$ is affinely independent, this probability measure τ is unique.*

Further Results on Properties of $\hat{u}^S$ and sender-preferred zones

Definition 4. *$S_a^{a'} \subset R_a$ denotes the region where the sender preferred action a' is taken in region R_a. Formally $S_a^{a'} \subset R_a$ is defined as $S_a^{a'} := \{\mu \in \Delta(\Omega) : \mu \in R_a \text{ and } a' \in \hat{A}(\mu)\ \hat{u}^S(a', \mu) \geqslant \hat{u}^S(\tilde{a}, \mu)\ \forall \tilde{a} \in \hat{A}(\mu)\}$.*

Remark**.** *Observe that by definition we have that $\forall a, a' \in A$ we have that $S_a^{a'} \subseteq S_{a'}^{a'}$.*

Lemma 9. *$\forall a, a' \in A$ $S_a^{a'}$ is closed and convex.*

Proof. We can define

$$S_a^{a'} = \left(\cap_{a' \neq a} \left\{ \mu \in R_a : \sum_{i < 0 \leqslant \Omega} \mu(\omega) \left(u^S(a, \omega) - u^S(a', \omega) \right) \geqslant 0 \right\}_{a' \in A(\mu)} \right),$$

which is intersection of finitely many half-spaces and closed, convex set R_a. ■

Lemma 10. *$\forall a, a' \in A$, $\hat{u}^S$ is an affine function over $S_a^{a'}$.*

Proof. For every posterior $\mu \in \Delta(\Omega)$ the receiver is indifferent between taking actions $a \in \hat{A}(\mu)$. For every $\mu \in S_a^{a'}$ receiver takes action a', by definition of sender preferred equilibrium. Given a fixed action a', $\hat{u}^S(a') = \mathbb{E}_\mu(u^S(a, \omega))$, which is affine over the simplex. ■

Corollary 4. *$\forall a \in A$, $\hat{u}^S$ is a continuous function over $\boldsymbol{int}(R_a)$.*

Remark. *$\hat{u}^S$ has jump discontinuities only at $\mu \in \Delta(\mu)$ such that $\mu \in R_a \cap R_{a'}$ with $R_a \cap R_{a'} = \boldsymbol{Bd}(R_a) \cap \boldsymbol{Bd}(R_{a'})$.*

Further properties of $\hat{u}^R$ and implications

Lemma 11. *In finite persuasion games, receiver utility in equilibrium: $\max_{a \in A} \hat{u}^R(a, \omega)$ is convex over $\Delta(\Omega)$. In fact, it is a polyhedral convex function.*

Proof. Observe that $\max_{a \in A} \hat{u}^R(a, \omega) = \max_{a \in A} \left\{ \{\mathbb{E}_\mu u^R(a', \omega)\}_{a' \in A} \right\}$. $\mathbb{E}_\mu u^R(a', \omega)$ denotes the expected utility for a fixed action $a' \in A$, which is an affine function over $\Delta(\Omega)$, and therefore convex. Then we have that epigraph of $\max_{a \in A} \hat{u}^R(a, \omega)$ is a polyhedral convex set. [20] ■

An immediate implication is the following.

Corollary 5. *Let τ be the optimal information structure with k-signals and τ' be the optimal information structure with with $k+1$ signals. If τ and τ' are Blackwell comparable we have that receiver prefers τ' over τ.*

The corollary follows from the definition of Blackwell comparability, and the fact that the receiver preferences must be convex.

B Proofs of Statements in the Main Text

Proof of Proposition 2

Let τ be the optimal information structure solving the sender's maximization problem, and suppose for a contradiction, $\sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} \neq \mathbb{E}_\tau \hat{u}^S$.

For the first case, let $\sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} < \mathbb{E}_\tau \hat{u}^S$. However, taking the beliefs in $\mathsf{supp}\,(\tau) = \{\mu_1, \ldots, \mu_k\}$, we know that by the feasibility of τ, $\exists \{\tau(\mu_1), \ldots, \tau(\mu_k)\} \in \Delta(\Delta(\Omega))$ such that $\sum_{i \leqslant k} \tau(\mu_i)\mu_i = \mu_0$ and $\sum_{i \leqslant k} \tau(\mu_i) = 1, 1 \geqslant \tau(\mu_i) \geqslant 0$. Thus, by definition 1, $(\mu_0, \mathbb{E}_\tau \hat{u}^S) \in \mathbb{CH}_k(\hat{u}^S)$. Therefore, we cannot have $\sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} < \mathbb{E}_\tau \hat{u}^S$.

[20] f is a polyhedral convex function if and only if its epigraph is polyhedral, as defined in Rockafellar (1970).

For the other case, let $\sup\{z|(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} > \mathbb{E}_\tau \hat{u}^S$. Since $(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)$, take the set of points $\{\hat{u}^S(\mu_1), \ldots, \hat{u}^S(\mu_k)\}$ and convex weights $\{\alpha_1, \ldots, \alpha_k\}$ with $\sum_{i\leqslant k} \alpha_i \mu_i = \mu_0$ and $\sum_{i\leqslant k} \alpha_i \hat{u}^S(\mu_i) = z$, also satisfying $\sum_{i\leqslant k} \alpha_i = 1, 1 \geqslant \alpha \geqslant 0$. We know these points and weights must exist by definition 1. Now observe that $\tau' = \{\mu_1, \ldots, \mu_k\}$ must be a feasible solution to the sender's maximization problem. We know that τ' satisfies Bayes plausibility by the definition given with the weights α_i. Therefore τ' could have been picked instead of τ in the sender's maximization problem, contradicting the optimality of τ.

□

Proof of Lemma 2

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

Proof of Lemma 3

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

Proof of Lemma 4

Let $\mathsf{supp}\,(\tau) = \{\mu_1, \ldots, \mu_k\}$ be affinely dependent. Then, there must exist $\{\lambda_1, \ldots, \lambda_k\}$ such that $\sum_{i\leqslant k} \lambda_i = 0$ and $\sum_{i\leqslant k} \lambda_i \mu_i = 0$. 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 $\forall i \neq j^*$, $\left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_j)}{\lambda_{j*}}\right) \geqslant 0$ and $\forall i \neq p^*$, $\left(\tau(\mu_i) - \lambda_i \frac{\tau(\mu_k)}{\lambda_{p*}}\right) \geqslant 0$.
If $\lambda_i = 0$, the inequalities hold trivially.
If $\lambda_i > 0$, the inequalities are equivalent to $\frac{\tau(\mu_i)}{\lambda_i} \geqslant \frac{\tau(\mu_{j*})}{\lambda_{j*}}$ and $\frac{\tau(\mu_i)}{\lambda_i} \geqslant \frac{\tau(\mu_{p*})}{\lambda_{p*}}$. 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 $\frac{\tau(\mu_i)}{\lambda_i} \leqslant \frac{\tau(\mu_{j*})}{\lambda_{j*}}$ and $\frac{\tau(\mu_i)}{\lambda_i} \leqslant \frac{\tau(\mu_{p*})}{\lambda_{p*}}$. 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:

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

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

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

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

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:

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

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. □

Proof of Lemma 5

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 $\mu_1^{''}, \mu_2'$ that satisfy the following condition:

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

With any $\mu_1^{''}, \mu_2'$ 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 $(\mu_1^{''}, \mu_2')$ that satisfy equation 2, 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:

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

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

$$\iff$$

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

$$\iff$$

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

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 $\mu_1', \mu_1^{''}, \mu_2', \mu_2^{''}$:

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

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

$$\Longleftrightarrow$$

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

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

$$\Longleftrightarrow$$

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

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

We can further prove the following related claim:

Claim 1. *Let $|\Omega| = n$ and $|A| = k$. Suppose we have an information structure τ with supp(τ) $= \mu = \{\mu_1, \ldots, \mu_k\}$ 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 4, 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 $q \geqslant (n-k)$. 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:

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

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

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 4, 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 2

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 $\mu_0\in\text{co}(\mu_{-k})$ 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 $(k-1)$ signals.

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

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 $M=\{\mu_0=\sum_{i=1}^{k-1}(\alpha_i\mu_i+\mu_0)|\alpha_i\geqslant 0\forall i\in\{1,\ldots,k-1\}\}$ 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 4 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 5 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 5, 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 $\{\mathcal{R}_a\}_{a \in A} = \{R_a \cap S\}_{a \in A}$.

Proof of Proposition 3

Suppose τ_k is the optimal information structure with k signals, and τ_{k-1} is the optimal information structure with $k-1$ signals. Denote by $V^*(k), V^*(k-1)$ the utilities obtained using these information structures.

Let $\mathsf{supp}\,(\tau_k) = \{\mu_1, \dots, \mu_k\}$. Observe that we can create a $k-1$ dimensional information structure that maintains Bayes plausibility by choosing two posteriors, say μ_1, μ_2, and define a new posterior as their mixture:

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

And define the new information structure with $\mathsf{supp}\,(\tau'_{12}) = \{\mu_{12}, \mu_3, \dots, \mu_k\}$, which maintains Bayes plausibility with the new weights $\{(\tau_k(\mu_1) + \tau_k(\mu_2)), \tau(\mu_3), \dots, \tau(\mu_k)\}$.

Now, we can define k different information structures containing $k-1$ posteriors each, denoted $\mu_{12}, \mu_{23}, \dots, \mu_{k-1,k}, \mu_{k1}$ where we mix the consecutive posteriors μ_l, μ_{l+1} and use the weights defined above to satisfy Bayes plausibility. By the optimality of τ_{k-1} among the information structures with $k-1$ signals, we must have the following k inequalities:

$$\begin{aligned}
V^*(k-1) \geqslant (\tau_k(\mu_1) + \tau_k(\mu_2))u^S\left(\frac{\tau_k(\mu_1)}{\tau_k(\mu_1) + \tau_k(\mu_2)}\mu_1 + \frac{\tau_k(\mu_2)}{\tau_k(\mu_1) + \tau_k(\mu_2)}\mu_2\right) \\
+ \tau_k(\mu_3)u^S(\mu_3) + \dots + \tau_k(\mu_k)u^S(\mu_k), \\
V^*(k-1) \geqslant \tau_k(\mu_1)u^S(\mu_1) + (\tau_k(\mu_2) + \tau_k(\mu_3))u^S\left(\frac{\tau_k(\mu_2)}{\tau_k(\mu_2) + \tau_k(\mu_3)}\mu_2 + \frac{\tau_k(\mu_3)}{\tau_k(\mu_2) + \tau_k(\mu_3)}\mu_3\right) \\
+ \dots + \tau_k(\mu_k)u^S(\mu_k), \\
\vdots \\
V^*(k-1) \geqslant \tau_k(\mu_1)u^S(\mu_1) + \dots + \\
(\tau_k(\mu_{k-1}) + \tau_k(\mu_k))u^S\left(\frac{\tau_k(\mu_{k-1})}{\tau_k(\mu_{k-1}) + \tau_k(\mu_k)}\mu_{k-1} + \frac{\tau_k(\mu_k)}{\tau_k(\mu_{k-1}) + \tau_k(\mu_k)}\mu_k\right),
\end{aligned}$$

$$\begin{aligned}
V^*(k-1) \geqslant \tau_k(\mu_2)u^S(\mu_2) + \tau_k(\mu_3)u^S(\mu_3) + \dots + \\
(\tau_k(\mu_1) + \tau_k(\mu_k))u^S\left(\frac{\tau_k(\mu_1)}{\tau_k(\mu_1) + \tau_k(\mu_k)}\mu_1 + \frac{\tau_k(\mu_k)}{\tau_k(\mu_1) + \tau_k(\mu_k)}\mu_k\right)
\end{aligned}$$

Dividing all inequalities by k and summing up, we have:

$$V^*(k-1) \geqslant \frac{k-2}{k}V^*(k) + \frac{2}{k}V' \geqslant \frac{k-2}{k}V^*(k)$$

Where V' is the utility gained from the k dimensional information structure consisting of the posteriors $\{\mu_{12}, \mu_{23}, \ldots, \mu_{k-1,k}, \mu_{k1}\}$. This implies the following upper bound on the value of an additional signal at $k-1$ signals:

$$V^*(k) - V^*(k-1) \leqslant \frac{2}{k} V^*(k)$$

Equivalently, the following relationship must hold between the maximum utilities attainable between k and $k-1$ signals:

$$\frac{k-2}{k} V^*(k) \leqslant V^*(k-1) \leqslant V^*(k)$$

□

When the sender utility u^S can be negative and has the infimum $\underline{u^S}$, the above inequalities can be equivalently stated as follows:

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

$$\frac{k-2}{k} V^*(k) + \frac{2}{k} \underline{u^S} \leqslant V^*(k-1) \leqslant V^*(k).$$

Proofs of the statements in section 4.1

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 $\mu_s(\omega_i) \geqslant T$, 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 $\mathbb{E}_{\mu_s}[u^R(a_i,\omega)] \geqslant \mathbb{E}_{\mu_s}[u^R(a_j,\omega)]$ 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:

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

where $\cdot$ denotes the Euclidean dot product.

Proof of Lemma 6

Let us first characterize the set Δ_c. We have[21] $\Delta_c = \Delta(\Omega)\backslash \mathrm{co}_2(R_1 \cup R_2 \cup R_3))$. We note that:

$$\begin{aligned}\mathrm{co}(R_1 \cup R_2) =& \mathrm{co}(\{\omega_1, (T, 1-T, 0), (T, 0, 1-T), \omega_2, (1-T, T, 0), (0, T, 1-T)\}) \\ =& \mathrm{co}\{\omega_1, (T, 0, 1-T), \omega_2, (0, T, 1-T)\} \end{aligned} \tag{3}$$

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

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

$$\mathrm{co}(R_2 \cup R_3) = \mathrm{co}\{\omega_2, (1-T, 0, T), \omega_3, (1-T, 0, T)\} \tag{5}$$

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 $\mathrm{co}(\{\omega_1, (T, 1-T, 0), (T, 0, 1-T), \omega_2, (1-T, T, 0), (0, T, 1-T)\})$. Similarly using equation (3), (4) and (5), $\mathrm{co}(R_i \cup R_j)$ can be identified as the inter-

[21]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.

section of a half space and the simplex i.e.

$$\text{co}(R_1 \cup R_2) = \Delta(\Omega) \cap \{(\mu - (T, 0, 1-T)) \cdot (-T, T, 0) \geqslant 0 | \mu \in \mathbb{R}^3\} \quad (6)$$
$$\text{co}(R_1 \cup R_3) = \Delta(\Omega) \cap \{(\mu - (T, 1-T, 0)) \cdot (-T, 0, T) \geqslant 0 | \mu \in \mathbb{R}^3\} \quad (7)$$
$$\text{co}(R_2 \cup R_3) = \Delta(\Omega) \cap \{(\mu - (1-T, T, 0)) \cdot (0, -T, T) \geqslant 0 | \mu \in \mathbb{R}^3\} \quad (8)$$

So we can define $\Delta_c \subset \Delta(\Omega)$ as $\Delta_c = \Delta(\Omega) \backslash \text{co}_2(R_1 \cup R_2 \cup R_3)$. By (6), (7) and (8) we can see that Δ_c is defined as

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

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

Proof of Lemma 7

We can identify the upper bounds through the following problem:

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

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 $\mu_0 \notin \Delta_c$ 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:

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

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 3

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 $\mu_0 \in \Delta_c$, $V(2,\mu_0)$ 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 signal and $V(2,\mu_0) < \frac{1}{2}$ implies increasing marginal value of a signal. □

Proof of Proposition 4

We will first establish a series of lemmas that shows the connection between choosing k-dimensional dimensional information structures in a belief space in R^n, and optimally compressing n states to k states, and then solving a Bayesian persuasion problem in the new belief space in R^k. After showing this result for Bayesian persuasion games, the statement in proposition 4 immediately follows as a corollary.

Lemma 12.

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

achieves the same optimal value with the problem:

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

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

Proof. $\mathcal{T}_k$ is homemorphic to the space that parameterizes all k-dimensional linear subspaces of the n-dimensional vector space which is called the Grassmannian space, which we will denote $\mathcal{G}_k(\mathbb{R}^n)$. The homeomorphism is obtained by subtracting μ_0 from each line equation.

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 $\mathbb{R}^n$. Define the Steifel manifold $\mathcal{V}_k(\mathbb{R}^n)$ as the set of all orthonormal k-frames[22] of $\mathbb{R}^n$. 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[23] i.e. $(\mathbb{S}^{n-1})^k$. Using this representation, we can use the inherited topology from $R^{n \times k}$ when discussing the compactness of $\mathcal{V}_k(\mathbb{R}^n)$. Noting that it is a closed subspace of a compact space, we can easily conclude the Steifel manifold $\mathcal{V}_k(\mathbb{R}^n)$ is compact.

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) 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.[24]

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 $C(X)$ de-

[22] 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

[23] $\mathbb{S}^{n-1} = \{x \in \mathbf{R}^n : \|x\| = 1\}$.

[24] 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.

notes 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 $\mathcal{G}_k(n)$. ■

We call the projections from $\Delta(\Omega)$ onto the flat $T \in T_k$ 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 14. *The optimal value of the maximization problem:*

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

is upper semi-continous in T.

Proof.

We will start with discussing some preliminary facts. The maximum and hence the value function $V(T)$ exists by the results in Kamenica and Gentzkow (2011), since the sender is solving a full dimensional Bayesian persuasion problem over T which is shown to be homeomorphic to $\mathbb{R}^k$.

Let μ_T be the optimal information structure on the flat T that is represented with the parallel subspace W and projection matrix X_T. Let $\mu_{T'}$ be the optimal information structure on the flat T' represented with the parallel subspace W' and projection matrix $X_{T'}$. The statement of the lemma is formally $\forall \epsilon > 0$, there exists a $\delta > 0$ such that whenever we have $|X_T - X_{T'}| < \delta$, we get $V(\mu_{T'}) \leqslant V(\mu_T) + \varepsilon$.

Finally, 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 $V(\mu') \leqslant V(\mu) + \epsilon$.

Now observe that:

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

Define M_T and $M_{T'}$ as information structures consisting of vectors $\{m_T | m_T \in T \cap \mathbf{Bd}\Delta(\Omega)\}$

and $\{m_{T'} | m_{T'} \in T' \cap \mathbf{Bd}\Delta(\Omega)\}$. We will show:

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

Let us first show that $||(X_T - X_{T'})\tilde{\mu}|| \geqslant \gamma ||M_T - M_{T'}||$. First, by definition of matrix norm $||(X_T - X_{T'})\tilde{\mu}|| \geqslant ||M_T - M_{T'}||_{\max} = \max_{r \in R} ||m_T^r - m_{T'}^r||_2$. By equivalence of finite dimensional norms, there exists a constant γ such that $||M_T - M_{T'}||_{\max} \geqslant \gamma ||M_T - M_{T'}||$. Hence, we obtain that $||(X_T - X_{T'})\tilde{\mu}|| \geqslant \gamma ||M_T - M_{T'}||$.
Now let us turn to the last inequality $\gamma ||M_T - M_{T'}|| \geqslant \gamma ||\mu_T - \mu_{T'}||$. 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 $V(\mu') \leqslant V(\mu) + \epsilon$. Then for each $\varepsilon > 0$ one can pick $\delta = \frac{1}{\gamma}\delta_\varepsilon$ to ensure that

$$\frac{1}{\gamma}\delta_\varepsilon > ||X_T - X_{T'}|| \geqslant ||\mu_T - \mu_{T'}||.$$

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 $V(\mu_{T'}) \leqslant V(\mu_T) + \varepsilon$. ■

By above lemmas, the existence of the optimal for the second maximization problem in lemma 12 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 12, it is straightforward to show that the two maximization problems yield the same maximum. Let μ_1 be the maximizer of equation (9) and μ_2 be the maximizer of equation (10). We show that $V(\mu_1) = V(\mu_2)$ where V is the value function. Suppose not, let $V(\mu_1) > V(\mu_2)$. But then in the second problem, we could have picked $T_{\mu_1} = \text{aff}(\mu_1)$ where aff denotes affine hull, and $\mu = \mu_1$ to get a higher value, contradicting the optimality of μ_2. Now suppose $V(\mu_1) < V(\mu_2)$, but then directly picking μ_2 in the first problem yields a better payoff, contradicting to the optimality μ_1 in the first problem. ■

Having established all these results, the proof of proposition 4 follows from Lipnowski and Ravid (2020) and the result in lemma 12. 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 $\{\tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_{\mu\sim\tau}\mu = \mu_0 \text{ and } |\mathsf{supp}\,\tau| \leqslant k\}$ is a closed subset of $\{\tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_{\mu\sim\tau}\mu = \mu_0\}$. The equivalence follows from lemma 12 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 $\hat{u}^s$.

References

Akyol, E., C. Langbort, and T. Basar (2016): "Information-theoretic approach to strategic communication as a hierarchical game," *Proceedings of the IEEE*, 105, 205–218.

Alonso, R. and O. Camara (2016): "Bayesian persuasion with heterogeneous priors," *Journal of Economic Theory*, 165, 672–706.

Aumann, R. J. and S. Hart (2003): "Long cheap talk," *Econometrica*, 71, 1619–1660.

Aumann, R. J. and M. Maschler (1995): *Repeated Games with Incomplete Information*, MIT Press.

Aybas, Y. and S. Callander (2022): "Efficient Cheap Talk in Complex Environments," .

Ball, I. and J.-A. Espín-Sánchez (2021): "Experimental Persuasion," .

Bergemann, D. and S. Morris (2019): "Information design: A unified perspective," *Journal of Economic Literature*, 57, 44–95.

Bloedel, A. W. and I. R. Segal (2018): "Persuasion with Rational Inattention," *Available at SSRN 3164033*.

Blume, A. (2000): "Coordination and learning with a partial language," *Journal of Economic Theory*, 95, 1–36.

Blume, A. and O. Board (2013): "Language barriers," *Econometrica*, 81, 781–812.

Chakraborty, A. and R. Harbaugh (2010): "Persuasion by cheap talk," *American Economic Review*, 100, 2361–82.

De Jaegher, K. (2003): "A game-theoretic rationale for vagueness," *Linguistics and Philosophy*, 26, 637–659.

Dughmi, S., D. Kempe, and R. Qiang (2016): "Persuasion with limited communication," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, ACM, 663–680.

Gentzkow, M. and E. Kamenica (2014): "Costly persuasion," *American Economic Review*, 104, 457–62.

Hagenbach, J. and F. Koessler (2020): "Cheap talk with coarse understanding," *Games and Economic Behavior*, 124, 105–121.

Ichihashi, S. (2019): "Limiting Sender's information in Bayesian persuasion," *Games and Economic Behavior*, 117, 276–288.

Jager, G., L. P. Metzger, and F. Riedel (2011): "Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals," *Games and economic behavior*, 73, 517–537.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kolotilin, A. (2015): "Experimental design to persuade," *Games and Economic Behavior*, 90, 215–226.

Le Treust, M. and T. Tomala (2019): "Persuasion with limited communication capacity," *Journal of Economic Theory*, 104940.

Lipnowski, E. and L. Mathevet (2017): "Simplifying Bayesian Persuasion," Tech. rep., mimeo.

——— (2018): "Disclosure to a psychological audience," *American Economic Journal: Microeconomics*, 10, 67–93.

Lipnowski, E. and D. Ravid (2020): "Cheap talk with transparent motives," *Econometrica*, 88, 1631–1660.

Rayo, L. and I. Segal (2010): "Optimal information disclosure," *Journal of political Economy*, 118, 949–987.

Rockafellar, R. T. (1970): *Convex analysis*, vol. 28, Princeton university press.

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

Tardella, F. (2008): "Existence and sum decomposition of vertex polyhedral convex envelopes," *Optimization Letters*, 2, 363–375.

Tsakas, E. and N. Tsakas (2018): "Noisy persuasion," *Available at SSRN 2940681.*

Volund, R. T. (2018): "Bayesian Persuasion on Compact Subsets," Tech. rep., Aarhus University.

Wei, D. (2018): "Persuasion Under Costly Learning," .