Persuasion with Coarse Communication*

Yunus C. Aybas[†] ⓡ Eray Turkel[‡]

February 1, 2021

Abstract

Communication between parties is an important part of persuasion. The difficulty of communication has a first order effect on the outcomes of strategic persuasion games and welfare of agents. We study a model of Bayesian persuasion in which the communication between the sender and the receiver is coarse, which we model by allowing the cardinality of the signal space to be less than the cardinality of the action space and the state space. This limits the number of action recommendations that the sender can make. We define the set of *simple* Bayes plausible information structures and show that there exists an optimal information structure that is simple for a signal space with arbitrary finite cardinality. We characterize the set of highest attainable payoffs for any cardinality of the signal space k, by what we call k-concavification of the sender utility. Under coarse communication, sender's willingness to pay for an additional signal can be interpreted as the value of precise communication for the sender. We provide an upper bound on the value of precision. While increased precision is always better for the sender, we show that the receiver may prefer coarse communication. We show this by analyzing a game of advice seeking, where the receiver can choose the cardinality of the signal space and ask for simple recommendations.

Keywords: Bayesian Persuasion; Information Design; Value of Precision; Coarse Communication; Information Structures
JEL Classification: D82, D83

*We are grateful to Avidit Acharya, Steven Callander, Mine Su Erturk, Françoise Forges, Matthew Gentzkow, Edoardo Grillo, Matthew Jackson, Semih Kara, Tarik Kara, Emin Karagozoglu, Cem Tutuncu, Robert Wilson, Kemal Yildiz, and seminar participants in Stanford University and Bilkent University 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

Communication is difficult. This is especially true when the content is very complicated, and the messages are relayed through coarse or imperfect channels. Financial analysts make recommendations to their clients about taking long or short positions on assets. Some firms give simple recommendations such as 'buy', 'sell' or 'hold', while others give more fine-grained advice, including 'strong buy' or 'strong sell'. Credit rating agencies use letter grades with the goal of communicating the riskiness of an investment. In general, the set of messages available to the sender may prove insufficent to describe all states of the world or actions. What is the effect of this coarseness in communication on agents who are interacting strategically? When is it possible for one person to persuade another to change her action by communicating through coarse channels? Under what conditions would it be optimal to send or receive information through coarse channels?

Motivated by these questions, this paper studies an information design problem between two rational agents, a sender and a receiver, in a setting with possibly coarse communication. We say that communication is coarse when the underlying space of possible states of the world is large relative to the set of signals that can be used to describe it.

The framework of Bayesian persuasion (Kamenica and Gentzkow 2011), and in general, information design (Bergemann and Morris 2016), analyzes strategic communication between agents who might have misaligned preferences. In the canonical model, the sender designs the informational environment of the receiver through a signaling scheme, creating beliefs that will induce desirable actions. The ability of the sender to commit to a signaling scheme distinguishes this framework from the literature on cheap talk, where similar restrictions on signal spaces have been studied by Jager et al. (2011), among others.

We will start by giving an overview of our contributions to the rapidly expanding field of information design and Bayesian persuasion, and then provide a review of related work to explain how our results complement the existing literature.

First, we generalize the Bayesian persuasion framework to settings where the signal space is limited in its cardinality. The standard model assumes the existence of a signal space that is large enough to describe all states of the world perfectly, or induce all possible actions, depending on which one of the constraints is binding. We show the existence a sender-optimal information structure. We then define the concept of *simple* information structures which (i) have an affinely independent support and (ii) make the receiver indifferent between multiple actions or give zero weight to some states of the world. We show that there exists an optimal information structure that is simple. Additionally, we show that optimal information structures that are simple can be found by maximizing over the set of posteriors, instead of the set of distributions over posteriors, since for simple information structures, the probability distribution that makes them Bayes plausible is uniquely determined.

Second, we extend the concavification approach in the literature. Aumann and Maschler (1995) define the concavification of a function f as the supremum of the convex hull of the hypograph of f. Under the assumption that the signal space is large enough, Kamenica and Gentzkow (2011) show that concavification of sender utility describes the highest payoff sender can achieve for every prior. We define the related concepts of *k-concavification* of a function and *k-convex hull* of a set. The *k-convex hull* of a set Λ is defined as all the elements that can be represented as the convex combination of at most k elements in Λ. Whenever $\Lambda \subseteq \mathbb{R}^d$ and $k \geqslant d+1$, the k-convex hull coincides with convex hull by Caratheodory Theorem. We define k-concavification of sender utility as the supremum of the k-convex hull of the hypograph of sender utility. We show that the set of achievable utilities for a sender in a game where the signal space has cardinality $|S| = k$ is given by the k-concavification of the sender utility. Additionally, we show that k-concavification of sender utility is quasiconcave.

Third, we analyze the effects of coarse communication for the sender and the receiver. We show that a larger signal space always weakly improves the sender's utility, so a sender would be willing to pay to get access to an additional signal. We call the sender's willingness to pay for an additional signal *the value of precision*, and provide an upper bound for it. The upper bound result is derived by using a novel insight linking higher and lower dimensional information structures: more precisely, given a higher dimensional information structure, we can combine some of the induced posteriors while still maintaining Bayes plausibility, and create lower dimensional information structures. Doing this in a systematic way, we can show an upper bound on the gap in achievable utilities under different cardinality constraints. We also analyze how the price of precision depends on the location of the prior, and the difficulty (for the sender) of inducing beneficial actions while maintaining Bayes plausibility. We show that the value of precision can be non-monotonic: e.g. the second signal can be more valuable than the third one, or vice versa.

Next, we show that the effect of additional signals on receiver's utility is ambiguous in general. We analyze a game of *optimal advice seeking*, where the receiver can choose the size of the signal space. Intuitively, this is a setting where the receiver can ask for simple or complicated recommendations from the sender. This framework can capture situations where the receiver has some power over the communication procedure. We show through an example that there exists equilibria where the receiver optimally chooses to ask for 'simple advice' with fewer action recommendations. Through an example, we also show that restricting the cardinality of the signal space might not lead to less informative information structures, in the sense of Blackwell informativeness.

Previous work on persuasion games has introduced costs for generating precise information structures, where the costs are usually motivated through information theoretic foundations, e.g. Gentzkow and Kamenica (2014) assume that the costs are proportional to the reduction in the entropy of prior beliefs. This approach still allows the sender to make arbitrarily many action recommendations subject to a cost, and the existence results rely on having a high dimensional signal space. Similarly, limitations to the informativeness of posteriors in a persuasion game can rise endogenously in a setting where the receiver has mental costs associated with processing more informative signals. This phenomenon has been analyzed under various specific preference structures, where the sender chooses to induce less informative posteriors due to increasing costs for paying attention to informative signals on the receiver side (Wei 2018; Bloedel and Segal 2018; Lipnowski and Mathevet 2018).

While we assume exogenous restrictions on the signal space to prove our main results, we provide multiple applications in section 4, where our model can be used to analyze settings in which limitations on the signal space can arise endogenously. Our analysis of advice seeking games, where a receiver determines the cardinality of the signal space, is similar to the setting with binary states and signals analyzed in Ichihashi (2019), in which the receiver limits the Blackwell informativeness of the signals. As we will see in one of our examples, optimal information structures under different cardinality constraints are not always Blackwell comparable. Hence, using Blackwell informativeness constraints and cardinality constraints will lead to different outcomes in general.

In a related paper to ours, Dughmi et al. (2016) examine the properties of a persuasion game with a restricted number of signals, but in the specific context of bilateral trade with assumptions on the underlying preference structure. They also prove the NP-hardness of approximating optimal sender utility in general persuasion games with coarse communication. Our focus is on characterizing the properties of the sender-optimal information structures and analyzing the various implications of coarse communication rather than the computational complexity of calculating the equilibrium sender utility.

Two recent papers analyze noisy persuasion games with similar motivating questions. Le Treust and Tomala (2019) study a repeated game of persuasion, where the sender has limited opportunities to intervene and send information through a noisy and cardinality-constrained channel. While they don't prove the existence of a maximum, their main result is an upper bound on achievable utilities by the sender. They also show that this bound is reached in the limit where the number of repetitions of the underlying game approaches infinity. Their result can be modified to apply to our setting with noiseless channels and a coarse signal space, giving an upper bound on the achievable utility of the sender. This asymptotic result is shown by making an elegant connection to Shannon's coding Theorem. Similarly, Tsakas and Tsakas (2018) focus on persuasion through noisy communication channels in a single persuasion game. They show that the effect of noise on sender utility is ambiguous in general, and within the class of symmetric noisy communication channels, more noise makes the sender worse off.

Our theoretical results complement the asymptotic framework of Le Treust and Tomala (2019): we focus on a single game and show the existence of an optimal information structure and characterize its properties. This result simplifies the search for optimal information structures and also enables us to solve for optimal signals that recommend the fewest possible actions. We also provide an upper bound result on achievable utilities with cardinality constraints, which provides a bound on the loss of utility due to coarseness in communication that applies to all finite persuasion games. Coarse communication always makes the sender worse off, as opposed to the case with noise where the effect is ambiguous, as is shown in Tsakas and Tsakas (2018). Our analyses of the value of precision and games of advice seeking also provide substantive applications for constrained persuasion games in various market settings.

While noisy channels make communication between parties more difficult, the restrictions on implementable information structures are different compared to cardinality constraints. Le Treust and Tomala (2019) show that in the case where the same persuasion game is repeated infinitely many times, all that matters for sender utility is the channel's capacity, which is affected both by the inherent noise in communication and the cardinality of the signal space. However, noisy and coarse signals have substantively different implications on the optimal information structure that will be chosen and achievable utilities for the sender in finite games. Noise prevents the sender from inducing posteriors where the receiver is certain about the state of the world, and there are no explicit restrictions on the number of inducable actions. Thus, the receiver can never be perfectly informed and there is always residual uncertainty in beliefs. With cardinality constraints, while the sender can induce informative posteriors, it's never possible to perfectly inform the receiver about *all* states of the world at the same time. Thus, the sender has to prioritize some of the actions that can be induced with its limited capabilities while also maintaining Bayes plausibility, which leads to different outcomes.

Mathematically, with noisy channels, the sender's choice is restricted to information structures in which posteriors are not too close to the extreme points of the simplex. With cardinality constraints, there are no restrictions on the locations of the posteriors, but the sender's problem reduces to optimally choosing a lower dimensional object embedded in a higher dimensional probability simplex (i.e., a line segment within the 3-simplex, or a triangle within the 4-simplex). This is also why we cannot use the intuitive concavification approach directly in our setting. Suppose the signal space is constrained to have cardinality k. It will not be possible to achieve all utility levels on the convex hull of the epigraph of sender utility. Specifically, if a utility level can only be achieved as a convex combination of more than k points, it will not be implementable in our setting. This insight will be clarified when we define the concept of a k-convex hull.

Finally, in a recent paper, Lipnowski and Mathevet (2017) show a way to simplify the Bayesian persuasion problem. Their results imply that there exists an optimal Bayes plausible information structure that is simple when signal space is rich i.e. $k \geqslant \min\{n, m\}$. However, their result critically relies on $k \geqslant \min\{n, m\}$, which enables them to use representation theorems of Krein-Milman and Caratheodory. There are two main differences between our results and theirs. First, we extend their result to settings with coarse communication, i.e. $k < \min\{n, m\}$. Second, we provide constructive proof that gives an algorithm to (weakly) improve any information structure that is not simple.

The rest of the paper is organized as follows. Section 2 provides a simple example and highlights some of the insights that will be analyzed in the rest of the paper. We introduce our model, provide the existence results, and the concept of k-concavification in section 3. Section 4 provides applications for our model, where we analyze the value of precise communication and optimal advice seeking. We conclude in section 5. All proofs and additional results appear in the appendix. In appendix we also provide additional result pertaning the existence of optimal information structure for continuum of states, equivalent full dimensional bayes persuasion games and equivalent chep talk games.

2 A simple example: Financial Advice

We begin by analyzing a simple example with 3 states and 3 actions. The sender is a financial institution and the receiver is a risk neutral customer, looking for advice on a financial position. The customer has a fixed budget to invest, and can take a long or short position on a risky asset, or avoid the risky asset and invest in a risk free bond. The value of the asset can increase, in which case the optimal action is to take a long position, it can decrease, in which case the optimal action is a short position, or it could hold steady, in which case the optimal action is avoiding the risky asset and investing in a risk free bond. Suppose for simplicity, that the value of the risky asset can increase or decrease by 1. The risk free bond provides a minimal return of $r = 0.3$. In addition, the institution can charge commissions on the transactions of the risky asset to the customer, denoted by c, which can be any real number. The payoff of the institution is the commission it can charge. The payoff for the customer from investing in the risky asset is 1 if the correct position is taken, and -1 if the incorrect position is taken, minus the commissions. The payoff from avoiding the asset and investing in the risk free bond is r.

Let p_+, p_0, p_- denote the common beliefs that the asset's value will increase, hold steady, or decrease, respectively, where $p_+ + p_0 + p_- = 1$. The sender (financial institution) and the receiver (customer) share a prior μ_0 which is in the interior of the three dimensional simplex. Sender commits to a signaling mechanism, using signals from a finite set S, where $|S| = 3$. As is usual in the Bayesian persuasion literature, by a signaling mechanism, we mean a collection of probability measures over S, one for each realization of the (uncertain) state of the world. The sender commits to this strategy prior to the realization of the state, and cannot change it afterwards. The receiver observes the signal (not the state of the world), and uses Bayesian updating to obtain posterior probabilities of each state conditional on the observed signal. It should be noted that signals $s \in S$ do not have an intrinsic meaning, but obtain their meaning in equilibrium via the announced signaling mechanism. After the signal is realized and the posterior beliefs are formed, the sender decides on the commission that will be charged. Finally, the receiver chooses their action.

Formally, the receiver's expected payoff will be $p_+ - p_- - r - c$ when taking a long position, and $p_- - p_+ - r - c$ when taking a short position, and r when investing in the risk free bond. For any given belief, the receiver will choose to take a position over avoiding the risky asset if and only if $|p_- - p_+| - r - c \geqslant 0$. The sender can therefore extract all the surplus by optimally setting $c = |p_- - p_+| - r$. Note that the commissions will be higher if the posterior beliefs approach the extreme points of the simplex at which the receiver chooses to take a position. Intuitively, the financial institution can charge higher commissions for inducing more precise posteriors, from customers that will have very optimistic or very pessimistic beliefs about the risky asset.

With $|S| = 3$, finding the optimal information structure and calculating the maximum sender utility achievable is easily done by inspecting the concavification of sender utility. The optimal information structure will induce beliefs on the extreme points of the simplex: the customer is absolutely sure about what will happen to the asset when he receives a signal. Letting the prior be $\mu_0 = (p_+, p_0, p_-) = (0.3, 0.4, 0.3)$, the optimal information structure will induce $(1, 0, 0)$ with probability 0.3, $(0, 1, 0)$ with probability 0.4, and $(0, 0, 1)$ with probability 0.3. In this equilibrium, we can intuitively interpret the signals leading to these posteriors as: strong buy, avoid and strong sell, as each signal perfectly reveals the state.

With $|S| = 2$, solving for the optimal information structure is not straightforward: we can no longer use concavification. It turns out that the optimal information structure in this case induces the belief $(1, 0, 0)$ with probability 0.3, and induces the belief $(0, 0.57, 0.43)$ with probability 0.7. In this equilibrium, we can interpret the signals leading to these posteriors as: strong buy and sell, as the first signal perfectly reveals the first state (the value of the risky asset is increasing) and the second signal pools the second and third states (the value will remain constant or decrease). The customer is still willing to take a short position after observing the second signal, but the beliefs are now less extreme compared to the case with three signals. The resulting equilibrium level of the commission and expected utility for the institution is therefore lower.

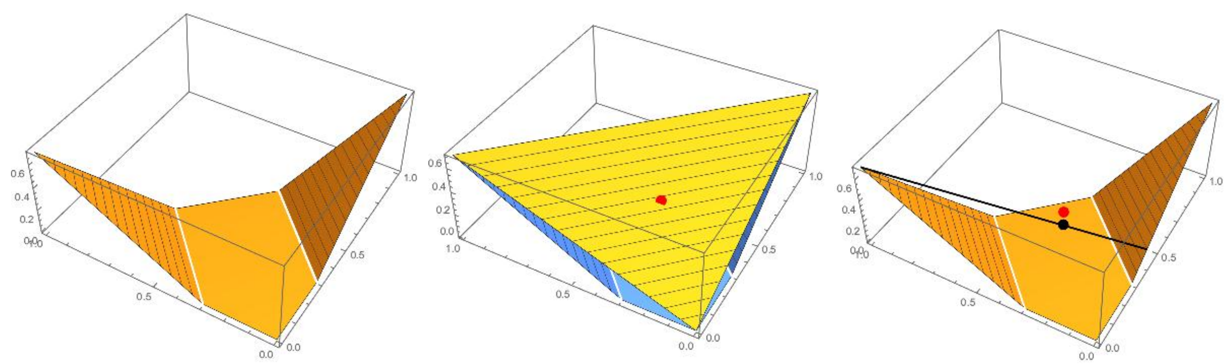

Figure 1: The financial advice example. The first plot shows the sender utility function over the simplex. The second plot shows the concavification of sender utility, where the red dot corresponds to the maximum utility achievable by the sender. The third plot shows the optimal information structure for the sender using two signals. The black dot corresponds to the maximum utility achievable with 2 signals, shown together with the maximum utility achievable with 3 signals.

The example demonstrates some of the key insights that will be generalized in this paper. First, the search for an optimal information structure is equivalent to searching for the highest value achievable by taking the convex combination of two points from the graph of the sender utility function. We formalize this insight in section 3.2. Second, the optimal information structures in this example are simple. The induced beliefs are located on the boundaries of the regions where the receiver's action is fixed and are the set of induced posteriors are affinely independent. This is developed further in section 3.3. Third, the utility achievable by the sender is lower with two signals, hence the sender would be willing to pay to get access to additional signals. The loss in sender utility will be defined as the 'value of precision' in section 4.1, where we analyze it in more detail and provide upper bounds.

3 Model and Results

3.1 Setup

There are two agents, a sender and a receiver, who are communicating about an uncertain state of the world. The state of the world ω can take values from a finite set Ω, which has cardinality $|\Omega| = n$. Actions the receiver can take are denoted $a \in A$ where A is the finite action space with $|A| = m$. The sender and the receiver have utility functions which depend on the state of the world and the receiver's action, respectively denoted by: $u^S, u^R : \Omega \times A \to \mathbb{R}$. The agents share a prior belief about the state of the world, μ_0, which is assumed to be in the interior of $\Delta(\Omega)$ that is denoted by $\mathbf{int}(\Delta(\Omega))$, and it is common knowledge that the agents hold a shared prior.[1]

Let S be the finite set of signal realizations. Critically, we focus on the case where $k < \min\{m, n\}$, but all of our results are valid for any $k \in \mathbb{N}$. With this restriction, we aim to capture coarseness in communication. The sender chooses a signaling policy which is a collection of conditional probability mass functions $\{\pi(.|\omega)\}_{\omega\in\Omega}$ over the signal space S with cardinality $|S| = k$. Thus, the sender cannot induce all possible actions or describe the state of the world perfectly, and has to decide which actions to induce through coarse communication while also maintaining Bayes plausibility. In the literature $\pi : \Omega \to \Delta(S)$ is sometimes framed as an experiment, hence our setting can be interpreted as analyzing persuasion via constrained experimentation. As a general fact, signals don't carry any meaning ex-ante, and obtain a meaning via the signaling policy at the equilibrium. We denote the set of all signaling policies $\pi : \Omega \to \Delta(S)$ with Π.

Given that the sender chooses $\pi \in \Pi$, and a signal realization $s \in S$ is observed, the receiver forms a posterior $\mu_s \in \Delta(\Omega)$ by Bayes' Rule. More explicitly:

$$\mu_s \in \Delta(\Omega) = \frac{\mu_0 \pi(s|\omega)}{\sum_{\omega'\in\Omega} \pi(s|\omega')\mu_0(\omega')} = \frac{\mu_0 \pi(s|\omega)}{\langle \mu_0, \pi(s|\cdot)\rangle} \quad , \forall s \in S \ , \forall \omega \in \Omega,$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product, and $\pi(s|\cdot)$ denotes the n-dimensional vector of probabilities for signal s under different states of the world. As the signal realization s induces a posterior μ_s, the signaling strategy $\pi \in \Pi$ leads to a set of posteriors $\mu = (\mu_1, \ldots, \mu_k) \in \Delta(\Omega)^k$, where $\mu_i := \mu_{s_i}$ i.e. μ_i corresponds to the posterior formed by the i^{th} signal. Similarly $\pi : S \to \Delta(\Omega)$ induces distribution over posteriors $\tau \in \Delta(\Delta(\Omega))$ with $\mathsf{supp}\,(\tau) = \mu = \{\mu_s\}_{s\in S}$ defined by:[2]

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

After forming the posterior μ_s, the receiver chooses an action from the set $\hat{A}(\mu_s) =$

[1] $\Delta(\Omega)$ denotes the simplex over $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$.

[2] $\mathsf{supp}\,(\tau)$ 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 $s \in S$.

$\arg\max_{a \in A} \mathbb{E}_{\omega \sim \mu_s} u^R(a, \omega)$.[3] The existence of 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. 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 $\hat{a}(\mu_s)$.

We can then characterize the sender's expected utility from $\pi \in \Pi$ as:

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

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

$$u^* = \max_{\pi \in \Pi} U^S(\pi) \tag{1}$$

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 $|\mathsf{supp}\,(\tau)| \leqslant k$. 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. The sender's utility when the posterior μ_s is induced will be $\hat{u}^S(\mu_s) = \mathbb{E}_{\omega \sim \mu_s} u^S(\hat{a}(\mu_s), \omega)$. Similarly, receiver's utility will be $\hat{u}^R(\mu_s) = \mathbb{E}_{\omega \sim \mu_s} u^R(\hat{a}(\mu_s), \omega)$. Expected utility of the sender under the information structure τ is denoted by $\mathbb{E}_{\mu_s \sim \tau} \hat{u}^S(\mu_s) : \Delta(\Delta(\Omega)) \to \mathbb{R}$. We similarly define the expected receiver utility under τ by $\mathbb{E}_{\mu_s \sim \tau} \hat{u}^R(\mu_s)$. Throughout the paper, τ will be called an information structure (induced by π). 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 $|\mathsf{supp}| \leqslant k$, and when $k \geqslant \min\{m, n\}$ this 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, as shown before. One can see that $\sum_{s \in S} \tau(\mu_s)\mu_s = \mu_0$. 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 $\mu_s \in \mathsf{supp}\,(\tau)$.

So 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 the only restriction imposed by Bayes plausibility (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 requirement $\mathsf{supp}\,(\tau) \leqslant k$ due to coarse communication. The sender's goal is therefore finding the optimal τ, which is described by the problem:

$$\max_{\tau \in \Delta(\Delta(\Omega))} \mathbb{E}_{\mu_s \sim \tau} \hat{u}^S(\mu_s) \text{ subject to } |\mathsf{supp}\,(\tau)| \leqslant k \text{ and } \mathbb{E}_\tau(\mu_s) = \mu_0 \tag{2}$$

[3] 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.

Finally, let us define *beneficial* information structures as τ with $\mathbb{E}_\tau(\hat{u}^S) \geqslant \hat{u}^S(\mu_0)$. These are information structures that give the sender higher utility compared to the default action, which can be achieved by sending no information. Throughout the paper, we will focus on the case where beneficial information structures exist: the other case is trivial and the sender always prefers sending no information.

Proposition 1. *An optimal information structure τ exists.*

We provide the following brief explanation of the proof for the interested reader. If we characterize the sender's maximization as described by equation 1 in which the sender picks $\pi \in \Pi$, existence follows from the fact that $U^S(\pi)$ is upper semi-continuous, as discussed in Tsakas and Tsakas (2018). This follows from the fact that we focus on sender-preferred equilibria. Similarly, if we characterize the sender's problem as picking Bayes plausible information structures as described by equation 2, existence follows directly from Kamenica and Gentzkow (2011) who show that $\hat{u}^S$ is upper semi-continuous and attains a maximum over all Bayes plausible information structures $\mathcal{T}$. Our problem with coarse communication also must attain a maximum, since the set of Bayes plausible information structures whose support has cardinality at most k, $T \subset \Delta(\Delta(\Omega))$ is defined by $T = \{\tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_\tau \mu_s = \mu_0 \text{ and } |\mathsf{supp}\,(\tau)| \leqslant k\}$ is a closed subset of all Bayes plausible information structures $\mathcal{T} = \{\tau \in \Delta(\Delta(\Omega)) | \mathbb{E}_\tau \mu_s = \mu_0\}$. We provide a constructive algorithm on how to find the utility maximizing information structure for the sender in the appendix.

The constraint on the signal space makes the set of payoffs attainable by the sender harder to identify. The achievable set of utilities can shrink considerably for the sender, compared to the baseline model with unrestricted communication. We focus on characterizing this set of achievable utilities for a given prior in the next section.

3.2 Achievable utilities and concavification

We proceed by showing the geometric characterization of the highest achievable payoff for each prior μ_0 with a signal space cardinality k. We call this characterization the k-concavification of sender utility. This will connect our solution technique to the concavification approach widely used in the Bayesian persuasion literature.

Let $\mathbb{CH}(\hat{u}^S)$ denote the convex hull of the hypograph of $\hat{u}^S$, in the space $\mathbb{R}^n$. With unrestricted communication, the point $(\mu_0, z) \in \mathbb{CH}(\hat{u}^S) \subset \mathbb{R}^n$ represents a sender payoff z which can be achieved by an information structure when the prior is μ_0.[4] This is the foundation of the concavification technique, first used in repeated games and then applied to Bayesian persuasion (Aumann and Maschler 1995; Kamenica and Gentzkow 2011). In canonical persuasion games, the existence of an optimal signal is usually proven by referencing extremal representation theorems from convex analysis. 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 prevents us from using this theorem in our setting.

[4] 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 $(\mu, z) \in \mathbb{R}^n$.

With restricted communication, the point $(\mu, z) \in \mathbb{CH}(\hat{u}^S)$ might not be feasible if the construction of (μ, z) requires a convex combination of more than k points from the hypograph of $\hat{u}^S$. A prior belief-utility pair (μ, z) will only be feasible if it can be contained in the convex hull of k or fewer points from the hypograph of $\hat{u}^S$. To represent achievable utilities, therefore, we need the following definition. Given a set $A \subseteq \mathbb{R}^n$ and an integer $0 < k \leqslant n$, define the set of points that can be represented as the convex combination of at most k points in A as the k-convex hull of A, denoted $\text{co}_k(A)$. Formally we provide the following definition:

Definition 1. *A given point $a \in co_k(A)$ if and only if there exists a set of **at most k points** $\{a_1, \ldots, a_k\} \subseteq A$ and a set of weights $\{\gamma_1, \ldots, \gamma_k\}$ which satisfy $\sum_{i \leqslant k} \gamma_i = 1$ and $\forall i, 1 > \gamma_i > 0$ such that $a = \sum_{i \leqslant k} \gamma_i a_i$. Therefore, we can write:*

$$co_k(A) = \{a \in \mathbb{R}^n : \exists \{a_1, \ldots, a_k\} \subseteq A, \exists \{\gamma_1, \ldots, \gamma_k\} \text{ with } \gamma_i \in \mathbb{R} \text{ s.t. } \sum_{i \leqslant k} \gamma_i = 1 \text{ and } 1 \geqslant \gamma_i \geqslant 0, a = \sum_{i \leqslant k} \gamma_i a_i\}$$

Let $\mathbb{CH}_k(\hat{u}^S)$ denote the k-convex hull of the hypograph of $\hat{u}^S$, in the space $\mathbb{R}^n$. 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 $\mathbb{E}_\tau(\hat{u}^S) = z$. Defining $V(\mu_0) = \sup\{z | (\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\}$, we get the largest payoff the sender can achieve when the prior is μ_0. 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.

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

Going back to the financial advice example, we can see in figure 2 that the optimal payoff for the sender given μ_0 can be observed by inspecting the 2-convex hull of the sender utility. The comparison with the regular convex hull (3-convex hull) reveals that the achieved utility must be lower. The optimal information structure can thus be determined by inspecting $\mathbb{CH}_k(\hat{u}^S) \subset \mathbb{R}^n$.

3.3 Simple Information Structures

We will now focus on a subset of Bayes plausible information structures what we call *simple*. In a similar paper, Lipnowski and Mathevet (2017) provide a way to simplify the Bayesian persuasion problem. Their result shows the existence of an optimal information structure that is simple when signal spaces are rich, i.e. $k \geqslant \min\{m, n\}$. We generalize their result that simplifies the search for optimal information to any cardinality of the signal space, especially when $k < \min\{m, n\}$. Moreover our generalization provides an algorithmic way to (weakly) improve any information structure that is not simple.

We start by using the underlying preference structure of the receiver to partition the space of posteriors. This will allow us to define the set of simple information structures.

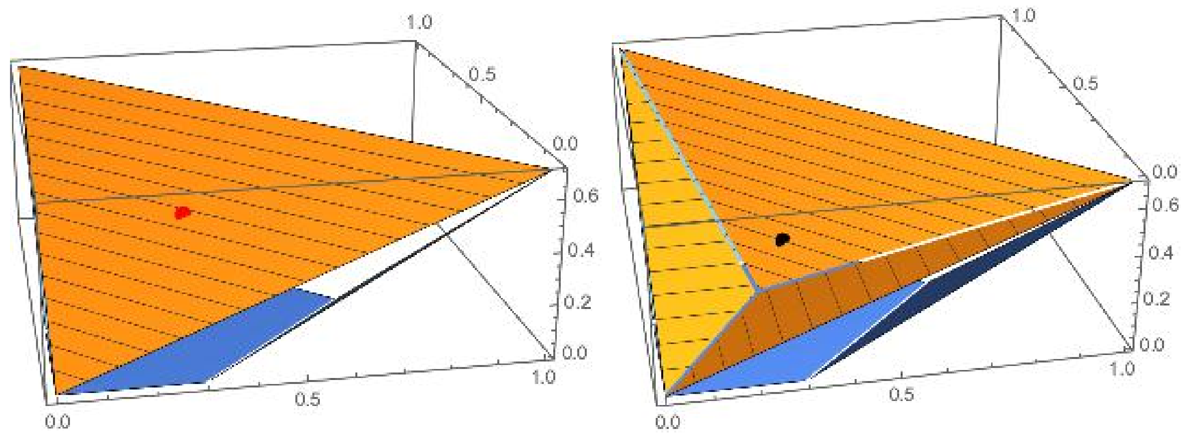

Figure 2: The supremum of the 3-convex hull and the 2-convex hull of sender utility from the financial advice example. The left figure shows the maximum achievable utility with 3 signals, and the right figure shows the maximum achievable utility with 2 signals as a function of the prior beliefs. The dots correspond to the prior belief given in the example ($\mu_0 = (0.3, 0.4, 0.3)$).

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.[5]

Definition 2. *The set $R_a \subseteq \Delta(\Omega)$ is the set of beliefs where the action a is receiver-optimal: $R_a = \{\mu_i \in \Delta(\Omega) : a \in \hat{A}(\mu_i)\}$ and $R = \{R_a\}_{a \in A}$ is the collection consisting of these sets for every action $a \in A$.*[6]

We call a Bayes plausible information structure *simple* if (i) it induces posteriors that are affinely independent and (ii) each induced posterior either makes the receiver indifferent between multiple actions or gives zero weight to some states of the world. Formally, we can state this as follows.

Definition 3. *We say that a Bayes plausible information structure $\tau \in \Delta(\Delta(\Omega))$ with $|\mathsf{supp}(\tau)| \leqslant k$ is **simple** if $\mathsf{supp}(\tau)$ is affinely independent and, for each $\mu_s \in \mathsf{supp}(\tau)$ $\mu_s \in \boldsymbol{Bd}(R_a)$ for some $a \in A$. We denote set of all Bayes plausible information structures that are simple by ζ_k.*

We will now show that for any Bayes plausible information structure $\tau \in \Delta(\Delta(\Omega))$ that is not simple, there exists a simple information structure $\tilde{\tau} \in \zeta_k \subset \Delta(\Delta(\Omega))$ such that $\mathbb{E}_{\tilde{\tau}} \hat{u}^S \geqslant \mathbb{E}_{\tau} \hat{u}^S$. This will be done in two steps. First, we will show that every information structure that induces a posterior in the interior of R_a can be (weakly) improved. Second, we will show that every information structure that induces affinely dependent posteriors can be (weakly) improved. These facts combined with the existence of an optimal information structure will imply that there exists an optimal information structure that is simple.

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

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

Before proceeding, a discussion of what is meant by simple information structures is warranted. We use this terminology for multiple reasons. First, for every information structure τ that is not affinely independent, there is a simple information structure τ' that induces fewer actions and the sender's expected utility is weakly higher: $\mathbb{E}_{\tau'}\hat{u}^S \geqslant \mathbb{E}_{\tau}\hat{u}^S$. Hence, optimal information structures which induce fewest possible actions are simple. Second, for every information structure τ that is simple, the probability distribution τ is uniquely identified from its support μ. This means that senders problem reduces to a problem of choosing posteriors, instead of choosing a distribution over posteriors, when the search space is restricted to simple information structures. Third, the set of simple information structures is vastly smaller than the set of all Bayes plausible information structures. Overall the listed properties of simple information structures both vastly simplify the search for an optimal information structure and ensure that the induced information structures are the ones that have the lowest dimension possible. We present a model of Bayesian persuasion with preferences for simplicity in appendix B.5.

The properties presented in Lemmas 2 and 3 have been applied in (i) the context of persuasion games where the receiver has psychological preferences over different posterior beliefs (Lipnowski and Mathevet 2018; Volund 2018) and (ii) general games of persuasion to simplify the search for optimal information structures.

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.*

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 points 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 points. 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.

These Lemmas presented above show us that in the subset of posteriors where the receiver's action is fixed, sender prefers inducing mean-preserving spreads in beliefs. In the model with unrestricted communication, these properties reduce the search for an optimal information structure to a more tractable optimization problem, since the optimal information structure must be supported by the outer points of the sets $R = \{R_a\}_{a \in A}$ as described in Lipnowski and Mathevet (2017). With coarse communication, we can prove a similar result.

The next Lemma formally states that an information structure can always be weakly improved by changing it in a way that maintains Bayes Plausiblity and moving all posteriors to the boundaries of an action region. In other words, the sender can restrict their search to posteriors that make the receiver indifferent between multiple actions and posteriors that assign zero probability to some states of the world (located on the boundary of the simplex $\Delta(\Omega)$), with no loss in utility. This result reduces the size of our search space considerably, and provides tractability in the optimal information design problems, similar to Lipnowski and Mathevet (2017).

Lemma 4. *Let τ be a feasible distribution of posteriors satisfying Bayes plausibility, that is also beneficial for the sender. Suppose that $\exists \mu_a \in \textsf{supp}\,(\tau)$ such that $\mu_a \in \textbf{int}(R_a)$ for some $R_a \in R$. Then, there exists a $\mu_k \in \textbf{Bd}(R_a)$ and a Bayes plausible $\tau' \neq \tau$ where $\textsf{supp}\,(\tau') = (\textsf{supp}\,(\tau)/\{\mu_a\}) \cup \{\mu_k\}$ such that $\mathbb{E}_{\tau'}\hat{u}^S \geqslant \mathbb{E}_{\tau}\hat{u}^S$.*

The proof explicitly constructs the information structure τ', and uses the convexity of $\hat{u}^S$ within each R_a. The outline of the argument is the following. Let τ be our original Bayes plausible information structure, with the corresponding probabilities $\{\tau(\mu_i)\}_{i\leqslant k}$ where $\sum_{i\leqslant k}\tau(\mu_i)\mu_i = \mu_0$. Let $\mu_a \in \textsf{supp}\,(\tau)$ be in the interior of some R_a and define the ray originating from μ_0 and passing through μ_a. This ray will intersect the boundary of R_a at two points μ' and μ'', since R_a is compact and convex. By convexity of $\hat{u}^S$ within R_a, sender utility must be weakly higher at one of those two points. First, we show that we can replace μ_a with one of these two points and still maintain Bayes plausibility. Since we're changing μ_a along the ray defined above, we can change $\{\tau(\mu_i)\}_{i\leqslant k}$ in a way that maintains Bayes plausibility. Note that greedily replacing μ_a with the point that provides higher utility within R_a might not always improve the expected sender utility $\mathbb{E}_{\tau}(\hat{u}^S)$, and the overall effect of this change depends on the relative positions of μ_0, μ_a, μ' and μ'', and the change in the probabilities $\{\tau(\mu_i)\}_{i\leqslant k}$ that will maintain Bayes plausibility. A carefully constructed argument relying on convexity shows that replacing μ_a with either μ' or μ'' will always yield higher expected utility, where the decision on which point to choose depends on the changes in $\{\tau(\mu_i)\}_{i\leqslant k}$.

While this simplifies the search, to solve and characterize the sender maximization problem in a tractable way, we proceed by showing that we can restrict our search of optimal information structures to the set of affinely independent information structures without any loss. The next Lemma shows that any affinely dependent information structure can be modified by dropping some beliefs to reach affine independence, weakly increasing sender utility and maintaining Bayes plausibility at every step. The proof is independent of Lemma 4 and holds for a general class of information design problems with or without constrained signal spaces.

Lemma 5. *Let τ be a distribution of posteriors satisfying Bayes plausibility. Suppose that $\textsf{supp}\,(\tau)$ is not affinely independent. Then, there must exist a Bayes plausible $\tau' \neq \tau$ such that $\textsf{supp}\,(\tau')$ is affinely independent and $\mathbb{E}_{\tau'}\hat{u}^S \geqslant \mathbb{E}_{\tau}\hat{u}^S$.*

The proof is again done constructively. Intuitively, for the sender, inducing affinely dependent beliefs is not a good use of the signals because some beliefs are 'redundant': these beliefs can be represented as affine combinations of each other, and the sender can always drop one of them and still maintain Bayes Plausability. Our proof outlines the details on how we can always find a belief that is optimal to drop from the affinely dependent information structure. We use the relationship between the convex weights characterizing μ_0 which are $\{\tau(\mu_i)\}_{i\leqslant k}$, and the set of affine weights that allows us to characterize beliefs in terms of each other to find the posterior that weakly increases sender utility when dropped.

Our analysis of cardinality-constrained signal spaces is also useful for Bayesian persuasion games with rich signal spaces where agents have preferences for simplicity. In a standard

Bayesian persuasion where $k \geqslant \min\{|A|, |\Omega|\}$, suppose the sender cares about the simplicity of the induced information structures, in addition to the utility received. In appendix B.5, we analyze this setting by defining an intuitive preference structure for the sender, and show that affinely independent information structures will be chosen at an equilibrium.

Theorem 1. *There exists a simple information structure that is optimal for the sender.*

This can be shown as follows: By Proposition 1 there exists an optimal Bayes plausible information structure. Then, either it is simple, or by Lemma 4 and Lemma 5 there exists another Bayes plausible information structure that attains a weakly better payoff. This result applies to finite Bayesian persuasion problems with signal spaces of arbitrary (finite) cardinality.

Lemma 5 states that we can restrict our search to affinely independent information structures, or in other words, lower dimensional simplices contained in the n-simplex $\Delta(\Omega)$. This gives us the uniqueness of the probability measure $\{\tau(\mu_i)\}_{i \leqslant k}$ representing μ_0 given that $\mu = (\mu_1, \ldots, \mu_k)$ via the Choquet's Theorem. The statement of this well known result (e.g., see Alfsen (1965)) is can be found in appendix B.[7] This ensures that when searching for affinely independent posteriors μ, $\tau(\mu)$ is uniquely determined by the choice of μ. Additionally, given a set of posteriors μ and the prior μ_0, we can describe the probability distribution over μ that maintains Bayes Plausability $\tau : \Delta(\Omega) :\to \Delta(\Delta(\Omega))$ with a matrix operation.

Lemma 6. *Let $\mu_0 \in \boldsymbol{int}(\Delta(\Omega))$, for any $j \in \mathbb{N}$ $\forall \mu = (\mu_1, \ldots, \mu_j) \in \Delta(\Omega)^j$ such that $\mu_0 \in co(\mu)$ and μ is affinely independent, there exists a unique probability distribution $\tau \in \Delta(\Delta(\Omega))$ with support μ and $\sum_i \tau(\mu_i)\mu_i = \mu_0$. The vector $\tau(\mu) = [\tau(\mu_1), \ldots, \tau(\mu_j)]$ is defined by $\tau(\mu) = T^L(\mu)\tilde{\mu}_0$ where T^L is the left inverse of T which is the matrix of posteriors μ with an added row of 1's and $\tilde{\mu}_0$ is the column vector μ_0 with an added entry 1.*

From Lemma 6, we can conclude that the choice of affinely independent posteriors which the sender decides to induce, uniquely determines the probability of each posterior being induced. With this result, we can formulate the sender's problem as a search over posteriors $\{\mu_j\}_{j \leqslant k}$ that are affinely independent, with the added constraint that $\forall \mu_i \in \mathsf{supp}(\tau)$, μ_i is in the boundary of some R_a. Formally, the optimal simple information structures can be found by solving the maximization problem:

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

which is equivalent to the following maximization problem by Lemma 6:

$$\max_{\mu \in \zeta_k} \mathbb{E}_{\mu_s \sim \tau(\mu)} \hat{u}^S(\mu_s) \text{ s.t. } \tau(\mu) = T^L(\mu)\tilde{\mu}_0.$$

[7] Existence, smoothness and uniqueness of this probability measure can be analyzed through barycentric coordinates by making use of existing work on generalized barycentric coordinates on convex sets (Warren 1996, 2003; Warren et al. 2007), but we use Choquet Theory as a more convenient tool.

4 Applications

4.1 The Value of Precision

We can further analyze the implications of restricting the signal space on sender's utility. Let $V^*(k, \mu_0)$ be the value the sender objective function attains with prior μ_0 when the signal space is restricted to have k elements. Then $V^*(k+1, \mu_0) - V^*(k, \mu_0)$ is what the sender would be willing to pay to increase the dimensionality of the signal space by one, given the fixed prior μ_0. This can be intuitively interpreted as the value of precision for the sender. Note that when $k \geqslant \min\{|\Omega|, |A|\}$, the value of precision will be equal to zero by the results in Kamenica and Gentzkow (2011). Therefore we focus exclusively on the coarse communication setting in which $k < \min\{|\Omega|, |A|\}$.

The value of precision depends on the sender and receiver utility functions, and the location of the prior belief μ_0. It critically depends on what restricted set of actions the sender can induce while still maintaining Bayes plausibility. If maintaining Bayes plausibility with lower dimensional signals requires inducing actions with lower payoffs, or 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 value of precision, or equivalently, a lower bound on the utility achievable with $k-1$ signals which applies to any finite game of Bayesian persuasion. Because of the geometric structure of the problem and the subtle relationship between $V^*(k, \mu_0)$ and $V^*(k-1, \mu_0)$, the loss in utility cannot be too high.

Theorem 2. *Suppose $|S| = k \geqslant 3$, and the sender utility function u^S is positive everywhere. Then, the following upper bound must hold for the value of precision at $k-1$ signals:*

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

or equivalently, the payoff with $k-1$ signals must lie between:

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

This provides a lower bound on the utility loss from using smaller signal spaces, as a function of utility achievable with unrestricted communication. Let τ_k^* and τ_{k-1}^* be the optimal information structures using k and $k-1$ signals, respectively. The proof relies on the observation that τ_k^* can be 'collapsed' to get an information structure with $k-1$ signals by combining two posteriors in a way that maintains Bayes plausibility. These new signals 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^*. The resulting inequalities yield the lower bound in Theorem 2.

4.1.1 Example: Non-Monotonicity of the Value of Precision

We will show that the value of precision can be non-monotone in general. We analyze an example with 3 states of the world to demonstrate how the behavior of the value of precision can depend on the location of the prior. In our example we will see that $V^*(2, \mu_0) - V^*(1, \mu_0)$ can be greater or less than $V^*(3, \mu_0) - V^*(2, \mu_0)$. We will also demonstrate how this difference depends on the difficulty of inducing beneficial 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:

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

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 $\bar{\pi}$. 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 $\mu_s^i \geqslant \bar{\pi}$, 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 $\mu_s^i > \bar{\pi}$. The action zones for these receiver preferences can be represented as:

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

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 aims to induce the non-default actions. The parameter $\bar{\pi}$ can be interpreted as the difficulty of inducing the beneficial actions for the sender.

Given this structure, it should be obvious that sender can attain a payoff of 1 by using 3-signal information structures. This follows from the fact that for every prior $\mu_0 \in \Delta(\Omega)$ with $\mu_0 = (\mu_0^1, \mu_0^2, \mu_0^3)$ the sender can use the information structure $(1, 0, 0)$ with probability μ_0^1, $(0, 1, 0)$ with probability μ_0^2 and $(0, 0, 1)$ with probability μ_0^3. This information structure corresponds to $\tau(\mu_s) \in \Delta(\Delta(\Omega))$ with $\tau((1, 0, 0)) = \mu_0^1$, $\tau((0, 1, 0)) = \mu_0^2$, $\tau((0, 0, 1)) = \mu_0^3$. We have that $\mathbb{E}_\tau u^s(a(\omega), \omega) = 1$. 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 1-signal information structures (i.e. no information transmission at all), we have that the payoff sender can achieve is

$$\mathbb{E}_{\mu_0} u^s(a(\mu_0), \omega) = \begin{cases} 1 & \text{if } \mu_0 \in R_i \quad \forall i \in \{1, 2, 3\} \\ 0 & \text{if otherwise} \end{cases}$$

Our goal is analyzing the non-trivial case of 2 signals. We precisely 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 define Δ_c as the set where two-signal information structures attain lower payoff than three-signal information structures. The following Lemma states the values of $\bar{\pi}$ such that this set is non-empty.

Lemma 7. *$\Delta_c \neq \varnothing$ if and only if $\bar{\pi} \geqslant \frac{2}{3}$.*

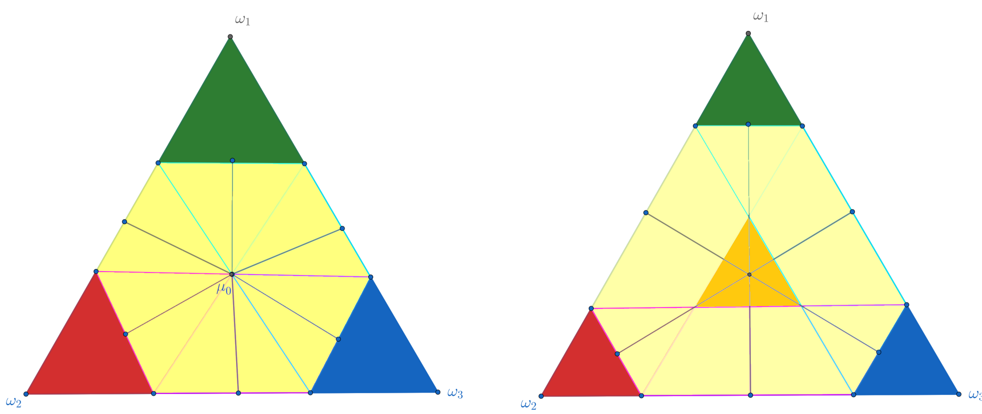

Figure 3: On the left, we have the action threshold $\bar{\pi} = \frac{2}{3}$ so it is possible to maintain Bayes plausibility when inducing non-default actions for every prior. On the right, $\bar{\pi} > \frac{2}{3}$, so for some beliefs, we have to mix the default action and the non-default action when constrained to 2 signals. The dark red, blue and green regions are the beneficial action regions. The yellow middle region is the default action region. Orange region in the right figure corresponds to Δ_c.

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

Lemma 8. *If $\bar{\pi} > \frac{2}{3}$ we have that $V(2, \mu_0) < V(3, \mu_0) = 1$ for every $\mu_0 \in \Delta_c$ and $V(2, \mu_0) = V(3, \mu_0) = 1$ for every $\mu_0 \notin \Delta_c$. Moreover, $\forall \mu_0 \in \Delta_c$, $\overline{V(2, \mu_0)} > V(2, \mu_0) > \underline{V(2, \mu_0)}$, where $\overline{V(2, \mu_0)} = \frac{2\bar{\pi}-1}{\bar{\pi}}$ and $\underline{V(2, \mu_0)} = \frac{1}{3\bar{\pi}}$.*

In figure 4, we plot $\underline{V(2, \mu_0)}$ and $\overline{V(2, \mu_0)}$ as a function of the action threshold $\bar{\pi}$. The following is an immediate implication of Lemma 8. Fixing the preferences of the sender and

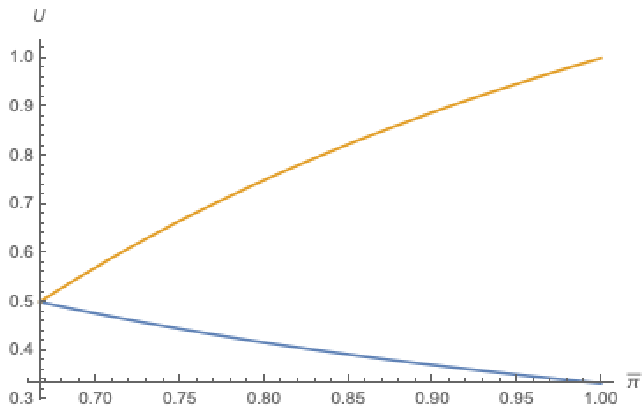

Figure 4: Achievable utilities with two signals for $\mu_0 \in \Delta_c$, as a function of the action thresholds $\bar{\pi}$. Blue line depicts the minimum of the equilibrium sender utility among all $\mu_0 \in \Delta_c$, and yellow line denotes the maximum value.

the receiver, for some prior beliefs, the value of an additional signal is an increasing function, and for others, it is decreasing.

Corollary 1. *Depending on the location of the prior inside Δ_c the value of precision can be increasing or decreasing with respect to additional signals. That is $\overline{V(2,\mu_0)} > \frac{1}{2}$ and $\underline{V(2,\mu_0)} < \frac{1}{2}$.*

The priors for which the value of precision is increasing are the ones that are the furthest away from the beneficial 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 value of precision is decreasing. 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. This is not necessarily the case for the receiver, as we will see in our next application.

4.2 Optimal Advice Seeking

Our model also can be used to analyze the optimal advice seeking behavior of a receiver. Suppose, before the game described in section 3.1 takes place, the receiver can choose the cardinality of the signal space $|S| = k$. Letting the receiver decide the cardinality of the signal space allows them to change the outcome in their favor. The receiver can choose to ask for 'simple advice' consisting of fewer action recommendations rather than a more complicated one. We will show through an example that the receiver will not always prefer using rich signal spaces.

First, observe that if there is perfect alignment between the receiver and the sender's utilities, so that $\hat{u}^R = \hat{u}^S$, the receiver will always pick the maximum number of signals possible. This is because the sender's utility (and therefore the receiver's utility) is weakly increasing in the number of signals available.

Let us now turn to the more interesting case of misalignment. Receiver's preferences over the number of signals will depend on the location of the prior and the degree of the misalignment between the sender and the receiver. We will make this more clear with the following example.

4.2.1 Example: Seeking Simple Advice

Suppose there are three states $\{\omega_1, \omega_2, \omega_3\} = \Omega$ and five actions $A = \{a_0, a_1, a_2, a_3, a_4\}$ where a_0 denotes the default action taken at the prior belief $\mu_0 = (1/3, 1/3, 1/3)$. For simplicity, suppose the sender's utility depends only on the actions taken, and the default action is the worst outcome. The sender prefers inducing the actions a_1, a_2, a_3 and a_4 over a_0, and a_2 and a_3 are preferred over a_1 and a_4.

Receiver preferences are such that the optimal actions are a_1, a_2, a_3 whenever the beliefs are certain enough, meaning that upon observing signal $s \in S$ it is the case that $\mu_s(\omega_i) > \bar{\pi}$ for $i \in \{1, 2, 3\}$. Moreover, whenever $\mu_s(\omega_2) < \bar{\pi}$ and $\mu_s(\omega_3) < \bar{\pi}$ but $\mu_s(\omega_2) + \mu_s(\omega_3) > \bar{\pi}$ the receiver takes action a_4. This means that there are two different actions that the receiver optimally takes when the beliefs are uncertain, which are a_0 (uncertain but leaning ω_1) and a_4 (uncertain but leaning ω_2 or ω_3). Figure 5 plots these preferences along with the optimal 2 and 3-signal information structures. The details about the utility function for the receiver is given in the appendix B.4.

We consider the following game: the receiver will pick the cardinality of the signal space $k = |S|$ first, sender observes this choice and picks the optimal Bayes plausible information structure with k signals. By sequential rationality and our previous calculations we can characterize the sender's behavior using our results. Namely, the sender will pick the optimal information structure for the k-constrained Bayesian persuasion game, given the choice of k by the receiver. Hence, the receiver will pick $k = |\Omega|$ such that expected receiver utility is maximized.

It is easy to verify that for every equilibrium (PBE) of this game the receiver will pick $k = 2$, as plotted in figure 5.[8] For the receiver's choice of $k = 2$, the sender will pick the information structure described by the red line in the lower left box in figure 5, inducing a_2 and a_1. Off path, for the choice of $k = 1$ by the receiver, the sender will pick a trivial information structure and for the choice of $k = 3$, the sender will pick the information structure shown with the blue triangle in the upper right corner in figure 5, inducing a_1, a_2, a_3. Hence, receivers picks k to be equal to 2. The lower right plot in figure 5 shows how the

[8] See appendix B.4 for utility functions.

two-signal information structure, three-signal information structure and the single-signal information structure compare in terms of expected utility for the receiver.

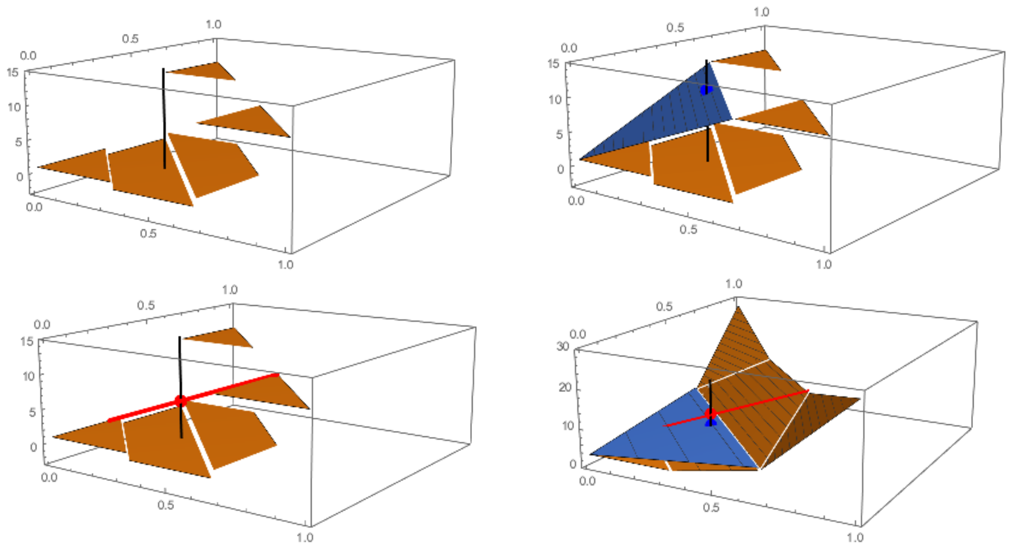

Figure 5: Partial misalignment and optimal advice seeking. In all figures, the black perpendicular line at (1/3,1/3,1/3) represents the location of the prior. The top left figure depicts the sender's utility over the simplex, which depends only on the actions taken by the receiver. The top right figure shows the optimal 3-signal information structure, and the bottom left figure shows the optimal 2-signal information structure. The bottom right figure depicts the receiver's utility, with the optimal 2-signal (red line) and 3-signal (blue surface) information structures. For the receiver, utility with 2 signals (red point) is higher than the utility with 3 signals (blue point).

We see that there is a misalignment between the receiver and the sender preferences. The receiver prefers outcomes that are more certain about ω_2 and ω_3 whereas the sender only wants to induce actions a_2 and a_3 and does not care about certainty in beliefs. The sender ideally wants to induce uncertain posteriors leading to action a_2 and a_3 with high probability.

Limiting the sender to two signals, the receiver can force the sender to induce more certain beliefs about ω_2. This is because under the Bayes plausibility constraint, imposing $k = 2$ forces the sender to induce a more certain posterior about ω_2. With $k = 3$, the sender optimally induces vague posteriors about ω_2 and ω_3.

This example shows that the receiver might prefer to limit senders ability to communicate and opt for simple advice. More elaborately, in the game considered above, we see that the receiver prefers the expected outcome with two signals ($k = 2$) over three signals ($k = 3$), and the expected outcome with three signals ($k = 3$) over no communication at all ($k = 1$). Thus, in the optimal advice seeking game with partially misaligned preferences, the receiver will profitably choose to limit the communication capacity of the sender, but not so much that no useful information can be transmitted at all.

The example also demonstrates an interesting property of cardinality constrained persuasion games. We see that the optimal information structures chosen with $k = 2$ and $k = 3$

are not Blackwell comparable. This observation implies that allowing the sender to use more signals does not guarantee that a more informative information structure will be chosen. If they were to be Blackwell comparable, we would have concluded that the receiver prefers the more informative one since their preferences have to be convex over the entire simplex.[9]

This presents an interesting avenue for studying coarse communication. The applied literature on Bayesian persuasion games has a variety of examples -e.g. Judge-Prosecutor communication (Kamenica and Gentzkow 2011)- where the receiver has some power on the communication procedure. One way to reflect this power is letting the receiver pick the cardinality of the sender's signal space. The example above shows that the receiver may prefer to pick k to be less than the cardinality of the action and state space. Our framework can be used to analyze these interactions in detail by characterizing the solutions to Bayesian persuasion problems with coarse signal spaces.

5 Conclusion

We set out to analyze the effect of coarseness in strategic communication, which was left unexplored by previous literature. The value of precise communication in a game where a sender is trying to persuade a receiver is characterized and an upper bound for this value which applies to all finite persuasion games is presented. We proved the existence of an optimal information structure that is *simple*. This greatly simplifies the search for an optimal information structure. Our paper hence complements the simplification result of Lipnowski and Mathevet (2017) by generalizing it to signal spaces of arbitrary finite cardinality, including the setting with coarse communication.

Our work also complements the asymptotic upper bound results of Le Treust and Tomala (2019) on infinitely repeated persuasion games with noisy and coarse channels, and the results in Tsakas and Tsakas (2018) on finite persuasion games over noisy channels. We show that constrained signal spaces create non-trivial difficulties for the sender in a persuasion game and demonstrate how we can analyze the outcomes using k-convex hulls. In settings where a receiver is asking for advice from a sender with misaligned preferences, we show that it might be optimal to ask for simple recommendations. This gives us a better understanding of 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 can apply our framework to various settings where we would like to analyze how much a sender would be willing to pay for more precise messages. Some of the most important questions studied using persuasion games can now be analyzed from this new viewpoint. How much would a politician be willing to spend to design a more detailed policy experiment to convince voters? How much would a lobbyist be willing to pay to send a more precise action recommendation to the politician that they are trying to persuade? How much would a firm trying to send product information to a potential customer be willing to pay for a longer, more detailed ad?

[9]By Corollary 3 in appendix B.2.

Our model can also be used to study competition between senders who have access to signal spaces with different degrees of complexity, or the problem of a sender trying to persuade a heterogeneous set of agents using public or private signals with different dimensionalities. We leave these questions for future work.

Appendices

A Proofs

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 $\sup\{z|(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} < \mathbb{E}_\tau \hat{u}^S$.

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. $\{\mu_1, \ldots, \mu_k\}$ must be elements of some facets, because otherwise by Lemma 4, we can show the existence of another information structure with higher expected utility, contradicting the fact that (μ_0, z) is a supremum. It must also be the case that $\{\mu_1, \ldots, \mu_k\}$ are affinely independent, because otherwise by Theorem 5, we can contradict (μ_0, z) being a supremum again. We know that τ' satisfies Bayes plausibility by the definition given above. Therefore $\tau' \in \zeta_F$ for some facet combination F, and it could have been picked instead of τ in the maximization problem, contradicting the optimality of τ.

Finally, note that for $V(\mu_0) = \sup\{z|(\mu_0, z) \in \mathbb{CH}_k(\hat{u}^S)\} = \sup_{\mu \in \mathcal{M}(\mu_0)} \hat{u}^s(\mu)$ where $\mathcal{M} = \{\mu \in \Delta(\Omega)^k | co(z) = \mu_0\}$. Then for all $\mu \in \mathcal{M}$ we have that $V(\mu_0)$ is concave. Concave functions are quasiconcave and pointwise maximum of quasiconcave functions is quasiconcave. Hence we conclude that $V(\mu_0)$ is quasiconcave. This completes the proof of Proposition 2.

□

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

We prove this claim by explicitly constructing τ'. Using the convexity of $\hat{u}^S$ within R_a, we can find two alternative beliefs μ', μ'' in $\mathbf{Bd}(R_a)$, such that replacing μ with one of these two beliefs maintains Bayes plausibility and weakly increases $\mathbb{E}_\tau \hat{u}^S$.

Let $\textsf{supp}\,(\tau) = \{\mu_1, \mu_2, \ldots, \mu_k\}$. Since τ satisfies Bayes plausibility, 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 $\forall i \in \{1, \ldots k\}\ 1 > \tau(\mu_i) > 0$. We want to show that we can construct τ' which satisfies Bayes plausibility and $\mathbb{E}_{\tau'} \hat{u}^S \geqslant \mathbb{E}_\tau \hat{u}^S$.

Without loss of generality, let $\mu_k \in \textsf{supp}\,(\tau)$ be the belief ($\mu_k \neq \mu_0$) such that for some $R_k \in R$, $\mu_k \in \textbf{int}(R_k)$. Consider the ray from μ_0 passing through μ_k, parameterized as $\{\mu_0 + s(\mu_k - \mu_0), s \in \mathbb{R}^+\}$.

First, assume $\mu_0 \notin R_k$. Since $\mu_k \in \textbf{int}(R_k)$, and R_k is closed, bounded, and convex, the line segment passing through the interior point μ_k intersects $\textbf{Bd}(R_k)$ at two points (Yaglom and Boltyansky 1961). Let these two points be denoted as μ_k' and μ_k''. Since these two points also lie on the ray passing through μ_k originating from μ_0, they can be written in a parametric form. Therefore, for some $\delta > 0$ and $1 > \gamma > 0$ we have:

$$\mu_k' = \mu_0 + (1+\delta)(\mu_k - \mu_0) = \mu_k + \delta(\mu_k - \mu_0),$$
$$\mu_k'' = \mu_0 + (1-\gamma)(\mu_k - \mu_0) = \mu_k - \gamma(\mu_k - \mu_0).$$

Moreover, we can write our original point μ_k as a convex combination of these two points as $\frac{\gamma}{\gamma+\delta}\mu_k' + \frac{\delta}{\gamma+\delta}\mu_k'' = \mu_k$. Note that by convexity of $\hat{u}^S$ within R_k, we get:

$$\frac{\gamma}{\gamma+\delta}\hat{u}^S(\mu_k') + \frac{\delta}{\gamma+\delta}\hat{u}^S(\mu_k'') \geqslant \hat{u}^S(\mu_k). \tag{3}$$

Now, let us define two new information structures, τ' and τ'', by replacing μ_k with μ_k' and μ_k'', respectively. We will now show that we maintain Bayes plausibility with these new information structures.

Lemma 9. *The new information structures τ' and τ'', constructed as described above, are Bayes plausible.*

Proof. Start with comparing τ' with τ. We have $\textsf{supp}\,(\tau') = \{\mu_1, \ldots, \mu_k'\}$ and $\textsf{supp}\,(\tau) = \{\mu_1, \ldots, \mu_k\}$. We know that τ is Bayes plausible, so 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 $\forall i, 1 > \tau(\mu_i) > 0$.

We know that $\mu_k' = \mu_0 + (1+\delta)(\mu_k - \mu_0) = \mu_k + \delta(\mu_k - \mu_0)$. Let us define a new probability distribution $\tau' \in \Delta(\Delta(\Omega))$ representing μ_0 i.e. $\mu_0 = \sum_{i<k} \tau'(\mu_i)\mu_i + \tau'(\mu_k')\mu_k'$. Simple algebra reveals that this equality will hold for τ':

$$\tau'(\mu_i) = \frac{\tau(\mu_i)(1+\delta)}{1+\delta-\tau(\mu_k)\delta} \text{ for } i < k \text{ and } \tau'(\mu_k) = \frac{\tau(\mu_k)}{1+\delta-\tau(\mu_k)\delta}.$$

Note that $1+\delta-\tau(\mu_k)\delta > 0$, and $\tau(\mu_k) < 1+\delta-\tau(\mu_k)\delta$, so $1 > \tau'(\mu_k) > 0$. Also note that $\tau(\mu_i)(1+\delta) < 1+\delta-\tau(\mu_k)\delta$ since $\frac{\tau(\mu_i)-1}{1-\tau(\mu_k)-\tau(\mu_i)} < 0 < \delta$. Therefore $\forall i, 1 > \tau'(\mu_i) > 0$.

Finally:

$$\sum_{i\leqslant k}\tau'(\mu_i) = \frac{(1+\delta)\sum_{i<k}\tau(\mu_i)}{1+\delta-\tau(\mu_k)\delta} + \frac{\tau(\mu_k)}{1+\delta-\tau(\mu_k)\delta} = \frac{1+\delta-\tau(\mu_k)\delta}{1+\delta-\tau(\mu_k)\delta} = 1.$$

Similarly, take τ''. We have $\mathsf{supp}\,(\tau'') = \{\mu_1,\ldots,\mu_k''\}$ and $\mathsf{supp}\,(\tau) = \{\mu_1,\ldots,\mu_k\}$. We know that $\mu_k'' = \mu_0 + (1-\gamma)(\mu_k-\mu_0) = \mu_k - \gamma(\mu_k-\mu_0)$. Let us define a new probability distribution $\tau'' \in \Delta(\Delta(\Omega))$ representing μ_0 i.e. $\mu_0 = \sum_{i<k}\tau''(\mu_i)\mu_i + \tau''(\mu_k'')\mu_k''$. Simple algebra reveals that this equality will hold for τ'':

$$\tau''(\mu_i) = \frac{\tau(\mu_i)(1-\gamma)}{1-\gamma+\tau(\mu_k)\gamma} \text{ for } i<k, \text{ and } \tau''(\mu_k'') = \frac{\tau(\mu_k)}{1-\gamma+\tau(\mu_k)\gamma}.$$

Note that $1-\gamma+\tau(\mu_k)\gamma > 0$ since $\gamma < 1$. Also, $\tau(\mu_i)(1-\gamma) < 1-\gamma+\tau(\mu_k)\gamma$ since $\forall i, \tau(\mu_i) < 1$. Therefore, $\forall i, 1 > \tau''(\mu_i) > 0$. Finally:

$$\sum_{i\leqslant k}\tau''(\mu_i) = \frac{(1-\gamma)\sum_{i<k}\tau(\mu_i)}{1-\gamma+\tau(\mu_k)\gamma} + \frac{\tau(\mu_k)}{1-\gamma+\tau(\mu_k)\gamma} = \frac{1-\gamma+\tau(\mu_k)\gamma}{1-\gamma+\tau(\mu_k)\gamma} = 1.$$

■

We are now ready to prove the main Theorem. Let $\mathbb{E}_{\mu_s\sim\tau'}\hat{u}^S(\mu_s)$ and $\mathbb{E}_{\mu_s\sim\tau''}\hat{u}^S(\mu_s)$ be the sender's utility under the new information structures τ' and τ''. Using the definitions of τ', τ'', we can calculate the difference between these new values and the sender's utility under τ, which is simply $\mathbb{E}_{\mu_s\sim\tau}\hat{u}^S(\mu_s) = \sum_{i\leqslant k}\tau(\mu_i)\hat{u}^S(\mu_i)$. We drop the argument of $\mathbb{E}_\tau\hat{u}^S(\mu_s)$ hereinafter to reduce notation. Simple algebra shows the following:

$$\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S = \hat{u}^S(\mu_k') - \hat{u}^S(\mu_k) + \delta\left(\left(\sum_{i\leqslant k}\tau(\mu_i)\hat{u}^S(\mu_i)\right) - \hat{u}^S(\mu_k)\right),$$

$$\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S = \hat{u}^S(\mu_k'') - \hat{u}^S(\mu_k) - \gamma\left(\left(\sum_{i\leqslant k}\tau(\mu_i)\hat{u}^S(\mu_i)\right) - \hat{u}^S(\mu_k)\right).$$

For a contradiction, suppose that both $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S < 0$ and $\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_\tau\hat{u}^S < 0$. Rearranging the above terms and multiplying with γ and δ respectively, we get:

$$\gamma\hat{u}^S(\mu_k') + \gamma\delta\left(\left(\sum_{i\leqslant k}\tau(\mu_i)\hat{u}^S(\mu_i)\right) - \hat{u}^S(\mu_k)\right) < \gamma\hat{u}^S(\mu_k), \textit{ and,}$$

$$\delta\hat{u}^S(\mu_k'') - \delta\gamma\left(\left(\sum_{i\leqslant k}\tau(\mu_i)\hat{u}^S(\mu_i)\right) - \hat{u}^S(\mu_k)\right) < \delta\hat{u}^S(\mu_k).$$

Which implies:

$$\delta\hat{u}^S(\mu_k'') + \gamma\hat{u}^S(\mu_k') < (\delta+\gamma)\hat{u}^S(\mu_k).$$

However, by inequality 3 implied by convexity, we have :

$$\frac{\gamma}{\gamma+\delta}\hat{u}^S(\mu_k') + \frac{\delta}{\gamma+\delta}\hat{u}^S(\mu_k'') \geqslant \hat{u}^S(\mu_k)$$

$$\Leftrightarrow \delta\hat{u}^S(\mu_k'') + \gamma\hat{u}^S(\mu_k') \geqslant (\delta+\gamma)\hat{u}^S(\mu_k)$$

Therefore, we get 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. We must have either $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S \geqslant 0$ or $\mathbb{E}_{\tau''}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S \geqslant 0$.

For the remaining case, assume $\mu_0, \mu_k \in R_k$. Since $\mu_k \in \mathbf{int}(R_k)$, and R_k is closed, bounded, and convex, the ray originating from μ_0 passing through μ_k intersects $\mathbf{Bd}(R_k)$ at a single point, which we will denote by μ_k'. Since μ_k' lies on this line, for some $\delta > 0$, we will have:

$$\mu_k' = \mu_0 + (1+\delta)(\mu_k - \mu_0) = \mu_k + \delta(\mu_k - \mu_0)$$

Moreover, we can write μ_k as a convex combination of μ_k' and μ_0, where $\frac{\mu_k'}{1+\delta} + \frac{\delta\mu_0}{1+\delta} = \mu_k$.

Consider a new information structure τ', where we replace μ_k with μ_k' in τ, implying $\mathsf{supp}\,(\tau') = \{\mu_1, \ldots, \mu_k'\}$. Similar to the first part of the proof, we construct a probability distribution $\tau' \in \Delta^2(\Omega)$ that represents μ_0 i.e. we need $\{\tau'(\mu_i)\}_{i \leqslant k}$ to satisfy $\mu_0 = \sum_{i<k} \tau'(\mu_i)\mu_i + \tau'(\mu_k')\mu_k'$. Simple algebra reveals that this equality will hold for τ':

$$\tau'(\mu_i) = \frac{\tau(\mu_i)(1+\delta)}{1+\delta-\tau(\mu_k)\delta} \text{ for } i<k, \text{ and } \tau'(\mu_k') = \frac{\tau(\mu_k)}{1+\delta-\tau(\mu_k)\delta}.$$

Since the original information structure is assumed to be beneficial, we know that the payoff is better than the payoff under receiver's default action. This implies:

$$\hat{u}^S(\mu_0) \leqslant \sum_{i \leqslant k} \tau(\mu_i)\hat{u}^S(\mu_i)$$

Also by the convexity of $\hat{u}^S$ within R_k, we know that:

$$\frac{\hat{u}^S(\mu_k')}{1+\delta} + \frac{\delta\hat{u}^S(\mu_0)}{1+\delta} \geqslant \hat{u}^S(\mu_k).$$

Now, let us calculate the difference in expected sender payoff between τ' and τ. We find that:

$$\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S = \hat{u}^S(\mu_k') - \hat{u}^S(\mu_k) + \delta\left(\left(\sum_{i \leqslant k} \tau(\mu_i)\hat{u}^S(\mu_i)\right) - \hat{u}^S(\mu_k)\right).$$

For a contradiction, suppose that $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S < 0$. This implies the following:

$$\hat{u}^S(\mu_k') - \hat{u}^S(\mu_k) + < \delta\left(\hat{u}^S(\mu_k) - \left(\sum_{i\leqslant k}\tau(\mu_i)\hat{u}^S(\mu_i)\right)\right) \leqslant \delta\left(\hat{u}^S(\mu_k) - \hat{u}^S(\mu_0)\right)$$

$$\Leftrightarrow \frac{1}{1+\delta}\hat{u}^S(\mu_k') + \frac{\delta}{1+\delta}\hat{u}^S(\mu_0) < \hat{u}^S(\mu_k).$$

This contradicts $\frac{\hat{u}^S(\mu_k')}{1+\delta} + \frac{\delta\hat{u}^S(\mu_0)}{1+\delta} \geqslant \hat{u}^S(\mu_k)$, which we know to be true from the convexity of $\hat{u}^S$ within R_k. Therefore $\mathbb{E}_{\tau'}\hat{u}^S - \mathbb{E}_{\tau}\hat{u}^S \geqslant 0$ must hold.

Replacing each $\mu_k \in \mathsf{supp}(\tau)$ that is not on the boundary of a region R_a through the steps described above, we can reach a τ' that yields weakly higher utility for the sender. This completes the proof.

Proof of Lemma 5
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. $\square$

Proof of Lemma 6

Existence and uniqueness comes from Choquet's Theorem because $\mu = (\mu_1, \ldots, \mu_k)$ is a simplex, by the affine independence condition. Now given the convex weights over μ, $\tau = (\tau(\mu_1), \ldots, \tau(\mu_k))$ one can transform them to the Cartesian coordinates for μ_0 by using the following matrix:

$$T(\mu) = \begin{bmatrix} \mu_{1,1} & \mu_{2,1} & \ldots & \mu_{k,1} \\ \mu_{1,2} & \mu_{2,2} & \ldots & \mu_{k,2} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{1,n} & \mu_{2,n} & \ldots & \mu_{k,n} \\ 1 & 1 & \ldots & 1 \end{bmatrix}$$

where $\mu_{i,j}$ is the j'th coordinate of i'th posterior in μ. $T(\mu)$ is a matrix with dimensions $(n+1, j)$, with linearly independent columns, which is guaranteed by the affine independence of μ. Let us denote $\tilde{\mu_0} = (\mu_{0,1}, \ldots, \mu_{0,n}, 1)\}$ which is the $(1, n+1)$ vector of cartesian coordinates of μ_0 with an added 1 for the $n+1$'st coordinate. Now we can write:

$$T(\mu)\tau(\mu) = \tilde{\mu}_0$$

We also know the left inverse of $T(\mu)$ exists by affine independence, denoted $T^L(\mu)$, which has dimensions $(k, n+1)$. Then we have that:

$$\tau(\mu) = T^L(\mu)\tilde{\mu}_0.$$

$\square$

Proof of Theorem 2

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 $\mathsf{supp}\,(\tau_k) = \{\mu_1, \ldots, \mu_k\}$. Observe that we can create a $k-1$ dimensional information structure that maintains bayes plausability 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 $\mathsf{supp}\,(\tau'_{12}) = \{\mu_{12}, \mu_3, \ldots, \mu_k\}$, which maintains Bayes plausibility with the new weights $\{(\tau_k(\mu_1) + \tau_k(\mu_2)), \tau(\mu_3), \ldots, \tau(\mu_k)\}$.

Now, we can define k different information structures containing $k-1$ posteriors each, denoted $\mu_{12}, \mu_{23}, \ldots, \mu_{k-1,k}, \mu_{k1}$ 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:

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

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

Dividing all inequalities by k and summing up, we 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:

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

□

Proofs of the statements in section 4.1.1

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 $\mu_s(\omega_i) \geqslant \bar{\pi}$, 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 $\mu_s(\omega_i) > \bar{\pi}$. Define $\delta_1 = (0, 1-\bar{\pi}, -(1-\bar{\pi}))$, $\delta_2 = (1-\bar{\pi}, 0, -(1-\bar{\pi}))$ and $\delta_3 = (1-\bar{\pi}, -(1-\bar{\pi}), 0)$ and $\Gamma_1 = (\bar{\pi}, 0, 1-\bar{\pi})$, $\Gamma_2 = (0, \bar{\pi}, 1-\bar{\pi})$ and $\Gamma_3 = (0, 1-\bar{\pi}, \bar{\pi})$. The action zones will become:

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

where $\cdot$ denotes the Euclidean dot product.

Proof of Lemma 7

Let us first characterize the set Δ_c. We have[10] $\Delta_c = \Delta(\Omega) \backslash \text{co}_2(R_1 \cup R_2 \cup R_3))$. We note that:

$$\begin{aligned} \text{co}(R_1 \cup R_2) =& \text{co}(\{\omega_1, (\bar{\pi}, 1-\bar{\pi}, 0), (\bar{\pi}, 0, 1-\bar{\pi}), \omega_2, (1-\bar{\pi}, \bar{\pi}, 0), (0, \bar{\pi}, 1-\bar{\pi})\}) \\ =& \text{co}\{\omega_1, (\bar{\pi}, 0, 1-\bar{\pi}), \omega_2, (0, \bar{\pi}, 1-\bar{\pi})\} \end{aligned} \tag{4}$$

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

$$\text{co}(R_1 \cup R_3) = \text{co}\{\omega_1, (\bar{\pi}, 1-\bar{\pi}, 0), \omega_3, (0, 1-\bar{\pi}, \bar{\pi})\} \tag{5}$$

$$\text{co}(R_2 \cup R_3) = \text{co}\{\omega_2, (1-\bar{\pi}, 0, \bar{\pi}), \omega_3, (1-\bar{\pi}, 0, \bar{\pi})\} \tag{6}$$

The second line follows from the first line since the $\{\omega_1, (\bar{\pi}, 0, 1-\bar{\pi}), \omega_2, (0, \bar{\pi}, 1-\bar{\pi})\}$ corresponds to the extreme points of $\text{co}(\{\omega_1, (\bar{\pi}, 1-\bar{\pi}, 0), (\bar{\pi}, 0, 1-\bar{\pi}), \omega_2, (1-\bar{\pi}, \bar{\pi}, 0), (0, \bar{\pi}, 1-\bar{\pi})\})$. Similarly using equation (4), (5) and (6), $\text{co}(R_i \cup R_j)$ can be identified as the intersection

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

of a half space and the simplex i.e.

$$\mathrm{co}(R_1 \cup R_2) = \Delta(\Omega) \cap \{(\mu - (\bar{\pi}, 0, 1-\bar{\pi})) \cdot (-\bar{\pi}, \bar{\pi}, 0) \geqslant 0 | \mu \in \mathbb{R}^3\} \tag{7}$$

$$\mathrm{co}(R_1 \cup R_3) = \Delta(\Omega) \cap \{(\mu - (\bar{\pi}, 1-\bar{\pi}, 0)) \cdot (-\bar{\pi}, 0, \bar{\pi}) \geqslant 0 | \mu \in \mathbb{R}^3\} \tag{8}$$

$$\mathrm{co}(R_2 \cup R_3) = \Delta(\Omega) \cap \{(\mu - (1-\bar{\pi}, \bar{\pi}, 0)) \cdot (0, -\bar{\pi}, \bar{\pi}) \geqslant 0 | \mu \in \mathbb{R}^3\} \tag{9}$$

So we can define $\Delta_c \subset \Delta(\Omega)$ as $\Delta_c = \Delta(\Omega) \backslash \mathrm{co}_2(R_1 \cup R_2 \cup R_3)$. By (7), (8) and (9) we can see that Δ_c is defined as

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

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

Proof of Lemma 8

We can identify the upper bounds through the following problem:

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

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 $\mu_0 \in \mathrm{co}(\mu_i, \mu_4)$ 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 $\mu_0^* = (1-\bar{\pi}, 1-\bar{\pi}, 2\bar{\pi}-1)$, $\mu_1^* = (\frac{1-\bar{\pi}}{2}, \frac{1-\bar{\pi}}{2}, \bar{\pi})$ $\mu_4^* = (0, \frac{1}{2}, \frac{1}{2})$. 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:

$$||(\bar{\pi}, \frac{1-\bar{\pi}}{2}, \frac{1-\bar{\pi}}{2}) - (2\bar{\pi}-1), 1-\bar{\pi}, 1-\bar{\pi}|| = \frac{\sqrt{6}}{2}(1-\bar{\pi})$$

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

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

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

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 $\bar{\pi}$.

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 $\mu_1^* = (\frac{1-\bar{\pi}}{2}, \frac{1-\bar{\pi}}{2}, \bar{\pi})$.

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 $\underline{V(2, \mu_0)} = \frac{1}{3\bar{\pi}}$. □

Proof of corollary 1

Observe that with fixed $\bar{\pi} = 2/3$, we have $\overline{V(2, \mu_0)} = \frac{1}{2} = \underline{V(2, \mu_0)}$. Also, $\overline{V(2, \mu_0)} = \frac{2\bar{\pi}-1}{\bar{\pi}}$ is increasing in $\bar{\pi}$ and $\underline{V(2, \mu_0)} = \frac{1}{3\bar{\pi}}$ is decreasing in $\bar{\pi}$. 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 of value of precision, $V(2, \mu_0) > \frac{1}{2}$ implies decreasing value of precision and $V(2, \mu_0) < \frac{1}{2}$ implies increasing value of precision. □

B Additional Results and Details

B.1 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.*

B.2 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 $S_a^{a'} := \{\mu \in \Delta(\Omega) : \mu \in R_a$ 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 10. *$\forall a, a' \in A$ $S_a^{a'}$ 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 11. *$\forall a, a' \in A$, $\hat{u}^S$ 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 2. *$\forall a \in A$, $\hat{u}^S$ 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 $R_a \cap R_{a'} = \boldsymbol{Bd}(R_a) \cap \boldsymbol{Bd}(R_{a'})$.*

B.3 Properties of $\hat{u}^R$ and receivers preferences for signal space cardinality

Lemma 12. *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.[11] ■

An immediate implication is the following.

Corollary 3. *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.4 Formal preferences for example in section 4.2.1 (Optimal Advice Seeking)

We say that the sender's utility only depends on the action, and a_2 and a_3 are preferred over a_1 and a_4, with the default action being the least preferred action, which we call a_0. For the parametric example drawn in figure 5, we set $u^s(a_0) = 0$,$u^s(a_1) = u^s(a_4) = 1$,$u^s(a_2) = u^s(a_3) = 10$.

The receiver has preferred actions when their beliefs are certain about the states. When the beliefs on ω_1, ω_2 or ω_3 are high enough, they prefer a_1, a_2, a_3 respectively. The default action is a_0, which is taken when the beliefs are 'leaning towards' ω_1, and there is another action a_4, which is taken when the beliefs are 'leaning away from' ω_1 but are not sufficiently close to ω_2 or ω_3. Formally, for the example in the figure, we define receiver utility as follows:

$$u^r(\omega_1, a_0) = -1, u^r(\omega_2, a_0) = 12, u^r(\omega_3, a_0) = 12$$
$$u^r(\omega_1, a_1) = 10/3, u^r(\omega_2, a_1) = 10/3, u^r(\omega_3, a_1) = 10/3$$
$$u^r(\omega_1, a_2) = -100/3, u^r(\omega_2, a_2) = 83/3, u^r(\omega_3, a_2) = -100/3$$
$$u^r(\omega_1, a_3) = -100/3, u^r(\omega_2, a_3) = -100/3, u^r(\omega_3, a_3) = 83/3$$
$$u^r(\omega_1, a_4) = -50/3, u^r(\omega_2, a_4) = 58/3, u^r(\omega_3, a_4) = 58/3$$

B.5 Simplicity in Persuasion

In the main text, we have shown that we can restrict attention to affinely independent structures while searching for the optimal information structure. The goal of this section is to clarify the connection between affine independence of information structures, preferences towards simplicity and cognitive costs arising from complexity. We formalize cognitive costs by making the sender not only care about the payoffs of the persuasion game, but also the complexity of the information structures implemented. Our approach and definition of complexity is motivated by the seminal paper of Rubinstein (1986) who studies complexity

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

of automata strategies in repeated games. We opt for a similar simple formalization that defines complexity of an information structure by the number of different posteriors induced i.e. the cardinality of the support of $\tau \in \Delta(\Delta(\Omega))$. This can be analogously thought as having a mental cost for each posterior induced by a signaling strategy. We work on the limiting case of infinitesimal costs. Thus, the sender primarily cares about the payoff, and cares lexicographically, only secondarily, about the number of posteriors induced. Formally, we can define the preference relation $\succ$ of the sender by defining

$$\tau \succ \tau' \text{ if } (\mathbb{E}_\tau \hat{u}^s, -|\mathsf{supp}\,(\tau)|) >_L (\mathbb{E}_{\tau'} \hat{u}^s, -|\mathsf{supp}\,(\tau')|)$$

where $>_L$ is the usual lexicographic[12] order on $\mathbb{R}^2$. This notion of complexity is fairly simple and intuitive, and captures some important considerations. The simplest way to motivate the cost of an additional signal is by assuming that generating higher dimensional signals is costly, and committing to an information structure with more signals and more action recommendations implies that the sender should invest in more capacity to send each different signal that is sent with positive probability.

Given a standard persuasion game with no limitations on the signal space and a sender who has preferences for simplicity, we can extend the result of Theorem 2. We can now state affine independence as a necessary condition of optimality and state that for every information structure τ whose support μ is not affinely independent there exists a strictly better information structure that is preferred by the sender. The result follows from the construction provided in the proof of Theorem 2. Existence of the optimal information structure is again established by Theorem 3.

These observations present an additional property of affinely independent information structures, as they also happen to be the simplest (in the sense of the lexicographic order defined above) possible information structures, within the set of information structures that achieve the same utility level. Hence, our analysis of Bayesian persuasion with coarse communication yields a general solution to Bayesian persuasion games where the agents have preferences for simplicity.

The lexicographic preference order defined above is analogous to having infinitesimal costs for additional signals. In general, using our definition for the value of precision, the sender can decide whether it is worth incurring the cost of an additional signal when costs are non-trivial.

[12] $(x_1, x_2) >_L (y_1, y_2)$ if and only if $x_1 > y_1$ or $x_1 = y_1$ and $x_2 > y_2$. That is to say that $\tau \succ \tau'$ if and only if $\mathbb{E}_\tau \hat{u}^s > \mathbb{E}_{\tau'} \hat{u}^s$ or $\mathbb{E}_\tau \hat{u}^s = \mathbb{E}_{\tau'} \hat{u}^s$ and $|\mathsf{supp}\,(\tau)| < |\mathsf{supp}\,(\tau')|$.

C Extension: Continuum of States

In this section, we will extend our results to the case where the state of the world ω can take values in a continuum i.e. $\Omega = [a, b]$. Without loss of generality, set $a = 0, b = 1$. Sticking to our usual notation, let τ be a signal or an information structure, and the signal space be S with cardinality K. The general setting is akin to Gentzkow and Kamenica (2016).

Suppose the action of the receiver only depends on the expected value of the state variable, $E_\mu(\omega)$, where μ is a posterior belief (a probability distribution) over Ω. Let F_0 be the CDF of the prior belief, with the mean m_0. A signal realization $s \in S$ will induce a posterior belief with CDF μ_s.

Each signal or information structure τ will induce at most K different posterior CDF's, denoted $\{\mu_1, \ldots, \mu_k\}$ with corresponding means $\{m_1, \ldots, m_k\}$. Note that τ will now induce a probability distribution over *posterior means*. Denote CDF of this distribution of posterior means by G.

We make the following assumptions: The set of actions A has cardinality and that there exists cutoffs $\gamma_0, \ldots, \gamma_m$ such that when $E_\mu(\omega) \in [\gamma_i, \gamma_{i+1}]$, the action a_i is optimal for the receiver. Additionally we assume that the sender's utility depends only on receiver's action and that u is an affine-closed function, and satisfies regularity conditions, defined in Dworczak and Martini (2019). Further, assume that the prior CDF, F_0, be continuous and have full support over Ω. These assumptions ensure that the optimal signal creates a distribution of posterior means which is a *monotone partitional signal*.

A monotone partitional signal partitions the state space into at most K continuous intervals such that for any interval in $\{[x_i, x_{i+1}]\}_{i=0}^{K}$, all the mass of G is on $E(X|X \in [x_i, x_{i+1}])$.

Let c_0 be the integral of the posterior mean function for the completely uninformative signal, which will be equal to 0 below the prior mean, and a linear function with slope 1 above the prior mean. Similarly, let c_1 be the integral of the posterior mean function for the fully revealing signal (which will use infinitely many signals). This signal reveals the state exactly. Therefore it will be equal to the integral of the prior.

It is shown by Gentzkow and Kamenica (2016) that the function c for any form of signal must lie between c_0 and c_1. Note that both of these depend on the prior. Now, note the following observation: the cardinality of the signal space K, determines how many 'kinks' the function c will have.

It is straightforward to observe that , with k monotone partitional signals, we will have k 'kinks' and a $k + 1$ piecewise linear functions as c. This follows from the fact that we are interested in the integral of G. Therefore the sender's problem reduces to choosing the location of these k kinks and the slope of the function c at each kink, subject to the constraint that c lies between c_0 and c_1. Remember our assumption of the existence of action cutoffs $\gamma_0, \ldots, \gamma_m$ such that when $E_\mu(\omega) \in [\gamma_i, \gamma_{i+1}]$, the action a_i is optimal for the receiver. The relationship between $\gamma_0, \ldots, \gamma_m$ and the signal partitions will not be obvious when $K < M$.

More precisely, let c_G denote the integral of G, $c_G(x) = \int_0^x G(t)dt$. c_G is a convex function and we can analyze c instead of analyzing signal distributions as in Gentzkow and Kamenica (2016). This definition also makes our focus on piecewise linear functions more clear. Gentzkow and Kamenica (2016) shows that each function in this interval can be represented by a signaling policy and vice versa. We will focus on solving the problem by choosing a function between c_0 and c_1 instead of finding signaling policies for tractability purposes. Let $\gamma_1, \ldots, \gamma_m$ be the action cutoffs, and let $c(x)$ be the chosen c function, with c' representing the left derivative. Action one is taken when $c_G(x) < \gamma_1$. Let U_1 be the sender utility when action 1 is taken. Action two is taken when $\gamma_1 \geqslant c_G(x) < \gamma_2$, let U_2 be sender utility when action 2 is taken, and so forth. The sender's utility is then

$$U(c') = \sum_{k=1}^{m} \left(c'(\gamma_k) - c'(\gamma_{k-1})\right) U_k$$

with the convention that $\gamma_0 = 0$ and $c'(\gamma_0) = 0$.

The set of possible functions c as:

$$\begin{aligned} \mathscr{F}_k = \{f \in C[0,1] | \exists \text{ a partitioning of [0,1] into k intervals: } \{s_l\}_{l=1}^k = \{(0, \overline{x_1}], (\overline{x_1}, \overline{x_2}], \ldots, (\overline{x_{k-2}}, \overline{x_{k-1}}], (\overline{x_{k-1}}, 1]\} \\ \text{and } \{\phi_l \in \mathbb{R}\}_{l=1}^k \text{such that: } k < K, \exists M \in \mathbb{N}\ \forall l \in \{1, \ldots, k\}\ 0 \leqslant \phi_l < M, \phi_l \leqslant \phi_{l+1}, \\ \text{and each } s \text{ is connected and has non-zero measure, where } f \text{ can be written as:} \\ f(x) = \mathbb{1}_{x \in s_1}(\phi_1 x) + \sum_{l=2}^{k} \mathbb{1}_{x \in s_l} \left(\phi_l x - \sum_{j=2}^{l} (\phi_j - \phi_{j-1}) \overline{x_{j-1}}\right) \} \end{aligned}$$

Given the definitions and the signal space of focus we establish existence of an optimal information structure for the sender.

Theorem 3. *$U(c_G)$ attains its maximum over $\mathscr{F}_k$.*

Proof. The proof proceeds by a series of Lemmas:

Lemma 13. *$\mathscr{F}_k$ is pre-compact*

Proof. By *Arzela-Ascoli* Theorem, proving pre-compactness suffices to showing *equi-continuity* and *equi-boundedness*. Note that the way that $\mathscr{F}_k$ defined ensures that its elements are Lipschitz continuous. Then we have that *equi-boundedness* trivially. For *equi-continuity* pick $M \in \mathbb{N}$ that is the largest Lipschitz constant for the set of functions in $\mathscr{F}$ and a set of functions with bounded Lipschitz constant forms an equicontinuous set. ■

Lemma 14. *$\mathscr{F}_k$ is closed.*

Proof.

Suppose there exists a sequence of functions where $\forall n \in N, f_n \in \mathscr{F}_k$ and $f_n \to f$ uniformly. We will show that $f \in \mathscr{F}_k$.

First, observe that all f_n are Lipschitz continuous, and therefore f must be Lipschitz continuous, in addition to being convex. Therefore f is differentiable almost everywhere. Let the set $D \subset [0, 1]$ represent the set of points where f is differentiable.

Since $f_n \to f$ uniformly and f_n, f are convex, we have that $\forall x \in D, f_n'(x) \to f'(x)$. We proceed by proving the following claim.

$(0, 1)\backslash D$ can have at most cardinality K.

Suppose not. Pick $K + 1$ elements from $(0, 1)/D$ and call this set X. By subclaim 2, $\forall x \in X$, we can find $h(x) > 0$ such that $\forall h \in [0, h(x))$, there exists some $N_{h(x)} \in N$ such that $\forall n > N_{h(x)}$, $f_n(x)$ and $f_n(x - h)$ are on the same linear piece. Similarly, we can also find $q(x) > 0$ such that $\forall q \in [0, q(x))$, there exists some $N_{q(x)} \in N$ such that $\forall n > N_{q(x)}$, $f_n(x)$ and $f_n(x + q)$ are on the same linear piece. Since there are $K + 1$ elements in X, we can compute $q^* = \min_{x \in X}(q(x))$, $h^* = \min_{x \in X}(h(x))$, and $N^* = \max_{x \in X}(\max(N_{h(x)}, N_{q(x)}))$.

Since f is differentiable almost everywhere, for every $x \in X \subseteq ((0, 1)/D)$, there must exist $\epsilon_1(x) > 0$ such that f is differentiable in the interval $(x - \epsilon_1(x))$ and also $\epsilon_2(x) > 0$ such that f is differentiable in the interval $(x + \epsilon_2(x))$. Let $\epsilon^* = \min_{x \in X}(\min(\epsilon_1(x), \epsilon_2(x)))$.

Define $\epsilon = \min(h^*, q^*, \epsilon^*)$. Now, $\forall x \in X$, and $\forall n > N^*$, we have that $f_n' = c_1(x)$ within the interval $(x - \epsilon, x)$, and $f_n' = c_2(x)$ within the interval $(x, x + \epsilon)$, for some constants $c_1(x), c_2(x)$. The intervals $(x - \epsilon, x)$ and $(x, x + \epsilon)$ are contained by the set D for every value of x, by definition. By the fact that within the set $D, f_n' \to f'$, we must have $f' = c_1(x)$ within $(x - \epsilon, x)$ and $f' = c_2(x)$ within $(x, x + \epsilon)$.

Since f is continuous and convex, and $\forall x \in X$, $f'(x)$ doesn't exist, we must have that $\forall x, c_1(x) < c_2(x)$. However, this implies that $\forall n > N^*$, f_n' also takes at least $K + 2$ unique values, which contradicts the fact that $f_n \in \mathscr{F}_k$, i.e., f_n cannot be K-piecewise linear. This completes the proof that $(0, 1)/D$ can have at most cardinality K.

Without loss of generality, suppose the set has cardinality K. The case where the cardinality is less than K will be analogous. Let us order the elements of $(0, 1)/D$ as $0 < x_1 < x_2 \cdots < x_K < 1$. Take the collection of intervals whose union is $[0, 1]$ as $\{s_l\}_{l=1}^K = \{[0, x_1], (x_1, x_2], \ldots, (x_K, 1]\}$. Within the interior of each interval, f is differentiable, hence we must have $f_n' \to f'$. Observe that f' can take at most $K + 1$ unique different values, because otherwise the convergence of f_n' cannot hold. Moreover, f' must be constant within the interior of each interval, since otherwise the cardinality of $(0,1)/D$ would exceed K.

Therefore, we can write $\forall l < K : \forall x \in \textbf{int}(s_l), \phi_l = f'(x)$, and hence $f(x) = \phi_l x + c_l$ for some c. Moreover, since $\forall n \in N, f_n$ and f are continuous, $\forall x \in (0,1)/D$, we must have:

$$\lim_{\epsilon\to 0} f(x+\epsilon) = \lim_{\epsilon\to 0} f(x-\epsilon) = \lim_{\epsilon\to 0} \phi_{l+1}(x+\epsilon) + c_{l+1} = \lim_{\epsilon\to 0} \phi_l(x-\epsilon) + c_l$$

Therefore to preserve continuity we must have $c_{l+1} - c_l = -(\phi_{l+1} - \phi_l)x$. Also, observe that within the first interval $[0, x_1]$, we have $f_n' \to f' = \phi_1$ and $f_n(x) = \phi_{1,n}x \to f(x) = \phi_1 x + c_1$. It follows that we must have $c_1 = 0$ to have convergence in derivatives and in values within the interval.

This shows that for $l \geqslant 2, c_l = -\sum_{i=2}^{l}(\phi_i - \phi_{i-1})x_{i-1}$. Therefore, f must have the desired form and $f \in \mathscr{F}_k$. This completes the proof that it is closed. ∎

Corollary 4. *$\mathscr{F}_k$ is compact.*

Proof. Follows from two Lemmas above and the definition of a pre-compact set. ∎

Lemma 15. *U(c) is continuous over $\mathscr{F}_k$.*

Proof. Let $f_n \in C$ be a sequence of convex functions such that $f_n \to f$ uniformly. This implies : $d(f_n, f) = \sup\{|f_n(x) - f(x)|, x \in [0,1]\} \to 0$ as $n \to \infty$. We need to show $U(f_n) \to U(f)$.

By above Lemma, since U only depends on the left derivatives on fixed and exogenous points $\gamma_1, \ldots, \gamma_k$, then we will have $U(f_n) \to U(f)$. Uniform convergence implies pointwise convergence, therefore f is convex.

Since f is convex, there will exist left and right derivatives at every point. For any γ value, and for any $\epsilon > 0$, we need to show $\exists N \in \mathbb{N}$ such that $\forall n > N$, $|f_n'(\gamma) - f'(\gamma)| < \epsilon$ where we write the left derivative at γ as:

$$f'(\gamma) = \lim_{h\to 0-} \frac{f(\gamma+h) - f(\gamma)}{h}$$

We proceed by proving two useful claims.

Claim 1. *$\exists h_1 > 0$ such that $\forall 0 \leqslant h < h_1$, $f(\gamma - h)$ and $f(\gamma)$ are on the same linear piece, meaning that: $f(\gamma - h) = \beta(\gamma - h)$ and $f(\gamma) = \beta\gamma$ for some $\beta > 0$. This implies $f'(\gamma - h) = f'(\gamma)$, $\forall 0 \leqslant h < h_1$.*

Proof. Follows from the fact that in our definition each linear piecewise interval is connected and has strictly non zero measure. ■

Claim 2. *$\exists h_2 > 0$ that satisfies the following : $\forall h \in [0, h_2)$, there exists some $N_h \in \mathbb{N}$ for which it holds that $\forall n > N_h$, $f_n(\gamma - h)$ and $f_n(\gamma)$ are on the same linear piece.*

Proof. Suppose not. For any given $h_2 > 0$, for all $h \in [0, h_2)$, there exists no N_h. Meaning that, $\forall n \in \mathbb{N}$, $f_n(\gamma - h)$ and $f_n(\gamma)$ are not on the same linear piece. Implying that, for any $0 \leqslant h < h_2$, for any n: there must be some β_n, θ_n where $f_n(\gamma - h) = \beta_n(\gamma - h)$ and $f_n(\gamma) = \theta_n \gamma$ where $\beta_n < \theta_n$ by convexity. Thus, $|f_n(\gamma) - f_n(\gamma - h)| = |(\theta_n - \beta_n)\gamma + \beta_n h|$. However, each f_n is also continuous, by convexity. This implies that, at the point γ: $\forall \epsilon > 0$, $\exists \delta > 0$ such that if $|x - \gamma| < \delta$, then $|f_n(x) - f_n(\gamma)| < \epsilon$.

For any f_n, choose $\epsilon = (\theta_n - \beta_n)\gamma$. Then, there exists some δ such that $|x - \gamma| < \delta$ implies $|f_n(x) - f_n(\gamma)| < (\theta_n - \beta_n)\gamma$ But then we can choose h where $h < h_2$ and $h < \delta$ is satisfied. Which means that we will have: $|f_n(\gamma) - f_n(\gamma - h)| = |(\theta_n - \beta_n)\gamma + \beta_n h| = (\theta_n - \beta_n)\gamma + \beta_n h$ from the first argument, and$|f_n(\gamma) - f_n(\gamma - h)| < (\theta_n - \beta_n)\gamma$ from the second argument. Therefore we have reached a contradiction. This completes the proof of claim 2. ■

Proceeding with the proof of Lemma 13, we have that uniform convergence implies pointwise convergence, therefore f is convex. Since f is convex, there will exist left and right derivatives at every point. For any γ value, and for any $\epsilon > 0$, we need to show $\exists N \in \mathbb{N}$ such that $\forall n > N$, $|f_n'(\gamma) - f'(\gamma)| < \epsilon$. Where we write the left derivative at γ as:

$$f'(\gamma) = \lim_{h \to 0+} \frac{f(\gamma - h) - f(\gamma)}{-h}$$

Suppose an $\epsilon > 0$ is given. By claim 1 and claim 2, pick an $h < \min\{h_1, h_2\}$. We have that:

$$f'(\gamma) - \epsilon < \frac{f(\gamma - h) - f(\gamma)}{-h} = f'(\gamma) = f'(\gamma - h) < f'(\gamma) + \epsilon$$

For the picked number h, by claim 2, let N_h be the number where $\forall n > N_h$, $f_n(\gamma - h)$ and $f_n(\gamma)$ are on the same linear piece.

Since f_n converges to f, there exists $N_c \in \mathbb{N}$ such that $\forall n > N_c$:

$$f'(\gamma) - \epsilon < \frac{f_n(\gamma - h) - f_n(\gamma)}{-h} < f'(\gamma) + \epsilon$$

Let $N > \{N_h, N_c\}$. Then, $\forall n > N$, the convergence result holds, and $f_n(\gamma - h)$ and $f_n(\gamma)$ are on the same linear piece. The following argument holds for all $n > N$: Since $f_n(\gamma - h)$ and $f_n(\gamma)$ are on the same linear piece, we must have that the left derivatives are the same at these two points and $f_n'(\gamma) = \frac{f_n(\gamma - h) - f_n(\gamma)}{-h}$. By direct substitution to the inequality above:

$$f'(\gamma) - \epsilon < f_n'(\gamma) < f'(\gamma) + \epsilon$$

$$\Leftrightarrow -\epsilon < f_n'(\gamma) - f'(\gamma) < \epsilon$$
$$\Leftrightarrow |f_n'(\gamma) - f'(\gamma)| < \epsilon$$

Therefore the left derivatives converge and $U(f_n) \to U(f)$, which completes the proof that U(c) is continuous over $\mathscr{F}_k$.

■

With all the Lemmas, the proof of Theorem 4 follows immediately by topological extreme value Theorem[13]. We have proved the existence of an optimal monotone partitional information structure. ■

[13]Let (S, d_S) and $(\mathbb{R}, d)$ be metric spaces d is the usual Euclidean metric defined for all x, y $d(x, y) = |x-y|$. Also let $X \subseteq S$ be a compact subset of S $f : S \to T$ be continuous on all of X. Then $f(X)$ is closed and bounded in T and f achieves its supremum and infimum on X, that is, there exists $p, q \in X$ such that $f(p) = \sup\{f(x) : x \in X\}$ and $f(q) = \inf\{f(x) : x \in X\}$

D Equivalent Full Dimensional Games

Recall that our that our setting is the following:

- State Space: $\omega \in \Omega$ with $|\Omega| = n$,
- Action Space: $a \in A$ with $|A| = m$,
- Signal Space: $s \in S$ with $|S| = k$.
- $k \leqslant \min\{m, n\}$

We make the following definitions to ease readiblity:

- $A_R^* : \Delta\Omega \rightrightarrows A$ with $\mu \mapsto \arg\max_{a \in A} \sum_{\omega \in \Omega} \mu(\omega) u^R(a, \omega)$
- $A_S^* : \Delta\Omega \rightrightarrows A$ with $\mu \mapsto \arg\max_{a \in A_R^*} \sum_{\omega \in \Omega} \mu(\omega) u^S(a, \omega)$
- $\hat{u}^S : \Delta\Omega \to \mathbb{R}$ with $\mu \mapsto \max_{a \in A_S^*} \sum_{\omega \in \Omega} \mu(\omega) u^s(a, \omega)$
- $R_a = \{\mu \in \Delta\Omega | a \in A_R^*\}$.
- $S_a = \{\mu \in \Delta\Omega | a \in A_S^*\}$

We have proven the following result in appendix B:

Lemma 16. *R_a and S_a are convex and closed sets. Moreover $\hat{u}^S$ is affine over S_a.‘*

Proposition 3. *There exists optimal information structure τ that induces posterior beliefs $\{\mu_i\}_{i \leqslant k}$ such that $A^S(\mu_i) \cap A^S(\mu_j) = \varnothing$ for each i, j such that $i \neq j$.*

Proof. Suppose not. Then there exists beliefs μ_i and μ_j induced under the optimal information structure such that $A^S(\mu_i) \cap A^S(\mu_j) = a'$. Then define the collapse of μ_i and μ_j by $\mu_{ij} = \tau(\mu_i)\mu_i + \tau(\mu_j)\mu_j$. Define the information structure τ' as $\tau'(\mu_r) = \tau(\mu_r)$ and $\tau'(\mu_{ij}) = \tau(\mu_i) + \tau(\mu_j)$.

First it should be obvious that this collapse preserves Bayes plausibility. Now we will show that it preserves sender payoff. By Lemma above, S_a are convex and closed sets and $\hat{u}^S$ is affine over S_a. So we have that $\mu_{ij} \in S_a$ and $\mathbb{E}_\tau \hat{u}^S(\mu) = \mathbb{E}_{\tau'} \hat{u}^S(\mu)$, since

$$\frac{\tau(\mu_i)}{\tau(\mu_i + \mu_j)} \hat{u}^S(\mu_i) + \frac{\tau(\mu_j)}{\tau(\mu_i + \mu_j)} \hat{u}^S(\mu_j) = \hat{u}^s(\tau(\mu^i)\mu^i + \tau(\mu_j)\mu_j) = \hat{u}^s(\mu_{ij})$$

Hence we established that

$$\tau(\mu_i)\hat{u}^s(\mu_i) + \tau(\mu_j)\hat{u}^s(\mu_j) = (\tau(\mu_i) + \tau(\mu_j))\hat{u}^s(\mu_{ij}) = \tau'(\mu_{ij})\hat{u}^s(\mu_{ij})$$

Doing this iteratively, we establish an optimal information structure that induces beliefs that induce different actions. ■

Let $l = C(m, k)$ and $\{A_1, \ldots, A_l\}$ be the set of k element subsets of A. Define $\mathscr{S}^i = \{S_a\}_{a \in A_i}$. As a Corrolary of our Proposition we have this:

Corollary 5. *Optimal information structure induces at most k different actions. Hence, for each μ_0 we have that*

$$\max_{\tau} \mathbb{E}_{\tau}\hat{u^s}(\mu) \text{ subject to } \mathbb{E}_{\tau}(\mu) = \mu_0 \text{ and } |supp(\tau)| \leqslant k$$

is equivalent to

$$\max_{\tau} \mathbb{E}_{\tau}\hat{u^s}(\mu) \text{ subject to } \mathbb{E}_{\tau}(\mu) = \mu_0 \text{ and } \mu \in \mathcal{S}_i \text{ for some } i \in \{1, \ldots, l\}$$

Define a ficticious sender utility to replicate the case using actions in $\mathscr{S}_i$. Start by defining concavification using only Z as:

$$\mathbb{CH}(\hat{u^s}|Z) : \text{co}(Z) \to \mathbb{R} \text{ with } \mu \mapsto \sup\{(\mu, z)|(\mu, z) \in \text{co}(\text{hypo}(\hat{u}^s|_Z))\}$$

Now we can define the equivalent sender utility of the ficticious game as

$$\overline{u_i^s} = \begin{cases} \hat{u^s}(\mu) & \text{if } \mu \in \mathscr{S}_i \\ \mathbb{CH}(\hat{u}_i^s|\mathbf{Bd}(\text{co}(\mathscr{S}_i)\backslash\mathscr{S}_i)) & \text{if } \mu \in \text{co}(\mathscr{S}_i)\backslash\mathscr{S}_i \\ -\infty & \text{if } \mu_0 \in \Delta(\Omega)\backslash \text{co}(\mathscr{S}_i) \end{cases}$$

Finally define

$$\overline{\mathbb{CH}(\hat{u^s}|\mathscr{S}_i)} : \begin{cases} \mathbb{CH}(\hat{u^s}|\mathscr{S}_i) & \text{if } \mu_0 \in \text{co}(\mathscr{S}_i) \\ -\infty & \text{if if } \mu_0 \in \Delta(\Omega)\backslash \text{co}(\mathscr{S}_i) \end{cases}$$

Corollary 6.

$$\overline{\mathbb{CH}(\hat{u^s}|\mathscr{S}_i)} = \mathbb{CH}(\overline{u_i^s})(\mu_0)$$

and hence

$$V(\mu_0) = \mathbb{CH}_k(\hat{u^s})(\mu_0) = \max_{i:\mu_0 \in \mathscr{S}_i} \mathbb{CH}(\hat{u^s}|\mathscr{S}_i)(\mu_0) = \max_{i:\mu_0 \in \mathscr{S}_i} \overline{\mathbb{CH}(\hat{u^s}|\mathscr{S}_i)} = \max_{i:\mu_0 \in \mathscr{S}_i} \mathbb{CH}(\overline{u^s}(\mathscr{S}_i))(\mu_0)$$

Equivalent Cheap Talk Games

Now we can define the equivalent sender utility of the ficticious game as

$$\overline{u_i^s}(\mu) = \begin{cases} \hat{u^s}(\mu) & \text{if } \mu \in \mathscr{S}_i \\ \mathbb{CH}(\hat{u}_i^s|\mathbf{Bd}(\text{co}(\mathscr{S}_i)\backslash\mathscr{S}_i)) & \text{if } \mu \in \text{co}(\mathscr{S}_i)\backslash\mathscr{S}_i \\ \alpha\cdot\mu_0+\beta & \text{if } \mu_0 \in \Delta(\Omega)\backslash\,\text{co}(\mathscr{S}_i) \end{cases}$$

Finally define

$$\overline{\mathbb{CH}(\hat{u^s}|\mathscr{S}_i)} : \begin{cases} \mathbb{CH}(\hat{u^s}|\mathscr{S}_i) & \text{if } \mu_0 \in \text{co}(\mathscr{S}_i) \\ \alpha\cdot\mu_0+\beta & \text{if if } \mu_0 \in \Delta(\Omega)\backslash\,\text{co}(\mathscr{S}_i) \end{cases}$$

Lemma 17. *There exists $\alpha, \beta \in \mathbb{R}^\Omega$ such that*

$$\mathbb{CH}(\overline{u_i^s})) = \overline{\mathbb{CH}(\hat{u^s}_i)}$$

Proof. The proof is ommited. The proof follows from looking each combination using extreme points in $\Delta\Omega\backslash\mathscr{S}_i$, as optimal information structures are supported on extreme points. ■

Corollary 7.

$$V(\mu_0) = \mathbb{CH}_k(\hat{u^s})(\mu_0) = \max_{i:\mu_0\in\mathscr{S}_i} \mathbb{CH}(\hat{u^s}|\mathscr{S}_i)(\mu_0) = \max_{i:\mu_0\in\mathscr{S}_i} \overline{\mathbb{CH}(\hat{u^s}_i)} = \mathbb{CH}(\overline{u_i^s})$$

Theorem 4. *K-limited bayes persuasion game $\Gamma_{BP} = (\{Receiver, Sender\}, \{u^R, u^S\}, \Omega, (A, S), \mu_0)$ is equivalent to the cheap talk game with $\Gamma_{CT} = \{Receiver, Sender\}, \{u^R, \overline{u_i^s}\}, \Omega, (A, S), \mu_0)$ for $i \in \arg\max_{i:\mu_0\in\mathscr{S}_i} \mathbb{CH}(\overline{u_i^s})(\mu_0)$ whenever $i \in \arg\max_{i:\mu_0\in\mathscr{S}_i} \mathbb{CH}(\overline{u_i^s})(\mu_0)$ we have that $\cap_{j:A_j\in\mathscr{S}_i} A_j = \varnothing$.*

Proof. This follows from Lipnowski (2020) and the fact that $\overline{u^s}_i$ is continious given the conditions. ■

References

ALFSEN, E. M. (1965): "On the geometry of Choquet simplexes," *Mathematica Scandinavica*, 15, 97–110.

AUMANN, R. J. AND M. MASCHLER (1995): *Repeated Games with Incomplete Information*, MIT Press.

BERGEMANN, D. AND S. MORRIS (2016): "Information design, Bayesian persuasion, and Bayes correlated equilibrium," *American Economic Review*, 106, 586–91.

BLOEDEL, A. W. AND I. R. SEGAL (2018): "Persuasion with Rational Inattention," *Available at SSRN 3164033*.

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.

DWORCZAK, P. AND G. MARTINI (2019): "The simple economics of optimal persuasion," *Journal of Political Economy*, 127, 000–000.

GENTZKOW, M. AND E. KAMENICA (2014): "Costly persuasion," *American Economic Review*, 104, 457–62.

——— (2016): "A Rothschild-Stiglitz approach to Bayesian persuasion," *American Economic Review*, 106, 597–601.

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.

LE TREUST, M. AND T. TOMALA (2019): "Persuasion with limited communication capacity," *Journal of Economic Theory*, 104940.

LIPNOWSKI, E. (2020): "Equivalence of Cheap Talk and Bayesian Persuasion in a Finite Continuous Model," .

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.

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

Rubinstein, A. (1986): "Finite automata play the repeated prisoner's dilemma," *Journal of economic theory*, 39, 83–96.

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.

Warren, J. (1996): "Barycentric coordinates for convex polytopes," *Advances in Computational Mathematics*, 6, 97–108.

——— (2003): "On the uniqueness of barycentric coordinates," *Contemporary Mathematics*, 334, 93–100.

Warren, J., S. Schaefer, A. N. Hirani, and M. Desbrun (2007): "Barycentric coordinates for convex sets," *Advances in computational mathematics*, 27, 319–338.

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

Yaglom, I. M. and V. G. Boltyansky (1961): *Convex figures*, New York: Holt, Rinehart and Winston.