Private Bayesian Persuasion

Itai Arieli, Yakov Babichenko*

February 17, 2016

Abstract

We consider a multi-agent Bayesian persuasion problem where an informed sender tries to persuade a group of agents to adopt a certain product. The sender is allowed to commit to a signalling policy where she sends a *private* signal to every agent. The payoff to the sender is a function of the subset of adopters. We characterize an optimal signalling policy and the maximal revenue to the sender for three different types of payoff functions: supermodular, symmetric submodular, and a supermajority function. Moreover, using tools from cooperative game theory we provide a necessary and sufficient condition under which public signalling policy is optimal.

1 Introduction

Due to the recent technological advances and the raising of online social networks, new forms of advertising have been developed, viral marketing being one of them. These forms of advertising use the network structure and the fact that customers often share information with others via the network channels to advertise their product and increase its awareness.

This brings two main challenges to marketers that want to increase their revenue by maximizing their number of customers. The first is to understand who

*Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, iarieli@tx.technion.ac.il.

are the "opinion leasers" in the network and what would be the implications, in terms of revenue, if a given subgroup of opinion leaders adopted their product. The second is how one should approach the opinion leaders and what sort of information should be revealed to them in order to achieve a profit maximization. This work focuses on the second question and tries to study the optimal information revelation under the assumption that the marketer can estimate its expected revenue from any subgroup of initial adopters.

The fundamentals of our model have the following features. A group of opinion leaders, henceforth agents, should choose whether to adopt a certain product. A firm, henceforth the sender, that knows all relevant information about the product is allowed to communicate privately with the customers to try to persuade them to purchase the product. Each agent has his own incentives and the sender's utility depends on the group of agents that adopt the product. This defines a multiple-agent problem of information revelation to the sender.

We assume that the sender can commit in advance to a revelation policy in terms of signal distribution. That is the sender commits in advance to a distribution over an abstract set of signals conditional on every state of the world. Our main goal in this paper is to provide an optimal revelation policy for the sender when private communication is allowed. Clearly, such an optimal policy may depend on the utility function of the sender. We restrict attention to utility functions that are increasing in the size of the set of adopters.

The first observation is that the payoff function of the sender is a *coalition function* that specifies a payoff to every subset of agents. We provide an optimal policy for three different types of payoff functions of the sender. The first is a supermodular (convex) function, the second sumodular (concave) symmetric function, and the third is a supermajority function. The convex case represents a sender that his marginal utility from persuading an additional agent is increasing in the size the current group adopters. This may hold due to, for example, decreasing production cost. The concave symmetric case represents a risk averse sender. Finally the supermajority function corresponds to the case where the revenue of the sender is a threshold function. That is, the sender achieves a utility of 1 if the group of adopters comprises at least k agents and 0 otherwise.

1.1 Related Literature

Most of the theoretical literature on viral marketing and advertising in networks such as Kempe et al. [7], and Domingos [5] try to estimate the optimal seeding strategy of the marketer in a given social network. The underlying assumption is that any subset of agents form a contagious process that spreads throughout the network. Here we take a complementary approach and assume that the marketer can accurately estimate his revenue from any subset of potential seeders. We try to focus on the inherent asymmetry between the information that is available to the marketer versus the information of the customers. These assumptions therefore translate the problem to a *Bayesian persuasion* problem for the marketer.

The recent literature on Bayesian persuasion, starting with Kamenica and Gentzkow [6], extends the cheap talk model of Crawford and Sobel [4] by allowing an informed sender to commit in advance to a signalling policy that reveals partial information to the receiver. Building upon the classical work by Aumann and Maschler [2], Kamenica and Gentzkow relate the optimal policy and the revenue of the sender to the concavification of the function that corresponds to the revenue of the sender for every given prior distribution, where communication is forbidden. Another recent paper by Rayo and Segal [9] provides the optimal signalling policy where a realized prospect, which specifies a profit to the sender and to the receiver, is through private information to the sender.

Schnakenberg [10] studies Bayesian persuasion problem where a sender can publicly disclose information to a set of voters prior to a vote over whether or not to implement a proposal. He studies conditions under which the receiver can use information to manipulate collective choices. He shows that this may reduce the ex ante expected utilities of all voters. Another recent paper by Alonso and Camara [1] characterizes, using a novel geometric tool, an optimal public signalling policy in the voting framework. The public signal is determined as a function of the set of *approval beliefs* that comprises all posterior beliefs for which the proposal is accepted by the voters. In addition they show that under a simple-majority rule, a majority of voters might be strictly worse off due to the politician's intervention. Our work is different from these works in three main aspects. First, we allow the sender to communicate privately with the agents. This provides the sender more flexibility to manipulate the agent, by forming either negative or positive correlation among agents' private information. Second we consider the case where there is no payoff externalities among the agents. Third, our analysis is restricted to a binary state space. In light of these results, we also characterize necessary and sufficient conditions under which the optimal policy is achieved via public signals.

Several previous works incorporate private signals in a Bayesian persuasion environment. Wang [12] considers a persuasion model in a voting setup where he restricts attention to private signal distributions that are conditionally independent among the agents given the realised state. We, however, allow for general private signal distribution. Another work by Taneva [11] considers a multi-agent model of information revelation by a sender where the agents have payoff externalities. She characterizes the optimal information structure for the sender subject to the agents playing a Bayes Nash equilibrium in the special case of 2×2 games.

Finally Chan, Li, and Wang [3] study a multi-agent Bayesian persuasion model when the sender has a supermajority function and private signalling policy is allowed. They construct the optimal signalling policy for a sender that is restricted to signal distribution which satisfies a monotone likelihood ratio property. They show that the optimal policy follows a cutoff rule where the sender only tries to influence the beliefs of those voters who are sufficiently easy to persuade. This cutoff property is not guaranteed by our optimal policy. In contrast, we allow for a general signalling policy and provide a precise expression for the optimal revenue of the sender. The main distinction of these works with ours is that we consider the case where there are no payoff externalities among the agents.

2 Preliminaries

Our Bayesian persuasion model comprises a group of agents $N = [n] = \{1, 2, ..., n\}$ and a sender. Each agent $i \in N$ has a binary action set $\{0, 1\}$ and a utility function $u_i : \Omega \times \{0, 1\} \to \mathbb{R}$, where $\Omega = \{\omega_0, \omega_1\}$ is a binary state space. The players share a common prior distribution where $0 < \gamma < 1$ is the probability of state ω_1, and $1 - \gamma$ of state ω_0. The sender's utility $V : \Omega \times \{0, 1\}^N \to \mathbb{R}$ is a function of the state and the group of adopters, i.e., those that choose action 1. We assume throughout that the senders' payoff is monotonically increasing with the group of adopters. That is, $V(\omega, T) \leq V(\omega, S)$ for every $S \subseteq T$ and $\omega \in \Omega$.

For clarity of the exposition, we make two simplifying assumptions. First we assume that all agents prefer action 1 at ω_1 and action 0 at ω_0. Namely, $u_i(\omega_1, 1) > u_i(\omega_1, 0)$ and $u_i(\omega_0, 0) > u_i(\omega_0, 1)$. Second we assume that the Sender's utility is independent of the state. That is $V : 2^N \to \mathbb{R}$. We show in Section 4 that all the analyses can be applied to the general case, where the utility of the agents is any general function and the sender's utility is state dependent.

We assume that the sender is informed of the realized state and the agents are not. As in Kamenica and Gentzkow [6] we allow the sender to commit in advance to an information revelation policy. In this work, however, we allow the sender to reveal the information to every agent privately. This translates to a state-dependent signalling distribution. Formally a policy of the informed sender comprises n finite sets $\{\Theta_i\}_{i=1,...,n}$ where Θ_i is the private signal set of agent i, and a mapping $F : \Omega \to \Delta(\Theta_1 \times \cdots \times \Theta_n)$. The sender can commit to a policy that is known to the agents prior to a stage where the state ω is realized. In contrast to the related literature we allow the sender to communicate privately with the agents, namely to produce different signals to different agents.

The timing of the interaction is as follows. First the sender commits to a signalling policy F. Then a state $\omega \in \Omega$ is realized in accordance with the prior $(\gamma, 1 - \gamma)$. Then a profile of signals $\theta = (\theta_1, \ldots, \theta_n)$ is generated according to F_ω. Every agent i observes his private signal realisation $\theta_i \in \Theta_i$. She then forms a posterior $\mathbf{P}_F(\omega_1|\theta_i) = p(\theta_i)$ and plays 1 if and only if

$$p(\theta_i)u_i(\omega_1, 1) + (1 - p(\theta_i))u_i(\omega_0, 1) \geq p(\theta_i)u_i(\omega_1, 0) + (1 - p(\theta_i))u_i(\omega_0, 0).$$

We assume that in case of indifference, agents take action 1. Agent i's best-reply action when he observes the signal θ_i is denoted by $g_i(\theta_i) \in \{0, 1\}$. We denote by $g(\theta)$ the action profile of the agents when the realized vector of signals is θ. Let $\Theta = \Theta_1 \times \cdots \times \Theta_n$. Let $F_1 \in \Delta(\Theta)$ be the signal distribution conditional on state ω_1 and $F_0 \in \Delta(\Theta)$ be the signal distribution conditional on state ω_0. The sender's utility from the policy (Θ, F) is denoted by

$$s(F) := \gamma \mathbb{E}_{\theta \sim F_1}[V(g(\theta))] + (1 - \gamma)\mathbb{E}_{\theta \sim F_0}[V(g(\theta))]. \tag{1}$$

A signalling policy (Θ, F) is *optimal* if it maximizes sender's utility (among all possible signal sets Θ, and all possible signals $F : \Omega \to \Theta$). Our goal is to characterize the optimal policy for the sender as a function of the properties of V.

2.1 Finding the Optimal Policy

In this section we shall show that in order to find the optimal policy for the sender we can restrict attention to a class of simple policies. In the spirit of Kamenica and Gentzkow we call a signalling policy (F, Θ) *straightforward* if $\Theta_i = \{0, 1\}$ for every agent $i \in N$ and $g_i(\theta_i) = \theta_i$. The following lemma generalizes Proposition 1 in Kamenica and Gentzkow and is fundamental to our analysis. It shows that one can restrict attention to a particular class of straightforward policies and it allows us to translate the optimization problem faced by the sender to a neat linear programming optimization problem.

Lemma 1. *There exists an optimal signalling policy (F, Θ) such that for every player i*

1. $\Theta_i = \{0, 1\}$.

2. $F_1(\theta_i = 1) = 1$ *for every agent i.*

3. $F_0(\theta_i = 1) \leq a_i$ *where* $a_i = \min(\frac{\gamma}{1-\gamma} \frac{u_i(\omega_1,1) - u_i(\omega_0,1)}{u_i(\omega_0,0) - u_i(\omega_0,1)}, 1)$.

Moreover, every policy that satisfies the above conditions is straightforward.

We note that since every policy (F, Θ) that satisfies the above conditions is straightforward, condition (2) implies that the utility of the sender from any such policy is:

$$s(F) = \gamma V(N) + (1-\gamma)\mathbb{E}_{\theta \sim F_0} V(\theta).$$

Since the utility of the sender from any of the above policies is fixed and equals $V(N)$ conditional on state ω_1 we can translate the optimization problem of the sender to finding a policy that satisfies the above conditions and which maximizes the utility of the sender conditional on state ω_0. That is, Lemma 1 translates the original optimization problem of the sender to a simpler problem where the sender should find a signalling policy that maximizes his payoff conditional on state ω_0 subject to the constraint $F_0(\theta_i = 1) \leq a_i$. We can therefore identify $F_0 = \mu = \{\mu_S\}_{S \subset N}$ with a vector μ where μ_S is the probability of the signal's profile $\theta = S$ where S is the indicator vector with $S_i = 1$ iff $i \in S$. We get the following corollary from the lemma.

Corollary 1. *There exists an optimal policy (F_0, F_1) in which F_1 sends the signal N with probability 1 and $F_0 = \mu$ maximizes the following linear optimization problem.*

$$\begin{aligned} \max_{\mu} \quad & \sum_{S \subseteq N} \mu_S V(S) \\ s.t. \quad & \mu_S \geq 0 \ \forall S \subset N \\ & \sum_{S \subseteq N} \mu_S = 1 \\ & \mathrm{marg}(\mu)_i = \sum_{\{S: i \in S\}} \mu_S \leq a_i \ \forall i \in N. \end{aligned} \tag{2}$$

Note that the number a_i which we henceforth call *the persuasion level* of agent i has a very simple interpretation when we view the signals of the senders as a recommendations to every agent. The optimal policy always recommends action 1 at state ω_1 to every agent i. In contrast, at state ω_0, in order to keep the policy straightforward he cannot always recommend agent i to play action 1. However, the sender can "lie" with a certain probability and recommend 1 also in state 0. a_i represents the maximal probability that he can lie so that the policy is still straightforward. The higher the persuasion level a_i, the more the sender can bluff agent i.

Without loss of generality we assume throughout that the persuasion level are decreasing with i. That is,

$$a_1 \geq a_2 \geq \ldots \geq a_n.$$

3 Main Results

Using Lemma 1 and Corollary 1 we construct the optimal policy and provide an exact account for the sender's utility for three different important classes of utility functions. We start with a motivating example.

Example 1. To gain more intuition on the nature of our result consider the following setup. There are two agents $N = \{1, 2\}$ with common utility $u : \Omega \times \{0, 1\} \to \mathbb{R}$ such that 0 is a safe action that guarantees 0 at every state. Action 1 is a risky action that yields a payoff of 1 at state ω_1 and -2 at state ω_0. The prior is $\gamma = \frac{1}{2}$. The utility for the sender is 1 if at least one of the agents takes action 1 and $c \geq 1$ if both of them take 1. By Lemma 1 and Corollary 1 the optimization problem of the sender is equivalent to maximize the utility conditional on state 0 subject to the probability that every signal for every player is $\frac{1}{2}$. That is to maximize

$$\mu_{\{1\}}V(\{1\}) + \mu_{\{2\}}V(\{2\}) + \mu_{\{1,2\}}V(\{1, 2\}) = \mu_{\{1\}} + \mu_{\{2\}} + \mu_{\{1,2\}}c,$$

subject to the constraint $\mu_{\{i\}} + \mu_{\{1,2\}} \leq \frac{1}{2}$ for $i = 1, 2$. We claim that if $c \geq 2$ then letting $\mu_{\{1,2\}} = \frac{1}{2} = \frac{1}{2}$ is optimal for the sender. To see this note that such a policy yields a conditional utility of $\frac{c}{2}$. If the sender decreases the $\mu_{\{1,2\}}$ to $\frac{1}{2} - \epsilon$, his maximal revenue is at most $(2\epsilon) + (\frac{1}{2} - \epsilon)c$. Since $c \geq 2$ this is not larger than c. This entails that if $c \geq 2$ sending public signal of 1 with probability $\frac{1}{2}$ in state ω_0 is an optimal policy for the sender.

In contrast, consider the case where $c < 2$. We claim that letting $\mu_{\{i\}} = \frac{1}{2}$ for $i = 1, 2$ is optimal for the sender. This yields a utility of 1. To see this note that letting $\mu_{\{1,2\}} = \epsilon$ yields a utility of at most $(1 - 2\epsilon) + \epsilon c$ which is smaller than 1 when $c < 2$. The policy $\mu_{\{i\}} = \frac{1}{2}$ for $i = 1, 2$ corresponds to sending a signal $(1, 0)$ with probability $\frac{1}{2}$ and $(0, 1)$ with probability $\frac{1}{2}$. This minimizes the correlation between the signals of agents 1 and 2. This reflects the fact that roughly speaking when the game is convex the sender wishes to increase the correlation among the signals of the agents and when the game is concave the sender wishes to decrease the correlation among the signal distribution of the agents.

3.1 Supermodular utility

As common in the literature a utility function $V : 2^N \to \mathbb{R}$ is *supermodular* if for every two subsets $S, T \subseteq N$ it holds that,

$$V(S) + V(T) \leq V(S \cup T) + V(S \cap T).$$

It can be easily shown that a function V is supermodular iff for every player i the function $V(S \cup i)$ is increasing with respect to inclusion as a function of the subset $S \subset N \setminus \{i\}$. The economic interpretation of supermodularity is that the marginal revenue of the sender from persuading agent i is increasing with the size of the other adopters. If for example the sender is a seller who tries to persuade the agents to buy a certain product supermodularity can be explained by decreasing production costs as a function of the group of adopters.

We start by describing the optimal policy $F_0 = \mu$. The policy assigns positive probability only to the sets $\emptyset$ and $\{1, 2, ..., k\}$ for $k \leq n$. The exact probabilities are given by

$$\mu_S = \begin{cases} a_n & \text{if } S = N, \\ a_k - a_{k+1} & \text{if } S = \{1, \dots, k\} \text{ for some } k < n, \\ 1 - a_1 & \text{if } S = \varnothing \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The policy has a very simple characterization. Conditional on state ω_0 the sender first sends a *public* signal

$$\sigma_1 = \begin{cases} 1 & \text{with probability } a_1 \\ 0 & \text{with probability } 1 - a_1 \end{cases}$$

If $\sigma_1 = 0$, then all agents see this and know that the true state is ω_0 and hence all play action 0. If the signal is 1, then the sender sends an additional signal

$$\sigma_2 = \begin{cases} 1 & \text{with probability } \frac{a_2}{a_1} \\ 0 & \text{with probability } 1 - \frac{a_2}{a_1} \end{cases}$$

that can be observed only by agents $\{2, \ldots, n\}$. Again, if $\sigma_2 = 0$ than all agents in $\{2, \ldots, n\}$ know that the state is ω_0 and therefore play action 0. This happens with probability $\mu_{\{1\}} = a_1 \times (1 - \frac{a_2}{a_1}) = a_1 - a_2$. If $\sigma_2 = 1$ then he proceeds and sends the signal

$$\sigma_3 = \begin{cases} 1 & \text{with probability } \frac{a_3}{a_2} \\ 0 & \text{with probability } 1 - \frac{a_3}{a_2} \end{cases}$$

to agents $\{3, \ldots, n\}$ only, and so forth.

This policy can be viewed also as a "greedy" policy that persuades as many agents as possible under the marginal constraints, and then does the same with the remaining agents whose marginal constraints are still positive.

Theorem 1. *The policy μ is optimal for every sender with a supermodular utility V.*

In particular Theorem 1 implies that the sender's optimal revenue in the supermodular case is given by

$$\max_F s(F) = \gamma V(N) + (1-\gamma)[a_n V(N) + \sum_{k=1}^{n-1} (a_k - a_{k+1}) V(\{1, ..., k\})].$$

Proof Outline of Theorem 1. We consider first a linear programming problem which is less constrained than the one introduced in equation (6). The new linear programming problem is identical to (6) but we omit the constraint that

$\sum_{S\subseteq N} \mu_S = 1$. We then show that μ is an optimal policy for the less constrained problem. This in particular shows that μ solves (6).

We show this in three steps. First in Lemma 4 we show, using the supermodularity condition, that there exists an optimal policy ν whose support is a chain, namely, if S, Q are two subsets with $\nu_S, \nu_Q > 0$ then either $Q \subset S$ or $S \subset Q$. Second, we show further that there exists an optimal $\nu = \{\nu_S\}_{S\subseteq N}$ such that the supported chain is $\emptyset \subset \{1\} \subset \{1,2\} \subset ... \subset N$. Then using stochastic dominance considerations we show that μ achieves larger utility than any ν that satisfies the above two conditions. The formal proof is relegated to the appendix.

3.2 Submodular utility

We recall that a utility function $V : 2^N \to \mathbb{R}$ is *submodular* if for every two subsets $S, T \subset N$ it holds that,

$$V(S) + V(T) \leq V(S \cup T) + V(S \cap T).$$

Analogously to the supermodular case, a function V is submodular iff for every player i the function $V(S \cup i)$ is decreasing with respect to inclusion as a function of the subset $S \subset N \setminus \{i\}$. The interpretation of submodularity is that the marginal revenue of the sender from persuading agent i is decreasing with the size of the other adopters. Thus submodularity represents decreasing marginal gains from any individual i.

In this section we consider the optimal revenue for a sender with *anonymous* submodular functions V; i.e., $V(S) = f(|S|)$ for a concave increasing function $f : \mathbb{N} \to \mathbb{R}$.

We first present a natural upper bound for the optimal revenue of the sender and then we show using constructive proof that this value is indeed achievable and therefore it is optimal. We denote by $\overline{f} : \mathbb{R} \to \mathbb{R}$ the linear interpolation of f. That is, if $x \in [0, n]$ satisfy $x = \alpha l + (1-\alpha)(l+1)$ for some $0 \leq l < n$ and some $0 \leq \alpha \leq 1$ then

$$\overline{f}(x) = \alpha f(l) + (1-\alpha) f(l+1).$$

We let $a = \sum_{i=1}^{n} a_i$ be the sum of the persuasion levels of all players. Our second main theorem states

Theorem 2. *Sender's utility in an optimal policy is* $\gamma\overline{f}(n) + (1-\gamma)\overline{f}(a)$.

Moreover, we show (see Lemma 5) that in an optimal policy, conditional on state ω_0 the sender always persuades either $\lfloor a \rfloor$ or $\lceil a \rceil$ agents.

The proof comprises two parts. First we show below why $\gamma\overline{f}(n) + (1-\gamma)\overline{f}(a)$ is an upper bound for the senders utility. This part is straightforward. Second we provide a constructive proof for an optimal policy that achieve this bound. This part is more involved, and is relegated to the appendix.

Lemma 2 (First part of Theorem 2)**.** *For every policy, conditional on state* ω_0 *sender's utility is at most* $\overline{f}(a)$.

Proof. We denote by $\mathcal{S}_k$ the subsets of N of size k. Consider the optimization problem (6) introduced in Corollary 1. Under the corollary the optimization problem translates to maximizing $\sum_{S\subseteq N} \mu_S V(S)$ subject to the constraints $\text{marg}(\mu)_i \leq a_i$ in (6) over all $i \in N$. By summing these constraints, we can deduce that for every feasible $\mu = \{\mu_S\}_{S\subseteq N}$ that satisfies the constraints it holds that

$$\sum_{S\subseteq N} \mu_S \cdot |S| = \sum_{i=1}^{n} \sum_{S:i\in S} \mu_S \leq a.$$

Or equivalently, $\sum_{j=1}^{n} \mu(\mathcal{S}_j) \cdot j \leq a$. Note that $\overline{f}$ is concave, because f is concave. Therefore, by Jensen's inequality we have

$$\sum_{S\subseteq N} \mu_S V(S) = \sum_{j=1}^{n} \mu(\mathcal{S}_j) f(j) = \sum_{j=1}^{n} \mu(\mathcal{S}_j)\overline{f}(j) \leq \overline{f}\left(\sum_{j=1}^{n} \mu(\mathcal{S}_j) \cdot j\right) \leq \overline{f}(a).$$

□

3.2.1 Proof outline of Theorem 2: the constructive part

We next outline the construction of the optimal policy that achieves the above upper bound. Consider the optimization problem (6) of Corollary 1. The following key lemma provides a tight characterization of the maximal "mass" that can be placed on subsets of size k under the constraints $\text{marg}(\mu)_i \leq a_i$. This characterization turns useful in the construction of the optimal policy. For any positive measure μ on $\{0,1\}^n$ we let $|\mu| = \sum_{S \subset N} \mu_S$ be the total mass of μ.

Lemma 3. *Let $1 \geq a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$ be a monotonic sequence. For every $1 \leq k \leq n$ define*

$$\beta_k = \min_{0 \leq m < k} \frac{1}{k-m}(a_{m+1} + \ldots + a_n).$$

There exists a positive measure μ (not necessarily a probability measure) that assigns positive values only to elements of $\mathcal{S}_k$ such that $|\mu| = \beta_k$, and μ satisfies the marginal constraints: $\text{marg}(\mu)_i \leq a_i$ *for every $1 \leq i \leq n$. Moreover, for any positive measure ν over $\mathcal{S}_k$ that satisfies the marginal constraints* $\text{marg}(\nu)_i \leq a_i$ *it holds that $|\nu| \leq \beta_k$.*

The proof of the lemma is constructive and is relegated to the appendix.

We then show the following corollary of Lemma 3.

Corollary 2. *Let $a = \sum_{i=1}^n a_i$. If $0 \leq a_i \leq 1$ and $a = k$ is an integer, then there exists a probability measure (i.e., $|\mu| = 1$) μ over $\mathcal{S}_k$ such that* $\text{marg}_i(\mu) \leq a_i$ *for every i.*

Corollary 2 constitutes a significant step in proving Theorem 2. For instance, in the particular case where a turns out to be an integer, Corollary 2 proves an existence of the required distribution μ. The case where a is not an integer is more involved. We denote $k < a < k+1$ for $k \in \mathbb{N}$. The idea is to construct the desired distribution μ as a convex combination of two probability distributions $\mu = (k+1-a)\kappa + (a-k)\nu$, where $\kappa(\mathcal{S}_k) = 1$ and $\nu(\mathcal{S}_{k+1}) = 1$. We do it by splitting the persuasion levels a_i into two numbers $a_i = (k+1-a)b_i + (a-k)c_i$ such that $\sum_i b_i = k$, $\sum_i c_i = k+1$, $0 \leq b_i \leq 1$, and $0 \leq c_i \leq 1$. This allows us to use Corollary 2 to construct κ and ν.

3.3 Supermajority utility

A leading application of Bayesian persuasion problems is for the case of voting. In this case the sender is a politician that wishes to persuade voters to vote for him or pass a certain law. The recent paper by Alonso and Camara [1] considers a setting of persuading voters via public signal. In their setting, unlike here, the state space is any finite set. Their first main result characterizes the optimal signalling policy as a function of the *approval set* which comprises all posterior beliefs for which the law passes.

In this case we assume that the sender has a supermajority utility $V_k : 2^N \to \{0,1\}$ that yields a utility of 1 if at least k voters take action 1 and utility 0 otherwise. It is important to note that V_k is neither supermodular nor submodular and hence this case doesn't follow directly from our previous two main theorems. Moreover in our setup there are no payoff externalities among the agents, and the the utility of every agent depends only on his action and the realised state of the world.

The following theorem characterizes the optimal utility of the politician as a function of k using Lemma 3.

Theorem 3. *For every $1 \leq k \leq n$ the optimal utility v_k of the politician with a utility function V_k is given by the following expression.*

$$v_k = \gamma + (1-\gamma)\max\{1, \beta_k\}.$$

It is worth noting that when the number of agents increases one can construct an example wherein the ratio between the optimal utility of a politician that is restricted to public signals versus his optimal utility with private signals, approaches 0. This demonstrates, as already noted by Chan, Li, and Wang [3], that in some cases allowing private signals may improve the welfare of the sender significantly.

Proof of Theorem 3. Note that every distribution μ over $\{0,1\}^N$ can be converted to a distribution ν over $\mathcal{S}_k \cup \{(0,...,0)\}$ with equal utility for the sender, and weakly smaller marginals in all coordinates (i.e., $\text{marg}(\mu)_i \geq \text{marg}(\nu)_i$). This can be done simply by removing all mass from $\mathcal{S}_m$ for $m < k$ to $(0,...,0)$, and removing all mass from $\mathcal{S}_m$ for $m > k$ to $\mathcal{S}_k$ by letting some signals be 0 (in an arbitrary manner) in some of the coordinates in the support of $\mathcal{S}_m$ such that the number of 1s becomes exactly k. Therefore, we can assume that the support of the optimal policy lies over $\mathcal{S}_k \cup \{(0, ..., 0)\}$. Now the theorem follows immediately from Lemma 3. □

3.4 Public vs. private Signals

This work integrates private signals to the multi-agent Bayesian persuasion problem. A natural question to ask is under which conditions does there exist a public optimal signalling policy where agents share the same signal. Indeed in some instances the only channel to deliver information is public. In this section we try to characterize the cases where communicating publicly with the agents is indeed optimal. Surprisingly this question has a strong connection to the theory of cooperative games.

Note that the utility of the sender V defines a *cooperative game* (N, V). We recall that the *core* of the game is the subset $C(N, V) \subset \mathbb{R}^n$ of all efficient payoff vectors where no coalition can profit by deviating from the grand coalition. Formally,

$$C(N, V) = \{x \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = V(N) \text{ and } \sum_{i \in S} x_i \geq V(S) \text{ for every } S \subseteq N\}.$$

Our characterization relates to the non-emptiness of the core.

Theorem 4. *Consider a Bayesian persuasion problem for which* $a_n < 1$*. Under the following conditions an optimal public signalling policy exists.*

1. *All agents have the same persuasion level. I.e.,* $a_1 = a_2 = \ldots = a_n$*.*

2. *The core of the game* (N, V) *is nonempty.*

Moreover, if V *is strictly increasing with respect to inclusion, then the above conditions are also necessary.*

Proof. We first show that if the above two conditions hold, then an optimal public signalling policy exists. We let $\alpha = a_i$ and define a policy using public signal by sending the signal 1 in state ω_0 with probability α and 0 with probability $1 - \alpha$. This public signal guarantees a utility of $\gamma V(N) + (1 - \gamma)\alpha V(N)$ to the sender.

We show first that this policy is optimal. To see this let $\mu = \{\mu_S\}_{S \subset N}$ be any positive measure such that $\text{marg}(\mu)_i \leq \alpha$ for every agent i. We contend that,

$$\sum_{S \subset N} \mu_S V(S) \leq \alpha V(N).$$

To see this we note that we can assume that $\text{marg}(\mu)_i = \alpha$ for every player i. If $\text{marg}(\mu)_i < \alpha$ we can redefine μ by increasing the weight of the singleton set $S = \{i\}$ and make $\text{marg}(\mu)_i = \alpha$. By monotonicity of V this can only increase the expectation $\sum_{S \subseteq N} \mu_S V(S)$. Since the core of the game (N, V) is nonempty it follows from the Bondareva-Shapley theorem (see Maschler et al. [8]) that

$$\sum_{S \subset N} \mu_S V(S) \leq \alpha V(N).$$

For the converse consider the case where V is strictly increasing. We show that if one of the conditions (1) or (2) is not satisfied, then public signal cannot constitute an optimal policy.

Assume that (1) is not satisfied. In that case $a_1 > a_n$. We let F be an optimal policy for the sender. If F constitutes a public signal then we can assume, by Lemma 1 and Corollary 1, that F_1 assigns probability 1 to $(1, \ldots, 1)$ and $F_0 = \mu = \{\mu_S\}_{S \subseteq N}$ has the property that $\mu_N = b$, $\mu_\varnothing = 1 - b$, and $\mu_S = 0$ for every nonempty $S \subsetneq N$. Again by Lemma 1 we must have that $\text{marg}(\mu)_i \leq a_i$ for every agent i; hence $b < a_1$. Define a probability measure ν as follows: $\nu_N = b$, $\nu_{\{1\}} = a_1 - b$ and $\nu_{\{\varnothing\}} = 1 - a_1$. Note that $\text{marg}(\nu)_i \leq a_i$ for every agent i and since $V(\{1\}) > 0$ by strict monotonicity we also have that ν achieves a strictly higher expected utility than μ. This stands in contradiction to the optimality of μ.

Assume now that condition (1) is satisfied but not condition (2). Assume by way of contradiction that $\mu = \{\mu_N, \mu_\varnothing\}$ is an optimal policy. We clearly have that $\mu_N = \alpha$ and ,$\mu_\varnothing = 1 - \alpha$. Again by the Bondareva-Shapley theorem there exists a vector $\nu = \{\nu_S\}_{\varnothing \neq S \subset N}$ for which $\text{marg}(\nu)_i = 1$ for every agent i such that,

$$\sum_{S \subset N} \nu_S V(S) > V(N).$$

Note that since $\text{marg}(\nu)_i = 1$ for every agent i we must have that $\sum_{S \subseteq N} \nu_S = K \geq 1$. Let $\epsilon = \min\{\frac{1-\alpha}{K}, \alpha\}$. Note that $\epsilon > 0$ since $0 < \alpha < 1$. Define a probability measure $\lambda = \{\lambda_S\}_{S \subseteq N}$ as follows:

$$\lambda_S = \begin{cases} \epsilon \nu_S & \text{if } S \notin \{\varnothing, N\}, \\ \alpha - \epsilon + \epsilon \nu_N & \text{if } S = N, \\ 1 - \alpha - (K-1)\epsilon & \text{if } S = \varnothing. \end{cases} \tag{4}$$

We claim first that λ is a probability measure. To see this note first that by the definition of ϵ it holds that $\lambda_S \geq 0$ for every $S \subseteq N$. Moreover

$$\sum_{S \subseteq N} \lambda_S = \alpha - \epsilon + (1 - \alpha - K\epsilon + \epsilon) + \epsilon \sum_{S \subset N} \nu_S = \alpha - \epsilon + (1 - \alpha - K\epsilon + \epsilon) + \epsilon + K\epsilon = 1.$$

Note further that for every agent i it holds that $\text{marg}(\lambda)_i = \alpha - \epsilon + \epsilon \text{marg}(\nu)_i = \alpha$. Finally,

$$\sum_{S \subseteq N} \lambda_S V(S) = (\alpha - \epsilon) V(N) + \epsilon \sum_{S \subset N} \nu_S V(S) > \alpha V(N).$$

Hence λ achieves a strictly higher utility than μ. This concludes the proof of Theorem 4. □

4 Extensions

In this section we explain how our results can be easily generalized to the case where the utility of the sender is state dependent and the case where the agents do not have identical interest conditional on the realised state. Consider first a general weakly increasing utility function $V : \Omega \times 2^N \to \mathbb{R}$ for the sender. Lemma 1 and Corollary 1 remain valid also for this case. Hence the utility for the sender from any policy F that satisfies the condition of Lemma 1 is:

$$s(F) = \gamma V(\omega_1, N) + (1-\gamma)\mathbb{E}_{\theta \sim F_0} V(\omega_0, \theta).$$

Therefore in order to generalize Theorem 2 for example one should only require that $V(\omega_0, \cdot) : 2^N \to \mathbb{R}$ is supermodular. This is also valid for the rest of the results.

We now turn to the case where agents do not have identical interest conditional on ω. In this case we can divide the group of agents into four disjoint subsets N_1, N_2, N_3, N_4. The subset N_1 (N_2) comprises those agents who prefer action 1 (0) at state ω_1 and action 0 (1) at state ω_0. The subset N_3 (N_4) comprises of those agents who prefer action 0 (1) at both states. Note that an agent $i \in N_3 \cup N_4$ is not persuadable, and therefore we may ignore these agents in the analysis.

Let

$$
\begin{aligned}
a_i &= \min(\frac{\gamma}{1-\gamma}\frac{u_i(\omega_1,1)-u_i(\omega_0,1)}{u_i(\omega_0,0)-u_i(\omega_0,1)},1) \text{ for every } i \in N_1,\\
b_j &= \min(\frac{1-\gamma}{\gamma}\frac{u_j(\omega_0,1)-u_j(\omega_1,1)}{u_j(\omega_1,0)-u_j(\omega_1,1)},1) \text{ for every } j \in N_2.
\end{aligned}
$$

By similar considerations as in Lemma 1 we can restrict attention to a policy with binary signal distribution $F : \Omega \to \Delta(\{0,1\}^n)$ such that satisfy the following conditions:

1. $F_1(\theta_i = 1) = 1$ for every $i \in N_1$,
2. $F_0(\theta_i = 1) \leq a_i$ for every $i \in N_1$,
3. $F_0(\theta_j = 1) = 1$ for every $j \in N_2$,
4. $F_1(\theta_j = 1) \leq b_j$ for every $j \in N_2$.

Under these conditions the policy is straightforward.

The optimization problem of the sender is therefore naturally divided into two distinct problems for each of the two states as follows. The optimization problem associated with state ω_0:

$$
\begin{aligned}
\max_{\mu} \quad & \sum_{S \subset N_1} \mu_S V(S \cup N_2)\\
\text{s.t.} \quad & \mu_S \geq 0 \ \forall S \subset N\\
& \sum_{S \subseteq N_1} \mu_S = 1\\
& \text{marg}(\mu)_i \leq a_i \ \forall i \in N.
\end{aligned}
\tag{5}
$$

The optimization problem associated with state ω_1:

$$\begin{aligned} \max_{\nu} \quad & \sum_{S \subseteq N_2} \mu_S V(S \cup N_1) \\ \text{s.t.} \quad & \mu_S \geq 0 \ \forall S \subset N \\ & \sum_{S \subseteq N_1} \mu_S = 1 \\ & \operatorname{marg}(\mu)_i \leq a_i \ \forall i \in N. \end{aligned} \tag{6}$$

These problems are solved separately for each of the two states. Note that if $V : 2^{N_1 \cup N_2} \to \mathbb{R}$ is supermodular/submodular/supermajority function then both functions $V_1 : 2^{N_1} \to \mathbb{R}$ and $V_2 : 2^{N_2} \to \mathbb{R}$ that are defined by $V_1(S) = V(S \cup N_2)$ and $V_2(S) = V(S \cup N_1)$ are supermodular/submodular/supermajority functions. Therefore we can get a similar characterization of optimal policies in all three cases: supermodular, symmetric submodular, and supermajority.

5 Conclusion

In this work we study a multi-agent Bayesian persuasion problem where a sender that is interested in persuading a designated group of agents to adopt a certain product is allowed to send private signals. The sender extracts a utility from any subgroup of adopters. We study the optimal policy and the revenue to the sender, as a function of his utility, for three different types of utilities: supermodular, submodular, and supermajority. Moreover, we provide a necessary and sufficient condition under which the optimal policy is achievable via public signals.

There is one natural question that arises from our results. Throughout we restrict attention to the binary state space. This facilitates our analysis and allows us to translate the problem to a linear programming problem. We believe that our analysis together with [1] can serve as a starting point for studying a private Bayesain persuasion problem in a general state space. This question is beyond the scope of this paper and we leave it open for future work.

A Appendix

A.1 Proof of Lemma 1

Proof of Lemma 1. Note first that, given a signalling policy $F = (F_0, F_1, (\Theta_i)_{i \in N})$, agent i plays action 1 for the signal θ_i iff his posterior belief $p(\theta_i)$ lies above $q_i = \frac{u_i(\omega_0,0)-u_i(\omega_0,1)}{u_i(\omega_1,1)-u_i(\omega_0,1)+u_i(\omega_0,0)-u_i(\omega_1,0)}$. Hence a binary signalling policy F is straightforward ($g_i(\theta_i) = \theta_i$) iff $\mathbf{P}(\theta_i = 1) \geq q_i$ and $p(\theta_i = 0) \geq q_i$

We claim first that for every policy G there exists a straightforward policy K defined over $\{0,1\}^n$ that yields the same utility for the sender as G. That is K satisfies $\mathbf{P}_K(\omega_1|\theta_i = 1) \geq q_i$, and $\mathbf{P}_K(\omega_1|\theta_i = 0) < q_i$. This follows directly from Proposition 1 in [6]. Instead of showing agent i his realised signal θ_i one can show him only $g_i(\theta_i)$. This yields a binary straightforward signalling policy K that yields the same utility as G.

Let K be a a straightforward signalling policy with binary signals. Given such K we shall show that there exists another straightforward policy which is better for the sender, for which:

$$H_1(\theta_i = 1) = 1.$$

To see this note first that for every agent i,

$$\mathbf{P}_H(\omega_1|\theta_i = 1) = \frac{\gamma H_1(\theta_i = 1)}{\gamma H_1(\theta_i = 1) + (1-\gamma) H_0(\theta_i = 1)}.$$

Therefore $\mathbf{P}_H(\omega_1|\theta_i = 1)$ is increasing in $H_1(\theta_i = 1)$. Hence we can increase $H_1(\theta_i = 1)$ to 1 without changing the fact that $\mathbf{P}_H(\omega_1|\theta_i = 1) \geq q_i$. Formally, we define $M : \Omega \to \{0,1\}^n$ as follows. $M_0 = H_0$ and let M_1 assign probability one to the vector $(1, \ldots, 1)$. We then have that $E_M(V|\omega_0) = E_H(V|\omega_0)$ and $E_M(V|\omega_1) \geq E_H(V|\omega_1)$.

The last thing that should be noted is that a binary signalling policy H for which $H_1(\theta_i = 1) = 1$ is straightforward iff $p(\theta_i = 1) \geq q_i$ for every agent i. Since the prior probability of state ω_1 is γ it follows from Bayes rule that this holds iff

$$H_0(\theta_i = 1) \leq \frac{\gamma(1-q_i)}{q_i(1-\gamma)} = \frac{\gamma}{1-\gamma} \frac{u_i(\omega_1,1) - u_i(\omega_0,1)}{u_i(\omega_0,0) - u_i(\omega_0,1)}.$$

Hence when $a_i = \min(\frac{\gamma}{1-\gamma}\frac{u_i(\omega_1,1)-u_i(\omega_0,1)}{u_i(\omega_0,0)-u_i(\omega_0,1)}, 1)$ such H is straightforward iff $H_0(\theta_i = 1) \leq a_i$. This concludes the proof of Lemma 1.

□

A.2 Proof of Theorem 1

It is straightforward to see that μ satisfies the constraints of Equation (6). We consider, instead, the following less constraint optimization problem.

$$\begin{aligned} \max_{\kappa} \quad & \sum_{S \subseteq N} \kappa_S V(S) \\ \text{s.t.} \quad & \kappa_S \geq 0 \ \forall S \subset N \\ & \text{marg}(\kappa)_i \leq a_i \ \forall i \in N. \end{aligned} \tag{7}$$

We will show that μ solves the optimization problem (7), and therefore, obviously μ solves the (more restrictive) optimization problem (6).

The distinction between the two problems is that in problem (7) the variables $\{\mu_S\}_{S \subseteq N}$ are not required to sum up to one and hence clearly the optimal value in problem (7) is at least as large as in problem (6).

We first show that an optimal solution to the maximization problem (7) assigns positive probability only to sets such that one contains the other.

Lemma 4. *Let $\kappa = \{\kappa_S\}_{S \subseteq N}$, be any feasible vector of the optimization problem (7). Assume $Q, T \subseteq N$ are two sets such that $Q \setminus T \neq \varnothing$, $T \setminus Q \neq \varnothing$, and $\kappa_Q \geq \kappa_T > 0$. There exists a feasible vector $\nu = \{\nu_S\}_{S \subseteq N}$ such that $\nu_T = 0$, and*

$$\sum_S \kappa_S V(S) \leq \sum_S \nu_S V(S). \tag{8}$$

The idea is that, given κ, we can "move mass" from the pair of sets Q and T to the pair of sets $Q \cup T$ and $Q \cap T$, in a way that preserves the marginal constraints and increases the sender's utility (because of the supermodularity). Formally, the proof is as follows.

Proof of Lemma 4. We define ν as follows,

$$\nu_S = \begin{cases} 0 & \text{if } S = T \\ \kappa_Q - \mu_T & \text{if } S = Q \\ \kappa_{Q\cap T} + \mu_T & \text{if } S = Q \cap T \\ \kappa_{Q\cup T} + \mu_T & \text{if } S = Q \cup T \\ \kappa_S & \text{otherwise.} \end{cases} \tag{9}$$

We will show first that ν is feasible, namely, it satisfies the constraints of Equation (7). We clearly have that $\nu_S \geq 0$ for every $S \subseteq N$. We have left to show that for every player i, $\text{marg}(\nu)_i \leq a_i$. This is straightforward for players $i \notin T \cup Q$. Let $i \in T \cup Q$ and let

$$D_i = \{S : i \in S,\ S \neq Q, T, Q \cap T, Q \cup T\}.$$

We can write,

$$\begin{aligned} \sum_{S\subseteq N: i\in S} \nu_s &= \sum_{S\in D_i} \kappa_S + \nu_Q + \nu_T + \nu_{Q\cap T} + \nu_{Q\cup T} \\ &= \sum_{S\in D_i} \kappa_S + (\kappa_Q - \kappa_T) + (\kappa_{Q\cap T} + \kappa_T) + (\kappa_{Q\cup T} + \kappa_T) \\ &= \sum_{S\in D_i} \kappa_S + \kappa_Q + \kappa_T + \kappa_{Q\cap T} + \kappa_{Q\cup T} = \text{marg}(\kappa)_i \leq a_i. \end{aligned}$$

The last inequality follows since κ is feasible.

To see equation (8) we need to show that

$$\sum_{S\in D_i} \kappa_S V(S) + \kappa_Q V(Q) + \kappa_T V(T) + \kappa_{Q\cap T} V(Q\cap T) + \kappa_{Q\cup T} V(Q\cup T)$$

is smaller than

$$\sum_{S\in D_i} \kappa_S V(S) + (\kappa_Q - \kappa_T) V(Q) + (\kappa_{Q\cap T} + \kappa_T) V(Q\cap T) + (\kappa_{Q\cup T} + \kappa_T) V(Q\cup T)$$

or equivalently,

$$\begin{aligned} &\kappa_Q V(Q) + \kappa_T V(T) + \kappa_{Q\cap T} V(Q\cap T) + \kappa_{Q\cup T} V(Q\cup T) \\ &\leq (\kappa_Q - \kappa_T) V(Q) + (\kappa_{Q\cap T} + \kappa_T) V(Q\cap T) + (\kappa_{Q\cup T} + \kappa_T) V(Q\cup T). \end{aligned} \tag{10}$$

Rearranging (10) yields,

$$V(Q) + V(T) \leq V(Q \cap T) + V(Q \cup T).$$

This clearly holds by supermodularity of V. □

Corollary 3. *The optimal value of problem* (7) *is obtained by a chain vector* $\nu = \{\nu_S\}_{S \subseteq N}$ *that satisfies* $\nu_Q > 0$ *and* $\nu_T > 0$ *implies* $Q \subset T$ *or* $T \subset Q$.

The Corollary follows from repeated application of Lemma 4. We can therefore conclude that the optimal solution of problem (7) is obtained by a vector $\nu = \{\nu_{S_j}\}_{j=1,\ldots,k}$ such that $\nu_{S_j} > 0$ for every j and $S_l \subsetneq S_j$ iff $l < j$. The proof of Theorem 1 follows after noting that among policies that have the above chain structure in their support, the suggested one μ has a maximal utility (roughly speaking, because it uses the marginal constraints up to the maximal point).

Proof of Theorem 1. Let $\nu = \{\nu_{S_j}\}_{j=1,\ldots,k}$ for which the above inclusion property is satisfied and that $\text{marg}(\nu)_i \leq a_i$ for every agent i. We first show that we can assume, with no loss of generality, that if $S_j \neq \varnothing$ then $S_j = \{1, \ldots, l_j\}$ for some $l_j \geq 1$. That is, S_j comprises all agents i such that $1 \leq i \leq l_j$. To see this let S_j be a subset that doesn't have the above form. We let l_j be the maximal agent in S_j. We claim that we can replace any such subset with the set $\{1, \ldots, l_j\}$ without violating the marginal constraint. To see this let $\tilde{\nu} = \{\nu_{\tilde{S}_j}\}_{j=1,\ldots,k}$ be the new vector that is obtained after the replacement. Consider agent n first. It clearly holds for every j that $n \in S_j$ iff $n \in \tilde{S}_j$; hence $\text{marg}(\nu)_n = \text{marg}(\tilde{\nu})_n \leq a_n$. We proceed by induction. Let i be any agent and assume that $\text{marg}(\tilde{\nu})_l \leq a_l$ for every $l > i$. Let m be the maximal index such that $i \in S_m$ and let l_m be the maximal agent in S_m. Then we must have that $i \leq l_m$ by definition. By the inclusion property we have for every j that if $i \in S_j$ then $l_m \in S_j$. Hence in particular,

$$\text{marg}(\tilde{\nu})_i \leq \text{marg}(\tilde{\nu})_{l_m} \leq a_{l_m} \leq a_i.$$

The last inequality follows since $i \leq l_m$, and the second to last follows by the induction hypothesis. Clearly since the utility of the sender is monotonically increasing, the vector $\tilde{\nu}$ achieves (weakly) higher utility than ν. We claim that for every $m \leq n$ it holds that,

$$\sum_{S:|S|\geq m} \nu_S \leq a_m.$$

To see this let j_0 be the minimal index for which $|S_j| \geq m$. Since $|S_{j_0}| \geq m$ it must contain an agent $l \geq m$. By the inclusion property $l \in S_j$ for every $j \geq j_0$. Hence,

$$\sum_{S:|S|\geq m} \nu_S = \sum_{j=j_0}^{n} \nu_{S_j} \leq \text{marg}(\mu)_l \leq a_m.$$

The last equality follows since $l \geq m$. We note that $\sum_{S:|S|\geq m} \mu_S = \text{marg}(\mu)_m = a_m$. Let $\overline{\nu} : \{1, \dots, n\} \to \mathbb{R}$ be the vector that is defined by $\overline{\nu}(m) = \nu_{S_j}$ if there exists S_j such that $|S_j| = m$ and $\overline{\nu}(m) = 0$ otherwise. Define $\overline{\mu}$ similarly with respect to μ. By the calculation above $\overline{\mu}$ first order stochastic dominates $\overline{\nu}$. We let $f(m) = V(\{1, \dots, m\})$. Note that f is increasing with m. We can therefore write

$$\sum_{S\subseteq N} \nu_S V(S) = \sum_{m=1}^{n} \overline{\nu}(m) f(m),$$

and

$$\sum_{S\subseteq N} \mu_S V(S) = \sum_{m=1}^{n} \overline{\mu}(m) f(m).$$

By the stochastic dominance we must therefore have

$$\sum_{S\subseteq N} \mu_S V(S) \geq \sum_{S\subseteq N} \nu_S V(S).$$

This conclude the proof of Theorem 1. □

A.3 Construction of the optimal policy

Recall that $\mathcal{S}_k$ denotes the subsets of size k. The second part of Theorem 2 can be stated as follows.

Lemma 5 (Second part of Theorem 2)**.** *There exists a policy F for which the sender's utility is $\gamma\overline{f}(n) + (1-\gamma)\overline{f}(a)$. Moreover, the optimal policy $F_0 = \mu$ satisfies $\mu(\mathcal{S}_k) = 0$ for $k \notin \{\lfloor a \rfloor, \lceil a \rceil\}$.*

In this subsection we provide a constructive proof for this lemma.

We start with the proof of Lemma 3.

Lemma 3. Let $a_1 \geq a_2 \geq \ldots \geq a_n \geq 0$ be a monotonic sequence. For every $1 \leq k \leq n$ define

$$\beta_k = \min_{0 \leq m < k} \frac{1}{k-m}(a_{m+1} + \ldots + a_n).$$

There exists a positive measure μ (not necessarily a probability measure) over $\mathcal{S}_k$ such that $|\mu| = \beta_k$, and μ satisfies the marginal constraints: $\text{marg}(\mu)_i \leq a_i$ for every $1 \leq i \leq n$. Moreover, for any positive measure ν over $\mathcal{S}_k$ which satisfies the marginal constraints $\text{marg}(\nu)_i \leq a_i$ holds $|\nu| \leq \beta_k$.

Proof of Lemma 3. We start with proving that β_k is an upper bound on the mass. For every $0 \leq m < k$, every set of size k contains at least $(k-m)$ elements from $\{m+1, ..., n\}$. Therefore, for every measure ν over $\mathcal{S}_k$ holds $\sum_{i=m+1}^{n} \text{marg}(\nu)_i \geq (k-m)|\nu|$, because every unit of mass appears in $\sum_{i=m+1}^{n} \text{marg}(\nu)_i$ at least $k-m$ times. If ν satisfies the marginal constraints then it follows that $|\nu| \leq \frac{1}{k-m}(a_{m+1} + ... + a_n)$. Since the inequality holds for every m, it also holds for the minimal m, i.e., $|\nu| \leq \beta_k$.

For existence of such a measure, it is sufficient to construct a measure μ and an index $0 \leq m < k$ such that $|\mu| = \frac{1}{k-m}(a_{m+1} + \ldots + a_n)$. We consider the following recursive process for producing the measure μ.

We set the initial marginal constraints vector $(a_1^0, ..., a_n^0) = (a_1, ..., a_n)$ to be the original constraints.

During the process we preserve the monotonicity of the marginal constraints vector and therefore we can denote the marginal constraints vector at time $t-1$ by

$$(a_1^{t-1}, ..., a_n^{t-1}) = (b_1, ..., b_j, \underbrace{c, c, ..., c}_{l-j \text{ times}}, b_{l+1}, ..., b_n)$$

where $b_j > c > b_{l+1}$ and $j < k \leq l$. Note that if $a_k^{t-1} = a_{k+1}^{t-1} = ... = a_n^{t-1}$ then $l = n$ and for simplicity of notation we assume $b_{n+1} = 0$. Note that if $a_1^{t-1} = a_2^{t-1} = ... = a_k^{t-1}$ then $j = 0$, and for simplicity of notation we assume $b_0 = n$ is a large constant.

At step t, the idea is to distribute mass equally over the subsets S of size k that satisfy $[j] \subseteq S \subseteq [l]$ (we have $\binom{l-j}{k-j}$ such sets). If we do so, after we have distributed x units of mass the remaining marginal constraints vector will be

$$b(x) = (b_1 - x, ..., b_j - x, c - \frac{k-j}{l-j}x, ..., c - \frac{k-j}{l-j}x, b_{l+1}, ..., b_n) \tag{11}$$

because every element $i = j+1, j+2, ..., l$ appears in exactly $\frac{k-j}{l-j}$ fraction of the above subsets. Step t terminates at the moment when one of the following two happens:

(1) The jth coordinate becomes equal to the $(j+1)$th coordinate.

(2) The lth coordinate becomes equal to the $(l+1)$th coordinate.

We denote by α the amount of mass that has been distributed during step[1] t. We denote by

$$\mu_t(S) = \begin{cases} \frac{\alpha}{\binom{k-m}{l-m}} & \text{if } |S| = k, \text{ and } \{1, ..., m\} \subset S \subset \{1, ..., l\}, \\ 0 & \text{otherwise,} \end{cases}$$

the measure of the distributed mass at step t. We denote by $(a_1^t, ..., a_n^t) = b(\alpha)$ the marginal constraints vector after step t, where $b(\cdot)$ is defined in equation (11).

If (2) happens before (1) and $l = n$, then it must be the case that $(a_1^t, ..., a_n^t) = (a_1^t, ..., a_m^t, 0, ..., 0)$. In such a case we terminate the process, and we denote by $T = t$ the number of steps in the process. In any other case we proceed to step $t+1$.

Finally, we define our measure to be $\mu = \sum_{t=1}^{T} \mu_t$. Note that at the moment of termination

$$(a_1^T, ..., a_n^T) = (a_1^T, ..., a_m^T, 0, ..., 0) = (a_1 - |\mu|, ..., a_m - |\mu|, 0, ..., 0)$$

for some index $0 \leq m < k$, because during the entire process the marginals a_i^t for $i \leq m$ are reduced at the same rate as the amount of the distributed mass.

[1] Simple calculations show that $\alpha = \min(\frac{l-j}{k-j}(b_j - c), \frac{l-j}{k-j}(c - b_{l+1}))$, but this exact expression will not be needed for the proof.

Moreover, $|\mu| = \frac{1}{k-m}(a_{m+1} + ... + a_n)$ because during the process $(a^t_{m+1} + ... + a^t_n)$ is decreasing $k - m$ times faster than the amount of the distributed mass (since every subset S of size k that satisfies $\{1, ..., m\} \subset S$ has exactly $k-m$ elements in $\{m + 1, ..., n\}$). Therefore we found a measure μ and an index m as needed. $\square$

As claimed, the following is a corollary of Lemma 3.

Corollary 2. If $0 \leq a_i \leq 1$ and $\sum_{i=1}^{n} a_i = k$ is an integer, then there exists a probability measure (i.e., $|\mu| = 1$) μ over $\mathcal{S}_k$ such that $\text{marg}_i(\mu) \leq a_i$ for every i.

Proof of Corollary 2. To see this we first claim that for every $m < k$

$$\frac{1}{k-m}(a_{m+1} + \ldots + a_n) \geq 1$$

or equivalently that,

$$a_{m+1} + \ldots + a_n \geq k - m. \tag{12}$$

Adding $a_1 + \cdots + a_m$ to the two sides of equation (12) we have to show that,

$$a \geq k - m + a_1 + \cdots + a_m.$$

Since $a = k$ we have left to show that,

$$m \geq a_1 + \cdots + a_m,$$

which follows directly as $a_i \leq 1$ for every $1 \leq i \leq m$. Clearly, by letting $m = 0$ we can see that $\beta_k = 1$.

Hence by Lemma 3 there exists *a probability measure* $\mu \in \Delta(\{0, 1\}^n)$ such that $\mu_k = 1$ and $\text{marg}(\mu)_i \leq a_i$ for every $1 \leq i \leq n$. $\square$

Lemma 6. *Let* $(a_1, ..., a_n) \in [0, 1]^n$ *be such that* $a_i \geq a_{i+1}$ *and let* $a = \sum_i a_i$ *and* $k < a < k + 1$. *There exist two vectors* $(b_1, \ldots, b_n), (c_1, ..., c_n) \in [0, 1]^n$ *with the following properties:*

1. $a_i = (k + 1 - a)b_i + (a - k)c_i$.

2. $b = \sum_i b_i = k$.

3. $c = \sum_i c_i = k + 1$.

Proof. We claim first that we can restrict attention to finding an appropriate vector $(b_1, \dots, b_n)$ only. From property (1) the value of $c_i = \frac{a_i-(k+1-a)b_i}{a-k}$ is uniquely determined by b_i. Property (3) follows from property (1) and property (2) as follows:

$$a = (k+1-a)b + (a-k)c = (k+1-a)k + (a-k)c \Rightarrow c = k+1.$$

The only thing that should be noted is that we require $(c_1, ..., c_n) \in [0,1]^n$. This requirement translates to the following two inequalities on b_i

$$a_i - (a-k) \le b_i(k+1-a) \le a_i.$$

Therefore to prove Lemma 6 we should prove the existence of a vector $(b_1, \dots, b_n)$ that satisfies $b = k$ under the following constraints

$$0 \le b_i \le 1 \text{ and } a_i - (a-k) \le b_i(k+1-a) \le a_i. \tag{13}$$

Note that the constraints in (13) are linear. Hence it is sufficient to prove the existence of two vectors $(b'_1, \dots, b'_n)$ and $(b''_1, \dots, b''_n)$ that satisfy the constraints of (13) such that $\sum_i b'_i \le k$ and $\sum_i b''_i \ge k$. Given such two vectors we can choose $(b_1, ..., b_n)$ as an appropriate convex combination of $(b'_1, \dots, b'_n)$ and $(b''_1, \dots, b''_n)$ that satisfies $b = k$.

We set $b'_i = \max(0, \frac{a_i-(a-k)}{k+1-a})$. Note that b'_i satisfies the constraints in (13) because

$$a_i - (a-k) \le k+1-a \Rightarrow b'_i \le 1 \text{ and}$$
$$a_i - (a-k) \le a_i \Rightarrow b'_i(k+1-a) \le a_i.$$

We have

$$\sum_i b'_i = \sum_{i:a_i>a-k} \frac{a_i-(a-k)}{k+1-a} = \sum_{i=1}^{m} \frac{a_i-(a-k)}{k+1-a}$$

where m is the maximal index for which $a_i > a-k$. We argue that it must be the case that $m \le k$, because otherwise (if $m \ge k+1$) we have

$$a = \sum_i a_i \ge (k+1)(a-k) \Rightarrow k^2 - k \ge ak \Rightarrow k-1 \ge a,$$

which is a contradiction. Therefore the inequality $\sum_i b'_i < k$ follows trivially from the fact that $b'_i \leq 1$ (because we sum up at most k elements).

We set $b''_i = \min(1, \frac{a_i}{k+1-a})$. It is easy to check that (b''_i) satisfies the constraints (13). Note that

$$\sum_i b''_i = m + \sum_{i=m+1}^{n} \frac{a_i}{k+1-a}$$

where m is the minimal index such that $a_i < k+1-a$. We argue that it must be the case that $m \geq k$ because otherwise (if $m \leq k-1$) we have

$$a = \sum_i a_i \leq (k-1)(k+1-a) \Rightarrow ak \leq k^2 - 1 \Rightarrow a \leq k - \frac{1}{k},$$

which is a contradiction. Therefore the inequality $\sum_i b''_i \geq k$ follows immediately. □

Now the proof of Lemma 5 follows directly from Corollary 2 and Lemma 6.

Proof of Proposition 5. Let (b_i) and (c_i) be the vectors from Lemma 6. Let κ be a distribution over $\mathcal{S}_k$ with marginals $\text{marg}(\kappa)_i \leq b_i$. Let ν be a distribution over $\mathcal{S}_{k+1}$ with marginals $\text{marg}(\nu)_i \leq c_i$. Such κ and ν exist by Corollary 2 and properties (2),(3) in Lemma 6. We define $\mu = (k+1-a)\kappa + (a-k)\nu$. By property (1) in Lemma 6, μ satisfies the marginal constraints $\text{marg}(\mu)_i \leq a_i$. Note that every signal in κ achieves utility of $f(k)$ to the sender, while every signal in ν achieves utility of $f(k+1)$. Therefore the expected utility of the sender is given by $(k+1-a)f(k) + (a-k)f(k+1)$ which (by the definition of interpolation) is equal to $\overline{f}(a)$. □

References

[1] R. Alonso, O. Camara, "Persuading voters," mimeo, 2015.

[2] R.J. Aumann, M. Maschler, "Repeated Games with Incomplete Information," *MIT Press,* Cambridge, 1995.

[3] J. Chan F. Li and Y. Wang "Discriminatory Persuasion: How to Convince a Group," mimeo, 2015.

[4] V.P. Crawford, J.Sobel, "Strategic Information Transmission," *Econometrica,* 50(6): 1431–51, 1982.

[5] P. Domingos, M. Richardson, "Mining the Network Value of Customers," *Seventh International Conference on Knowledge Discovery and Data Mining,* 2001.

[6] E. Kamenica, M. Gentzkow, "Bayesian persuasion" *American Economic Review,* 101, 2590–2615, 2011.

[7] D. Kempe, J. Kleinberg, E. Tardos, "Maximizing the spread of influence in a social network," *In ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD),* 2003.

[8] M. Maschler, S. Zamir, E. Solan *Cambridge University Press,* 2013.

[9] L. Rayo, I. Segal, "Optimal Information Disclosure," *Journal of Political Economy,* 118(5): 949–87, 2010.

[10] K.E. Schnakenberg, "Expert advice to a voting body," *Journal of Economic Theory* 160, 102113, 2015.

[11] I. Taneva, "Information Design," mimeo, 2014.

[12] Y. Wang, "Bayesian Persuasion with Multiple Receivers," mimeo, 2013.