Persuasion with Unknown Beliefs*

Svetlana Kosterina†

First draft: Summer 2017

Abstract

A sender designs a signal structure to persuade a receiver to choose one action over another. The sender is maximally ignorant about the receiver's prior on the states where the sender and the receiver disagree about the best action and has additional information about the receiver's prior on other states. I characterize the optimal signal structures in this environment. The lack of knowledge of the receiver's prior causes the persuasion mechanism to never completely give up: the optimal signal recommends the high action with a strictly positive probability in all states. I show that the probability that the high action is recommended is continuous in the state and the optimal signal may reveal the state with some probability. Finally, I show that the solution to the problem of persuasion with unknown beliefs is the same as the solution to the problem of persuading all members of a large group with heterogeneous priors.

1 Introduction

When trying to persuade someone, one finds it useful to know the beliefs the target of persuasion holds. Yet often such beliefs are unknown to the persuader. How should persuasion be designed when knowledge of prior beliefs is limited?

The following example illustrates an application of the model. A pharmaceutical company commissions an experiment on the safety of a drug that it would like to persuade the Food and Drug Administration (the FDA) to approve. The company does not know the exact prior belief of the regulator about the safety of the drug and instead only knows that

*I am indebted to my advisor, Wolfgang Pesendorfer, for his encouragement and support in developing this project. I am also grateful to Faruk Gul and Germán Gieczewski for insightful conversations. For helpful comments I would like to thank David Dillenberger, Kris Ramsay, Juan Pablo Xandri, Leeat Yariv, the audience at the 29th International Conference on Game Theory at Stony Brook University (2018) and seminar audiences at Princeton.

†Svetlana Kosterina is a PhD student at Princeton University; email: svetlana.kosterina@princeton.edu.

the FDA believes that the probability that the safety of the drug exceeds α is at least β. How should the company design the experiment?

The present paper aims to answer this question by studying a model of Bayesian persuasion in a setting where the receiver's prior belief is not known to the sender. The sender believes that Nature chooses the receiver's prior from a set of priors to minimize the sender's payoff. The sender has a known prior over the states and designs an experiment to maximize her payoff in the worst case scenario.

The receiver can take one of two actions. For instance, the FDA can approve a drug or not. The sender wishes to convince the receiver to take the high action in all states of the world. Thus a pharmaceutical company aims to convince the FDA to approve the drug regardless of its quality. The receiver takes the action desired by the sender only if the receiver's expectation of the state given his information is above a threshold and takes the other action otherwise. We call this threshold a *threshold of doubt*. In line with this reasoning, the FDA only approves the drugs that it believes to be sufficiently safe.

I focus on the case where the sender is maximally ignorant, and yet is still able to design a persuasion mechanism inducing the receiver to act where he would not have acted in absence of persuasion. In particular, I assume that the sender only knows that the receiver believes that the probability the state is greater than α is at least β.

This set of the receiver's priors has an appealing interpretation: the sender knows that the receiver's belief puts no less than a certain mass on the optimistic region of the state space containing sufficiently high states but is maximally ignorant about the receiver's belief within the optimistic and the pessimistic regions of the state space. As I show in section 5, my results generalize to the case where the sender is maximally ignorant about the receiver's prior on the region of states where the sender and the receiver disagree on the best action and has some additional information about the receiver's prior on the region where they agree on the best action.

We say that the sender conducts an *investigation* in state ω if there is a signal realization that has positive probability if and only if the state is ω. The reason we interpret this as an investigation is that, conditional on seeing this signal realization, the receiver knows the state. We say that the receiver makes a *mistake* in state ω if, conditional on the state being ω, the receiver takes an action different from the one he would take if he knew the state.

In the standard Bayesian persuasion model, the sender and the receiver have a common prior belief about the state. An optimal signal in that model recommends the high action with probability one in all states above a threshold and recommends the low action with probability one in all states below the threshold. We call this threshold a *threshold of action*. The threshold of action is below the threshold of doubt, so that the receiver takes the high action on a greater range of states than he would under complete information. The characterization implies that mistakes happen on the interval of intermediate states between the threshold of action and the threshold of doubt.

The optimal signal in the standard model is not unique, but any optimal signal cannot feature investigations on any set of states with a positive measure above the threshold of doubt. The reason for this is that revealing states above the threshold of doubt reduces the sender's ability to gain credibility by pooling favorable states with unfavorable ones.

The main contribution of this paper is a characterization of the optimal signal structure in a model of persuasion when the receiver's beliefs are unknown. In particular, I provide a formula for the optimal signal. The persuasion mechanism that the sender designs recommends the high action with a strictly positive probability in every state of the world.

The results thus change the way we think about Bayesian persuasion: unlike the intuition in the standard model, it is not optimal to pool all sufficiently high states together and give up on persuading receiver in the lower states. Instead, the sender must allow persuasion to fail with some probability on some of the high states and is able to persuade the receiver with a positive probability in the low states. This reveals the nature of persuasion to be fundamentally local: for small increases in the state the sender is able to increase the probability with which the receiver is persuaded only by a small amount.

The signal robust to the sender's uncertainty over the receiver's prior has three distinct features: the incidence of *mistakes*, the *continuity* in the state of the probability with which the high action is recommended, and the incidence of *investigations*. The pattern of mistakes in the model with unknown beliefs differs from the one in the standard model. Because the optimal signal recommends the high action with a strictly positive probability in all states, mistakes happen with a strictly positive probability in all states below a certain threshold. This happens because in the model with unknown beliefs the sender's ability to gain credibility by pooling sufficiently favorable states together is limited: if a signal realization recommends the high action in states below the threshold of doubt with a probability that is too high, then Nature is able to choose the receiver's prior that causes the receiver to take the low action after observing this signal realization instead.

On the other hand, because in the standard model the sender pools sufficiently favorable states together, when the receiver's beliefs are known, mistakes only happen on the interval of intermediate states between the threshold of action and the threshold of doubt. We see that, when the receiver's beliefs are unknown, especially pernicious outcomes are possible. For instance, the FDA approves even the most unsafe drugs with a strictly positive probability, whereas if the receiver's prior is known, the probability that the most unsafe drugs are approved is zero. Thus a model of persuasion with unknown beliefs can rationalize the occurrence of adverse outcomes that cannot be explained by the standard model.

In the model where the receiver's beliefs are unknown the probability of the high action is *continuous* in the state. We can interpret the continuity as similar treatment of similar cases. Thus pharmaceutical firms which produce drugs adhering to similar safety standards face a similar probability that the drug is approved. In the model where the receiver's beliefs are known, on the other hand, the probability of the high action is discontinuous, jumping from zero to one at the threshold.

The patterns of investigations that may arise in a model with unknown beliefs differ from the patterns of investigations possible in the model where the receiver's beliefs are known. One optimal information policy in a model with unknown beliefs involves recommending the high action with some probability and revealing the state with a complementary probability. Importantly, some states above the threshold of doubt are revealed with a strictly positive probability. We see that, unlike in the model with known beliefs, when the receiver's beliefs are unknown, *investigations* are possible even in the states where the receiver would take the high action if he knew the state. The reason for this is that in the model with unknown beliefs the sender's ability to gain credibility by pooling favorable states with unfavorable ones is limited, so the sender's payoff is not reduced if she reveals some states above the threshold of doubt with a sufficiently small probability. Moreover, the probability that an investigation is conducted need not be monotone in the state and may be maximal when the state is intermediate.

An additional contribution of this paper is the analysis of the welfare implications of persuasion with unknown beliefs, including the value of information for the receiver, the comparative statics on the sender's welfare and the cost of sender's ignorance.

The value of information for the receiver is the difference between the receiver's payoff under persuasion and under no information. I show that the value of information is zero for receivers with some priors, while for receivers with other priors it is strictly positive. I find that making the receiver more optimistic in the sense of first order stochastic dominance may either increase or decrease the receiver's value of information, a result that relies on the receiver's prior not having full support. The finding that the value of information for the receiver can be strictly positive may help explain why receivers participate in persuasion even though this is costly in terms of time and effort. Because the value of information in the standard model of persuasion is zero, such behavior cannot be explained by the standard model.

How does the sender's payoff from persuasion change as the sender's prior, interpreted as a known distribution of the state, changes? I show that the sender benefits from the distribution of the state being more dispersed in the sense of a mean-preserving spread on the states below some threshold. This is because the probability of the high action induced by the optimal signal is convex on the states below the threshold. The result that the sender may prefer certain more risky distributions of the state implies that if the sender has a choice as to what to persuade the receiver about, the sender may choose the object of persuasion that is more risky. Thus, for instance, a pharmaceutical company may prefer to experiment with drugs that are more risky.

I define the cost of the sender's ignorance as the difference between the sender's payoff when the receiver's prior is known and the sender's payoff when the receiver's prior is unknown. I provide a formula for the cost of the sender's ignorance. Moreover, I show that the sender's cost of ignorance is lower if the distribution of the state is more dispersed either below some threshold that is lower than the threshold of action or between this threshold and the threshold of action. This is because the cost of the sender's ignorance as a function of the state is concave on both of these intervals.

I observe that the solution to the problem of persuasion with unknown beliefs coincides with the solution to the problem of persuading all members of a group with known heterogeneous beliefs to take an action. In particular, if the sender faces a large group of the receivers where all prior beliefs that put a mass of at least β on states above α are represented, it is optimal for her to commit to the same signal structure as she would had she faced a single receiver with unknown beliefs.

The final contribution of the paper lies in solving a mechanism design problem to which the revelation principle does not apply. Solving such problems tends to be challenging. I show that the model in the present paper can be solved by the means of using a fixed-point argument to define the receiver's prior chosen by Nature in response to the sender choosing a signal structure.

The rest of the paper proceeds as follows. Section 2 introduces the model. Section

3 presents the characterization of the optimal signal structure and discusses the distinct features of the optimal signal. Section 4 analyzes the welfare implications of persuasion with unknown beliefs. Section 5 generalizes the set of the receiver's priors. Section 6 discusses the persuasion of groups with diverse beliefs. Section 7 provides a sketch of the proof. Section 8 reviews the related literature. Section 9 concludes.

2 Model

2.1 Payoffs

The state space is an interval $\Omega = [a_0, \overline{a}]$ such that $a_0 > 0$. The sender's preferences are state-independent. The sender gets a utility of $u(a)$ if the receiver's expected value of the state given the receiver's prior and the signal realization is a. The sender's utility function u is

$$u(a) = \begin{cases} 0 \text{ if } a \in [a_0, a^*) \\ 1 \text{ if } a \in [a^*, \overline{a}] \end{cases}$$

The model can be interpreted as one where the receiver can take one of the two actions, 0 or 1, and he takes action 1 if and only if his expectation of the state is weakly greater than a^*.

2.2 Priors

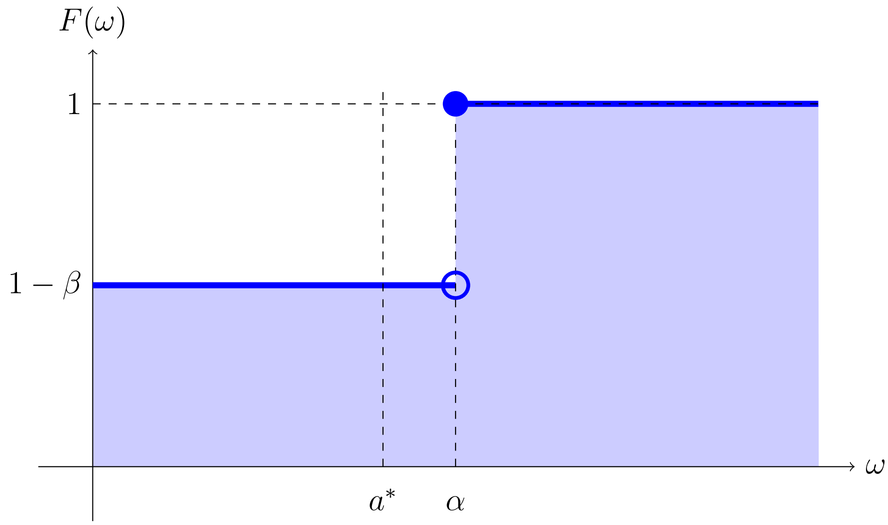
The sender has a known prior over the states, while the receiver has a set of priors. This is in contrast to the standard model of Bayesian persuasion, where the sender and the receiver have a known common prior. Let φ denote the set of all CDFs on Ω. The sender's prior is a CDF F_s.[1]

Given $\alpha \in \Omega$ and $\beta \in (0, 1)$, define a set of CDFs $C(\alpha, \beta)$ as follows:

$$C(\alpha, \beta) = \{F \in \varphi : 1 - \lim_{w \uparrow \alpha} F(w) \geq \beta\}$$

[1]Because the optimal signal in section 3 does not depend on the sender's prior F_s, the signal would remain optimal if we assumed that the sender does not know the distribution of the state and Nature chooses the sender's prior as well as the receiver's prior.

Figure 1: The Set of Receiver's Priors



The set of the receiver's priors is the set of CDFs $C(\alpha, \beta)$. Thus the sender knows that there exists a state $\alpha \in \Omega$ such that the mass that the receiver's prior puts on states above α is at least β for some $\beta \in (0, 1)$. Figure 1 illustrates the set of the receiver's priors: the set consists of all CDFs below the CDF drawn in bright blue in figure 1.

Observe that the fact that the highest CDF drawn in bright blue is constant on the intervals $[a_0, \alpha)$ and $[\alpha, \overline{a}]$ implies that the receiver's prior can put a mass of no more than $1 - \beta$ on the pessimistic region $[a_0, \alpha)$ and a mass of no less than β on the optimistic region $[\alpha, \overline{a}]$, but subject to this constraint, any allocation of mass within these regions is possible. Therefore, the fact that the highest CDF is constant on the intervals $[a_0, \alpha)$ and $[\alpha, \overline{a}]$ reflects the sender's maximal ignorance about the receiver's prior within the pessimistic region $[a_0, \alpha)$ and the optimistic region $[\alpha, \overline{a}]$.

I assume that $\alpha > a^*$, so that the receiver's prior puts a mass of at least β on states strictly above a^*. This assumption ensures that the solution to the sender's problem is non-trivial.

Note that the worst prior the receiver can have if the sender provides no information is $\mu_F = (1 - \beta)\delta_{a_0} + \beta\delta_{\alpha}$. I assume that

$$(1 - \beta)a_0 + \beta\alpha < a^*$$

This assumption ensures that the receiver with the worst prior takes action zero if he does not receive any information.

2.3 Information Structures

The order of moves is as follows. First, the sender commits to a signal structure P. Next, Nature chooses the receiver's prior $F \in C(\alpha, \beta)$ to minimize the sender's payoff. Then the state is realized (from the sender's perspective, the state is drawn from the distribution F_s). After this, the signal is realized according to the signal structure P. Then, having seen a signal realization σ, the receiver forms an expectation of the state given that the receiver's prior is F and that the signal structure is P.

We let the message space be $M = \Omega$. A signal structure is a kernel P. Informally, we can interpret $P(\omega, \sigma)$ as the probability of signal realization σ given that the state is ω. Formally, letting $\mathcal{B}(M)$ and $\mathcal{B}(\Omega)$ denote the Borel sigma-algebras on M and Ω respectively, a kernel P is defined as a mapping $P : \Omega \times \mathcal{B}(M) \to [0, 1]$ such that for every $\omega \in \Omega$, $B \mapsto P(\omega, B)$ is a probability measure on M and for every $B \in \mathcal{B}(M)$, $\omega \mapsto P(\omega, B)$ is $\mathcal{B}(\Omega)$-measurable. I restrict my attention to the signal structures that have a finite number of signal realizations given each state.[2]

Given a distribution F on Ω, a kernel P is a regular conditional probability. That is, given a kernel P and a distribution F, there exists a joint probability measure m_F over $\Omega \times M$ such that

$$m_F(A, B) = \int_A \int_B P(\omega, d\sigma) dF(\omega)$$

for all measurable $A \subseteq \Omega$, $B \subseteq M$.[3]

Define a probability measure $\mu_{G_{F,P}}$ by setting $\mu_{G_{F,P}}(B) = \int_\Omega \int_B P(\omega, d\sigma) dF(\omega)$ for each B in the Borel sigma-algebra on M. Let $M_{F,P} = \operatorname{supp} \mu_{G_{F,P}}$ denote the support of $\mu_{G_{F,P}}$.

Then there exists a kernel $Q_{M_{F,P}}$[4] such that

$$m_F(A, B) = \int_B \int_A Q_{M_{F,P}}(\sigma, d\omega) dG_{F,P}(\sigma)$$

for all measurable $A \subseteq \Omega$, $B \subseteq M_{F,P}$.[5]

[2] I conjecture, but do not have a proof, that my results hold for all signal structures.

[3] See Theorem 20 on page 86 in Pollard (2002).

[4] The kernel $Q_{M_{F,P}}$ is a mapping $Q_{M_{F,P}} : M_{F,P} \times \mathcal{B}(\Omega) \to [0, 1]$ such that for every $\sigma \in M_{F,P}$, $B \mapsto Q_{M_{F,P}}(\sigma, B)$ is $\mathcal{B}(\Omega)$-measurable and for every $B \in \mathcal{B}(\Omega)$, $\sigma \mapsto Q_{M_{F,P}}(\sigma, B)$ is a probability measure on Ω.

[5] Note that, for all $A \in \mathcal{B}(M)$, if $\mu_{G_{F,P}}(A) = 0$, then $\operatorname{marg}_M m_F(A) = 0$. That is, the M-marginal of

$Q_{M_{F,P}}$ is defined formally in footnote 4. Informally, we can interpret $Q_{M_{F,P}}(\sigma, \omega)$ as the probability of state ω given that the signal realization is σ.

2.4 Updating

Because the present model features priors with possibly non-overlapping supports, the issue of the receiver updating beliefs after observing a zero-probability event needs to be addressed.

I assume that, having seen a signal realization σ such that (ω, σ) is not in the support of the joint probability measure $m_F(\Omega \times M)$ for any $\omega \in \Omega$, the receiver with prior μ_F assigns probability one to the lowest state ω such that σ is in the support of $P(\omega, \cdot)$. That is, having seen a signal realization σ such that $(\omega', \sigma) \notin \operatorname{supp} m_F(\Omega \times M)$ for any $\omega' \in \Omega$, the receiver with prior μ_F puts probability one on

$$\underline{\omega}(\sigma) = \inf\{\omega : \sigma \in \operatorname{supp} P(\omega, \cdot)\}$$

This assumption can be justified by considering Nature choosing a prior that is a convex combination of a distribution μ supported on $[\underline{\omega}(\sigma), \underline{\omega}(\sigma) + \eta]$ for some $\eta \in (0, 1)$ and the distribution μ_F with weights $\epsilon \in (0, 1)$ and $1 - \epsilon$ respectively. Given this prior, Bayesian updating leads the receiver to assign probability one to the interval $[\underline{\omega}(\sigma), \underline{\omega}(\sigma) + \eta]$ after seeing the signal realization σ. Note that as $\eta \to 0$, a distribution μ on $[\underline{\omega}(\sigma), \underline{\omega}(\sigma) + \eta]$ converges to $\delta_{\underline{\omega}(\sigma)}$.[6] The receiver's behavior specified in the assumption can then be obtained by taking the limit of the receiver's beliefs as η and ϵ go to zero.

Observe that, because the sender's payoff is increasing in the receiver's expectation of the state, the limit belief is the receiver's belief that is the worst for the sender among the limits of sequences of beliefs that can be obtained through Bayesian updating after seeing a signal realization σ provided that the receiver's prior is a convex combination of μ_F and some other measure.

The assumption about updating is not important for the main results of the paper concerning the equilibrium distribution of actions in a model of persuasion with unknown beliefs. In particular, we could have assumed instead that upon seeing a signal realization

m_F is absolutely continuous with respect to $\mu_{G_{F,P}}$. Then the result follows from Theorem 1 on page 339 in Pollard (2002).

[6] $\delta_{\underline{\omega}(\sigma)}$ denotes the Dirac measure on $\underline{\omega}(\sigma)$.

impossible under the receiver's prior the receiver does not update his beliefs. In this case, the equilibrium distribution of actions would be the same.

The assumption about updating is used in obtaining the optimal signal that features revealing some of the states above the threshold of doubt with a positive probability. If instead we assumed that a receiver seeing a signal he believed to be impossible does not update his beliefs, a signal with investigations in states above the threshold of doubt would never be optimal.

To augment the receiver's posterior beliefs with beliefs the receiver holds after observing events to which his prior assigns zero probability, I define a kernel $Q_F : M \times \mathcal{B}(\Omega) \to [0,1]$ such that for all $\sigma \in M$ we have

$$Q_F(\sigma,\cdot) = \begin{cases} Q_{M_{F,P}}(\sigma,\cdot) \text{ if } \sigma \in M_{F,P} \\ \delta_{\underline{\omega}(\sigma)} \text{ if } \sigma \notin M_{F,P} \end{cases}$$

2.5 Evaluation of Payoffs

If the sender chooses a kernel P and a receiver with a prior F sees a signal realization σ, then the receiver's expectation of the state is

$$E_{F,P}[\omega|\sigma] = \int_\Omega w Q_F(\sigma, dw)$$

Then, if the sender chooses a kernel P and Nature chooses $F \in C(\alpha,\beta)$, the sender's payoff is

$$\int_\Omega \int_M \mathbb{1}_{E_{F,P}[\omega|\sigma]\geq a^*} P(w,d\sigma)dF_s(w)$$

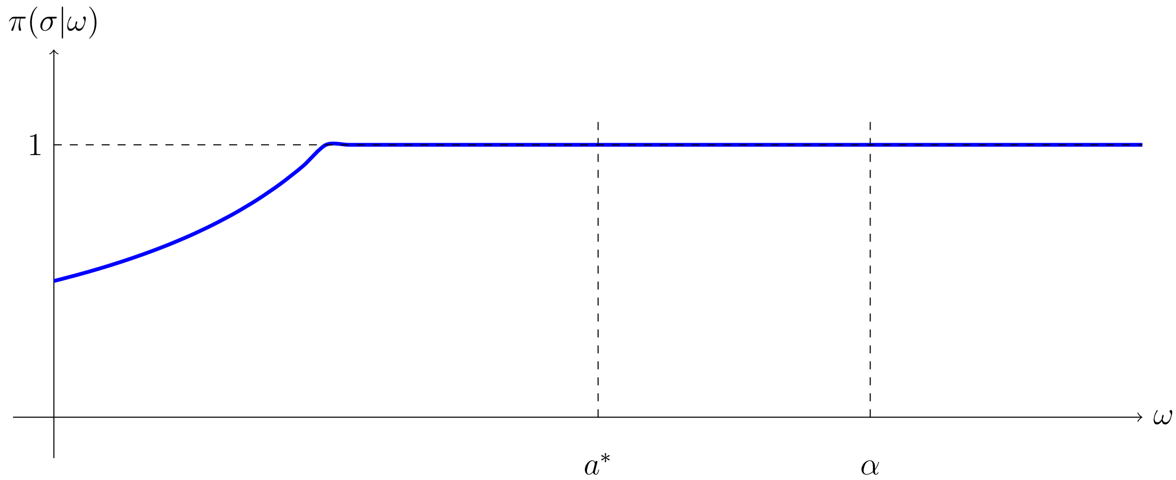
Thus the sender's equilibrium payoff is

$$\max_P \min_{F\in C(\alpha,\beta)} \int_\Omega \int_M \mathbb{1}_{E_{F,P}[\omega|\sigma]\geq a^*} P(w,d\sigma)dF_s(w) \tag{1}$$

3 The Optimal Signal

A signal is said to be optimal if it solves the sender's problem (1). Theorem 1 describes an optimal signal and the distribution over the receiver's actions induced by it.

Figure 2: Optimal Information Policy without Revelation



Theorem 1. *There exists an optimal signal that induces a distribution over the receiver's actions such that the probability of action* 1 *in state* ω *is given by*

$$s(\omega) = \begin{cases} 1 \text{ if } \omega \in (a^*, \overline{a}] \\ \min\left\{\frac{\beta(\alpha - a^*)}{(1-\beta)(a^* - \omega)}, 1\right\} \text{ if } \omega \in [a_0, a^*] \end{cases}$$

If there exists an optimal signal that induces a distribution over the receiver's actions where the probability of action 1 *is given by* s' *and* $s' \neq s$*, then* $s'(\omega) = s(\omega)$ μ_{F_s}*-a.e.*

An optimal signal inducing the distribution s *over the receiver's actions is given by* $\pi(\sigma|\omega) = s(\omega)$*,* $\pi(\sigma_0|\omega) = 1 - s(\omega)$ *for all* $\omega \in \Omega$*.*

Theorem 1 says that any signal that is optimal for the sender induces a distribution over the actions of the receiver that is unique up to sets that have measure zero under the sender's prior. If the state is in $(a^*, \overline{a}]$, the receiver takes action 1 with probability one, and if the state is in $[a_0, a^*]$, the receiver takes action 1 with probability $\min\left\{\frac{\beta(\alpha - a^*)}{(1-\beta)(a^* - \omega)}, 1\right\}$.

The optimal signal described in Theorem 1 features two signal realizations, σ and σ_0. The receiver takes action 1 after seeing signal σ and action zero after seeing signal σ_0. The sender recommends action 1 with probability one if the state is above a certain threshold. If the state is below this threshold, the sender recommends action 1 with some strictly positive probability that decreases continuously as the state decreases. Figure 2 illustrates this information policy.

In the standard Bayesian persuasion model where the sender knows the receiver's prior the signal optimal for the sender has a threshold structure, recommending action 1 with probability one if the state is above a threshold and recommending action 0 otherwise. When a prior common to the sender and the receiver is fixed, recommending action 1 in higher states below the threshold of doubt a^* yields a strictly greater benefit to the sender than recommending action 1 in lower states, so the sender recommends action 1 in all sufficiently high states such that the receiver's expectation upon seeing the signal realization σ is exactly a^*.

The threshold signal structure is not optimal when the receiver's beliefs are unknown. To gain an intuition as to why, consider a signal structure with two realizations, σ and σ_0, satisfying $\pi(\sigma|\alpha) = 1$ and the receiver's prior that puts a mass of β on α and a mass of $1-\beta$ on some state ω below the threshold a^*. Then the receiver's expectation conditional on seeing the signal realization σ is $E[\omega|\sigma] = \frac{\omega(1-\beta)\pi(\sigma|\omega)+\alpha\beta}{(1-\beta)\pi(\sigma|\omega)+\beta}$. In order for the receiver to take action 1 after seeing σ, we need $E[\omega|\sigma] \geq a^*$, which turns out to be equivalent to $\pi(\sigma|\omega) \leq \frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}$.

Note then that Nature moves after the sender chose the signal structure and can put a mass of $1-\beta$ on any state ω below the threshold a^*. Thus if the probability of signal realization σ exceeds the bound $\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}$ at any state below the threshold, then by putting a mass of $1-\beta$ on this state Nature can ensure that the receiver never takes action 1. Therefore, $\pi(\sigma|\omega)$ must be below the bound in all states. On the other hand, the sender's payoff is increasing in the probability that action 1 is taken, implying that it is best for the sender to maximize $\pi(\sigma|\omega)$ subject to the constraint that it be below the bound. Thus setting $\pi(\sigma|\omega)$ equal to the bound in all states yields the optimal signal when the receiver's beliefs are unknown.

If the state is below a^*, then $\pi(\sigma|\omega)$, the probability with which the optimal signal recommends action 1, is increasing in α and β. Thus if the sender knows that the receiver's prior puts a positive mass on states above a higher threshold or that the mass that the receiver's prior puts on states above a certain threshold is higher, the sender is able to induce the receiver to take action 1 with a higher probability.

3.1 Comparison with the Standard Model

In this section I introduce the standard model of Bayesian persuasion and compare my results to the optimal signal in the standard model. The standard model has the same payoff structure as the model with unknown beliefs but the sender and the receiver have a common prior $F_s \in C(\alpha, \beta)$ over the states. For simplicity, I assume that F_s admits a density.[7] I assume that the receiver takes action 0 if no information is provided, that is, that $E_{F_s}[\omega] < a^*$.

It is known that in this case the solution of the standard persuasion model has a threshold structure: there is an optimal signal p with two signal realizations, σ and σ_0, such that $p(\sigma|w) = 1$ for $w \in [a', \overline{a}]$, $p(\sigma|w) = 0$ for $w \in [a_0, a')$ for some action threshold $a' \in [a_0, a^*)$ and the receiver takes action 1 if and only if the realized signal is σ.

Recall that the equilibrium probability of action 1 in the model with unknown beliefs is given by $s(w) = \min\left\{\frac{\beta(\alpha - a^*)}{(1-\beta)(a^*-w)}, 1\right\}$ on states w such that $w < a^*$ and is given by $s(w) = 1$ on states w such that $w \geq a^*$. Thus the function $w \mapsto s(w)$ is strictly increasing on the states w such that $\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-w)} \leq 1$ and is constant on the states w such that $\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-w)} > 1$.

Let $\underline{\alpha}$ denote the state such that $w \mapsto s(w)$ is strictly increasing on the states below $\underline{\alpha}$ and is constant on states above $\underline{\alpha}$. That is, define

$$\underline{\alpha} = a^* - \frac{\beta}{1-\beta}(\alpha - a^*)$$

The optimal signal in the model with unknown beliefs recommends action 0 with a strictly positive probability on states below $\underline{\alpha}$ and recommends action 0 with probability zero on states above $\underline{\alpha}$.

Proposition 1. *$\underline{\alpha} > a'$.*

Proposition 1 shows that we must have $\underline{\alpha} > a'$, that is, that the state at which the robust signal starts recommending action 1 with probability one is higher than the state at which the signal in the standard persuasion model starts recommending action 1 with probability one.

This comparison highlights why the signal that is optimal in the standard model is not optimal in the model with unknown beliefs. Because $\underline{\alpha} > a'$, the interval $[a', \underline{\alpha}]$ is non-empty. The signal optimal in the standard model recommends action 1 with probability one on the interval $[a', \underline{\alpha}]$. On the other hand, the signal optimal in the model with unknown beliefs recommends action 1 with probability that is strictly less than one on this interval. If instead action 1 was recommended with a greater probability on this interval, then Nature would be able to choose the receiver's prior ensuring that the receiver's expectation given the signal

[7]In the general case, the optimal signal may involve randomization at the threshold.

realization σ is strictly less than a^*, causing the receiver to never take action 1.

3.2 A Simpler Model

The model in the present paper has Nature choose the receiver's prior after the sender commits to a signal structure but before the signal is realized. We could imagine a simpler model, where Nature chooses the receiver's prior after the signal is realized. This model is simpler because if there is a feasible prior such that the receiver with this prior takes action 0 after observing a signal, Nature can choose this prior after this signal. Thus if action 1 is taken after a signal realization in equilibrium, it must be that a receiver with any feasible prior takes action 1 after seeing the signal realization. This implies that there is no loss in pooling the signals after which action 1 gets taken into one signal. Therefore, there is an optimal signal with only two signal realizations in the simpler model.

I observe that the optimal signal in the model in the present paper is the same as in the simpler model described above. This follows from the fact that a signal that is optimal in the model of persuasion with unknown beliefs has two realizations, σ and σ_0, and a receiver with any feasible prior takes action 1 after seeing signal realization σ and action 0 after seeing signal realization σ_0. Thus Nature cannot do any better when choosing the receiver's prior after, as compared to before, the signal is realized – because no matter which prior Nature chooses, the receiver takes the same action after the signal realization, making Nature indifferent among all feasible priors.

The equivalence need not hold in general. In particular, it is possible that with a different set of priors and a different utility function of the sender the optimal signal and the prior that Nature chooses are such that there is more than two signal realizations and for some of the signal realizations there are different feasible priors such that receivers with these priors take action 0 after seeing the signal realization. In this case, when choosing the prior before the signal is realized, Nature may be unable to choose a prior ensuring action 0 after all of these signal realizations.

3.3 Mistakes

Persuasion with unknown beliefs has distinct implications for both the patterns of mistakes in the receiver's decision-making and for the strategy of evidence-gathering by the sender, which I highlight below.

The distinct features of robust persuasion are a strictly positive probability of the worst possible outcome and continuity of the probability of the high action in the state. Because the high action is recommended with a strictly positive probability given every state, the outcome given the state is less determined in a model of persuasion with unknown beliefs. With some probability, even the least safe drugs will be approved by the FDA. Thus the model of robust persuasion can rationalize the occurrence of exceptionally deleterious outcomes that the agent making the decision has the greatest incentive to prevent and that would not occur in absence of persuasion.

Continuity of the probability of the high action in the state implies that small changes in the state result in small changes in the probability that the high action is taken. Therefore, robustness concerns lead the FDA to treat the firms adhering to similar safety standards similarly.

3.4 Investigations

The optimal information policy described in Theorem 1 is not unique. The following is an example of a different optimal information policy:

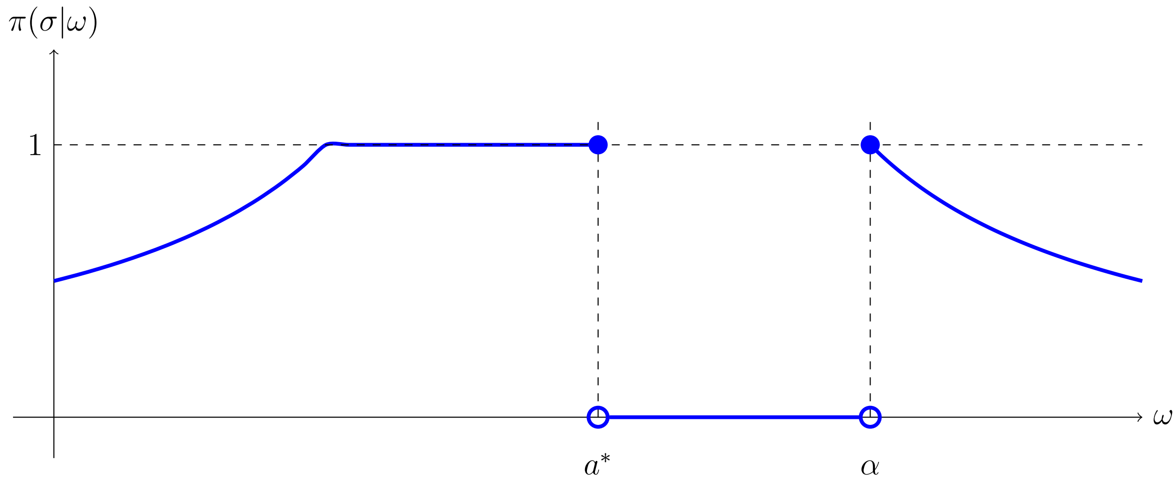
$$\pi(\sigma|\omega) = \begin{cases} \frac{\alpha - a^*}{\omega - a^*} \text{ if } \omega \in [\alpha, \overline{a}] \\ 0 \text{ if } \omega \in (a^*, \alpha) \\ \min\left\{\frac{\beta(\alpha - a^*)}{(1-\beta)(a^* - \omega)}, 1\right\} \text{ if } \omega \in [a_0, a^*] \end{cases}$$

$$\pi(\sigma_w|\omega) = \begin{cases} 1 - \frac{\alpha - a^*}{\omega - a^*} \text{ if } \omega = w \text{ and } \omega \in [\alpha, \overline{a}] \\ 1 \text{ if } \omega = w \text{ and } \omega \in (a^*, \alpha) \\ 1 - \min\left\{\frac{\beta(\alpha - a^*)}{(1-\beta)(a^* - \omega)}, 1\right\} \text{ if } w = \omega \text{ and } \omega \in [a_0, a^*] \end{cases}$$

Under this information policy, the signal reveals each state with some probability, and, with a complementary probability, the signal provides no information. The optimality of this information policy relies on our assumption about how the receiver updates his beliefs upon seeing a signal that is impossible under the receiver's prior: the receiver puts probability one on the lowest state in which the signal is sent. If instead the receiver did not update his beliefs in this situation, then revealing states above the threshold of doubt would not be optimal.

When the state is low, the state is revealed with a high probability, which decreases

Figure 3: Optimal Information Policy with Revelation



as the state increases. The probability that the state is revealed decreases to zero and stays at zero until the state is equal to the decision maker's threshold of doubt a^*. Then the probability that the state is revealed jumps to one and stays at one until the state is equal to the decision maker's threshold of knowledge α. After this, the probability that the state is revealed jumps to zero and then increases continuously to some number strictly less than one.

We can interpret the probability that the state is revealed as the incidence of transparency or as the probability that the pharmaceutical firm commissions a thorough investigation. Under the information policy with the maximal transparency the incidence of transparency is non-monotone in the state. In particular, the incidence of transparency has two troughs. Thus the pharmaceutical firm is most likely to practice transparency when the drug safety is intermediate, moderately likely to practice transparency when the drug is either very safe or very dangerous, and is unlikely to practice transparency when the drug safety is somewhat low or somewhat high. Figure 3 illustrates this information policy.

4 Welfare

4.1 The Value of Information

In this section I ask the following question: how does the provision of information by the sender affect the utility of the receiver under robust persuasion? I define the value of information for the receiver as the difference between the receiver's utility under persuasion and the receiver's utility under no information. The receiver's expected utility is evaluated according to his prior, so the question addressed is: what does the receiver *believe* the value of information is to him?

Consider the optimal information policy π with two signal realizations, σ and σ_0. Lemma 13 in the Appendix shows that a receiver with a prior μ_F, $F \in C(\alpha, \beta)$, strictly benefits from persuasion if and only if

$$E_F[\omega] < a^* \qquad E_F[\omega|\sigma] > a^*$$

or

$$E_F[\omega] \geq a^* \qquad \int_\Omega \pi(\sigma_0|w)dF(w) > 0$$

That is, the value of persuasion is strictly positive if and only if the receiver's expectation of the state given signal σ is strictly greater than his threshold of doubt provided that the receiver's expectation of the state under no information is strictly less than the threshold of doubt, and the probability of the signal realization σ_0 is strictly positive according to the receiver's prior provided that the receiver's expectation of the state under no information is weakly greater than the threshold of doubt.

It is shown in the Appendix that the optimal information policy is such that we have $E_F[\omega|\sigma] \geq a^*$ for all receiver's priors of the form $\mu_F = (1-\beta)\delta_\omega + \beta\delta_\alpha$, $\omega \in [a_0, \alpha)$. This implies that a sufficient condition to have $E_F[\omega|\sigma] > a^*$ for a prior with support $\{\omega, \alpha\}$ is that the receiver's prior puts a mass strictly greater than β on the state α.

Thus, for instance, a receiver with a prior $\mu_F = (1-\beta)\delta_\omega + \beta\delta_\alpha$ for $\omega \in [a_0, \alpha)$ such that $E_F[\omega] < a^*$ does not strictly benefit from persuasion because $E_F[\omega|\sigma] = a^*$, while a receiver with a prior $\mu_G = (1-\gamma)\delta_\omega + \gamma\delta_\alpha$ for some $\gamma \in (\beta, 1)$ satisfying $E_G[\omega] < a^*$ does because $E_G[\omega|\sigma] > a^*$. This shows that making a receiver *more* optimistic in the sense of first order stochastic dominance can change the receiver's value of information from zero to strictly positive.

On the other hand, a receiver with a prior $\mu_F = (1-\beta)\delta_\omega + \beta\delta_\alpha$ for $\omega \in [a^*, \alpha)$ does not strictly benefit from persuasion because $\int_\Omega \pi(\sigma_0|w)dF(w) = 0$, while a receiver with a prior $\mu_G = (1-\beta)(\gamma\delta_{\omega'} + (1-\gamma)\delta_\omega) + \beta\delta_\alpha$ for some ω' such that $\pi(\sigma|\omega') < 1$ and $\gamma \in (0, 1)$ sufficiently small does because $\int_\Omega \pi(\sigma_0|w)dG(w) > 0$ (note that we have $E_F[\omega] > a^*$ and, for $\gamma \in (0, 1)$ sufficiently small, $E_G[\omega] > a^*$). This shows that making a receiver *less* optimistic in the sense of first order stochastic dominance can also change the receiver's value of information from zero to strictly positive.

In the standard model of persuasion the receiver's value of information is zero because the sender is able to choose the signal structure such that, upon seeing the signal recommending the high action, the receiver with a known prior is indifferent between taking the high and the low action. In contrast, in a model of persuasion with unknown beliefs there is a set of priors Nature can choose from. We can rank some of the feasible priors by first-order stochastic dominance. Because the sender must persuade receivers with all feasible priors to take the high action, if a receiver with some prior is indifferent between the high and the low action, there is a feasible prior dominating it in the FOSD sense such that a receiver with this prior strictly prefers to take the high action. This explains why greater optimism can increase the value of information.

The receiver in the standard model of persuasion takes the low action if no information is provided. Because there is a set of feasible priors in a model of persuasion with unknown beliefs and, in equilibrium, Nature is indifferent among all feasible priors, Nature may choose a prior such that the receiver with this prior takes the high action if no information is provided. Moreover, the receiver with this prior may believe that he will always take the high action – if he believes that the probability of seeing the signal recommending the low action is zero. In this case, the value of information is zero for this receiver: he believes that he will take the same action under persuasion and under no information. We can find a feasible prior such that the first prior dominates it in the FOSD sense and the receiver with this prior believes that there is a positive probability of seeing the signal recommending the low action. In this case, the value of information is strictly positive for the receiver: he believes that under persuasion he will sometimes take the low action, whereas he would always take the high action under no information. This explains why greater optimism can decrease the value of information. Note that the receiver's prior not having full support on $[a_0, \overline{a}]$ is necessary for this result.

4.2 The Sender's Welfare

Consider the optimal signal with two realizations, σ and σ_0. The sender's payoff is the probability that the optimal signal recommends action 1 averaged across states according to the sender's prior F_s:

$$\int_{\Omega} \pi(\sigma|w) dF_s(w)$$

Recall from section 3.1 that $\underline{\alpha} = a^* - \frac{\beta}{1-\beta}(\alpha - a^*)$ denotes the state such that $w \mapsto \pi(\sigma|w)$ is strictly increasing on the states below $\underline{\alpha}$ and is constant on the states above $\underline{\alpha}$.

Given probability measures μ and m on Ω, we say that μ is a *local mean-preserving spread* of m on $[a_0, \underline{\alpha}]$ if

$$\mu = \gamma\mu_1 + (1-\gamma)\mu_2 \qquad m = \gamma m_1 + (1-\gamma)m_2$$

for some $\gamma \in [0,1]$, $\mu_1(\Omega \setminus [a_0, \underline{\alpha}]) = m_1(\Omega \setminus [a_0, \underline{\alpha}]) = 0$, $\mu_2([a_0, \underline{\alpha}]) = m_2([a_0, \underline{\alpha}]) = 0$, $\mu_2 = m_2$ and μ_1 is a mean-preserving spread of m_1.[8]

Thus μ is a local mean-preserving spread of m on $[a_0, \underline{\alpha}]$ if μ and m have the same distribution conditional on the state being in $\Omega \setminus [a_0, \underline{\alpha}]$, μ and m put the same mass on $\Omega \setminus [a_0, \underline{\alpha}]$, and the distribution of μ conditional on the state being in $[a_0, \underline{\alpha}]$ is a mean-preserving spread of the distribution of m conditional on the state being in $[a_0, \underline{\alpha}]$. A local mean-preserving spread is a mean-preserving spread, but not all mean-preserving spreads are local mean-preserving spreads.

Proposition 2. *If μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0, \underline{\alpha}]$, then*

$$\int_\Omega \pi(\sigma|w)dF_s(w) \geq \int_\Omega \pi(\sigma|w)dG_s(w)$$

Consider two priors, F_s and G_s, such that μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0, \underline{\alpha}]$. Proposition 2 states that the sender's payoff is higher under the prior F_s. The reason for this is that the function $w \mapsto \pi(\sigma|w)$ is strictly convex on $[a_0, \underline{\alpha}]$.

Therefore, when the sender is motivated by robustness concerns, the sender benefits when the distribution of states is more dispersed on the states below $\underline{\alpha}$. Because the sender can only induce the receiver to take the high action with probability less than one in states below $\underline{\alpha}$, these states are unfavorable for the sender. Thus Proposition 2 implies that, conditional on the average unfavorable state being the same, it is better for the sender if unfavorable states have similar probability than if some unfavorable states are more likely than others.

[8]Recall that μ_F is a mean-preserving spread of μ_G if they have the same mean and G second order stochastically dominates F. Recall also that G second order stochastically dominates F if $\int_0^x G(y)dy \leq \int_0^x F(y)dy$ for all $x \in [0, \infty)$.

4.3 The Cost of Ignorance

In this section I compare the sender's welfare in the model where the sender does not know the receiver's beliefs to the sender's welfare in the model where the receiver's beliefs are known. In particular, I compare the sender's welfare in the robust model conditional on $F_s \in C(\alpha, \beta)$ being both the true distribution of the state of the world and the receiver's belief about this distribution to the sender's welfare in a model where both the sender and the receiver know that F_s is the true distribution of the state. As in section 3.1, I assume that the receiver would take action 0 if no information is provided ($E_{F_s}[\omega] < a^*$) and, for simplicity, I assume that F_s admits a density.

As shown in section 3.1, the solution of the standard persuasion model has a threshold structure: action 1 is recommended if and only if the state is above the threshold a'. Then $\int_{a'}^{\overline{a}} dF_s(w)$ is the sender's payoff in the standard model, and $\int_{a_0}^{\overline{a}} \pi(\sigma|w)dF_s(w)$ is the sender's payoff in the robust model. Proposition 1 in section 3.1 established that we must have $\underline{\alpha} > a'$, that is, that the state at which the robust signal starts recommending action 1 with probability one is higher than the state at which the signal in the standard persuasion model starts recommending action 1 with probability one.

Let $D(w)$ denote the difference between the sender's payoffs in the standard model and in the robust model conditional on the state being w. Then, because $\underline{\alpha} > a'$, we have

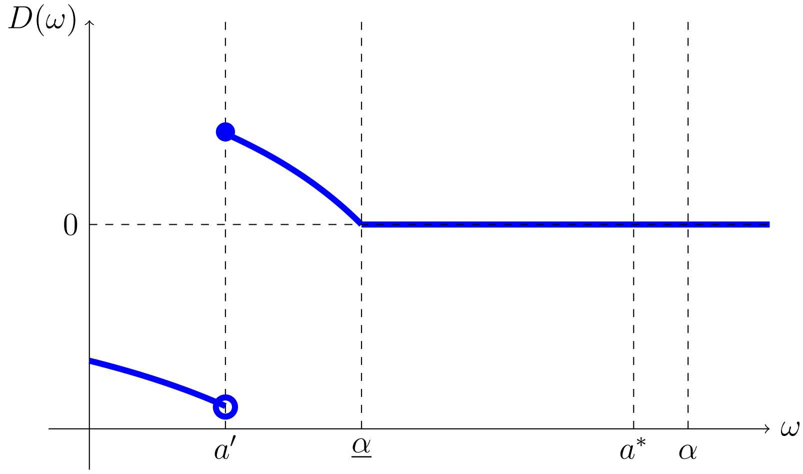
$$D(w) = \begin{cases} -\pi(\sigma|w) \text{ if } w \in [a_0, a') \\ 1 - \pi(\sigma|w) \text{ if } \omega \in [a', \underline{\alpha}) \\ 0 \text{ if } \omega \in [\underline{\alpha}, \overline{a}] \end{cases}$$

Note that the difference in the payoffs $D(\cdot)$ is strictly concave on the interval $[a_0, a')$, strictly concave on the interval $[a', \underline{\alpha})$ and is constant on the interval $[\underline{\alpha}, \overline{a}]$. Moreover, $D(\cdot)$ is decreasing on the interval $[a_0, a')$, decreasing on the interval $[a', \overline{a}]$ and is discontinuous at exactly one point, a', where it jumps up. Figure 4 illustrates the difference in the sender's payoffs.

Then the cost of the sender's ignorance is the average difference between the sender's payoffs in the standard model and in the robust model evaluated according to the prior F_s:

$$\int_{\Omega} D(w)dF_s(w)$$

Figure 4: The Difference between the Sender's Payoffs in the Standard Model and in the Robust Model



Proposition 3 provides comparative statics on the sender's cost of ignorance with respect to the distribution of the state F_s.

Proposition 3. *If μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0, a')$ or μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a', \underline{\alpha})$, then*

$$\int_\Omega D(w)dF_s(w) \leq \int_\Omega D(w)dG_s(w)$$

Consider two priors, F_s and G_s, such that either μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0, a')$ or μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a', \underline{\alpha})$. Proposition 3 states that the difference in the payoffs between the standard and the robust models is smaller under the prior F_s.

Observe that ignorance must be costly, that is, the difference must be positive for both priors. Thus making the distribution of states more dispersed on the states below a' or on the states between a' and $\underline{\alpha}$ reduces the cost of ignorance.

5 Generalizing the Set of Priors

In this section I show that the solution to the sender's problem obtained in this paper remains valid with a more general set of priors. In particular, as long as the worst prior that the receiver can have[9] puts a mass of exactly β on the state $\alpha > a^*$, the sender may have additional information about the receiver's prior on the region of states where the sender and the receiver agree about the optimal action.

Given $\alpha \in (a^*, \overline{a}]$ and $\beta \in (0,1)$, consider a CDF G satisfying $G(\omega) = 1-\beta$ if $\omega \in [a_0, \alpha)$ and $G(\omega) = 1$ if $\omega \in [\alpha, \overline{a}]$. G is the CDF of the measure that puts a mass of $1-\beta$ on state a_0 and a mass of β on state α.

Define the following set:

$$\underline{A} = \{F : [a_0, a^*) \to [0, 1-\beta] \text{ such that } F \text{ is increasing and right-continuous}\}$$

$\underline{A}$ is the set of all increasing and right-continuous functions on the interval $[a_0, a^*)$ with the codomain $[0, 1-\beta]$.

Consider a set of priors A satisfying the following conditions:

1. $G \in A$,
2. $A \subseteq C(\alpha, \beta)$,
3. for all $\tilde{F} \in \underline{A}$ there exists $F \in A$ such that $F(w) = \tilde{F}(w)$ for all $w \in [a_0, a^*)$.

A is a subset of the set of priors $C(\alpha, \beta)$ such that for all functions $\tilde{F}$ in the set $\underline{A}$ there exists a CDF F in A that is equal to $\tilde{F}$ on the interval $[a_0, a^*)$.

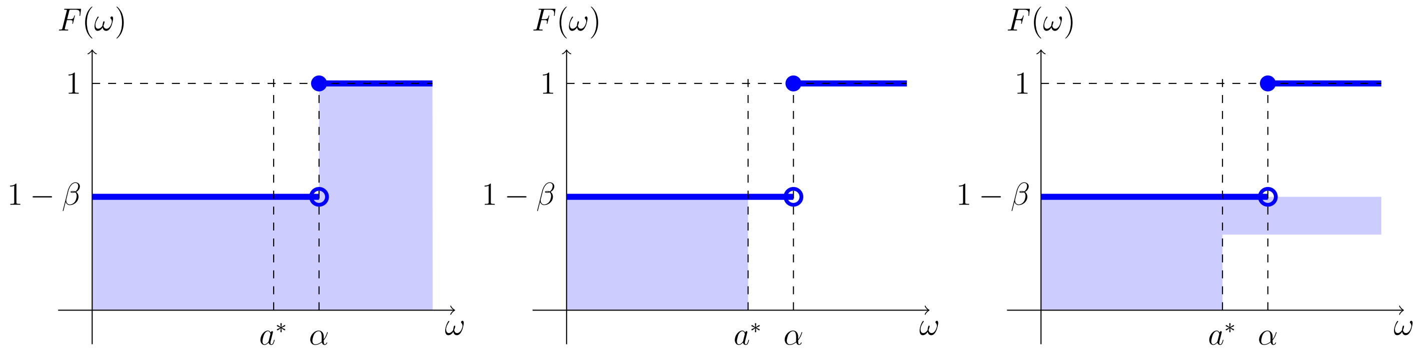
It can be shown that the equilibrium distribution of the induced actions and the signal described in Theorem 1 remain optimal if the set of the receiver's priors is A.

We can interpret this generalization as follows. The sender knows that the receiver's prior puts a mass of $1-\beta$ on the disagreement region $[a_0, a^*)$ and may have more information about the receiver's prior on the agreement region $[a^*, \overline{a}]$. The sender is maximally ignorant about the receiver's prior on the disagreement region, believing that any allocation of mass $1-\beta$ or less on the disagreement region is possible.

Figure 5 illustrates possible sets of the receiver's priors. The plot of the left shows $C(\alpha, \beta)$, the set of the receiver's prior that put a mass of at least β on states above α. The plot in the center shows the set of priors consisting of all measures such that their CDFs are below the CDF drawn in bright blue on the interval $[a_0, a^*)$ and coincide with the CDF drawn in bright blue on the interval $[a^*, \overline{a}]$. Finally, the plot on the right shows the set of

[9] By the worst prior of the receiver I mean the feasible prior of the receiver that minimizes the receiver's expectation of the state under no information.

Figure 5: Possible Sets of Receiver's Priors



priors that contains all measures in the set of priors shown in the center, as well as some additional measures, so that now the CDFs of some measures in this set do not coincide with the CDF drawn in bright blue on the interval $[a^*, \overline{a}]$.

6 Convincing Groups with Diverse Beliefs

Consider a problem of convincing a large group to take an action or to refrain from taking an action. The members of the group hold diverse prior beliefs. It is known that all members of the group have beliefs that put a mass of at least β on states above α. Every such belief is represented in the group, so that for every prior that puts a mass of at least β on states above α, we can find a group member which holds this belief.

The following is an example of a situation to which the above logic may apply. In 1995, the Clinton administration introduced the Project Excellence in Leadership (XL) that provided opportunities for firms to engage in experimentation with the goal of developing technological innovations leading to greater environmental benefits. The firms were to reach an agreement on their experimental projects with the Environmental Protection Agency (EPA) and stakeholders. The firms' concern about litigation by environmental protection groups was a salient feature of the implementation of Project XL: the firms were conscious of the need to design their experiments in such a way as to prevent costly litigation (Fiorino 2006: 141).

This strategic situation can be described as follows. A company wishes to convince environmental NGOs that the technology it plans to use is sufficiently environmentally friendly. The company can design an experiment that will reveal the characteristics of the technology with some probability. The NGOs hold diverse prior beliefs about environmental friendliness of the technology. If an NGO believes that the technology is sufficiently harmful, the NGO will challenge the company's use of the technology in court. In order to prevent litigation, the company needs to convince all the NGOs that the technology is sufficiently safe.

While in the model of persuasion with unknown beliefs Nature chooses the receiver's prior after the sender commits to a signal structure but before the signal is realized, the model where the sender aims to convince all members of a group with diverse beliefs can be interpreted as a model where Nature chooses the receiver's prior after the signal is realized. It follows from the discussion in section 3.2 that the optimal signal is the same in both models.

7 Sketch of the Proof

In this section I sketch the proof of Theorem 1.

Given a prior F with a finite support and a signal structure π, I denote by $R(F, \pi)$ the finite set of indices of signal realizations such that the receiver with prior F takes action 1 after seeing these signal realizations. I show in lemma 1 that for all signal structures π there exists a feasible receiver's prior F such that the sum of the probabilities of signal realizations with indices in $R(F, \pi)$ given each state $\omega \in [a_0, a^*)$ is bounded above by $\min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\}$. The proof of the lemma uses a fixed-point argument outlined in greater detail below.

Next, I show in lemma 2 that if there exists a signal π that achieves this bound in all states $\omega \in [a_0, a^*)$ for the receiver's prior F and, moreover, if Nature prefers to choose the receiver's prior F whenever the sender chooses π, then π maximizes the sender's payoff among all signals.

Then I show in lemmas 3 - 11 that the signal π defined in Theorem 1 satisfies the conditions in lemma 2 and thus maximizes the sender's payoff among all signals.

7.1 Complicated Signals do not Benefit the Sender

In the remainder of the section I provide an outline of the proof of the statement that for all signals π (with a finite number of realizations conditional on each state) there exists a feasible receiver's prior F such that the sum of the probabilities of signal realizations with indices in $R(F, \pi)$ given each state $\omega \in [a_0, a^*)$ is bounded above by $\min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\}$. Because this bound is achieved by a simple signal that has exactly one signal realization which recommends action 1, the lemma shows that complicated signals do not benefit the sender.

The proof consists in constructing, given a signal structure π, a feasible prior F satisfying the requirement above.

7.2 The Support of the Receiver's Prior

I first construct a finite support for the receiver's prior as follows. I denote by R the finite set of indices of signal realizations that have a strictly positive probability under the signal structure π given that the state is α. I define Y as the set of states in $[a_0, a^*)$ such that there exists at least one signal realization with an index in R that has a strictly positive probability given the state.

For each state $\omega \in [a_0, a^*)$, I define $C(\omega)$, the set of all subsets A of indices in R such that the sum of the probabilities of signals with indices in A is strictly greater than $\min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\}$. I let the sets $Y_1, \ldots, Y_\kappa$ denote the coarsest partition of Y such that for each set in the partition the signal realizations with indices in R that have strictly positive probability under π are the same and and the sets $C(\omega)$ are the same for each $\omega \in Y_i$.

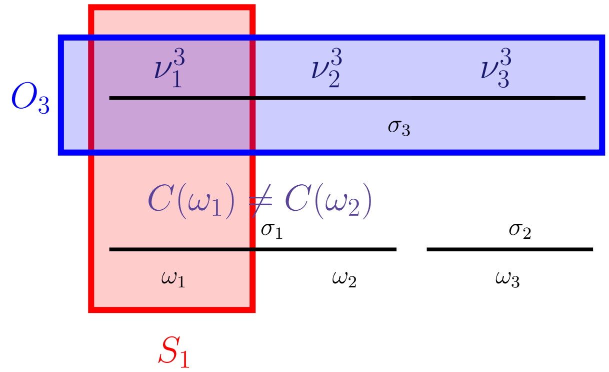
I choose exactly one state ω_i from each element Y_i of the partition and denote the resulting set of states by W. Finally, I let O_i denote the set of states in W such that the probability of signal realization σ_i given each such state is strictly positive.

$\cup_{\omega \in W} \omega \cup \alpha$ is the support of the receiver's prior.

7.3 The Receiver's Prior

I next construct the receiver's prior. Given a state ω_l, I let S_l denote the set of signal realization indices such that the signal realizations with these indices have a strictly positive probability under π given that the state is ω_l.

Figure 6: The Construction of the Receiver's Prior



I suppose that Nature chooses the receiver's prior F that places a mass of

$$(1-\beta) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha) \right]$$

on each state $\omega_l \in W$ and places a mass of β on α. The weights $\nu = \{\nu_m^i\}_{i \in R, m \in O_i}$ satisfy $\nu_l^k \in [0,1]$ and $\sum_{l:\omega_l \in O_k} \nu_l^k = 1$.

The weights $\nu = \{\nu_m^i\}_{i \in R, m \in O_i}$ are defined as the fixed point of the mapping $H : \times_{i \in R} \Delta(O_i) \to \times_{i \in R} \Delta(O_i)$

$$H_{im}(\nu) = \frac{\pi(\sigma_i|\omega_m)(a^* - \omega_m)\left[\sum_{k \in S_m \cap R} \nu_m^k \pi(\sigma_k|\alpha)\right]}{\sum_{l:\omega_l \in O_i} \pi(\sigma_i|\omega_l)(a^* - \omega_l)\left[\sum_{k \in S_l \cap R} \nu_l^k \pi(\sigma_k|\alpha)\right]}$$

Figure 6 illustrates the construction of the prior. The black intervals denote the supports of different signal realizations. ω_1, ω_2 and ω_3 is the set W. The red selection in figure 6 shows S_1, the set of all signal indices i such that $\pi(\sigma_i|\omega_1) > 0$. The blue selection in figure 6 shows O_3, the set of all state indices j such that $\pi(\sigma_3|\omega_j) > 0$. ν_j^i is the weight corresponding to the state ω_j and signal σ_i. Figure 6 shows the weights ν_1^3, ν_2^3 and ν_3^3 corresponding to the signal σ_3 and states ω_1, ω_2 and ω_3 respectively.

The partition induced by intersecting the supports of the signals has two elements, one containing ω_1 and ω_2, and the other one containing ω_3. We obtain a finer partition by grouping the states ω for which $C(\omega)$ is the same in one partition element. In the example in figure 6 the finer partition has three elements, each containing ω_1, ω_2 and ω_3 respectively.

8 Related Literature

The present paper is related to two strands of literature: the literature on Bayesian persuasion and the literature on robust mechanism design.

Early papers on Bayesian persuasion include Brocas and Carillo (2007), Ostrovsky and Schwarz (2010) and Rayo and Segal (2010). Kamenica and Gentzkow (2011) introduce a general model of Bayesian persuasion and provide a characterization of the sender's value in this model. Alonso and Camara (2016*a*) consider a model of Bayesian persuasion where the sender and the receiver have heterogeneous priors. Alonso and Camara (2016*b*) study the persuasion of voters.

Kolotilin (2017) and Kolotilin et al. (2017), among others, consider models of Bayesian persuasion with an infinite state space. Kolotilin et al. (2017) prove an equivalence result concerning public and private persuasion of a privately informed receiver. Perez-Richet (2014) and Hedlund (2017) consider a Bayesian persuasion model with a privately informed sender. Ely (2017), among others, studies a dynamic Bayesian persuasion model.

In all Bayesian persuasion models to date, either the prior of the receiver is known or, if private information is present, there is a known distribution over the receiver's beliefs. In contrast, in the present paper, there is a set of possible priors of the receiver but the sender does not have a distribution over the receiver's beliefs. Instead, the sender acts as if Nature is choosing the receiver's prior from this set in order to minimize the sender's payoff.

The literature on robust mechanism design studies the design of optimal mechanisms in the environments where the designer does not know the distribution of agents' types and designs a mechanism to maximize his utility in the worst case scenario.

Carroll (2015) shows that linear contracts have attractive robustness properties. Carroll (2017) shows that the robust multidimensional screening mechanism sells objects separately. Carrasco et al. (2017) study robust selling mechanisms under moment conditions on the distribution of the buyer's types. Chassang (2013) considers robust dynamic contracts. To the best of the author's knowledge, the present paper is the first one to consider a model of robust Bayesian persuasion.

9 Conclusion

This paper analyzes a model of Bayesian persuasion with unknown beliefs and characterizes the optimal signal. The optimal signal recommends the high action with a strictly positive probability in every state, with the probability that the high action is recommended changing gradually as the state changes. The optimal signal may also reveal the state with some probability.

These distinct features of the robust signal can potentially be useful for empirically distinguishing between the senders who know the prior of the receiver and those who do not have this knowledge.

Appendix

B Notation and Definitions

B.1 Measures

When a measure is referred to in this paper, a measure should be understood to mean a Borel probability measure. Moreover, any measurable spaces referred to in this paper should be understood to be equipped with a Borel sigma-algebra.

By Theorem 12.4 in Billingsley (2012: 185), if F is a non-decreasing, right-continuous real-valued function on $\mathbb{R}$, then there exists a unique Borel measure μ such that $\mu((a, b]) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$. Given a CDF F, I use μ_F to denote the unique Borel probability measure that corresponds to the CDF F.

B.2 Signal Structures

Given Borel subsets A, $B \subseteq \Omega$, let $\mathcal{K}(A, B)$ denote the set of kernels $P : A \times \mathcal{B}(B) \to [0, 1]$ such that for every $a \in A$, $C \mapsto P(a, C)$ is a probability measure on B and for every $C \in \mathcal{B}(B)$, $a \mapsto P(a, C)$ is $\mathcal{B}(A)$-measurable.[10]

Let $\mathcal{D}$ denote the set of signal structures that have a finite number of signal realizations conditional on each state. That is, let

$$\mathcal{D} = \{P \in \mathcal{K}(\Omega, M) : |\operatorname{supp} P(\omega, \cdot)| < \infty \text{ for all } \omega \in \Omega\}$$

B.3 Priors with Finite Support

Define

$$\mathcal{G}_\gamma = \{F \in \varphi : \mu_F = (1 - \gamma)\delta_\omega + \gamma\delta_{\omega'}, \omega \in [a_0, \alpha), \omega' \in [\alpha, \overline{a}]\}$$

Given $\gamma \in [0, 1]$, $\mathcal{G}_\gamma$ is the set of all CDFs F such that μ_F is a convex combination of a point mass on ω for some $\omega \in [a_0, \alpha)$ and a point mass on ω' for some $\omega' \in [\alpha, \overline{a}]$ with

[10] Recall that $\mathcal{B}(A)$ and $\mathcal{B}(B)$ denote the Borel sigma-algebras on A and B respectively.

weights $1-\gamma$ and γ.

Moreover, let

$$\mathcal{F} = \{\mathcal{G}_\gamma\}_{\gamma\in[\beta,1]}$$

$\mathcal{F}$ is the set of all CDFs F such that μ_F is a convex combination of a point mass on ω for some $\omega \in [a_0, \alpha)$ and a point mass on ω' for some $\omega' \in [\alpha, \overline{a}]$ with weights $1-\gamma$ and γ respectively for $\gamma \in [\beta, 1]$.

Define

$$\mathcal{F}_\alpha = \{F \in \varphi : \mu_F = (1-\beta)\mu_G + \beta\delta_\alpha, \operatorname{supp}\mu_G \subseteq [a_0, a^*), |\operatorname{supp}\mu_G| < \infty\}$$

$\mathcal{F}_\alpha$ is the set of all CDFs F that are a convex combination of a measure on $[a_0, a^*)$ with a finite support and a point mass on α with weights $1-\beta$ and β respectively.

B.4 Signal Indices and the Bound

Given a signal structure $\pi \in \mathcal{D}$ and a state ω_j, define

$$S_j = \{k : \pi(\sigma_k|\omega_j) > 0\}$$

S_j is the set of indices of signal realizations which have a strictly positive probability in the state ω_j under the signal structure π.

Define $b(\omega) = \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\}$ if $\omega \in [a_0, a^*)$ and $b(\omega) = 1$ if $\omega \in [a^*, \overline{a}]$.

B.5 Expectations

$E_{F,P}[\omega|\sigma]$ denotes the receiver's expectation of the random variable ω with distribution F after the receiver observes signal realization σ given that the signal structure (the kernel) is P. When no confusion can result, I suppress the dependence on P and denote this expectation by $E_F[\omega|\sigma]$.

B.6 Convergence

A sequence of measures $\{\mu^N\}_{N\in\mathbb{N}}$ is said to *converge weakly* to a measure μ if

$$\lim_{N\to\infty}\int_\Omega g(w)d\mu^N(w) = \int_\Omega g(w)d\mu(w)$$

for every bounded and continuous function g.

C Proofs

Lemma 1. *Fix a signal structure $\pi \in \mathcal{D}$. Denote by $\sigma_1, \ldots, \sigma_r$ the finite number of signal realizations that have a strictly positive probability given that the state is α. Define $R = \{1, \ldots, r\}$.*

Given $F \in \mathcal{F}_\alpha$, let $R(F,\pi) \subseteq R$ denote the finite number of signal indices such that $E_{F,\pi}[\omega|\sigma_i] \geq a^*$ for all $i \in R(F,\pi)$.

Then there exists $F \in \mathcal{F}_\alpha$ such that

$$\sum_{i\in R(F,\pi)} \pi(\sigma_i|\omega_m) \leq \frac{\beta(\alpha - a^*)}{(1-\beta)(a^* - \omega_m)}$$

for all $\omega_m \in [a_0, a^*)$.

Proof of lemma 1.

Without loss of generality, suppose that, for all $i \in R$, we have $\pi(\sigma_i|\omega) > 0$ for some $\omega \in [a_0, a^*)$ (if for some $I \subseteq R$ we have $\pi(\sigma_i|\omega) = 0$ for all $i \in I$ and for all $\omega \in [a_0, a^*)$, then it is enough to show that $\sum_{j\in R\setminus I} \pi(\sigma_j|\omega_m) \leq \frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega_m)}$ for all $\omega_m \in [a_0, a^*)$).

Step 1: Partitioning the set $[a_0, a^*)$.

For each $\omega \in [a_0, a^*)$, define

$$C(\omega) = \{A \subseteq R : \sum_{i\in A} \pi(\sigma_i|\omega) > b(\omega)\}$$

$C(\omega)$ is the set of all sets A of indices in R such that the sum of the probabilities of signals with indices in A is strictly greater than $b(\omega)$.

Define a finite number[11] of sets $Y_1, \ldots, Y_\kappa$ such that $Y_i \subseteq [a_0, a^*)$ for all $i \in \{1, \ldots, \kappa\}$ as follows:

1. ω, $\omega' \in Y_i$ implies that, for all $j \in R$, if $\pi(\sigma_j|\omega) > 0$, then $\pi(\sigma_j|\omega') > 0$;
2. ω, $\omega' \in Y_i$ implies that $C(\omega) = C(\omega')$;
3. $\omega_i \in Y_i$, $\omega_j \in Y_j$ for $i \neq j$ implies that
   (a) either there exists a signal realization σ_l such that $l \in R$ and either $\pi(\sigma_l|\omega_i) > 0$, $\pi(\sigma_l|\omega_j) = 0$ or $\pi(\sigma_l|\omega_j) > 0$ and $\pi(\sigma_l|\omega_i) = 0$ or
   (b) $C(\omega_i) \neq C(\omega_j)$

Define

$$Y = \{\omega \in [a_0, a^*) : \text{ there exists } i \in R : \pi(\sigma_i|\omega) > 0\}$$

Y is the subset of $[a_0, a^*)$ such that for every state ω in Y there exists a signal realization σ_i with an index $i \in R$ that has a strictly positive probability in state ω given the signal structure π. The sets $Y_1, \ldots, Y_\kappa$ form the coarsest partition of Y such that for each set Y_i in the partition the signal realizations in R that have a strictly positive probability under π given $\omega \in Y_i$ are the same and the sets $C(\omega)$ are the same for each $\omega \in Y_i$.

Next, we choose a finite set of states as follows. From each Y_i, $i \in \{1, \ldots, \kappa\}$, choose exactly one ω_i. Let W denote the set of states $\{\omega_i\}_{i=1}^{\kappa}$ chosen in this manner.

Step 2: Some notation and reference states.

Define

$$J = \{j : \omega_j \in W\}$$

J is the set of indices of the states in W.

Define also

$$O_i = \{\omega \in W : \pi(\sigma_i|\omega) > 0\}$$

O_i is the set of states in W such that the probability of signal realization σ_i given each such state is strictly positive.

[11] The number of sets is finite because R is finite.

Step 3: Nature's strategy.

Suppose that Nature chooses the receiver's prior μ_F that, for all $i \in R$, places a mass of $\nu_l^i \pi(\sigma_i|\alpha)(1-\beta)$ on each $\omega_l \in O_i$ such that $\nu_l^i \in [0,1]$ for all $l \in J$ satisfying $\omega_l \in O_i$ and $\sum_{l:\omega_l \in O_i} \nu_l^i = 1$ and places a mass of β on α.

Note that, by construction, $O_i \neq \varnothing$ for all $i \in R$. Then we have

$$\sum_{i=1}^{r} \sum_{l:\omega_l \in O_i} \nu_l^i \pi(\sigma_i|\alpha)(1-\beta) = \sum_{i=1}^{r} \pi(\sigma_i|\alpha)(1-\beta) = 1-\beta$$

so Nature places the mass of $1-\beta$ on states in $[a_0, a^*)$. Observe that we have $\mu_F \in \mathcal{F}_\alpha$.

Observe also that the Nature's strategy as defined above implies that, given $l \in J$, Nature puts a mass of $(1-\beta)\sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]$ on ω_l.

Step 4: $E_F[\omega|\sigma_i] \geq a^*$ given Nature's strategy.

Fix $i \in R$. Then, because Nature puts a mass of β on state α and puts a mass of $(1-\beta)\sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]$ on each ω_l such that $l \in J$, we have

$$\int_\Omega w\pi(\sigma_i|w)dF(w) = \sum_{\omega_l \in O_i} \omega_l \pi(\sigma_i|\omega_l)(1-\beta) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right] + \alpha\beta\pi(\sigma_i|\alpha)$$

$$\begin{aligned}\int_\Omega \pi(\sigma|w)dF(w) = \quad & \beta\pi(\sigma_i|\alpha) + \int_{[a_0,\alpha)} \pi(\sigma_i|w)dF(w) = \\ & \beta\pi(\sigma_i|\alpha) + \sum_{\omega_l \in O_i} \pi(\sigma_i|\omega_l)(1-\beta) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]\end{aligned}$$

$E_F[\omega|\sigma_i] \geq a^*$ is equivalent to

$$\frac{\alpha\beta\pi(\sigma_i|\alpha) + \sum_{\omega_l \in O_i} \omega_l \pi(\sigma_i|\omega_l)(1-\beta) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]}{\beta\pi(\sigma_i|\alpha) + \sum_{\omega_l \in O_i} \pi(\sigma_i|\omega_l)(1-\beta) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]} \geq a^*$$

Equivalently,

$$1 \leq (\alpha - a^*) \frac{\beta}{1-\beta} \frac{\pi(\sigma_i|\alpha)}{\sum_{l:\omega_l \in O_i} \pi(\sigma_i|\omega_l)(a^* - \omega_l) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]}$$

Given $\omega_m \in O_i$, we multiply both sides by $\pi(\sigma_i|\omega_m)$ to obtain

$$\pi(\sigma_i|\omega_m) \leq (\alpha - a^*) \frac{\beta}{1-\beta} \frac{\pi(\sigma_i|\alpha)\pi(\sigma_i|\omega_m)}{\sum_{l:\omega_l \in O_i} \pi(\sigma_i|\omega_l)(a^* - \omega_l) \sum_{k \in S_l \cap R} \left[\nu_l^k \pi(\sigma_k|\alpha)\right]} \tag{2}$$

Step 5: Defining the receiver's prior.

Given $m \in J$, let

$$g_m = \sum_{k \in S_m \cap R} \nu_m^k \pi(\sigma_k|\alpha)$$

Choose ν_m^i such that

$$\frac{\pi(\sigma_i|\omega_m)(a^* - \omega_m)}{\sum_{l:\omega_l \in O_i} \pi(\sigma_i|\omega_l)(a^* - \omega_l)} \frac{\pi(\sigma_i|\alpha)}{g_l} = \frac{\nu_m^i \pi(\sigma_i|\alpha)}{g_m} \tag{3}$$

This implies that we set

$$\nu_m^i = \frac{\pi(\sigma_i|\omega_m)(a^* - \omega_m) g_m}{\sum_{l:\omega_l \in O_i} \pi(\sigma_i|\omega_l)(a^* - \omega_l) g_l} \tag{4}$$

Step 6: The receiver's prior is well-defined.

Next, we will show that the collection of weights $\{\nu_m^i\}_{i \in R, m \in O_i}$ is well-defined. In particular, we will show that there exists a collection of weights satisfying (4), $\nu_m^i \in [0,1]$ for all $i \in R$ and $m \in J$ such that $\omega_m \in O_i$, and satisfying $\sum_{m:\omega_m \in O_i} \nu_m^i = 1$.

Observe that

$$\sum_{m:\omega_m\in O_i} \nu_m^i = \sum_{m:\omega_m\in O_i} \frac{\pi(\sigma_i|\omega_m)(a^*-\omega_m)g_m}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_m)g_l} = \frac{\sum_{m:\omega_m\in O_i}\pi(\sigma_i|\omega_m)(a^*-\omega_m)g_m}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)g_l} = 1$$

This shows that $\sum_{m:\omega_m\in O_i}\nu_m^i = 1$.

Fix $m\in J$ and $i\in S_m\cap R$. Substituting in the formulas for g_m and g_l, (4) is equivalent to

$$\nu_m^i = \frac{\pi(\sigma_i|\omega_m)(a^*-\omega_m)\big[\sum_{k\in S_m\cap R}\nu_m^k\pi(\sigma_k|\alpha)\big]}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)\big[\sum_{k\in S_l\cap R}\nu_l^k\pi(\sigma_k|\alpha)\big]} \tag{5}$$

Let

$$T = \times_{i\in R}\Delta(O_i)$$

where $\Delta(O_i)$ denotes the simplex over O_i and $\times$ denotes the Cartesian product.

Let $\nu = \{\nu_m^i\}_{i\in R, m\in O_i}$. Define a mapping $H: T\to T$ by

$$H_{im}(\nu) = \frac{\pi(\sigma_i|\omega_m)(a^*-\omega_m)\big[\sum_{k\in S_m\cap R}\nu_m^k\pi(\sigma_k|\alpha)\big]}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)\big[\sum_{k\in S_l\cap R}\nu_l^k\pi(\sigma_k|\alpha)\big]}$$

Then we can write (5) as

$$\nu_m^i = H_{im}(\nu)$$

Thus the weights $\{\nu_m^i\}_{i\in R, m\in O_i}$ are defined by the equation

$$\nu = H(\nu) \tag{6}$$

Observe that $\Delta(O_i)$ is a compact and convex set. Because the Cartesian product of convex sets is convex and the Cartesian product of compact sets is compact, this implies that $T = \times_{i\in R}\Delta(O_i)$ is a compact and convex set. Thus $H(\cdot)$ is a continuous self-map on a compact and convex set T. Then the Brouwer fixed point theorem implies that H has a fixed point.

Step 7: A Useful Equality.

We have

$$
\begin{aligned}
\sum_{i\in S_m\cap R}\frac{\pi(\sigma_i|\alpha)\pi(\sigma_i|\omega_m)}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)g_l} &= \sum_{i\in S_m\cap R}\frac{\nu^i_m\pi(\sigma_i|\alpha)}{(a^*-\omega_m)g_m}\\
&= \frac{1}{a^*-\omega_m}\sum_{i\in S_m\cap R}\frac{\nu^i_m\pi(\sigma_i|\alpha)}{g_m}\\
&= \frac{1}{a^*-\omega_m}
\end{aligned}
\tag{7}
$$

where the first equality follows because dividing both sides of (3) by $a^*-\omega_m$ yields $\frac{\pi(\sigma_i|\alpha)\pi(\sigma_i|\omega_m)}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)g_l}=\frac{\nu^i_m\pi(\sigma_i|\alpha)}{(a^*-\omega_m)g_m}$ and the last equality holds because $\sum_{i\in S_m\cap R}\frac{\nu^i_m\pi(\sigma_i|\alpha)}{g_m}=1$ by definition of g_m.

Step 8: Conclusion of the proof.

Suppose that Nature chooses the weights defined by the fixed-point equation (6). By construction, since $E_F[\omega|\sigma_i]\geq a^*$ for all $i\in R(F,\pi)$, we have that (2) holds for all $m\in J$ and for all $i\in R(F,\pi)$.

Observe that (2) is equivalent to

$$
\pi(\sigma_i|\omega_m)\leq(\alpha-a^*)\frac{\beta}{1-\beta}\frac{\pi(\sigma_i|\alpha)\pi(\sigma_i|\omega_m)}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)g_l} \tag{8}
$$

Then we have

$$
\begin{aligned}
\sum_{i\in S_m\cap R(F,\pi)}\pi(\sigma_i|\omega_m) &\leq (\alpha-a^*)\frac{\beta}{1-\beta}\sum_{i\in S_m\cap R(F,\pi)}\frac{\pi(\sigma_i|\alpha)\pi(\sigma_i|\omega_m)}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)g_l}\\
&\leq (\alpha-a^*)\frac{\beta}{1-\beta}\sum_{i\in S_m\cap R}\frac{\pi(\sigma_i|\alpha)\pi(\sigma_i|\omega_m)}{\sum_{l:\omega_l\in O_i}\pi(\sigma_i|\omega_l)(a^*-\omega_l)g_l}
\end{aligned}
\tag{9}
$$

where the first inequality follows from summing (8) over $R(F,\pi)$ and the second inequality follows from the fact that $R(F,\pi)\subseteq R$.

Thus (9) and (7) imply that

$$
\sum_{i\in R(F,\pi)}\pi(\sigma_i|\omega_m)\leq(\alpha-a^*)\frac{\beta}{1-\beta}\frac{1}{a^*-\omega_m} \tag{10}
$$

Suppose for the sake of contradiction that for some $w_l \in [a_0, a^*)$ we had $\sum_{i \in R(F,\pi)} \pi(\sigma_i|w_l) > (\alpha - a^*)\frac{\beta}{1-\beta}\frac{1}{a^*-\omega_l}$. By definition of $C(\omega_l)$, this is equivalent to saying that $R(F, \pi) \in C(\omega_l)$.

Let Y_m denote the element of the partition $Y_1, \ldots, Y_\kappa$ such that $w_l \in Y_m$. Then, because $C(\omega) = C(\omega')$ for all ω, $\omega' \in Y_m$, it must be the case that $R(F, \pi) \in C(\omega_m)$. By definition of $C(\omega_m)$, this is equivalent to saying that $\sum_{i \in R(F,\pi)} \pi(\sigma_i|\omega_m) > (\alpha - a^*)\frac{\beta}{1-\beta}\frac{1}{a^*-\omega_m}$. However, this contradicts (10).

Therefore, we must have $\sum_{i \in R(F,\pi)} \pi(\sigma_i|\omega_m) \leq (\alpha - a^*)\frac{\beta}{1-\beta}\frac{1}{a^*-\omega_m}$ for all $\omega_m \in [a_0, a^*)$. ■

Lemma 2. *Consider a signal $\pi \in \mathcal{D}$. Let $R(\pi) = \{1, \ldots, r\}$ denote the finite subset of signal realization indices such that $\sigma_1, \ldots, \sigma_r$ have a strictly positive probability under π given that the state is α. Suppose that $\pi \in \mathcal{D}$ is such that*

1. *$E_{F,\pi}[\omega|\sigma_i] \geq a^*$ for all $i \in R(\pi)$,*
2. *for all $F \in \mathcal{F}_\alpha$, for all $\omega_m \in [a_0, a^*)$ we have $\sum_{i \in R(\pi)} \pi(\sigma_i|\omega_m) = \min\left\{\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega_m}, 1\right\}$,*
3. *if the sender chooses π, then Nature weakly prefers to choose $F \in \mathcal{F}_\alpha$.*

Then π maximizes the sender's payoff among all signals in $\mathcal{D}$.

Proof of lemma 2.

Given $\pi \in \mathcal{D}$ and $G \in \mathcal{F}_\alpha$, let $R(G, \pi)$ denote the subset of $R(\pi)$ such that $E_{G,\pi}[\omega|\sigma_i] \geq a^*$ for all $i \in R(G, \pi)$. Suppose for the sake of contradiction that there is another signal $\pi' \in \mathcal{D}$ giving the sender a payoff strictly higher than that of signal π.

Suppose that after the sender chooses π', Nature chooses $F \in \mathcal{F}_\alpha$ such that

$$\sum_{i \in R(F,\pi')} \pi'(\sigma_i|\omega_m) \leq \min\left\{\frac{\beta}{1-\beta}\frac{\alpha - a^*}{a^* - \omega_m}, 1\right\} \tag{11}$$

for all $\omega_m \in [a_0, a^*)$ (the fact that such F exists follows from lemma 1).

Note that if Nature chooses $F \in \mathcal{F}_\alpha$, then π guarantees a payoff of 1 to the sender in states $\omega \in [a^*, \overline{a}]$. Thus if the sender chooses π' and Nature chooses $F \in \mathcal{F}_\alpha$, then the sender's payoff U_1 is no greater than $\tilde{U}_1 = \int_{[a_0,a^*)} \left[\sum_{i \in R(F,\pi')} \pi'(\sigma_i|\omega_m)\right] dF_s(\omega_m) + \int_{[a^*,\overline{a}]} dF_s(\omega_m)$, while if the sender chooses π and Nature chooses $F \in \mathcal{F}_\alpha$, then the sender's payoff is $U_2 = \int_{[a_0,a^*)} \left[\sum_{i \in R(F,\pi)} \pi(\sigma_i|\omega_m)\right] dF_s(\omega_m) + \int_{[a^*,\overline{a}]} dF_s(\omega_m) = \int_{[a_0,a^*)} \left[\sum_{i \in R(\pi)} \pi(\sigma_i|\omega_m)\right] dF_s(\omega_m) +$

$\int_{[a^*,\overline{a}]} dF_s(\omega_m) = \int_{[a_0,a^*)} \min\left\{\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega_m}, 1\right\} dF_s(\omega_m) + \int_{[a^*,\overline{a}]} dF_s(\omega_m)$ (the second equality in the expression for U_2 follows from the fact that $E_{F,\pi}[\omega|\sigma_i] \geq a^*$ for all $i \in R(\pi)$, and the third equality in the expression for U_2 follows from the fact that $\sum_{i\in R(\pi)} \pi(\sigma_i|\omega_m) = \min\left\{\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega_m}, 1\right\}$ for all $\omega_m \in [a_0, a^*)$).

Observe that (11) implies that $\int_{[a_0,a^*)} \left[\sum_{i\in R(F,\pi')} \pi'(\sigma_i|\omega_m)\right] dF_s(\omega_m) \leq \int_{[a_0,a^*)} \left[\sum_{i\in R(\pi)} \pi(\sigma_i|\omega_m)\right] dF_s(\omega_m)$. Therefore, $\tilde{U}_1 \leq U_2$.

Moreover, if after the sender chooses π', Nature chooses $F' \neq F$, then it must be the case that the sender's payoff U_3 given π' and F' satisfies $U_3 \leq U_1$. If the sender chooses π and Nature chooses $F' \neq F$, then, because Nature weakly prefers to choose F given that the sender chooses π, the sender's payoff $\tilde{U}_2$ given π and F' must satisfy $U_2 \leq \tilde{U}_2$.

Then we have $U_3 \leq U_1 \leq \tilde{U}_1 \leq U_2 \leq \tilde{U}_2$. Thus π yields a weakly higher payoff to the sender than π', a contradiction. ■

Lemma 3. *Consider a signal $\pi \in \mathcal{D}$ such that π has exactly one realization σ satisfying $\pi(\sigma|\omega') > 0$ for $\omega' \in [\alpha, \overline{a}]$ and $\pi(\sigma|\omega) > 0$ for $\omega \in [a_0, a^*)$. Then we have $E_F[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{G}_\gamma$, $\gamma \in [\beta, 1]$, if and only if*

$$\pi(\sigma|\omega) \leq \pi(\sigma|\omega')\frac{\gamma}{1-\gamma}\frac{\omega'-a^*}{a^*-\omega}$$

for all $\omega \in [a_0, a^*)$, $\omega' \in [\alpha, \overline{a}]$.

Proof of lemma 3.

Fix $\gamma \in [\beta, 1]$. If Nature puts a mass of $1-\gamma$ on some $\omega \in [a^*, \alpha)$ and puts a mass of γ on $\omega' \in [\alpha, \overline{a}]$, then it is easy to see that we have $E_F[\omega|\sigma] \geq a^*$ for all π satisfying the conditions in the lemma.

Therefore, fix $\omega \in [a_0, a^*)$ and suppose that Nature places a mass of $1-\gamma$ on ω. Suppose also that Nature places a mass of γ on $\omega' \in [\alpha, \overline{a}]$. Then $\int_\Omega w\pi(\sigma|w)dF(w) = \omega\pi(\sigma|\omega)(1-\gamma) + \omega'\gamma\pi(\sigma|\omega')$ and $\int_\Omega \pi(\sigma|w)dF(w) = \gamma\pi(\sigma|\omega') + \int_{[a_0,a^*)} \pi(\sigma|w)dF(w) = \gamma\pi(\sigma|\omega') + \pi(\sigma|\omega)(1-\gamma)$.

$E_F[\omega|\sigma] \geq a^*$ is equivalent to $\frac{\int_\Omega w\pi(\sigma|w)dF(w)}{\int_\Omega \pi(\sigma|w)dF(w)} \geq a^*$, which, in turn, is equivalent to $\frac{\pi(\sigma|\omega')\omega'\gamma+\omega\pi(\sigma|\omega)(1-\gamma)}{\gamma\pi(\sigma|\omega')+\pi(\sigma|\omega)(1-\gamma)} \geq a^*$. Simplifying, we obtain $\pi(\sigma|\omega')\omega'\gamma + \omega\pi(\sigma|\omega)(1-\gamma) \geq a^*\gamma\pi(\sigma|\omega') + a^*\pi(\sigma|\omega)(1-\gamma)$, which is equivalent to $\pi(\sigma|\omega) \leq \pi(\sigma|\omega')\frac{\gamma}{1-\gamma}\frac{\omega'-a^*}{a^*-\omega}$. ■

Lemma 4. *Consider the signal $q \in \mathcal{D}$ described in lemma 5. We have $E_{F,q}[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{F}$.*

Proof of lemma 4.

By lemma 3, $E_F[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{G}_\gamma$, $\gamma \in [\beta, 1]$, is equivalent to $q(\sigma|\omega) \leq q(\sigma|\omega')\frac{\gamma}{1-\gamma}\frac{\omega'-a^*}{a^*-\omega}$ for all $\omega \in [a_0, a^*)$, $\omega' \in [\alpha, \overline{a}]$. Because $\gamma \in [\beta, 1]$ and $\gamma \mapsto \frac{\gamma}{1-\gamma}$ is increasing, if the preceding inequality holds for $\gamma = \beta$, then it holds for any $\gamma \in (\beta, 1]$. Therefore, we have $E_F[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{F}$ if and only if $E_F[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{G}_\beta$.

By lemma 3, $E_F[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{G}_\beta$ is equivalent to $q(\sigma|\omega) \leq q(\sigma|\omega')\frac{\beta}{1-\beta}\frac{\omega'-a^*}{a^*-\omega}$ for all $\omega \in [a_0, a^*)$, $\omega' \in [\alpha, \overline{a}]$.

Note that q described in lemma 5 satisfies $q(\sigma|\omega) = \min\left\{\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega}, 1\right\}$ for all $\omega \in [a_0, a^*)$. Then $q(\sigma|\omega) \leq q(\sigma|\omega')\frac{\beta}{1-\beta}\frac{\omega'-a^*}{a^*-\omega}$ for all $\omega \in [a_0, a^*)$, $\omega' \in [\alpha, \overline{a}]$ is equivalent to $\min\left\{\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega}, 1\right\} \leq \min\left\{q(\sigma|\omega')\frac{\beta}{1-\beta}\frac{\omega'-a^*}{a^*-\omega}, 1\right\}$ for all $\omega \in [a_0, a^*)$, $\omega' \in [\alpha, \overline{a}]$.

To show that the preceding inequality holds, it is enough to show that $\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega} \leq q(\sigma|\omega')\frac{\beta}{1-\beta}\frac{\omega'-a^*}{a^*-\omega}$ for all $\omega \in [a_0, a^*)$, $\omega' \in [\alpha, \overline{a}]$. This is equivalent to $q(\sigma|\omega') \geq \frac{\alpha-a^*}{\omega'-a^*}$ for all $\omega' \in [\alpha, \overline{a}]$. Therefore, we need $q(\sigma|\omega') \in \left[\frac{\alpha-a^*}{\omega'-a^*}, 1\right]$ for all $\omega' \in [\alpha, \overline{a}]$. Noting that q described in lemma 5 satisfies this concludes the proof. ■

Lemma 5. *Define*

$$q(\sigma|\omega) = \begin{cases} \frac{\alpha-a^*}{\omega-a^*} \text{ if } \omega \in [\alpha, \overline{a}] \\ 0 \text{ if } \omega \in (a^*, \alpha) \\ \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} \text{ if } \omega \in [a_0, a^*] \end{cases}$$

$$q(\sigma_w|\omega) = \begin{cases} 1 - \frac{\alpha-a^*}{\omega-a^*} \text{ if } \omega = w \text{ and } \omega \in [\alpha, \overline{a}] \\ 1 \text{ if } w = \omega \text{ and } \omega \in (a^*, \alpha) \\ 1 - \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} \text{ if } w = \omega \text{ and } \omega \in [a_0, a^*] \end{cases}$$

Suppose that π is as defined in lemma 12. For all $F \in \varphi$, if $E_{F,q}[\omega|\sigma] \geq a^*$, then $E_{F,\pi}[\omega|\sigma] \geq a^*$.

Proof of lemma 5.

We want to show that if $\frac{\int_{[a_0,a^*]} wq(\sigma|w)dF(w)+\int_{(a^*,\overline{a}]} wq(\sigma|w)dF(w)}{\int_{[a_0,a^*]} q(\sigma|w)dF(w)+\int_{(a^*,\overline{a}]} q(\sigma|w)dF(w)} \geq a^*$, then $\frac{\int_{[a_0,a^*]} w\pi(\sigma|w)dF(w)+\int_{(a^*,\overline{a}]} w\pi(\sigma|w)dF(w)}{\int_{[a_0,a^*]} \pi(\sigma|w)dF(w)+\int_{(a^*,\overline{a}]} \pi(\sigma|w)dF(w)} \geq a^*$. The first inequality is equivalent to $\int_{[a_0,a^*]} wq(\sigma|w)dF(w) - a^* \int_{[a_0,a^*]} q(\sigma|w)dF(w) \geq a^* \int_{(a^*,\overline{a}]} q(\sigma|w)dF(w) - \int_{(a^*,\overline{a}]} wq(\sigma|w)dF(w)$ and the second inequality is equivalent to $\int_{[a_0,a^*]} w\pi(\sigma|w)dF(w) - a^* \int_{[a_0,a^*]} \pi(\sigma|w)dF(w) \geq a^* \int_{(a^*,\overline{a}]} \pi(\sigma|w)dF(w) - \int_{(a^*,\overline{a}]} w\pi(\sigma|w)dF(w)$.

Observe that we can write these inequalities as $\int_{[a_0,a^*]} q(\sigma|w)(w - a^*)dF(w) \geq \int_{(a^*,\overline{a}]} q(\sigma|w)(a^* - w)dF(w)$ and $\int_{[a_0,a^*]} \pi(\sigma|w)(w - a^*)dF(w) \geq \int_{(a^*,\overline{a}]} \pi(\sigma|w)(a^* - w)dF(w)$ respectively.

We have

$$\begin{aligned} \int_{[a_0,a^*]} \pi(\sigma|w)(w - a^*)dF(w) &\geq \int_{[a_0,a^*]} q(\sigma|w)(w - a^*)dF(w) \\ &\geq \int_{(a^*,\overline{a}]} q(\sigma|w)(a^* - w)dF(w) \\ &\geq \int_{(a^*,\overline{a}]} \pi(\sigma|w)(a^* - w)dF(w) \end{aligned} \tag{12}$$

where the first inequality follows because $w - a^* \leq 0$ for $w \in [a_0, a^*]$ and $\pi(\sigma|w) \leq q(\sigma|w)$ for $w \in [a_0, a^*]$, the second inequality follows from the fact that $\int_{[a_0,a^*]} \pi(\sigma|w)(w - a^*)dF(w) \geq \int_{(a^*,\overline{a}]} \pi(\sigma|w)(a^* - w)dF(w)$, and the third inequality follows because $a^* - w \leq 0$ for $w \in (a^*, \overline{a}]$ and $\pi(\sigma|w) \geq q(\sigma|w)$ for $w \in (a^*, \overline{a}]$.

(12) implies that we have $\int_{[a_0,a^*]} \pi(\sigma|w)(w - a^*)dF(w) \geq \int_{(a^*,\overline{a}]} \pi(\sigma|w)(a^* - w)dF(w)$, as required. ■

Lemma 6. *If $a_i > 0$, $b_i > 0$ for all $i \in \{1, \ldots, N\}$ and $\frac{a_j}{b_j} \leq \frac{a_i}{b_i}$ for all $i \in \{1, \ldots, N\}$ and some $j \in \{1, \ldots, N\}$, then $\frac{a_j}{b_j} \leq \frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N b_i}$.*

Proof of lemma 6.

We have that $\frac{a_i}{b_i} \geq \frac{a_j}{b_j} = c$ for all $i \in \{1, \ldots, N\}$ and some $c > 0$. Then $a_i \geq cb_i$ for all $i \in \{1, \ldots, N\}$. Thus $\frac{\sum_{i=1}^N a_i}{\sum_{i=1}^N b_i} \geq \frac{\sum_{i=1}^N cb_i}{\sum_{i=1}^N b_i} = c\frac{\sum_{i=1}^N b_i}{\sum_{i=1}^N b_i} = c = \frac{a_j}{b_j}$. ■

Lemma 7. *Given $A_1 = [\alpha, \overline{a}]$, $A_2 = (a^*, \alpha)$, $A_3 = [a_0, a^*]$, for all μ_F such that $F \in C(\alpha, \beta)$, there exist probability measures μ_{G_1}, μ_{G_2} and μ_{G_3} such that $\mu_{G_1}(\Omega \setminus A_1) = 0$, $\mu_{G_2}(\Omega \setminus A_2) = 0$, $\mu_{G_3}(\Omega \setminus A_3) = 0$ and*

$$\mu_F = \alpha_1 \mu_{G_1} + \alpha_2 \mu_{G_2} + \alpha_3 \mu_{G_3}$$

for $\alpha_i = \mu_F(A_i)$, $i = 1, 2, 3$.[12]

[12]Moreover, because the measures μ_{G_i}, $i = 1, 2, 3$, are conditional probabilities, they are unique up to almost sure equality.

Proof of lemma 7.

Define the probability measures μ_{G_1}, μ_{G_2} and μ_{G_3} as follows. Fix $B_1 \in \mathcal{B}(A_1)$, $B_2 \in \mathcal{B}(A_2)$, $B_3 \in \mathcal{B}(A_3)$.[13] Observe that we must have $\mu_F(A_1) > 0$ because $F \in C(\alpha, \beta)$. Let $\mu_{G_1}(B_1) = \frac{\mu_F(B_1)}{\mu_F(A_1)}$. Let $\mu_{G_2}(B_2) = \frac{\mu_F(B_2)}{\mu_F(A_2)}$ if $\mu_F(A_2) > 0$ and let $\mu_{G_3}(B_3) = \frac{\mu_F(B_3)}{\mu_F(A_3)}$ if $\mu_F(A_3) > 0$. If $\mu_F(A_2) = 0$, let μ_{G_2} be arbitrary, and if $\mu_F(A_3) = 0$, let μ_{G_3} be arbitrary. ■

Lemma 8. *Any probability measure $\mu_F = \sum_{i=1}^{n} \gamma_i \delta_{\omega_i}$ such that $\gamma_i \in [0, 1]$, $\sum_{i=1}^{n} \gamma_i = 1$ and $\mu_F \in C(\alpha, \beta)$, can be written as $\mu_F = \sum_{i=1}^{N} \alpha_i \mu_i$ for some probability measures $\mu_i = (1 - \gamma)\delta_{\omega_i} + \gamma \delta_{v_i}$, $\gamma \in [\beta, 1]$, $\omega_i \in [a_0, \alpha)$, $v_i \in [\alpha, \overline{a}]$, and weights $\alpha_i \in [0, 1]$ such that $\sum_{i=1}^{N} \alpha_i = 1$ for some $N \in \mathbb{N}$.*

Proof of lemma 8.

Fix $\mu_F = \sum_{i=1}^{n} \gamma_i \delta_{\omega_i}$ such that $\gamma_i \in [0, 1]$, $\sum_{i=1}^{n} \gamma_i = 1$ and $\mu_F \in C(\alpha, \beta)$. An argument similar to the proof of lemma 7 implies that there exist measures μ_{G_1} and μ_{G_2} such that $\mu_{G_1}(\Omega \setminus [a_0, \alpha)) = 0$, $\mu_{G_2}(\Omega \setminus [\alpha, \overline{a}]) = 0$ and $\mu_F = (1 - \gamma)\mu_{G_1} + \gamma \mu_{G_2}$ for some $\gamma \in [\beta, 1]$.

Let $L = \{i : \omega_i \in \operatorname{supp} \mu_F, \omega_i \in [a_0, \alpha)\}$ and $H = \{i : \omega_i \in \operatorname{supp} \mu_F, \omega_i \in [\alpha, \overline{a}]\}$ denote the sets of the indices of the states that are in the support of μ_F and are respectively below and above α. For all $i \in L$, define $m_i = (1-\gamma)\delta_{\omega_i} + \gamma \mu_{G_2}$. Note that, because $\sum_{j \in L} \gamma_j = 1 - \gamma$ and $\sum_{i \in L} \frac{\gamma_i}{\sum_{j \in L} \gamma_j} = 1$, we have $\mu_F = \sum_{i \in L} \frac{\gamma_i}{\sum_{j \in L} \gamma_j} m_i$.

Given $i \in L$ and $j \in H$, let $\mu_{ij} = (1 - \gamma)\delta_{\omega_i} + \gamma \delta_{\omega_j}$. Note that, because $\sum_{k \in H} \gamma_k = \gamma$ and $\sum_{j \in H} \frac{\gamma_j}{\sum_{k \in H} \gamma_k} = 1$, we have $m_i = \sum_{j \in H} \frac{\gamma_j}{\sum_{k \in H} \gamma_k} \mu_{ij}$.

Thus $\mu_F = \sum_{i \in L} \frac{\gamma_i}{\sum_{l \in L} \gamma_l} \left(\sum_{j \in H} \frac{\gamma_j}{\sum_{k \in H} \gamma_k} \mu_{ij} \right)$, which implies that $\mu_F = \sum_{i \in L, j \in H} \alpha_{ij} \mu_{ij}$ with α_{ij} given by $\alpha_{ij} = \frac{\gamma_i}{\sum_{l \in L} \gamma_l} \frac{\gamma_j}{\sum_{k \in H} \gamma_k}$. ■

Lemma 9. *Let $A_1 = [\alpha, \overline{a}]$, $A_2 = (a^*, \alpha)$, $A_3 = [a_0, a^*]$. Consider a signal $y \in \mathcal{D}$ such that $y(\sigma|\cdot)$ is continuous on each interval A_1, A_2 and A_3 and $E_F[\omega|\sigma] \geq a^*$ for all $F \in \mathcal{F}$. Then $E_F[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha, \beta)$.*

Proof of lemma 9.

In the argument that follows, I suppress the dependence of the conditional expectation on y, using $E_F[\omega|\sigma]$ to denote $E_{F,y}[\omega|\sigma]$.

Claim 9.1. *$E_F[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha, \beta)$ such that $\mu_F = \sum_{i=1}^{n} \gamma_i \delta_{\omega_i}$ for $\gamma_i \in [0, 1]$, $\sum_{i=1}^{n} \gamma_i = 1$.*

[13] Recall that $\mathcal{B}(A_i)$ denotes the Borel sigma-algebra on A_i.

Proof of claim 9.1.

We know that $E_F[\omega|\sigma] \geq a^*$ is true for all $F \in \mathcal{F}$. That is, $E_F[\omega|\sigma] \geq a^*$ is true for all F such that μ_F is a convex combination of a Dirac measure on ω' for some $\omega' \in [\alpha, \overline{a}]$ and a Dirac measure on ω for some $\omega \in [a_0, \alpha)$, with weights γ and $1-\gamma$ for some $\gamma \in [\beta, 1]$ respectively. In particular, we have

$$\frac{\int_\Omega wy(\sigma|w)d\mu_i(w)}{\int_\Omega y(\sigma|w)d\mu_i(w)} \geq a^* \tag{13}$$

for any probability measure μ_i such that $\mu_i = (1-\gamma)\delta_\omega + \gamma\delta_{\omega'}$, $\gamma \in [\beta, 1]$, $\omega \in [a_0, \alpha)$ and $\omega' \in [\alpha, \overline{a}]$.

Lemma 8 shows that any $\mu_F = \sum_{i=1}^n \gamma_i \delta_{\omega_i}$ such that $\gamma_i \in [0,1]$, $\sum_{i=1}^n \gamma_i = 1$, $\mu_F \in C(\alpha, \beta)$, can be written as $\mu_F = \sum_{i=1}^N \alpha_i \mu_i$ for some probability measures $\mu_i = (1-\gamma)\delta_{\omega_i} + \gamma\delta_{v_i}$, $\gamma \in [\beta, 1]$, $\omega_i \in [a_0, a^*)$, $v_i \in [a^*, \overline{a}]$, and weights $\alpha_i \in [0,1]$ such that $\sum_{i=1}^N \alpha_i = 1$ for some $N \in \mathbb{N}$.

Let $F \in C(\alpha, \beta)$ such that $\mu_F = \sum_{i=1}^n \gamma_i \delta_{\omega_i}$, $\gamma_i \in [0,1]$, $\sum_{i=1}^n \gamma_i = 1$, be given. Then

$$\begin{aligned}\frac{\int_\Omega wy(\sigma|w)d\mu_F(w)}{\int_\Omega y(\sigma|w)d\mu_F(w)} = \;& \frac{\sum_{i=1}^n \gamma_i \int_\Omega wy(\sigma|w)d\delta_{w_i}}{\sum_{i=1}^n \gamma_i \int_\Omega y(\sigma|w)d\delta_{w_i}} = \\ & \frac{\sum_{i=1}^N \alpha_i \int_\Omega wy(\sigma|w)d\mu_i(w)}{\sum_{i=1}^N \alpha_i \int_\Omega y(\sigma|w)d\mu_i(w)} \geq \\ & \min_{i \in \{1,\ldots,N\}} \frac{\alpha_i \int_\Omega wy(\sigma|w)d\mu_i(w)}{\alpha_i \int_\Omega y(\sigma|w)d\mu_i(w)} = \\ & \min_{i \in \{1,\ldots,N\}} \frac{\int_\Omega wy(\sigma|w)d\mu_i(w)}{\int_\Omega y(\sigma|w)d\mu_i(w)} \geq a^*\end{aligned} \tag{14}$$

where the first inequality follows from lemma 6 and the second inequality follows from (13). ■

Claim 9.2. *Given $F \in C(\alpha, \beta)$, we have* $\frac{\int_\Omega wy(\sigma|w)d\mu_F(w)}{\int_\Omega y(\sigma|w)d\mu_F(w)} = \lim_{n\to\infty} \frac{\sum_{i=1}^n \gamma_i \int_\Omega wy(\sigma|w)d\delta_{w_i}}{\sum_{i=1}^n \gamma_i \int_\Omega y(\sigma|w)d_{w_i}}$.

Proof of claim 9.2.

By lemma 7, for each μ_F such that $F \in C(\alpha, \beta)$ there exist measures G_1, G_2, G_3 such that $\mu_F = \alpha_1\mu_{G_1} + \alpha_2\mu_{G_2} + \alpha_3\mu_{G_3}$, $\mu_{G_1}(\Omega \setminus A_1) = 0$, $\mu_{G_2}(\Omega \setminus A_2) = 0$ and $\mu_{G_3}(\Omega \setminus A_3) = 0$, $\alpha_i = \mu_F(A_i)$ for $i = 1, 2, 3$.

By Theorem 15.10 in Aliprantis and Border (2006: 513), the set of probability measures on Ω with finite support is dense in the set of all Borel probability measures on Ω (with the topology of weak convergence). Then, given any measure μ supported on A_i and any bounded continuous function g, we have that $\lim_{N\to\infty}\int_{A_i} g(w)d\mu^N(w) = \int_{A_i} g(w)d\mu(w)$ where, for all $N \in \mathbb{N}$, $\mu^N = \sum_{i=1}^N \eta_i\delta_{\omega_i}$, $\{\delta_{\omega_i}\}_{i=1}^N$ is a finite collection of Dirac probability measures, $\eta_i \in [0,1]$ for $i = 1,\dots,N$ and $\sum_{i=1}^N \eta_i = 1$.

Then

$$\begin{aligned}
\frac{\int_\Omega wy(\sigma|w)d\mu_F(w)}{\int_\Omega y(\sigma|w)d\mu_F(w)} = \quad & \frac{\sum_{i=1}^3 \alpha_i \int_{A_i} wy(\sigma|w)d\mu_{G_i}(w)}{\sum_{i=1}^3 \alpha_i \int_{A_i} y(\sigma|w)d\mu_{G_i}(w)} = \\
& \frac{\sum_{i=1}^3 \alpha_i \lim_{N\to\infty} \int_{A_i} wy(\sigma|w)d\mu_{G_i}^N(w)}{\sum_{i=1}^3 \alpha_i \lim_{N\to\infty} \int_{A_i} y(\sigma|w)d\mu_{G_i}^N(w)} = \\
& \frac{\sum_{i=1}^3 \alpha_i \lim_{N\to\infty} \int_{A_i} wy(\sigma|w)d(\sum_{j=1}^N \eta_j\delta_{\omega_j}(w))}{\sum_{i=1}^3 \alpha_i \lim_{N\to\infty} \int_{A_i} y(\sigma|w)d(\sum_{j=1}^N \eta_j\delta_{\omega_j}(w))} = \\
& \lim_{N\to\infty} \frac{\sum_{j=1}^N\sum_{i=1}^3 \alpha_i\eta_j \int_{A_i} wy(\sigma|w)d\delta_{\omega_j}(w)}{\sum_{j=1}^N\sum_{i=1}^3 \alpha_i\eta_j \int_{A_i} y(\sigma|w)d\delta_{\omega_j}(w)} = \\
& \lim_{n\to\infty} \frac{\sum_{i=1}^n \gamma_i \int_\Omega wy(\sigma|w)d\delta_{\omega_i}(w)}{\sum_{i=1}^n \gamma_i \int_\Omega y(\sigma|w)d\delta_{\omega_i}(w)}
\end{aligned}$$

where the first equality follows from the fact that $\mu_F = \alpha_1\mu_{G_1} + \alpha_2\mu_{G_2} + \alpha_3\mu_{G_3}$, the second equality follows from the fact that $\{\mu_{G_i}^N\}_{N\in\mathbb{N}}$ converges weakly to μ_{G_i} for $i = 1,2,3$ and the function $y(\sigma|\cdot)$ is bounded and continuous on each of the intervals A_1, A_2 and A_3, and the third equality follows from the definition of $\mu_{G_i}^N$. ■

Claim 9.3. *$E_F[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha,\beta)$.*

Proof of claim 9.3.

(14) implies that $\lim_{n\to\infty} \frac{\sum_{i=1}^n \gamma_i \int_\Omega wy(\sigma|w)d\delta_{w_i}}{\sum_{i=1}^n \gamma_i \int_\Omega y(\sigma|w)d_{w_i}} \geq a^*$. Because, by claim 9.2, given $F \in C(\alpha,\beta)$, we have $\frac{\int_\Omega wy(\sigma|w)d\mu_F(w)}{\int_\Omega y(\sigma|w)d\mu_F(w)} = \lim_{n\to\infty} \frac{\sum_{i=1}^n \gamma_i \int_\Omega wy(\sigma|w)d\delta_{w_i}}{\sum_{i=1}^n \gamma_i \int_\Omega y(\sigma|w)d_{w_i}}$, this implies that $\frac{\int_\Omega wy(\sigma|w)d\mu_F(w)}{\int_\Omega y(\sigma|w)d\mu_F(w)} \geq a^*$. ■ ■

Lemma 10. *Let π be the signal defined in lemma 12. Then $E_F[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha,\beta)$.*

Proof of lemma 10.

Consider the signal q defined in lemma 5. By lemma 5, for all $F \in \varphi$, if $E_{F,q}[\omega|\sigma] \geq a^*$, then $E_{F,\pi}[\omega|\sigma] \geq a^*$. We will show that $E_{F,q}[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha,\beta)$. By lemma 5, this will imply that $E_{F,\pi}[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha,\beta)$.

By lemma 4, $E_{F,q}[\omega|\sigma] \geq a^*$ is true for all $F \in \mathcal{F}$.

Let $A_1 = [\alpha, \overline{a}]$, $A_2 = (a^*, \alpha)$, $A_3 = [a_0, a^*]$. Because the signal $q \in \mathcal{D}$ is such that $q(\sigma|\cdot)$ is continuous on each interval A_1, A_2, A_3 and $E_{F,q}[\omega|\sigma] \geq a^*$ is true for all $F \in \mathcal{F}$, lemma 9 applies. By lemma 9, we have $E_{F,q}[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha, \beta)$.

Therefore, we have $E_{F,\pi}[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha, \beta)$, as required. ■

Lemma 11. *The signal π specified in lemma 12 satisfies the conditions in lemma 2.*

Proof of lemma 11.

Given $\pi \in \mathcal{D}$, let $R(\pi) = \{1, \ldots, r\}$ denote the finite subset of signal realization indices such that $\sigma_1, \ldots, \sigma_r$ have a strictly positive probability under π given that the state is α.

Then the statement of the lemma is equivalent to the following. $\pi \in \mathcal{D}$ specified in lemma 12 is such that $E_{F,\pi}[\omega|\sigma_i] \geq a^*$ for all $i \in R(\pi)$ and for all $F \in \mathcal{F}_\alpha$, for all $\omega_m \in [a_0, a^*)$, we have $\sum_{i \in R(\pi)} \pi(\sigma_i|\omega_m) = \min\left\{\frac{\beta}{1-\beta}\frac{\alpha - a^*}{a^* - \omega_m}, 1\right\}$ and whenever the sender chooses π, Nature prefers to choose some $F \in \mathcal{F}_\alpha$.

Lemma 10 implies that π specified in lemma 12 is such that $E_F[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha, \beta)$.

Moreover, by construction, given $\omega_m \in [a_0, a^*)$, we have $\sum_{i \in R(\pi)} \pi(\sigma_i|\omega_m) = \pi(\sigma|\omega_m) = \min\left\{\frac{\beta}{1-\beta}\frac{\alpha - a^*}{a^* - \omega_m}, 1\right\}$.

Observe that the signal structure π is such that if the signal realization is not σ, then the signal reveals the state. If the signal reveals that the state is ω, then the receiver's belief is δ_ω regardless of the receiver's prior, which implies that Nature's choice of the receiver's prior affects the sender's payoff only by determining $E_F[\omega|\sigma]$.

Note that if F is such that $E_F[\omega|\sigma] \geq a^*$, then $u(E_F[\omega|\sigma]) = u(a^*)$ because u is constant on $[a^*, \overline{a}]$. Therefore, because $E_F[\omega|\sigma] \geq a^*$ for all $F \in C(\alpha, \beta)$ and because, if the signal realization is not σ, then the signal reveals the state, Nature is indifferent when choosing among the receiver's priors $F \in C(\alpha, \beta)$. It follows that Nature weakly prefers to choose $F \in \mathcal{F}_\alpha$. ■

Lemma 12 proves the optimality of an information policy that is more general than the policy described in Theorem 1. Lemma 12 implies both the result in Theorem 1 and the result about an optimal information policy with revelation in section 3.4.

Lemma 12. *There exists an optimal signal that induces a distribution over the receiver's* *actions where the probability of action* 1 *in state* ω *is given by*

$$s(\omega) = \begin{cases} 1 \text{ if } \omega \in (a^*, \overline{a}] \\ \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} \text{ if } \omega \in [a_0, a^*] \end{cases}$$

If there exists an optimal signal that induces a distribution over the receiver's actions where the probability of action 1 *is given by* s' *and* $s' \neq s$*, then* $s'(\omega) = s(\omega)$ μ_{F_s}*-a.e.*

An optimal signal inducing the distribution s *over the receiver's actions is given by*

$$\pi(\sigma|\omega) = \begin{cases} p(\omega) \in \left[\frac{\alpha-a^*}{\omega-a^*}, 1\right] \text{ if } \omega \in [\alpha, \overline{a}] \\ t(\omega) \in \left[0, 1\right] \text{ if } \omega \in (a^*, \alpha) \\ \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} \text{ if } \omega \in [a_0, a^*] \end{cases}$$

$$\pi(\sigma_w|\omega) = \begin{cases} 1 - p(\omega) \text{ if } \omega = w \text{ and } \omega \in [\alpha, \overline{a}] \\ 1 - t(\omega) \text{ if } \omega = w \text{ and } \omega \in (a^*, \alpha) \\ 1 - \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} \text{ if } w = \omega \text{ and } \omega \in [a_0, a^*] \end{cases}$$

where $\omega \mapsto p(\omega)$ *and* $\omega \mapsto t(\omega)$ *are measurable functions.*

Proof of lemma 12.

We first show that an optimal signal is given by π specified in the lemma.

Given $\pi \in \mathcal{D}$, let $R(\pi) = \{1, \ldots, r\}$ denote the finite subset of signal realization indices such that $\sigma_1, \ldots, \sigma_r$ have a strictly positive probability under π given that the state is α.

By lemma 2, if there exists a signal $\pi \in \mathcal{D}$ such that $E_{F,\pi}[\omega|\sigma_i] \geq a^*$ for all $i \in R(\pi)$ and for all $F \in \mathcal{F}_\alpha$, if for all $\omega \in [a_0, a^*)$, we have $\sum_{i \in R(\pi)} \pi(\sigma_i|\omega_m) = \min\left\{\frac{\beta}{1-\beta}\frac{\alpha-a^*}{a^*-\omega}, 1\right\}$, and if it is the case that whenever the sender chooses π, Nature prefers to choose $F \in \mathcal{F}_\alpha$, then π maximizes the sender's payoff among all signals in $\mathcal{D}$.

Lemma 11 shows that the signal π specified in lemma 12 satisfies the conditions in lemma 2. Then lemma 2 implies that π maximizes the sender's payoff among all signals in $\mathcal{D}$.

Next, we show that if an optimal signal induces a distribution $s' \neq s$ over the receiver's actions, then $s'(\omega) = s(\omega)$ μ_{F_s}-a.e.

It is easy to see that π induces s. We will show that there does not exist an equilibrium in which the sender chooses a signal that induces action 1 with probability strictly greater than $s(\omega)$ for all $\omega \in A$, where $A \subseteq \Omega$ is some set satisfying $\mu_{F_s}(A) > 0$.

Given $\pi' \in \mathcal{D}$ and $F \in C(\alpha, \beta)$, let $R(F, \pi)$ denote the finite number of signal indices such that $E_{F,\pi'}[\omega|\sigma_i] \geq a^*$ and $\pi(\sigma_i|\alpha) > 0$ for all $i \in R(F, \pi')$. By lemma 1, for all $\pi' \in \mathcal{D}$ there exists a feasible receiver's prior $F \in C(\alpha, \beta)$ such that $\sum_{i \in R(F,\pi')} \pi'(\sigma_i|\omega) \leq \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} = s(\omega)$ for all $\omega \in [a_0, a^*)$.

Suppose for the sake of contradiction that there is an equilibrium in which the sender chooses $\pi' \in \mathcal{D}$ and Nature chooses $F' \neq F$ such that action 1 is induced with probability strictly greater than $s(\omega)$ for all $\omega \in A$ for some set $A \subseteq \Omega$ satisfying $\mu_{F_s}(A) > 0$.

Let U^* denote the sender's equilibrium payoff if the sender chooses the signal π. Because the sender can guarantee a payoff of U^* by choosing π, if the sender chooses π', then π' must yield a payoff of at least U^* to the sender.

Therefore, the sender who chooses π' receives a payoff of at least U^* if Nature chooses F or F'. The fact that the sender who chooses π' receives a payoff of at least U^* if Nature chooses F implies that $\sum_{i \in R(F,\pi')} \pi'(\sigma_i|\omega) = \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} = s(\omega)$ μ_{F_s}-a.e. on $[a_0, a^*)$. This implies that the induced distribution over actions cannot be such that action 1 is induced with probability strictly greater than $s(\omega)$ for all $\omega \in A$ with $\mu_{F_s}(A) > 0$, a contradiction.

Thus if an optimal signal induces a distribution $s' \neq s$ over the receiver's actions, then $s'(\omega) \leq s(\omega)$ μ_{F_s}-a.e. Moreover, if a signal π' induces a distribution $s' \neq s$ such that $s'(\omega) \leq s(\omega)$ μ_{F_s}-a.e. and $s'(\omega) < s(\omega)$ for all $\omega \in A$ for some set $A \subseteq \Omega$ satisfying $\mu_{F_s}(A) > 0$, then the signal π' gives the sender a strictly lower payoff than the signal π inducing the distribution s, so the signal π' cannot be optimal.

Therefore, if an optimal signal induces a distribution $s' \neq s$ over the receiver's actions, then we must have $s'(\omega) = s(\omega)$ μ_{F_s}-a.e., as required. ∎

Theorem 1. *There exists an optimal signal that induces a distribution over the receiver's actions such that the probability of action* 1 *in state* ω *is given by*

$$s(\omega) = \begin{cases} 1 \text{ if } \omega \in (a^*, \overline{a}] \\ \min\left\{\frac{\beta(\alpha-a^*)}{(1-\beta)(a^*-\omega)}, 1\right\} \text{ if } \omega \in [a_0, a^*] \end{cases}$$

If there exists an optimal signal that induces a distribution over the receiver's actions where the probability of action 1 *is given by* s' *and* $s' \neq s$, *then* $s'(\omega) = s(\omega)$ μ_{F_s}*-a.e.*

An optimal signal inducing the distribution s over the receiver's actions is given by $\pi(\sigma|\omega) = s(\omega)$, $\pi(\sigma_0|\omega) = 1 - s(\omega)$ *for all* $\omega \in \Omega$.

Proof of theorem 1.

Follows from lemma 12. ■

Lemma 13. *Consider the optimal information policy with two signal realizations, σ and σ_0, described in Theorem 1. Then a receiver with a prior μ_F, $F \in C(\alpha, \beta)$, strictly benefits from persuasion if and only if either*

$$E_F[\omega] < a^* \qquad E_F[\omega|\sigma] > a^*$$

or

$$E_F[\omega] \geq a^* \qquad \int_\Omega \pi(\sigma_0|w) dF(w) > 0$$

Proof of lemma 13.

Note that, because $F \in C(\alpha, \beta)$, we must have $\int_\Omega \pi(\sigma|w) dF(w) > 0$.

So far we have conducted the analysis by writing the utility of the sender in terms of the receiver's expectation of the state and assuming that the receiver takes the high action if and only if the receiver's expectation is above the threshold a^*. In order to analyze the receiver's welfare, it is now necessary to specify the utility function of the receiver. Let $x \in \{0, 1\}$ denote the actions available to the receiver and let $u_R(x, w) : \{0, 1\} \times \Omega \to \mathbb{R}$ denote the utility function of the receiver. Without loss of generality, we let $u_R(x, w) = w$ if $x = 1$ and let $u_R(x, w) = a^*$ if $x = 0$. Then $E_F[u_R(1, w)|\sigma] = E_F[w|\sigma]$ and $E_F[u_R(0, w)|\sigma] = a^*$, so a receiver seeing signal σ takes action $x = 1$ if and only if $E_F[w|\sigma] \geq a^*$.

Because in equilibrium under the optimal information policy with two signal realizations described in Theorem 1 the receiver takes action 1 if and only if he sees signal realization σ, the receiver's equilibrium payoff is

$$\int_\Omega \Big[\pi(\sigma|w) u_R(1, w) + \pi(\sigma_0|w) u_R(0, w)\Big] dF(w)$$

Suppose that $E_F[\omega] < a^*$ and $E_F[\omega|\sigma] > a^*$. Thus if no information is provided, the receiver takes action 0 and gets a payoff of $\int_\Omega u_R(0, w) dF(w)$.

Because $E_F[\omega|\sigma] > a^*$ and $\int_\Omega \pi(\sigma|w)dF(w) > 0$, it must be the case that

$$
\begin{aligned}
&\int_\Omega \Big[\pi(\sigma|w)u_R(1,w) + \pi(\sigma_0|w)u_R(0,w)\Big]dF(w) = \\
&\int_\Omega \pi(\sigma|w)wdF(w) + \int_\Omega \pi(\sigma_0|w)u_R(0,w)dF(w) = \\
&\int_\Omega \pi(\sigma|w)dF(w)E_F[w|\sigma] + \int_\Omega \pi(\sigma_0|w)dF(w)a^* > \\
&\int_\Omega \pi(\sigma|w)dF(w)a^* + \left(1 - \int_\Omega \pi(\sigma|w)dF(w)\right)a^* = \\
&a^* = \int_\Omega u_R(0,w)dF(w)
\end{aligned}
$$

Therefore, the receiver strictly benefits from persuasion.

Next, suppose that $E_F[\omega] \geq a^*$ and $\int_\Omega \pi(\sigma_0|w)dF(w) > 0$. Thus if no information is provided, the receiver takes action 1 and gets a payoff of $\int_\Omega u_R(1,w)dF(w)$.

Because $E_F[\omega|\sigma_0] < a^*$ and $\int_\Omega \pi(\sigma_0|w)dF(w) > 0$, it must be the case that

$$
\begin{aligned}
&\int_\Omega \Big[\pi(\sigma|w)u_R(1,w) + \pi(\sigma_0|w)u_R(0,w)\Big]dF(w) = \\
&\int_\Omega \pi(\sigma|w)wdF(w) + \int_\Omega \pi(\sigma_0|w)u_R(0,w)dF(w) = \\
&\int_\Omega \pi(\sigma|w)dF(w)E_F[w|\sigma] + \int_\Omega \pi(\sigma_0|w)dF(w)a^* > \\
&\int_\Omega \pi(\sigma|w)dF(w)E_F[w|\sigma] + \left(1 - \int_\Omega \pi(\sigma|w)dF(w)\right)E_F[w|\sigma_0] = \\
&\int_\Omega u_R(0,w)dF(w)
\end{aligned}
$$

Therefore, the receiver strictly benefits from persuasion.

Note that we must have $E_F[\omega|\sigma] \geq a^*$, so if it is not the case that $E_F[\omega|\sigma] > a^*$, then we must have $E_F[\omega|\sigma] = a^*$. Thus suppose that $E_F[\omega] < a^*$ and $E_F[\omega|\sigma] = a^*$. Then $E_F[\omega] < a^*$, so if no information is provided, the receiver takes action 0 and gets a payoff of $\int_\Omega u_R(0,w)dF(w)$.

Because $E_F[\omega|\sigma] = a^*$, it must be the case that

$$\int_\Omega \Big[\pi(\sigma|w)u_R(1,w) + \pi(\sigma_0|w)u_R(0,w)\Big]dF(w) =$$
$$\int_\Omega \Big[\pi(\sigma|w)u_R(0,w) + \pi(\sigma_0|w)u_R(0,w)\Big]dF(w) = a^* = \int_\Omega u_R(0,w)dF(w)$$

Therefore, the receiver's utility is the same under persuasion and when no information is provided.

If $\int_\Omega \pi(\sigma_0|w)dF(w) = 0$, then it is easy to see that the receiver's payoff under persuasion is the same as in the case when no information is provided because in this case the receiver takes the high action when no information is provided and believes that under persuasion only the signal realization σ inducing him to take the high action is possible. ■

Proof of Proposition 2.

Because $\pi(\sigma|w) = 1$ for all $w > \underline{\alpha}$, we have

$$\int_\Omega \pi(\sigma|w)dF_s(w) = \int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dF_s(w) + \int_{(\underline{\alpha},\overline{a}]} dF_s(w)$$

$$\int_\Omega \pi(\sigma|w)dG_s(w) = \int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dG_s(w) + \int_{(\underline{\alpha},\overline{a}]} dG_s(w)$$

Because μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0,\underline{\alpha}]$, we have $\mu_{F_s} = \gamma\mu_1 + (1-\gamma)\mu_2$, $\mu_{G_s} = \gamma m_1 + (1-\gamma)m_2$ for some $\gamma \in [0,1]$, $\mu_1((\underline{\alpha},\overline{a}]) = m_1((\underline{\alpha},\overline{a}]) = 0$, $\mu_2([a_0,\underline{\alpha}]) = m_2([a_0,\underline{\alpha}]) = 0$. Therefore, $\int_{(\underline{\alpha},\overline{a}]} dF_s(w) = \int_{(\underline{\alpha},\overline{a}]} dG_s(w) = 1-\gamma$.

Moreover, we have $\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dF_s(w) = \gamma\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)d\mu_1(w)$ and $\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dG_s(w) = \gamma\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dm_1(w)$

Then $\int_\Omega \pi(\sigma|w)dF_s(w) \geq \int_\Omega \pi(\sigma|w)dG_s(w)$ is equivalent to $\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dF_s(w) \geq \int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dG_s(w)$, which is equivalent to $\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)d\mu_1(w) \geq \int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dm_1(w)$. $\int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)d\mu_1(w) \geq \int_{[a_0,\underline{\alpha}]} \pi(\sigma|w)dm_1(w)$ holds because the function $w \mapsto \pi(\sigma|w)$ is strictly convex on $[a_0,\underline{\alpha}]$ and μ_1 is a mean-preserving spread of m_1. ■

Lemma 14. *Suppose that $F_s \in C(\alpha,\beta)$ admits a density and that $E_{F_s}[\omega] < a^*$. Consider the standard persuasion model such that F_s is the common prior of the sender and the receiver. Given $a \in \Omega$, let p_a denote the signal with two realizations, σ and σ_0, satisfying $p_a(\sigma|w) = 1$ if $w \geq a$ and $p(\sigma|w) = 0$ if $w < a$. Let a' denote the threshold that is optimal in the standard* *persuasion model, let π a signal with two realizations, σ and σ_0, that is optimal in the model with unknown beliefs, and let $\underline{\alpha} = a^* - \frac{\beta}{1-\beta}(\alpha - a^*)$. Then $\underline{\alpha} > a'$.*

Proof of lemma 14.

Suppose for the sake of contradiction that $\underline{\alpha} \leq a'$.

Note that, because $F_s \in C(\alpha, \beta)$, we have $\mu_{F_s}((a^*, \overline{a}]) > 0$. Then the signal $p_{a'}$ optimal in the standard persuasion model must be such that $E_{F_s, p_{a'}}[\omega|\sigma] = a^*$.

Next, note that the fact that $E_{F_s}[\omega] < a^*$ and $E_{F_s, p_{a'}}[\omega|\sigma] = a^*$ implies that $\mu_{F_s}([a_0, a')) > 0$.

We have

$$a^* = E_{F_s, p_{a'}}[\omega|\sigma] \geq E_{F_s, p_{\underline{\alpha}}}[\omega|\sigma] > E_{F_s, \pi}[\omega|\sigma]$$

where the equality follows from the form of the optimal solution in the standard persuasion model, the first inequality follows from the assumption that $\underline{\alpha} \leq a'$, and the second inequality follows from the formula for the signal π and from the fact that $\mu_{F_s}([a_0, a')) > 0$.

Therefore, we have $E_{F_s, \pi}[\omega|\sigma] < a^*$, which is a contradiction because the optimal signal π must satisfy $E_{F, \pi}[\omega|\sigma] \geq a^*$ for every feasible prior $F \in C(\alpha, \beta)$. ■

Proof of Proposition 1. Follows from lemma 14. ■

Proof of Proposition 3.

Without loss of generality, suppose that μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0, a')$.

We have

$$\int_\Omega D(w)dF_s(w) = \int_{[a_0, a')} D(w)dF_s(w) + \int_{\Omega \setminus [a_0, a')} D(w)dF_s(w)$$

$$\int_\Omega D(w)dG_s(w) = \int_{[a_0, a')} D(w)dG_s(w) + \int_{\Omega \setminus [a_0, a')} D(w)dG_s(w)$$

Because μ_{F_s} is a local mean-preserving spread of μ_{G_s} on $[a_0, a')$, we have $\mu_{F_s} = \gamma\mu_1 + (1-\gamma)\mu_2$, $\mu_{G_s} = \gamma m_1 + (1-\gamma)m_2$ for some $\gamma \in [0, 1]$, $\mu_1(\Omega \setminus [a_0, a')) = m_1(\Omega \setminus [a_0, a')) = 0$, $\mu_2([a_0, a')) = m_2([a_0, a')) = 0$ and $\mu_2 = m_2$. Therefore, we have $\int_{\Omega \setminus [a_0, a')} D(w)dF_s(w) = \int_{\Omega \setminus [a_0, a')} D(w)dG_s(w)$.

Moreover, we have $\int_{[a_0,a')} D(w)dF_s(w) = \gamma \int_{[a_0,a')} D(w)d\mu_1(w)$ and $\int_{[a_0,a')} D(w)dG_s(w) = \gamma \int_{[a_0,a')} D(w)dm_1(w)$.

Then $\int_\Omega D(w)dF_s(w) \geq \int_\Omega D(w)dG_s(w)$ is equivalent to $\int_{[a_0,a')} D(w)dF_s(w) \geq \int_{[a_0,a')} D(w)dG_s(w)$, which is equivalent to $\int_{[a_0,a')} D(w)d\mu_1(w) \geq \int_{[a_0,a')} D(w)dm_1(w)$. $\int_{[a_0,a')} D(w)d\mu_1(w) \geq \int_{[a_0,a')} D(w)dm_1(w)$ holds because the function $w \mapsto D(w)$ is strictly convex on $[a_0, a')$ and μ_1 is a mean-preserving spread of m_1. ■

References

Aliprantis, Charalambos D., and Kim C. Border. 2006. *Infinite Dimensional Analysis, third edition.* Berlin, Germany: Springer-Verlag.

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

Alonso, Ricardo, and Odilon Câmara. 2016*b*. “Persuading Voters.” *American Economic Review* 106(11): 3590–3605.

Billingsley, Patrick. 2012. *Probability and Measure.* Hoboken, New Jersey: John Wiley & Sons.

Brocas, Isabelle, and Juan D. Carillo. 2007. “Influence through ignorance.” *RAND Journal of Economics* 38(4): 931–947.

Carrasco, Vinicius, Vitor Farinha Luz, Nenad Kos, Matthias Messner, Paulo Monteiro, and Humberto Moreira. 2017. “Optimal Selling Mechanisms Under Moment Conditions.” Manuscript.

Carroll, Gabriel. 2015. “Robustness and Linear Contracts.” *American Economic Review* 105(2): 536–563.

Carroll, Gabriel. 2017. “Robustness and Separation in Multidimensional Screening.” *Econometrica* 85(2): 453–488.

Chassang, Sylvain. 2013. “Calibrated Incentive Contracts.” *Econometrica* 81(5): 1935–1971.

Ely, Jeffrey C. 2017. “Beeps.” *American Economic Review* 107(1): 31–53.

Fiorino, Daniel J. 2006. *The New Environmental Regulation.* Cambridge, MA: MIT Press.

Hedlund, Jonas. 2017. “Bayesian persuasion by a privately informed sender.” *Journal of Economic Theory* 167: 229–268.

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

Kolotilin, Anton. 2017. “Optimal Information Disclosure: A Linear Programming Approach.” *Theoretical Economics* forthcoming.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. 2017. "Persuasion of a privately informed receiver." *Econometrica* forthcoming.

Ostrovsky, Michael, and Michael Schwarz. 2010. "Information Disclosure and Unraveling in Matching Markets." *American Economic Journal: Microeconomics* 2(2): 34–63.

Perez-Richet, Eduardo. 2014. "Interim Bayesian Persuasion: First Steps." *American Economic Review: Papers & Proceedings* 104(5): 469–474.

Pollard, David. 2002. *A User's Guide to Measure Theoretic Probability.* New York, NY: Cambridge University Press.

Rayo, Luis, and Ilya Segal. 2010. "Optimal Information Disclosure." *Journal of Political Economy* 118(5): 949–987.