Abstract

A sender designs an information structure to persuade a receiver to take an action. The sender is ignorant about the receiver’s prior, and evaluates each information structure using the receiver’s prior that is the worst for the sender. I characterize the optimal information structures in this environment. I show that there exists an optimal signal with two realizations, characterize the support of the signal and provide a formula that the signal must satisfy on the support, showing that the optimal signal is a hyperbola. The lack of knowledge of the receiver’s prior causes the sender to hedge her bets: the optimal signal induces the high action in more states than in the standard model, albeit with a lower probability. Increasing the sender’s ignorance can hurt both the sender and the receiver.

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 and instead only knows that the FDA believes that the drug is sufficiently safe with a high enough probability.

The company is designing the experiment in an environment where 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. 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.

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. If the sender and the receiver have commonly known heterogeneous priors, the optimal signal is not necessarily threshold but is still partitional: the high action is recommended either with probability one or with probability zero given a state. As the results in this paper will establish, when the receiver’s beliefs are unknown, the optimal signal is very different.

Towards establishing the results, let us now describe the setting this paper focuses on. I model the sender’s ignorance by assuming that 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.

I focus on the case where the sender knows that the receiver’s prior assigns no less than a certain probability to each state but is otherwise ignorant about the receiver’s prior. Formally, the set of the receiver’s priors is all priors that put a mass of at least \((1 - a)g(w)\) on each state \(w\). This set of priors has the advantage of allowing me to flexibly model the lack of the sender’s knowledge about the receiver’s prior. Because the higher \(a\) is, the smaller the mass the receiver’s prior has to put on each state, \(a\) measures the sender’s ignorance or, equivalently, the flexibility that Nature has in choosing the receiver’s prior. If \(g\) is higher on one subset of states than on the other, then the sender has more information about the
receiver’s prior on the first subset.

The main contribution of this paper is a characterization of the optimal information structure in a model of persuasion when the receiver’s beliefs are unknown. The contribution has several parts. First, I show that there exists a sender-optimal information structure with only two signal realizations. Second, I characterize the support of the signal realization recommending the high action, showing that it consists of all states such that an importance index exceeds a threshold. Third, I provide a formula for the probability that the high action is recommended on the support of the signal, showing that the optimal signal is a hyperbola. Fourth, I analyze comparative statics on the optimal information structure with respect to the sender’s knowledge of the receiver’s prior.

I show that a state \( w \) is in the support of the signal realization recommending the high action if the importance index \( \frac{f_s(w)}{g(w)(w^* - w)} \) exceeds a threshold, where \( f_s \) is the density of the sender’s prior. Thus the optimal signal is more likely to induce the high action with a strictly positive probability in a state if the sender’s prior assigns a high probability to the state, Nature has more freedom in choosing the probability that the receiver assigns to this state, and the state is close to the threshold of doubt.

I provide results showing that full support is a robust property of the signal chosen by an ignorant sender. In particular, I establish that if either the sender is sufficiently ignorant or the sender’s knowledge is detail-free, then the optimal signal recommends the high action with a strictly positive probability in every state.

I provide comparative statics on the optimal information structure with respect to the sender’s knowledge of the receiver’s prior. I show that the more ignorant the sender is, the more she hedges her bets and spreads out on the states the probability with which the high action is recommended. Formally, if we increase \( a \), thereby decreasing the weight \( (1-a)g(w) \) that the receiver’s prior has to put on each state \( w \), then the support of the optimal signal expands, so that the high action is recommended in more states, but the probability with which the high action is recommended decreases.

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 the 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 on the low states. The model thus makes clear the impact of the sender’s lack of knowledge about the receiver’s prior on the sender-optimal signal: the
lack of knowledge causes her to hedge her bets and spread out the probability with which
the high action is recommended. This reveals the nature of persuasion to be fundamentally
local: oftentimes, 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.

I next consider the welfare implications of the sender’s ignorance. I show that the
impact of increasing the sender’s ignorance on the receiver’s welfare is ambiguous: it can
either benefit or hurt the receiver. Because greater ignorance always hurts the sender, this
implies that the sender’s ignorance about the receiver’s prior can hurt both the sender and
the receiver. I also show that the receiver strictly prefers to face an ignorant sender rather
than a sender that is perfectly informed about the receiver’s prior. Finally, I show that if
the sender-optimal signal recommends the high action with a strictly positive probability
in every state, then greater sender’s ignorance benefits the receiver. These results have
important implications for the optimal transparency requirements for the FDA, showing
that full transparency is never optimal and that if there are bounds on the transparency
requirements that can be enforced, the optimal transparency level may be interior.

My results imply that when the receiver’s beliefs are unknown especially pernicious
outcomes are possible. For instance, there are parameters under which the FDA approves
even the most unsafe drugs with a strictly positive probability, whereas if the receiver’s prior
is known (and coincides with the sender’s prior), 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.

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
an information structure. Importantly, the set of priors that Nature can choose from has
to be sufficiently rich to enable Nature to match the information structure chosen by the
sender, and my analysis highlights the correct notion of richness for this problem.

The rest of the paper proceeds as follows. Section 2 introduces the model. Section 3
presents the characterization of the optimal information structure. Section 4 provides a
sketch of the proof. Section 5 contains results on the comparative statics of the optimal
signal with respect to the sender’s ignorance and conducts welfare analysis. Section 6 shows
that full support is a robust property of the signal chosen by an ignorant sender and provides
examples of optimal information structures. Section 7 reviews the related literature. Section
concludes.

2 Model

2.1 Payoffs

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

\[
u(e) = \begin{cases} 
0 & \text{if } e \in [l, w^*) \\
1 & \text{if } e \in [w^*, h] 
\end{cases}
\]

The model can be interpreted as one where the receiver can take one of the two actions, 0 or 1, the payoff to taking action 0 is 0 and the payoff to taking action 1 is linear in the state. Here the receiver takes action 1 if and only if his expectation of the state is weakly greater than $w^*$.

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 $\varphi$ denote the set of all CDFs of probability measures on $\Omega$. The sender’s prior is a probability measure on $\Omega$ with a CDF $F_s$. I assume that $F_s$ admits a density $f_s$ that is $C^1$. For the sake of convenience, I also assume that $f_s$ has full support.

The set of the receiver’s priors is indexed by a reference prior $G$ and an ignorance index $\alpha \in [0, 1]$. The reference prior $G$ admits a $C^1$ density $g$. For convenience, in most of the

---

1. The assumption that $l > 0$ is without loss of generality and is made for convenience.
2. Some of the results in this paper do not require this assumption. In particular, the result that there exists an optimal signal with only two realizations does not depend on it, nor does the characterization in Corollary ?? of the optimal information structure in the case when the sender only knows that the receiver’s prior puts a mass of at least $\beta$ on states above $\alpha$.
3. All results generalize to the case when the sender’s prior does not have full support.
paper I also assume that $g$ has full support. The receiver’s set of priors is

$$C_{a,g} = \{F \in \varphi : \mu_F(A) \geq (1-a)\mu_G(A) \text{ for all } A \in B([l,h])\}$$

That is, the receiver’s set of priors is the set of all priors that put on each (measurable) set $A$ a mass that is at least as large as the mass that the reference prior $G$ puts on $A$, scaled down by $1-a$.

To understand the assumption on the set of priors, consider a version of the model in which the state space is discrete. Then the receiver’s set of priors consists of all priors that assign a probability of at least $(1-a)g(w)$ to each state $w$. Thus the sender knows that the receiver believes that the probability of each state $w$ is at least $(1-a)g(w)$, but does not know what exactly this probability is. The fact that $a \in (0, 1)$ implies that $\int_{[l,h]}(1-a)g(w)dw < 1$, which ensures that the receiver’s prior is not completely pinned down by this requirement.

Observe that $a = 1 - (1-a)\int_{[l,h]}g(w)dw$ is the difference between 1 and the mass put on the state space $[l, h]$ by the measure $(1-a)\mu_G$. Thus the ignorance index $a$ measures the flexibility that Nature has in choosing the receiver’s prior: the larger $a$ is, the more freedom Nature has. In particular, if $a = 0$, so that there is no ignorance, then the set of the receiver’s priors collapses to one just prior – and this is the reference prior $G$, while if $a = 1$, so that there is complete ignorance, then the set of the receiver’s priors includes all priors.

I assume that $\int_{[w^*,h]}g(w)dw > 0$ and $(1-a)\int_{[l,h]}wg(w)dw + al < w^*$. These assumptions ensure that the sender’s problem is non-trivial. In particular, the assumption that $\int_{[w^*,h]}g(w)dw > 0$ ensures that there exists a feasible information structure that induces the receiver to take action 1 with a strictly positive probability. The assumption that $(1-a)\int_{[l,h]}g(w)wdw + al < w^*$ ensures that if no information is provided, then the receiver will take action 0.

I next introduce special classes of sets of the receiver’s priors. Definition 1 says that the sender’s knowledge is detail-free if the reference prior $G$ puts zero mass on states below the receiver’s threshold – where the sender and the receiver disagree about the optimal action.

**Definition 1.** The sender’s knowledge is said to be detail-free if $\mu_G([l, w^*)) = 0$.

The reason that a set of priors satisfying the condition in definition 1 represents detail-free knowledge is that the reference prior $G$ quantifies the limited knowledge that the sender has about the receiver’s prior. If $g$ varies on the states where the sender and

\[\text{Here } B([l, h]) \text{ denotes the Borel sigma-algebra on } [l, h].\]
the receiver disagree, then the sender knows more about the receiver’s prior on some such states than on others: in particular, the sender knows more about the prior on the states where \( g \) is high, since Nature has less freedom in choosing the prior there. If, on the other hand, \( g \) is 0 on all states below the threshold, then Nature can move a given probability mass freely below the threshold, and the sender is equally ignorant about the receiver’s prior on all states in which there is a conflict of interest. Thus the sender only has a vague idea of how pessimistic the receiver can be – the receiver’s prior can put no more than a certain mass on states below the threshold, but the sender lacks knowledge of the details of how the receiver’s prior may vary on the states where there is disagreement.

2.3 Information Structures and Evaluation of Payoffs

The order of moves is as follows. First, the sender commits to an information structure \( \pi \). Next, Nature chooses the receiver’s prior \( F \in C_{a,g} \) 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 information structure \( \pi \). Then, having seen a signal realization \( \sigma \), the receiver forms an expectation of the state given that the receiver’s prior is \( F \) and that the information structure is \( \pi \).

I restrict my attention to information structures that have a finite number of signal realizations given each state. I conjecture, but do not have a proof, that my results hold for all information structures.

If the sender chooses an information structure \( \pi \) and a receiver with a prior \( F \) sees signal realization \( \sigma \), then the receiver’s expectation of the state is \( E_{F,\pi}[\omega|\sigma] \). Then, if the sender chooses an information structure \( \pi \) and Nature chooses \( F \in C_{a,g} \), the sender’s payoff is

\[
\int_{\Omega} \sum_{\sigma} 1_{E_{F,\pi}[\omega|\sigma] \geq w^*} \pi(\sigma|w) dF_s(w)
\]

---

5An information structure is a Markov kernel \( \pi \). Informally, we can interpret \( \pi(\sigma|\omega) \) as the probability of signal realization \( \sigma \) given that the state is \( \omega \). Formally, letting \( \mathcal{B}(M) \) and \( \mathcal{B}(\Omega) \) denote the Borel sigma-algebras on the message space \( M \) and the state space \( \Omega \) respectively, a Markov kernel \( \pi \) is defined as a mapping \( \pi : \Omega \times \mathcal{B}(M) \rightarrow [0, 1] \) such that for every \( \omega \in \Omega \), \( B \mapsto \pi(B|\omega) \) is a probability measure on \( M \) and for every \( B \in \mathcal{B}(M) \), \( \omega \mapsto \pi(B|\omega) \) is \( \mathcal{B}(\Omega) \)-measurable. See Pollard (2002) for more details.

6If \( g \) does not have full support, then we need to specify how the receiver updates his beliefs after observing signal realizations that have zero probability under the receiver’s prior. In this case, reasonable updating rules such as the receiver not changing his prior belief or putting mass one on the lowest state in the support of the signal realization ensure that the results in the present paper hold.
Recall that Nature chooses the prior of the receiver from the set $C_{a,g}$ in order to minimize the sender’s payoff. This approach is in the spirit of robust mechanism design, where we would like to design mechanisms that are robust to any possible type that the agent might have and are independent of the higher-order beliefs of the sender. Here the worst case analysis offers a simple benchmark. Moreover, there is evidence that people are ambiguity averse, especially when the process by which uncertainty resolves is unknown or poorly understood. In terms of the FDA example, the kind of uncertainty that pharmaceutical companies face about the FDA’s decision-making procedures is a plausible candidate for the kind of uncertainty that is likely to generate ambiguity aversion.

Thus the sender’s payoff from choosing an information structure $\pi$ is

$$U(\pi) = \min_{F \in C_{a,g}} \int_{\Omega} \sum_{\sigma} 1_{E_{F,\pi}[\omega|\sigma] \geq w^*} \pi(\sigma|w)dF_s(w)$$

and the sender’s equilibrium payoff is

$$\max_{\pi} \min_{F \in C_{a,g}} \int_{\Omega} \sum_{\sigma} 1_{E_{F,\pi}[\omega|\sigma] \geq w^*} \pi(\sigma|w)dF_s(w)$$

(1)

3 The Optimal Signal

This section characterizes the optimal signal. I show that there is an optimal signal with two realizations. Under this signal, the probability of the signal realization recommending the high action is $1$ above the threshold $w^*$ and is a hyperbola on the support below $w^*$. The support of this signal realization below $w^*$ is the set of all states such that an importance index exceeds a threshold $t$. There is a tradeoff between adding more states to the support (by increasing the threshold $t$) and recommending the high action with a greater probability (by increasing the constant $c$ scaling the hyperbola), and the optimal signal balances these considerations. I start by defining a class of distributions over the receiver’s actions that have the above form.

Letting $t = \min_{w \in [l, w^*)} \frac{f_s(w)}{q(w)(w^*-w)}$, I define a class of distributions $S^{tc}$ over the receiver’s actions as follows: given constants $t \geq t$, $c \geq 0$, the probability of action 1 in state $\omega$ is given
by

\[ S_{te}(\omega) = \begin{cases} 1 & \text{for } \omega \in [w^*, h] \\ \min \left\{ \frac{c}{w^* - \omega}, 1 \right\} & \text{for } \omega \in \Omega(t) \\ 0 & \text{for } \omega \in [l, w^*) \setminus \Omega(t) \end{cases} \]

where

\[ \Omega(t) = \left\{ w \in [l, w^*) : \frac{f_s(w)}{g(w)(w^* - w)} \geq t \right\} \]

Theorem 1 will show that, given a sender-optimal information structure with two realizations, the support below the threshold \( w^* \) of the signal realization recommending the high action is \( \Omega_1(t) \).

\( \Omega_1(t) \) consists of all states \( w \) below the threshold \( w^* \) such that the importance index \( \frac{f_s(w)}{g(w)(w^* - w)} \) is greater than some threshold \( t \). The importance index is the likelihood ratio of the sender’s prior and the reference prior normalized by the distance between the state and the receiver’s threshold. \( \frac{f_s(w)}{g(w)(w^* - w)} \) is the importance index because it puts higher weight on the states that are more important to the sender. Thus the sender is more likely to induce the receiver to take the high action in state \( w \) if state \( w \) is more important to the sender because the sender believes that this state is very likely, if the conflict of interest between the sender and the receiver at state \( w \) is not too large because the distance \( w^* - w \) between the state and the threshold of doubt \( w^* \) is small, and if Nature has greater freedom in choosing the probability that the receiver assigns to state \( w \) – because \( g(w) \) is small.

To ensure the essential uniqueness of the distribution of the high action induced by a sender-optimal information structure, I make use of the following assumption.

**Assumption 1.** \( \mu \left( \left\{ w \in [l, w^*) : \frac{f_s(w)}{g(w)(w^* - w)} = c_0 \right\} \right) = 0 \) for all \( c_0 > 0 \).

Assumption 1 says that the set of states for which the importance index \( \frac{f_s(w)}{g(w)(w^* - w)} \) is equal to a constant \( c_0 \) is of measure zero for all constants \( c_0 > 0 \).

A signal is said to be sender-optimal if it solves the sender’s problem. Theorem 1 describes a sender-optimal signal and the distribution over the receiver’s actions induced by it.

**Theorem 1.** There exist unique \( t \geq t_0 \), \( c \geq 0 \) such that any sender-optimal information

---

It can be shown that a sufficient condition for Assumption 1 to be satisfied is that \( f_s \) and \( g \) are real-analytic functions on \([l, h]\), and the importance index \( \frac{f_s(w)}{g(w)(w^* - w)} \) is not a constant function on \([l, w^*)\).
structure induces a distribution \( s \) over the receiver’s actions satisfying \( s = S^{\text{tc}} \mu_{F_s} \)-a.e.

A sender-optimal information structure inducing the distribution \( s \) is given by \( \pi(\sigma|\omega) = s(\omega) \), \( \pi(\sigma_0|\omega) = 1 - s(\omega) \) for all \( \omega \in [l, h] \).

Theorem 1 says that the distribution of the receiver’s actions induced by a sender-optimal information structure is unique. Moreover, there is a sender-optimal information structure with two realizations, \( \sigma \) and \( \sigma_0 \). The receiver takes action 1 after seeing signal \( \sigma \) and action zero after seeing signal \( \sigma_0 \). If the state is in \([w^*, h]\), the receiver takes action 1 with probability one. If the state is in a set \( \Omega(t) \), the receiver takes action 1 with probability \( \min\left\{ \frac{c}{w^* - \omega}, 1 \right\} \) for some constant \( c \). Note that on the support \( \Omega(t) \) the probability of action 1 follows a hyperbola. Finally, if the state is below \( w^* \) and not in \( \Omega(t) \), then the receiver takes action 1 with probability zero.

4 Sketch of the Proof

In this section I sketch the proof of Theorem 1. The proof has two parts. The first part establishes that there exists a sender-optimal information structure with only two realizations. The second part characterizes this information structure. There I show that the optimal signal has a hyperbolic functional form and reduce the potentially very complicated problem of characterizing the optimal information structure to the problem of computing two numbers: the threshold \( t \) and the constant of proportionality \( c(t) \) for the hyperbola.

4.1 Sketch of the Proof: Binary Optimal Signal

I show that for all information structures there exists a feasible receiver’s prior \( F \) such that the sum of the probabilities of signal realizations recommending the high action (provided that the receiver’s prior is \( F \)) given each state \( \omega \) below the threshold \( w^* \) is bounded above by an expression depending on \( \omega \). Then I show that there is an information structure with only two realizations that achieves this bound.

The key step in the proof is showing that there exists a feasible prior of the receiver with the abovementioned property. I prove this using a fixed point argument. In particular, I show that this prior is given by a fixed point of a certain mapping. In general, it may not be possible to compute this prior explicitly, but we know that it exists – because a fixed point theorem ensures existence.
Importantly, richness of the set of the receiver’s priors is needed for the above prior to be feasible. Otherwise the set of priors may not contain the prior given by the fixed point. The richness of the set of priors allows Nature to tailor the prior it chooses to the features of the possibly complicated information structure designed by the sender. I show that if the set of priors is not rich enough, then there may not exist an optimal information structure with only two realizations. Section C in the Appendix provides counterexamples establishing this.

4.2 Intuition for the Hyperbola

I next provide the intuition for the hyperbolic functional form of the optimal signal. In the standard Bayesian persuasion model where the sender knows the receiver’s prior the sender-optimal signal 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 \( w^* \) 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 \( \sigma \) is exactly \( w^* \).

The threshold information structure is not optimal when the receiver’s beliefs are unknown. The case in which the set of the receiver’s priors has a particularly simple form is convenient for providing the intuition for the non-optimality of a threshold signal. Suppose that the set of priors is such that any receiver’s prior has to puts a mass of \( 1 - a \) on some state \( \alpha > w^* \), and, subject to this constraint, any prior is allowed. Consider an information structure with two realizations, \( \sigma \) and \( \sigma_0 \), satisfying \( \pi(\sigma|\alpha) = 1 \) and the receiver’s prior that puts a mass of \( 1 - a \) on \( \alpha \) and a mass of \( a \) on some state \( \omega \) below the threshold \( w^* \). Then the receiver’s expectation conditional on seeing \( \sigma \) is \( E[w|\sigma] = \frac{\omega a \pi(\sigma|\omega) + a(1-a)}{\omega \pi(\sigma|\omega) + 1-a} \). In order for the receiver to take action 1 after seeing \( \sigma \), we need \( E[w|\sigma] \geq w^* \), which turns out to be equivalent to \( \pi(\sigma|\omega) \leq \frac{(1-a)(\alpha-w^*)}{a(w^*-\omega)} \).

Note that Nature moves after the sender chose the information structure and can put a mass of \( a \) on any state \( \omega \) below the threshold \( w^* \). Thus if the probability of \( \sigma \) exceeds the bound \( \frac{(1-a)(\alpha-w^*)}{a(w^*-\omega)} \) at any state below the threshold, then by putting a mass of \( a \) 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 signal that is optimal (in the class of all signals with two realizations) when the receiver’s beliefs are unknown.

### 4.3 How Do We Compute the Optimal \( c \) and \( t \)?

I have reduced the problem of computing the optimal information structure to computing two numbers, the threshold \( t \) and the constant of proportionality \( c \). Next I explain show these two numbers are computed.

Note that decreasing the threshold \( t \) or increasing the constant \( c \) decreases the expected state conditional on the approval recommendation. Then for each threshold \( t \) there exists a maximal constant of proportionality \( c(t) \) such that the receiver is willing to follow the approval recommendation. Next, for every pair \((t, c(t))\) I compute the marginal change in the sender’s payoff from expanding the support. This marginal change must be zero at the unique optimum: if it is positive, then the sender prefers to expand the support, while if it is negative, she prefers to shrink the support. Thus we set the marginal change to zero. Together with the equation providing the formula for the constant of proportionality \( c(t) \), these two equations allow us to solve for the two unknowns, \( t \) and \( c(t) \).

### 5 Comparative Statics and Welfare

Proposition \( \Box \) shows that if we take two sets of the receiver’s priors such that the reference prior is the same but \( 0 < a' < a < 1 \), so that the sender is more ignorant under the set of prior \( C_{a,g} \) than \( C_{a',g} \), then, provided that the sender-optimal signal for the second set of priors does not have full support, the support of the sender-optimal signal for the second set contains the support of the sender-optimal signal for the first set and the probability with which the high action is induced on the support of the signal is higher for the first set. Figure \( \Box \) illustrates the comparative statics described in Proposition \( \Box \). Here I let \( t_a \) denote the optimal \( t \) at the ignorance index \( a \).

**Proposition 1.** If \( 0 < a' < a < 1 \) and the optimal signal at \( a \) does not have full support, then

\[
\Omega_1(t_{a'}) \subset \Omega_1(t_a) \mu_{F_s} - a.e. \ \text{and} \ \ c(t_{a'}) > c(t_a)
\]

We can interpret Proposition \( \Box \) as saying that increasing the sender’s ignorance has two
effects: the support effect and the probability effect. The support effect is that the range of states on which the sender’s signal induces the receiver to take action 1 with a positive probability expands, and the probability effect is that the probability with which the receiver takes action 1 on these states decreases. Thus greater ignorance about the receiver’s prior makes the sender “hedge her bets” and spread out the probability with which the high action is recommended on the states below the threshold.

Next, consider two sets of priors, $C_{a,g}$ and $C_{a',g}$, for $0 < a' < a < 1$, so that the sender is more ignorant under $C_{a,g}$ than under $C_{a',g}$. Let $V(a)$ denote the equilibrium payoff of the receiver when the set of priors is $C_{a,g}$. Proposition 2 shows how the receiver’s equilibrium payoff changes as the sender becomes more ignorant.

**Proposition 2.** Suppose that the receiver’s prior is $F \in C_{a',g} \subset C_{a,g}$ for $0 < a' < a < 1$. Then

1. If the optimal signal recommends approval in all states with a strictly positive probability under $a$ and $a'$, then $V(a) \geq V(a')$. If, in addition, the reference prior $g$ has full support, then $V(a) > V(a')$. 

13
2. There exist $g$, $F_s$ such that $0 < V(a) < V(a')$.

Proposition 2 first considers the case in which the sender-optimal signal recommends the high action with a strictly positive probability in every state. The Proposition shows that in this case an increase in the sender’s ignorance weakly increases the receiver’s payoff. Moreover, if the receiver considers all states below the threshold possible, then the increase in the receiver’s payoff is strict.

The intuition for these results is as follows. An increase in the sender’s ignorance has two effects: it expands the support of the signal and lowers the probability with which the high action is recommended on the support. If the signal already has full support, then only the second effect is present: more ignorant senders recommend the high action with a lower probability below the threshold. Because the receiver does not want to take the action anywhere below the threshold, this is unambiguously good for the receiver.

The second part of Proposition 2 shows that the receiver’s equilibrium payoff can be strictly lower when the sender is more ignorant. Because ignorance always hurts the sender, an implication of this is that an increase in sender’s ignorance can hurt both the sender and the receiver.

The reason that the sender’s ignorance can hurt the receiver is that the sender designs the information structure to make the worst types in the set of priors just indifferent between taking the action and not. If a receiver with some prior strictly prefers to act upon seeing the signal recommending the action, so that the receiver’s expectation given the signal is strictly above the threshold of action, the sender’s payoff is not affected by the distance between the expectation and the threshold. Because of this, the impact of the change in the information structure due to the greater ignorance of the sender on the agents with priors that are not the worst in the set of priors that the sender considers possible can be ambiguous.

To better understand the intuition for the possible non-monotonicity of the receiver’s payoff, recall the two comparative statics effects that increasing the sender’s ignorance has, the support effect and the probability effect. The support effect is bad for the receiver because expanding support means that more bad drugs are approved. The probability effect is good for the receiver because a decreasing probability of approval on the support means that bad drugs are approved with a lower probability. My analysis shows that the support effect can dominate, causing greater ignorance of the sender to reduce the receiver’s payoff.

One reason why the support effect can dominate is that the priors of the sender and the receiver may differ, and the sender makes the decision as to which states to add to
the support based in part on her own prior $F_s$. Thus if, for example, the sender’s prior puts a sufficiently high weight on state $w$, the sender will add this state to the support no matter how much she has to reduce the probability with which approval is recommended in other states. If adding $w$ to the support reduces the probability of approval in other states a lot, then the probability effect dominates and the receiver’s payoff increases, while if it reduces the probability of approval only a little bit, then the support effect dominates and the receiver’s payoff decreases.

The final observation is that complete ignorance of the sender is the best possible circumstance for the receiver. This is because as the sender’s ignorance $a$ converges to 1, the sender becomes extremely cautious and recommends that the receiver takes the high actions only on the states where the sender and the receiver agree. Thus as a pharmaceutical company becomes very ignorant about the FDA’s prior, it recommends that the FDA approves a drug if and only if the FDA would approve this drug under complete information, which is the first-best outcome for the FDA.

The result has important implications for the design of the optimal transparency requirements for the FDA. The FDA can change the level of knowledge that pharmaceutical companies have about its prior belief by changing the rules it uses to assign to the review teams for new drug applications. If the companies know the likely composition of the review teams for their drugs, they would have a reasonably good idea of the prior beliefs of the agents making the decision, while if the composition of the review teams is hard to predict, there is significant uncertainty about the priors. Interpreting greater transparency requirements as a reduction in the uncertainty of the pharmaceutical companies about the FDA’s beliefs in this manner, the result implies that full transparency is never optimal. The result also implies that in a setting where the FDA can enforce transparency requirements within some bounds, the optimal level of transparency may be interior.

Figure 2 provides a numerical example for the case in which $f_s$ is uniform, and $g$ is uniform on the intervals $[l, w_0]$ and $[w_0, h]$, with $g$ having a higher density $H$ on the interval $[l, w_0]$. The graph on the left plots the constant of proportionality $c$ and the threshold $w'$. We see that the threshold $w'$ is strictly decreasing between 0.1 and 0.12, constant and equal to $w_0$ for $a$ from 0.12 to 0.7, strictly decreasing from 0.7 to $a = 0.9$, and is constant and equal to $l$ for $a$ greater than 0.9. The reason there is an interval on which the threshold is constant and strictly greater than $l$ is that $g$ in this example is discontinuous. The graph on the right plots the payoff of the receiver with prior $\mu_F = 0.9\mu_G + 0.1\delta_h$ as a function of the sender’s ignorance. The graph shows that the payoff of the receiver is non-monotone in the
Figure 2: Numerical Example: Dependence of the Optimal Signal on the Sender’s Ignorance

(a) Constant of proportionality and the threshold
(b) The value of the receiver

The parameters in the numerical example are $f_s(w) = \frac{1}{h-w}$ for all $w \in [l,h]$, $g(w) = H$ for $w \in [l,w_0]$, $g(w) = L$ for $w \in [w_0,h]$, $l = 0$, $h = 1000$, $w^* = 800$, $H = 1/550$ and $w_0 = 450$. The prior of the receiver is $\mu_F = 0.9\mu_G + 0.1\delta_h$.

sender’s ignorance: the payoff is increasing until $a = 0.7$, then decreasing until $a = 0.9$, then increasing again. The reason the payoff starts decreasing at $a = 0.75$ is that the sender starts adding to the support states with the high density $g(w) = H$ which are relatively cheap for the sender because Nature has less freedom in choosing the prior on them. The payoff starts increasing again at $a = 0.9$ because this is the point at which the signal attains full support, and further increases in the sender’s ignorance only cause the sender to recommend the high action with a lower probability.

6 Support of the Optimal Signal

In this section I show that recommending approval with a strictly positive probability in every state is a robust property of the signal chosen by an ignorant sender. I also consider the limits of the optimal information structure as the ignorance of the sender approaches its maximal and minimal values respectively and provide examples of the support for different parameters.

6.1 Detail-Free Knowledge and Full Support

Recall that the sender’s knowledge is detail-free if the reference prior $G$ puts mass zero on the states below the threshold $w^*$. Proposition 3 shows that, more generally,
the distribution over the receiver’s actions induced by an optimal information structure is independent of the sender’s prior if and only if the sender’s knowledge is detail-free. That is, the distribution is independent of the sender’s prior if and only if the receiver may consider all states on which he disagrees with the sender with regard to the optimal action impossible.

**Proposition 3.** A sender-optimal information structure induces a distribution over the receiver’s actions that is independent of the sender’s prior $F_s$ if and only if the sender’s knowledge is detail-free.

What is the intuition for this result? To understand the intuition, consider a simple set of priors where Nature has to put probability 0.1 on some state $w < w^*$ and can put a mass of no more than 0.5 on states below $w^*$. Then there is a benefit to giving up and not recommending approval in state $w$: if the sender gives up on $w$, then Nature can put a mass of no more than 0.4, rather than 0.5, on the support below $w^*$ of the signal realization recommending approval. The cost of giving up is that the sender would not collect the payoff from approval in state $w$. The sender will not want to give up on a state if her prior on that state is high enough and might want to give up otherwise. On the other hand, if the sender’s knowledge is detail-free, so that Nature can put a mass of no more than 0.5 on states below $w^*$ and there are no other restrictions, then there is no benefit to giving up on the states. This is because no matter which states the sender gives up on, Nature can still put a mass of 0.5 on states below $w^*$.

### 6.2 Maximal Ignorance and Full Support

The next Proposition describes how the support of the optimal signal behaves as the sender becomes very ignorant.

**Proposition 4.** As the ignorance index $a$ goes to 1, $\Omega_1(t)$ converges to $[l, w^*)$. Moreover, if $f_s$ is bounded away from zero, then $\Omega_1(t) = [l, w^*)$ for all $a$ sufficiently close to 1.

Proposition 4 shows that, as the ignorance of the sender approaches its maximal value of 1, in the limit the optimal signal recommends the high action with a strictly positive probability in all states. Moreover, if the density of the sender’s prior is bounded away from zero, then the signal has full support not just in the limit but for all sufficiently high levels of the sender’s ignorance. This clarifies the sense in which recommending the high action with a strictly positive probability in all states is a robust property of the signal chosen by an ignorant sender.
The intuition for the result is that as the set of the receiver’s priors converges to the whole set, the probability with which approval is recommended in the states in which the sender and the receiver disagree converges to zero. Intuitively, this is because only very convincing can guard against very pessimistic priors. As the probability of approval recommendation decreases, a receiver who was willing to approve before now strictly prefers to approve. Then we can expand the support of the signal realization recommending approval a little bit, and this receiver is still willing to approve. This explains why, as the probability of approval recommendation decreases to zero, the support converges to the whole state space.

6.3 Minimal Ignorance and Continuity

Here I ask how the optimal information structure behaves as the ignorance of the sender converges to zero. It can be shown that the sender-optimal information structure when the receiver has a commonly known prior with a density \( g \) (and the sender has a commonly known prior with a density \( f_s \)) is partitional (so, conditional on each state, approval happens with probability 0 or 1), the set of states where approval happens with probability 1 is \( \Omega(t) \), where the threshold \( t \) is pinned down by the requirement that the receiver is indifferent after she sees the recommendation to approve. Proposition 5 establishes a continuity result as the ignorance of the sender converges to zero.

**Proposition 5.** As \( a \to 0 \), the signal that is optimal under an unknown prior converges to \( \pi_0 \).

6.4 Examples of Support

Here I provide examples of the optimal information structures for different parameters. Naturally, this is not an exhaustive list of the possible cases – instead, it is meant to illustrate the possible features of the optimal signal.

The first question I ask here is: when is the support of the optimal signal connected? The support is connected if the importance index \( \frac{f_s(w)}{g(w)(w^* - w)} \) is increasing on \([l, w^*] \). Thus, for example, the sender’s prior \( f_s \) and the reference prior \( g \) being equal is sufficient to ensure that the support of the optimal signal is connected. The sender’s prior \( F_s \) dominating the prior \( G \) in the monotone likelihood ratio property sense is also sufficient for the support to

---

\^8 Assumption 1 guarantees that the signal \( \pi_0 \) is essentially unique.
Figure 3: Optimal Information Policy: Connected Support of the Signal, \( w \mapsto \frac{f_s(w)}{g(w)(w^*-w)} \) is Increasing

\[
\pi(\sigma|\omega)
\]

be connected. Figure 3 shows an example of the optimal information policy when \( \frac{f_s(w)}{g(w)(w^*-w)} \) is increasing.

Interestingly, the support of the optimal signal consists of exactly two intervals of the form \([l, w'] \cup [w^*, h]\) if the importance index \( \frac{f_s(w)}{g(w)(w^*-w)} \) is decreasing on \([l, w^*]\). That is, the optimal signal cuts out an interval in the middle if the sender perceives the lower states to be sufficiently more likely and the receiver perceives the lower states to be sufficiently less likely. Observe that the importance index \( \frac{f_s(w)}{g(w)(w^*-w)} \) being decreasing on \([l, w^*]\) requires that \( f_s \) converges to zero as \( w \) converges to \( w^* \), so that the sender’s prior puts very little mass on the states right below the threshold \( w^* \). Figure 4 shows an example of the optimal information policy when \( \frac{f_s(w)}{g(w)(w^*-w)} \) is decreasing.

As a final example of the support of the optimal signal, observe that the support consists of two intervals of the form \([l, w'] \cup [w'', h]\) for \( w'' \leq w^* \) if the importance index \( \frac{f_s(w)}{g(w)(w^*-w)} \) is single-dipped on \([l, w^*]\). Figure 5 shows an example of the optimal information policy when \( \frac{f_s(w)}{g(w)(w^*-w)} \) is single-dipped.
Figure 4: Optimal Information Policy: A Gap in the Middle, 
\[ w \mapsto \frac{f_s(w)}{g(w)(w^*-w)} \] is Decreasing

\[ \pi(\sigma|\omega) \]

\[ \omega \]

\[ w^* \]

Figure 5: Optimal Information Policy: A Gap in the Middle, 
\[ w \mapsto \frac{f_s(w)}{g(w)(w^*-w)} \] is Single-Dipped

\[ \pi(\sigma|\omega) \]

\[ \omega \]

\[ w^* \]

20
7 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 (2016a) consider a model of Bayesian persuasion where the sender and the receiver have heterogeneous priors. Alonso and Camara (2016b) and Chan et al. (2019) 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. Guo and Shmaya (2019) also consider persuasion of a privately informed receiver and show that the optimal mechanism takes the form of nested intervals. 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.

One strand of the Bayesian persuasion literature studies models where the distribution of the state is endogenous: after observing the information structure chosen by the sender, the agent can take an action affecting the distribution of the state. Thus Rosar (2017) studies the design of tests when privately informed agents can opt out of taking a test. Boleslavsky and Kim (2018) and Rodina (2017) consider models where the prior distribution of the state depends on the agent’s effort. Perez-Richet and Skreta (2018) characterize information structures optimal for the receiver when the persuader can manipulate the state. These papers are similar to the present paper in that there is an agent that can affect the distribution of the state, which changes the prior belief of the receiver (in my model, this agent is Nature). The papers differ in the objective that the agent has, the way in which the agent can change the prior, as well as the characterization of the optimal signal.

Several papers consider models related to ambiguity in Bayesian persuasion. Laclau and Renou (2017) consider a model of publicly persuading receivers with heterogeneous priors under the unanimity rule. Their model is equivalent to a model of persuading a single receiver who has multiple priors where, after the sender commits to an information structure and after a signal realization from this information structure is realized, Nature chooses the receiver’s prior to minimize the sender’s payoff. In contrast, in the present paper Nature
chooses the receiver’s prior after the sender commits to an information structure but before a signal realization from this information structure is realized. The difference has several important implications. First, the model of Laclau and Renou (2017) has a concave closure characterization, which means that standard methods can be applied to solve it. In contrast, the model in the present paper, in general, does not have a concave closure characterization. In order to solve my model, I develop a novel fixed-point argument. Second, the model in the present paper has the interpretation of the sender not knowing the receiver’s beliefs and designing an information structure that is robust to this lack of knowledge, while the model of Laclau and Renou (2017) does not have this interpretation.

A paper by Hu and Weng (2018) features a sender persuading a receiver who will receive private information unknown by the sender. The sender is ambiguity averse with regard to this private information and believes that Nature chooses the distribution of the private information to minimize the sender’s payoff. The differences from the present paper are as follows. First, in the paper by Hu and Weng (2018) Nature chooses a distribution of the receiver’s private information (that is, a distribution over the receiver’s posteriors subject to the constraint that the expectation is equal to the common prior), whereas in the present paper Nature chooses the prior of the receiver. This leads to significant differences in the results: because in Hu and Weng (2018) Nature engages in information design to maximize the negative of the sender’s payoff, unlike in the present paper, the solution to the Nature’s problem is a concave closure of the negative of the sender’s induced utility function (which is equal to the convex closure of the sender’s induced utility function). Second, Hu and Weng (2018) solve for the optimal information structure in the example with two states (and two actions), whereas I am able to characterize the optimal information structure in a setting with a continuum of states. Moreover, the nature of the optimal information structure in the present paper differs markedly from that in the example in Hu and Weng (2018). Finally, the proof techniques differ: the novel fixed-point technique is unique to the present paper.

Beauchêne et al. (2019) consider a model in which the sender and the receiver share a common prior, but the sender can commit to ambiguous information structures, and both the sender and the receiver are ambiguity averse. They show that the sender can benefit from committing to ambiguous signals. The model of Beauchêne et al. (2019) cannot be interpreted as one where the sender does not know the receiver’s beliefs.

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 in which the prior belief of the receiver is unknown to the sender.

8 Conclusion

This paper analyzes a model of Bayesian persuasion with unknown beliefs and characterizes the optimal signal. The paper finds that there is an optimal signal with two realizations. The support below $w^*$ of the signal realization recommending the high action is given by all states below $w^*$ with an importance index exceeding a threshold. On the support, the probability that the signal recommends the high action is a hyperbola. More ignorant senders induce the high action in more states but with a lower probability. An increase in the sender’s ignorance can hurt both the sender and the receiver. If the sender’s knowledge is detail-free, the optimal signal recommends the high action with a strictly positive probability in every state.

These distinct features of the robust signal can potentially be useful for empirically distinguishing between senders who know the prior of the receiver and those who do not have this knowledge. When the sender is ignorant, especially pernicious outcomes are possible: even the least safe drugs are approved with a positive probability. Moreover, the results have important implications for the optimal transparency requirements for the FDA. While full transparency is never optimal, maximal obfuscation may not be a good choice either: an intermediate level of transparency may be the best.
Appendix

(For Online Publication)

B Notation and Definitions

Define $\mathcal{D} = \{\pi : |\text{supp}\pi(\cdot|\omega)| < \infty \text{ for all } \omega \in [l, h]\}$. Define $\mathcal{F}_{a,g} = \{F \in \varphi : \mu_F = \mu_\varphi + (1 - a)\mu_G, \text{supp} \mu_\varphi \subseteq [l, w^*]\}$. Given $\omega \in [l, h]$, define $\mu_{F_\omega} = (1-a)\mu_G + a\delta_\omega$.

Given an information structure $\pi \in \mathcal{D}$ and a state $\omega$, define $S_\omega = \{k : \pi(\sigma_k|\omega) > 0\}$. Given an information structure $\pi \in \mathcal{D}$, define $R(\pi) = \{i : \int_{w^*}^h \pi(\sigma_i|w)g(w)dw > 0\}$. Given an information structure $\pi \in \mathcal{D}$ and a prior $F$, define $R(F, \pi) = \{i \in R(\pi) : E_{F,\pi}[\omega|\sigma_i] \geq w^*\}$.

Given $\pi \in \mathcal{D}$, define $\Omega^*_1 = \{w \in [l, w^*] : \pi(\sigma_i|w) > 0 \text{ for some } i \in R(\pi)\}$. Given $\pi \in \mathcal{D}$, define $\Omega^*_2 = \{w \in \Omega^*_1 : \sum_{i \in R(\pi)} \pi(\sigma_i|w) = 1\}$. Define a function $t \mapsto \Omega_1(t)$ as $\Omega_1(t) = \left\{w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} \geq t\right\}$. Define a function $(t, z) \mapsto \Omega_1(t, z)$ as $\Omega(t, z) = \left\{w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} = t \text{ and } w \geq z\right\}$. Define a correspondence $t \mapsto \overline{\Theta}_1(t)$ as $\overline{\Theta}_1(t) = \left\{w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} > t\right\} \cup \Omega_0$, $\Omega_0 \in B\left(\left\{w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} = t\right\}\right)$.

Given an information structure $\pi$ with two realizations, $\sigma$ and $\sigma_0$, such that $\pi(\sigma_0|w) = 0 \mu_{F^*}$-a.e. $w \in [w^*, h]$ and a state $w' \in [l, w^*]$, define $Z_\pi(w') = \text{min}\left\{\int_{[w^*, h]}(w-w')g(w)dw - \int_{[l,w^*]} g(w)\pi(\sigma|w)(w^*-w)dw, 1\right\}$. Define a function $\Omega_1 \mapsto \hat{c}(\Omega_1)$ as $\hat{c}(\Omega_1) = \frac{\int_{[w^*, h]}(w-w')g(w)(w^*-w)dw - \int_{\Omega_1 \cap [w^*-c, w^*]} g(w)(w^*-w)dw}{\int_{[w^*, h]}(w-w')g(w)(w^*-w)dw} \cdot \frac{\int_{[l,w^*]} g(w)\pi(\sigma|w)(w^*-w)dw}{\int_{[l,w^*]} g(w)\pi(\sigma|w)(w^*-w)dw}$. Define a function $t \mapsto c(t)$ as $c(t) = \frac{\int_{[w^*, h]}(w-w')g(w)(w^*-w)dw - \int_{\Omega_1 \cap [w^*-c, w^*]} g(w)(w^*-w)dw}{\int_{[w^*, h]}(w-w')g(w)(w^*-w)dw} \cdot \frac{\int_{[l,w^*]} g(w)\pi(\sigma|w)(w^*-w)dw}{\int_{[l,w^*]} g(w)\pi(\sigma|w)(w^*-w)dw}$. Define a function $(c, t) \mapsto x(c, t)$ as $x(c, t) = \int_{[w^*, h]}(w-w')g(w)dw - \int_{\Omega_1 \cap [w^*-c, w^*]} g(w)(w^*-w)dw - c\left(\frac{\alpha}{1- \alpha} + \int_{\Omega_1 \cap [l, w^*-c]} g(w)dw\right)$. Define functions $t \mapsto J(t)$ and $t \mapsto y(t)$ as $J(t) = \frac{\int_{\Omega_1 \cap [l, w^*-c]} f_s(w)dw}{\int_{[l, w^*-c]} g(w)dw} \quad \text{and} \quad y(t) = t\left(\frac{\alpha}{1- \alpha} + \int_{\Omega_1 \cap [l, w^*-c]} g(w)dw\right) - \int_{\Omega_1 \cap [l, w^*-c]} \frac{f_s(w)}{w^*-w}dw$. Define a correspondence $t \mapsto \overline{J}(t)$ as $\overline{J}(t) = \frac{\int_{\Omega_1 \cap [l, w^*-c]} f_s(w)dw}{\int_{[l, w^*-c]} g(w)dw}$, $\Omega_1 \in \overline{\Theta}_1(t)$. Define a correspondence $t \mapsto \overline{y}(t)$ as $\overline{y}(t) = t\left(\frac{\alpha}{1- \alpha} + \int_{\Omega_1 \cap [l, w^*-c]} g(w)dw\right) - \int_{\Omega_1 \cap [l, w^*-c]} \frac{f_s(w)}{w^*-w}dw$, $\Omega_1 \in \overline{\Theta}_1(t)$. Define $t = \text{min}_{w \in [l, w^*]} g(w)(w^*-w)$. 24
C Counterexamples

In this section I provide counterexamples, showing that there may not exist an optimal information structure with only two realizations if the assumptions on the sender’s utility function and on the set of the receiver’s priors made in the present paper fail.

I first show that there exist utility functions of the sender such that any sender-optimal information structure must have more than two realizations. Suppose that the state space \( \Omega = \{0, 1\} \) is binary, so that the receiver’s prior \( f \) is the probability that the state is 1. Suppose also that the sender’s utility \( u(E) \) as a function of the receiver’s posterior belief \( E \) is \( u(E) = 1 \) for \( E = w^* \) for some \( w^* \in (0, 1) \) and \( u(E) = 0 \) for \( E \in [0, 1] \setminus w^* \). The set of the receiver’s priors is a finite set \( \{f_1, \ldots, f_k\} \) satisfying \( 0 < f_1 < \ldots < f_k < 1 \).

Observe that the set of the signal realizations of a sender-optimal information structure in this setting must have the cardinality no smaller than the cardinality of the set of priors. In particular, letting \( \sigma_i \) denote a signal realization such that, upon observing this signal realization, a receiver with prior \( f_i \) has the posterior belief \( w^* \), we find that the set of signal realizations must contain the set \( \{\sigma_i\}_{i \in \{1, \ldots, k\}} \). To see why, consider an information structure that does not contain \( \sigma_i \) for some \( i \in \{1, \ldots, k\} \). Then if Nature chooses the receiver’s prior \( f_i \), the receiver’s posterior is not equal to \( w^* \) after any signal realization, implying that the sender obtains a payoff of zero. However, the sender can do strictly better by choosing an information structure that sends the signal realization \( \sigma_i \) with a strictly positive probability.

Next, I provide an example showing that in a setting with multiple states where the utility functions are the same as in the model in the present paper, if the receiver’s set of priors does not satisfy the assumptions in the present paper, then there may not exist an optimal signal with only two realizations.

Suppose that there are five states, 0, 1, 2, 3 and 4. The receiver takes the high action if and only if the receiver’s expectation given the signal realization is weakly greater than \( w^* = 2.5 \). Two priors of the receiver are possible: \( f_1 = \left( \frac{1}{4}, \frac{1}{24}, \frac{1}{2}, \frac{5}{24}, 0 \right) \) and \( f_2 = \left( \frac{1}{4}, \frac{1}{2}, \frac{3}{16}, 0, \frac{1}{16} \right) \). The sender’s prior \( f_s = f_2 \) coincides with the second possible prior of the receiver.

Consider an information structure \( \pi \) with three realizations, \( \sigma_1, \sigma_2 \) and \( \sigma_0 \), such that \( \pi(\sigma_2|1) = 1, \pi(\sigma_2|3) = 1 \) and \( \pi(\sigma_1|2) = 1, \pi(\sigma_1|4) = 1 \). Note that \( E_{F_1, \pi}[\omega|\sigma_2] \geq w^* \), which implies that the sender’s payoff given \( \pi \) and the receiver’s prior \( f_1 \) is greater than \( \pi(\sigma_2|1)f_s(1) + \pi(\sigma_2|3)f_s(3) = f_s(1) + f_s(3) = f_2(1) + f_2(3) = \frac{1}{2} \). On the other hand, \( E_{F_2, \pi}[\omega|\sigma_2] < w^* \) and \( E_{F_2, \pi}[\omega|\sigma_1] \geq w^* \), which implies that the sender’s payoff given \( \pi \) and
\( f_s \) is \( \pi(\sigma_1|2)f_s(2) + \pi(\sigma_1|4)f_s(4) = f_s(2) + f_s(4) = f_2(2) + f_2(4) = \frac{1}{4}. \)

Then, faced with the information structure \( \pi \), Nature strictly prefers to choose the second prior of the receiver \( f_2 \), which yields a payoff of \( \frac{1}{4} \) to the sender. It can be shown that there do not exist information structures with two realizations that guarantee a payoff of \( \frac{1}{4} \) to the sender. This shows that in this example there does not exist a sender-optimal information structure with two realizations. Importantly, the set of the receiver’s priors in the example is non-convex, whereas the conditions imposed in the present paper ensure the convexity of the set of priors.\[9\]

**D Proofs**

**Lemma 1.** \( \sum_{i \in R(\pi)} \int_{[w^*, h]} (w - w^*) g(w) \pi(\sigma_i|w) dw \leq \int_{[w^*, h]} (w - w^*) g(w) dw. \) Moreover, if \( \sum_{i \in R(\pi)} \pi(\sigma_i|w) = 1 \) for \( \mu_{F_s} \)-a.e. \( w \in [w^*, h] \) and \( f_s \) has full support, then \( \sum_{i \in R(\pi)} \int_{[w^*, h]} (w - w^*) g(w) \pi(\sigma_i|w) dw = \int_{[w^*, h]} (w - w^*) g(w) dw. \)

**Proof of lemma 1.** The first part follows from the fact that \( \sum_{i \in R(\pi)} \pi(\sigma_i|w) \leq 1 \) for all \( w \in [w^*, h] \), the second part is immediate. \[\blacksquare\]

**Lemma 2.** Suppose that there exists a information structure \( \pi \in \mathcal{D} \) such that

1. \( |R(\pi)| > 1 \)
2. \( \sum_{i \in R(\pi)} \pi(\sigma_i|w) = 1 \) \( \mu_{F_s} \)-a.e. \( w \in [w^*, h] \)
3. for any \( i \in R(\pi) \) there exists \( F \in C_{a, g} \) such that \( E_{F, \pi}[w|\sigma_i] \geq w^* \) and Nature finds it optimal to choose \( F \).

Then there exists \( F \in F_{a, g} \) such that for all \( w' \in [l, w^*] \) we have \( \sum_{i \in R(F, \pi)} \pi(\sigma_i|w') \leq \frac{\int_{[w^*, h]} (w - w^*) g(w) dw - \int_{[w^*, h]} g(w) (\sum_{i \in R(F, \pi)} \pi(\sigma_i|w)) (w^* - w) dw}{(w^* - w')^\frac{1}{2}}. \)

**Proof of lemma 2.**

Without loss of generality, suppose that for all \( i \in R(\pi) \) we have \( \pi(\sigma_i|\omega) > 0 \) for some \( \omega \in [l, w^*] \).

**Step 1: Partitioning the set \( \Omega^+_1. \)**

\[9\]I conjecture, but have not been able to prove, that if the set of the receiver’s prior beliefs is convex, then there exists an optimal information structure with only two realizations.
Given $\pi \in D$, $A \subseteq R(\pi)$ and $\omega \in [l, w^*)$, define $Z^A_\pi(\omega) = \frac{\int_{[w^*, h]}(w-w^*)g(w)dw - \int_{\Omega_1} g(w)(\sum_{i \in A} \pi(\sigma_i|w))(w-w^*)dw}{(w^*-w)\sum_{i \in A} \pi(\sigma_i|w)}$. For each $\omega \in [l, w^*)$, define $C(\omega) = \{A \subseteq R(\pi) : \sum_{i \in A} \pi(\sigma_i|\omega) > Z^A_\pi(\omega)\}$.

Define a finite number of sets $Y_1, \ldots, Y_J$ such that $Y_i \subseteq \Omega_1$ for all $i \in \{1, \ldots, J\}$ as follows:

1. $\omega, \omega' \in Y_i$ implies that, for all $j \in R(\pi)$, if $\pi(\sigma_j|\omega) > 0$, then $\pi(\sigma_j|\omega') > 0$;
2. $\omega, \omega' \in Y_i$ 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 $\sigma_l$ such that $l \in R(\pi)$ and either $\pi(\sigma_l|\omega_i) > 0$, $\pi(\sigma_l|\omega_j) = 0$ or $\pi(\sigma_l|\omega_j) > 0$ and $\pi(\sigma_l|\omega_i) = 0$
   
   (b) or $C(\omega_i) \neq C(\omega_j)$

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

**Step 2: Some notation and reference states.**

Define $O_i = \{\omega \in W : \pi(\sigma_i|\omega) > 0\}$. For all $w \in W$, define $b_w = w^* - w$. For each $i \in R(\pi)$, choose a reference state $\omega(i) \in W$ such that $\pi(\sigma_i|\omega(i)) > 0$. Given $i \in R(\pi)$ with a reference state $\omega(i)$ and $w \in W$, define $\lambda_{iw} = \frac{\pi(\sigma_i|w)}{\pi(\sigma_i|\omega(i))}$. Finally, for each $i \in R(\pi)$, we define $Q_i = \int_{\Omega_1} b_w \pi(\sigma_i|w)g(w)dw$, $Q = \sum_{i \in R(\pi)} Q_i$ and $\eta_i = \frac{Q_i}{Q}$.

**Step 3:** $\int_{[w^*, h]}(w-w^*)g(w)\pi(\sigma_i|w)dw - \eta_iQ \geq 0$ for all $i \in R(\pi)$.

**Step 4: Defining weights for signal realizations.**

In this step I define a collection of weights $\{\epsilon_i\}_{i \in R(\pi)}$ for signal realizations with indices in $R(\pi)$ satisfying $\epsilon_i \in [0, 1]$ for all $i \in R(\pi)$ and $\sum_{i \in R(\pi)} \epsilon_i = 1$. For each $i \in R(\pi)$, define a weight $\epsilon_i$ as $\epsilon_i = \frac{\int_{[w^*, h]} g(w)(w-w^*)\pi(\sigma_i|w)dw - \eta_iQ}{\int_{[w^*, h]} g(w)(w-w^*)dw - Q}$. Note that Step 3 implies that $\epsilon_i \geq 0$.

Lemma 1 implies that if $\sum_{i \in R(\pi)} \pi(\sigma_i|w) = 1$ for $\mu_{F_s}$-a.e. $w \in [w^*, h]$ and $f_s$ has full support, then $\sum_{i \in R(\pi)} \int_{[w^*, h]}(w-w^*)g(w)\pi(\sigma_i|w)dw = \int_{[w^*, h]}(w-w^*)g(w)dw$. Then, because

\footnotesize
\cite{10}The number of sets is finite because $R(\pi)$ is finite.

27
\[ \sum_{i \in R(\pi)} \pi(\sigma_i | w) = 1 \] for \( \mu_F \)-a.e. \( w \in [w^*, h] \) and, by assumption, \( f_* \) has full support, we have \( \sum_{i \in R(\pi)} \int_{[w^*, h]} (w - w^*) g(w) \pi(\sigma_i | w) dw = \int_{[w^*, h]} (w - w^*) g(w) dw \). This and the fact that \( \sum_{i \in R(\pi)} \frac{Q_i}{Q} = \frac{Q}{Q} = 1 \) imply that \( \sum_{i \in R(\pi)} \epsilon_i = 1 \).

**Step 5: Nature’s strategy.**

Suppose that Nature chooses the receiver’s prior \( \mu_F \) that places \( (1 - a) \mu_G \) on states \([l, h]\) and, in addition, for all \( i \in R(\pi) \), places the mass of \( \nu_i \epsilon_i a \) on each \( \omega \in O_i \) such that \( \nu_i \epsilon_i \in [0, 1] \) for all \( \omega \in W \) satisfying \( \omega \in O_i \) and \( \sum_{\omega \in O_i} \nu_i \epsilon_i = 1 \).

Note that, by construction, \( O_i \neq \emptyset \) for all \( i \in R(\pi) \). Then we have \( \sum_{i \in R(\pi)} \sum_{\omega \in O_i} \nu_i \epsilon_i a = \sum_{i \in R(\pi)} \epsilon_i a = a \), where the last equality follows from the fact that \( \sum_{i \in R(\pi)} \epsilon_i = 1 \). Thus Nature places the mass of \( a + (1 - a) \int_{[l, w^*]} g(w) dw \) on states in \([l, w^*]\).

Observe that \( \mu_F \in \mathcal{F}_{a, g} \).

**Step 6:** \( E_F[\omega|\sigma_i] \geq w^* \) is equivalent to

\[
\frac{\int_{[w^*, h]} g(w) w \pi(\sigma_i | w) dw + a \sum_{\omega \in O_i} \omega \pi(\sigma_i | \omega) \sum_{k \in S \cap R(\pi)} \frac{\nu_k \epsilon_k}{l + \alpha}}{\int_{[w^*, h]} g(w) \pi(\sigma_i | w) dw + a \sum_{\omega \in O_i} \pi(\sigma_i | \omega) \sum_{k \in S \cap R(\pi)} \frac{\nu_k \epsilon_k}{l + \alpha}} \geq w^*. 
\]

**Step 7:** \( E_F[\omega|\sigma_i] \geq w^* \) given Nature’s strategy is equivalent to \( \pi(\sigma_i | v) \leq \frac{\lambda_{w} b_{w} \sum_{k \in S \cap R(\pi)} \nu_{k} \epsilon_{k}}{1 - a \sum_{\omega \in O_i} \lambda_{w} b_{w} g_{\omega} - Q_{\omega}} \).

**Step 8: Defining the receiver’s prior.**

Given \( w \in W \), let \( z_w = \sum_{k \in S \cap R(\pi)} \nu_{w} \epsilon_{k} \). For each \( i \in R(\pi) \), choose \( \nu_{w} \) such that

\[
\frac{\epsilon_i \lambda_{w} b_{w} z_w}{\sum_{\omega \in O_i} \lambda_{w} b_{w} g_{\omega}} = \frac{\nu_{w} \epsilon_i}{z_w}. \]

This implies that we set \( \nu_{w} = \frac{\lambda_{w} b_{w} z_w}{\sum_{\omega \in O_i} \lambda_{w} b_{w} g_{\omega}} \).

**Step 9:** The receiver’s prior is well-defined.

Next, we will show that the collection of weights \( \{ \nu_{w} \}_{i \in R(\pi), w \in O_i} \) is well-defined. Observe that \( \sum_{w \in O_i} \nu_{w} = 1 \). We will show that there exists a collection of weights satisfying \( \nu_{w} = \frac{\lambda_{w} b_{w} z_w}{\sum_{\omega \in O_i} \lambda_{w} b_{w} g_{\omega}} \) such that \( \nu_{w} \in [0, 1] \) for all \( i \in R(\pi) \) and for all \( w \in W \) satisfying \( w \in O_i \).

Let \( T = \&_{i \in R(\pi)} \Delta(O_i) \) where \( \Delta(O_i) \) denotes the simplex over \( O_i \) and \( \& \) denotes the
Cartesian product. Let $\nu = \{\nu^i_w\}_{i \in R(\pi), w \in O_i}$. Define a mapping $H : T \rightarrow T$ by $H_{iw}(\nu) = \frac{\lambda_{iw} b_w \left( \sum_{k \in S_w \cap R(\pi)} \nu^i_k \epsilon_k \right)}{\sum_{O_i} \lambda_{iw} b_w \left[ \sum_{k \in S_w \cap R(\pi)} \nu^i_k \epsilon_k \right]}
$. Then we can write $\nu^i_w = \frac{\lambda_{iw} b_w \left( \sum_{k \in S_w \cap R(\pi)} \nu^i_k \epsilon_k \right)}{\sum_{O_i} \lambda_{iw} b_w \left[ \sum_{k \in S_w \cap R(\pi)} \nu^i_k \epsilon_k \right]}$ as $\nu^i_w = H_{iw}(\nu)$. Thus the weights $\{\nu^i_w\}_{i \in R(\pi), w \in O_i}$ are defined by the equation $\nu = H(\nu)$. 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 $T = \times_{i \in R(\pi)} \Delta(O_i)$ 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 10:** Bound on $\sum_{i \in S_w \cap R(\pi)} \left( \frac{\epsilon_i \lambda_{iw}}{\sum_{O_i} \lambda_{iw} b_w z_w} \right)$.

Define $\tau_{iw} = \frac{\epsilon_i \lambda_{iw}}{\sum_{O_i} \lambda_{iw} b_w z_w}$. Then $\frac{\epsilon_i \lambda_{iw} b_w}{\sum_{O_i} \lambda_{iw} b_w} = \frac{\nu^i_w \epsilon_i}{z_w}$ implies that $\tau_{iw} b_w = \frac{\nu^i_w \epsilon_i}{z_w}$, which is equivalent to $\tau_{iw} = \frac{\nu^i_w \epsilon_i}{z_w}$. Thus $\sum_{i \in S_w \cap R(\pi)} \tau_{iw} = \frac{\nu^i_w \epsilon_i}{z_w}$, which holds for all $w \in O_i$, for all $i \in R(F, \pi)$. This is equivalent to $\pi(\sigma_i | w) \leq \tau_{iw} \frac{1}{1-a} \left( \int_{[w^*, b]} (w - w^* g(w)) d\omega - Q \right)$.

Observe that the fact that $\int_{[w^*, h]} (w - w^* g(w)) d\omega (\sigma_i | w) dw - \eta_i Q \geq 0$ for all $i \in R(\pi)$ by Step 3 and the fact that $\int_{[w^*, h]} (w - w^* g(w)) d\omega - Q = \sum_{i \in R(\pi)} \left( \int_{[w^*, h]} (w - w^* g(w)) d\omega (\sigma_i | w) dw - \eta_i Q \right)$ imply that $\int_{[w^*, h]} (w - w^* g(w)) d\omega - Q \geq 0$. Then we have $\sum_{i \in S_w \cap R(F, \pi)} \pi(\sigma_i | w) = \frac{1}{b_w} \frac{1}{1-a} \left( \int_{[w^*, h]} (w - w^* g(w)) d\omega - Q \right)$.

Observe that $\sum_{i \in R(F, \pi)} \pi(\sigma_i | w) \leq \frac{\int_{[w^*, h]} (w - w^* g(w)) d\omega - Q}{\frac{1}{1-a} b_w}$ holds for all $w \in W$. Let $Q = \sum_{i \in R(F, \pi)} Q_i$. Then, because $R(F, \pi) \subseteq R(\pi)$ and $Q_i \geq 0$ for all $i \in R(\pi)$, we have $Q = \sum_{i \in R(F, \pi)} Q_i \leq \sum_{i \in R(\pi)} Q_i = Q$. Therefore, the above implies that for all $w \in W$ we have $\sum_{i \in R(F, \pi)} \pi(\sigma_i | w) \leq \frac{\int_{[w^*, h]} (w - w^* g(w)) d\omega - \overline{Q}}{\frac{1}{1-a} b_w}$.  

29
Step 12: Conclusion of the proof.

Suppose for the sake of contradiction that for some \( w \in [l, w^*] \) we had
\[
\sum_{i \in R(F, \pi)} \pi(\sigma_i | w) > \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \bar{Q}}{\frac{a}{1-a} b_w}.
\]
By definition of \( C(w) \), this is equivalent to saying that \( R(F, \pi) \in C(w) \).

Let \( Y(w) \) denote the element \( Y_i \) of the partition \( Y_1, \ldots, Y_J \) such that \( w \in Y_i \). Observe that, because \( \sum_{i \in R(F, \pi)} \pi(\sigma_i | w) > \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \bar{Q}}{\frac{a}{1-a} b_w} \), the fact that \( \sum_{i \in R(F, \pi)} \pi(\sigma_i | w) \leq \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \bar{Q}}{\frac{a}{1-a} b_w} \) by Step 11 implies that \( w \notin W \). Let \( w^* \) denote the element of \( W \) such that \( w^* \in Y(w) \). Then, because \( C(w) = C(\omega) \) for all \( \omega, \omega' \in Y(w) \), it must be the case that \( R(F, \pi) \in C(w^*) \). By definition of \( C(w^*) \), this is equivalent to \( \sum_{i \in R(F, \pi)} \pi(\sigma_i | w^*) > \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \bar{Q}}{\frac{a}{1-a} b_w} \). However, this contradicts \( \sum_{i \in R(F, \pi)} \pi(\sigma_i | w) \leq \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \bar{Q}}{\frac{a}{1-a} b_w} \) for all \( w \in [l, w^*] \), as required.

Lemma 3. Let \( \mathcal{F} \) denote the set of the extreme points of \( \mathcal{F}_{a,g} \). Suppose that \( \pi \) is an information structure with two realizations, \( \sigma \) and \( \sigma_0 \). If \( E_{F,\pi}[\omega|\sigma] \geq w^* \) for all \( F \in \mathcal{F} \), then \( E_{F,\pi}[\omega|\sigma] \geq w^* \) for all \( F \in \mathcal{F}_{a,g} \).

Proof of lemma 3. Fix \( F \in \mathcal{F}_{a,g} \). Because \( \mathcal{F} \) is the set of the extreme points of \( \mathcal{F}_{a,g} \), we have \( \mathcal{F} = \{ F_w : w \in [l, w^*] \} \). Then there exists a collection \( \{ \alpha_w \}_{w \in [l, w^*]} \) such that \( \int_{w \in [l, w^*]} \alpha_w dw = 1 \), \( \mu_F = \int_{w \in [l, w^*]} \alpha_w \mu_{F_w} dw \) and \( \alpha_w \in [0, 1] \), \( \mu_{F_w} = (1-a)\mu_{G} + a\delta_w \) for all \( w \in [l, w^*] \).

Because \( E_{F,\pi}[\omega|\sigma] \geq w^* \) for all \( F_w \in \mathcal{F} \), for all \( \omega \in [l, w^*] \), we have that
\[
\frac{\int_{\Omega} \pi(\sigma|w) d\mu_{F,w}(w)}{\int_{\Omega} \pi(\sigma|w) d\mu_{F,w}(w)} \geq \min_{\omega \in [l, w^*]} \frac{\int_{\Omega} \pi(\sigma|w) d\mu_{F,w}(w)}{\int_{\Omega} \pi(\sigma|w) d\mu_{F,w}(w)} \geq w^*.
\]

Lemma 4. If \( \pi \) is an information structure with two realizations satisfying \( \pi(\sigma|w) = 1 \) \( \mu_{F_w} \)-a.e. \( w \in [w^*, h] \), then, given \( w' \in [l, w^*] \), \( E_{F_{w'},\pi}[\omega|\sigma] \geq w^* \) is equivalent to \( \pi(\sigma|w') \leq \min \left\{ \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \int_{\Omega_{\pi}^1 \cup \Omega_{\pi}^2} g(w) \pi(\sigma|w)(w^*-w) dw}{(w^*-w') \frac{a}{1-a}}, 1 \right\} \). If \( \pi \) sender-optimal in the class of all information structures with two realizations, then it satisfies \( \pi(\sigma|w') \leq \min \left\{ \frac{\int_{[w^*, h]} (w-w^*) g(w) dw - \int_{\Omega_{\pi}^1 \cup \Omega_{\pi}^2} g(w) \pi(\sigma|w)(w^*-w) dw}{(w^*-w') \frac{a}{1-a}}, 1 \right\} \) for all \( w' \in \Omega_{\pi}^1 \) and \( \pi(\sigma|w) = 1 \) \( \mu_{F_w} \)-a.e. \( w \in [w^*, h] \).

Proof of lemma 4. The first part of the lemma is immediate. For the proof of the second part of the lemma observe that, if \( \pi \) is sender-optimal in the class of all information
structures with two realizations, we have \( \pi(\sigma|w) = 1 \mu_{F_s}\)-a.e. \( w \in [w^*, h] \) because the sender’s payoff \( \int_{[l,h]} \pi(\sigma|w)f_s(w)dw \) is increasing in \( \pi(\sigma|w) \). Moreover, because \( \pi \) is sender-optimal in the class of all information structures with two realizations, we must have \( E_{F_{w'},\pi}[w|\sigma] \geq w^* \) for all \( w' \in \Omega_1^* \).

**Lemma 5.** Suppose that an information structure \( \pi \) with two realizations satisfies \( \pi(\sigma|w') \leq Z_\pi(w') \) for all \( w' \in [l, w^*), \pi(\sigma|w') = 1 \mu_{F_s}\)-a.e. \( w' \in [w^*, h], \) and \( \pi(\sigma_0|w) = 1 - \pi(\sigma|w) \). Then \( U(\pi) = \int_{[l,h]} \pi(\sigma|w)f_s(w)dw \).

**Proof of lemma 5.** By lemma 4 an information structure \( \pi \) satisfying \( \pi(\sigma|w) = 1 \mu_{F_s}\)-a.e. \( w \in [w^*, h] \) satisfies \( \pi(\sigma|w') \leq Z_\pi(w') \) for all \( w' \in [l, w^*) \) if and only if \( \pi \) satisfies \( E_{F_{w'},\pi}[w|\sigma] \geq w^* \) for all priors \( F_{w'} \) of the form \( \mu_{F_{w'}} = (1 - a)\mu_G + a\delta_{w'} \) for some \( w' \in [l, w^*) \).

Observe that priors of the form \( \mu_{F_{w'}} = (1 - a)\mu_G + a\delta_{w'} \) for some \( w' \in [l, w^*) \) are the extreme points of the set of priors \( F_{a,g} \). Because \( E_{F_{w'},\pi}[\omega|\sigma] \geq w^* \) for all priors \( F \) in the set of the extreme points of \( F_{a,g} \), lemma 3 implies that \( E_{F,\pi}[\omega|\sigma] \geq w^* \) for all priors \( F \) in \( F_{a,g} \).

Observe next that if \( E_{F,\pi}[\omega|\sigma] \geq w^* \) for all \( F \in F_{a,g} \), then \( E_{F,\pi}[\omega|\sigma] \geq w^* \) for all \( F \in F_{a,0} \). Then the receiver takes action 1 after seeing the signal realization \( \sigma \) for all priors in the feasible set of priors \( F_{a,0} \). Therefore, \( U(\pi) = \int_{[l,h]} \pi(\sigma|w)f_s(w)dw \), as required.

**Lemma 6.** For all \( \pi_0 \in \mathcal{D} \) there exists an information structure \( \pi \) with two realizations satisfying \( U(\pi) \geq U(\pi_0) \). Moreover, if \( \pi_0 \) is sender-optimal, then \( \pi \) can be chosen such that \( U(\pi) = \int_{[l,h]} \left( \sum_{i \in R(\hat{F},\pi_0)} \pi_0(\sigma_i|w) \right) f_s(w)dw \) and \( \pi(\sigma|w) = \sum_{i \in R(\hat{F},\pi_0)} \pi_0(\sigma_i|w) \) for all \( w \in [l, h] \).

**Proof of lemma 6.** Observe that if \( \pi_0 \in \mathcal{D} \) does not satisfy \( \sum_{i \in R(\pi_0)} \pi_0(\sigma_i|w) = 1 \mu_{F_s}\)-a.e. \( w \in [w^*, h] \) or the requirement that for any \( i \in R(\pi_0) \) there exists \( \hat{F} \in F_{a,0} \) such that \( E_{\hat{F},\pi_0}[w|\sigma_i] \geq w^* \) and Nature finds it optimal to choose \( \hat{F} \), then there exists a \( \pi_1 \in \mathcal{D} \) that satisfies these requirements and gives a payoff to the sender that is at least as high as the payoff from \( \pi_0 \). Then it is enough to prove the lemma for the information structures satisfying the above requirements.

Fix \( \pi_0 \in \mathcal{D} \) satisfying \( |R(\pi_0)| > 1 \), \( \sum_{i \in R(\pi_0)} \pi_0(\sigma_i|w) = 1 \mu_{F_s}\)-a.e. \( w \in [w^*, h] \) and the requirement that for any \( i \in R(\pi_0) \) there exists \( \hat{F} \in F_{a,0} \) such that \( E_{\hat{F},\pi_0}[w|\sigma_i] \geq w^* \) and Nature finds it optimal to choose \( \hat{F} \). Then \( \pi_0 \) satisfies the conditions in the hypothesis of lemma 2. By lemma 2 there exists \( F \in F_{a,0} \) such that for all \( w' \in [l, w^*), \pi_0 \) satisfies \( \sum_{i \in R(F,\pi_0)} \pi_0(\sigma_i|w') \leq \frac{\int_{[w^*,h]} (g(w)dw - \int_{[w^*,h]} \sum_{i \in R(F,\pi_0)} \pi_0(\sigma_i|w)g(w)(w^*-w)dw)}{w^*-w'} \).
We define an information structure $\pi$ with two realizations, $\sigma$ and $\sigma_0$, as follows. We set $\pi(\sigma | w) = \sum_{i \in R(F, \pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [l, w^*)$, $\pi(\sigma | w) = 1$ for all $w \in [w^*, h]$, and set $\pi(\sigma_0 | w) = 1 - \pi(\sigma | w)$ for all $w \in [l, h]$.

Observe that $\pi$ satisfies $\pi(\sigma | w') \leq \int_{[w^*, h]} (w^* - w') g(w) dw - \int_{[l, w^*]} g(w) \pi(\sigma | w) (w^* - w) dw + \int_{l} g(w) (w^* - w) dw (w^* - w') \frac{w - w'}{w - w'}$ for all $w' \in [l, w^*)$. Then $U(\pi) = \int_{[l, h]} \pi(\sigma | w) f_s(w) dw \geq \int_{[l, h]} \left( \sum_{i \in R(F, \pi_0)} \pi_0(\sigma_i | w) \right) f_s(w) dw$, where the equality follows from lemma 5 because because $\pi(\sigma | w') \leq Z(\pi'(w')$ for all $w' \in [l, w^*)$ and $\pi(\sigma | w') = 1$ for all $w' \in [w^*, h]$, and the inequality follows from the way in which we defined $\pi$.

Observe that $U(\pi_0) \leq \min_{\tilde{F} \in C_a, g} \int_{[l, h]} \left( \sum_{i \in R(\tilde{F}, \pi_0)} \pi_0(\sigma_i | w) \right) f_s(w) dw$. In particular, $U(\pi_0) \leq \int_{[l, h]} \left( \sum_{i \in R(F, \pi_0)} \pi_0(\sigma_i | w) \right) f_s(w) dw$. Then the above implies that $U(\pi_0) \leq U(\pi)$, as required.

**Claim 6.1.** If $\pi_0$ is sender-optimal (in $D$), then $R(F, \pi_0) = R(\pi_0)$.

**Proof of claim 6.1.** Suppose for the sake of contradiction that $R(F, \pi_0) \subset R(\pi_0)$. Let $S_0 = R(\pi_0) \setminus R(F, \pi_0)$. Define an information structure $\pi_2$ with three signal realizations, $\sigma$, $\sigma_k$ and $\sigma_0$, as follows. Set $\pi_2(\sigma | w) = \pi(\sigma | w)$ for all $w \in [l, h]$. Set $\pi_2(\sigma_k | w) = \sum_{i \in S_0} \pi_0(\sigma_i | w)$ for $w \in [w^*, h]$ and $\pi_2(\sigma_k | w) = 0$ for $w \in [l, w^*)$.

By construction, $E_{\tilde{F}, \pi_2}[\omega | \sigma_k] \geq w^*$ for all $\tilde{F} \in C_a, g$. Moreover, the fact that $E_{\tilde{F}, \pi_2}[\omega | \sigma] \geq w^*$ for all $\tilde{F} \in C_a, g$ and $\pi_2(\sigma | w) = \pi(\sigma | w)$ for all $w \in [l, h]$ implies that $E_{\tilde{F}, \pi_2}[\omega | \sigma] \geq w^*$ for all $\tilde{F} \in C_a, g$.

Thus $U(\pi_2) = \int_{[l, h]} (\pi_2(\sigma | w) + \pi_2(\sigma_k | w)) f_s(w) dw > \int_{[l, h]} \pi_2(\sigma | w) f_s(w) dw = \int_{[l, h]} \pi(\sigma | w) f_s(w) dw = U(\pi)$. Because $\pi_0$ is sender-optimal and, by the first part of the lemma, $U(\pi) \geq U(\pi_0)$, $\pi$ is also sender-optimal. Then $U(\pi_2) > U(\pi)$ contradicts the sender-optimality of $\pi$. ■

**Claim 6.2.** If $\pi_0$ is sender-optimal, $\pi$ can be chosen such that $\pi(\sigma | w) = \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [l, h]$ for all $w \in [l, h]$.

**Proof of claim 6.2.** Note that, because $\pi(\sigma | w) = 1$ for all $w \in [w^*, h]$ and $\sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) = 1 \mu_{F_\pi}$-a.e. $w \in [w^*, h]$, we have $\pi(\sigma | w) = \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) \mu_{F_\pi}$-a.e. $w \in [w^*, h]$. Observe if $\tilde{\pi}$ satisfies $\tilde{\pi}(\sigma | w) = \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [w^*, h]$, then $U(\tilde{\pi}) = U(\pi)$. Then we may assume that $\pi(\sigma | w) = \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [w^*, h]$. Because $R(\pi_0) = R(F, \pi_0)$ by claim 6.1 we then have $\pi(\sigma | w) = \sum_{i \in R(F, \pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [w^*, h]$.

Thus, because $\pi(\sigma | w) = \sum_{i \in R(F, \pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [l, w^*)$, we have $\pi(\sigma | w) =
\[ \sum_{i \in R(F, \pi_0)} \pi_0(\sigma_i | w) \text{ for all } w \in [l, h]. \] Then the fact that \( R(F, \pi_0) = R(\pi_0) \) by claim \( 6.1 \) implies that \( \pi(\sigma | w) = \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) \) for all \( w \in [l, h] \).

Claim 6.3. If \( \pi_0 \) is sender-optimal, then \( R(\tilde{F}, \pi_0) = R(\pi_0) \) for all \( \tilde{F} \in C_{a,g} \).

Proof of claim 6.3. By claim 6.2 we have \( U(\pi) = \int_{[l, h]} \pi(\sigma | w) f_s(w) dw = \int_{[l, h]} \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) f_s(w) dw \). The fact that \( \pi_0 \) is sender-optimal and \( U(\pi) \geq U(\pi_0) \) imply that \( U(\pi_0) = U(\pi) \), so that \( U(\pi_0) = \int_{[l, h]} \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) f_s(w) dw \).

Suppose for the sake of contradiction that \( R(\tilde{F}, \pi_0) \subset R(\pi_0) \) for some \( \tilde{F} \in C_{a,g} \). Then if Nature chooses \( \tilde{F} \), the sender’s payoff from \( \pi_0 \) satisfies \( U(\pi_0) \leq \int_{[l, h]} \left( \sum_{i \in R(\tilde{F}, \pi_0)} \pi_0(\sigma_i | w) \right) f_s(w) dw < \int_{[l, h]} \left( \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) \right) f_s(w) dw = U(\pi_0) \).

Then \( U(\pi_0) < U(\pi_0) \), a contradiction.

Claim 6.4. If \( \pi_0 \) is sender-optimal, \( \pi \) can be chosen such that \( \pi(\sigma | w) = \sum_{i \in R(\tilde{F}, \pi_0) \text{ for all } \tilde{F} \in C_{a,g}} \pi_0(\sigma_i | w) \) for all \( w \in [l, h] \).

Proof of claim 6.4. By claim 6.2 \( \pi \) can be chosen such that \( \pi(\sigma | w) = \sum_{i \in R(\pi_0)} \pi_0(\sigma_i | w) \) for all \( w \in [l, h] \) for all \( w \in [l, h] \). By claim 6.3 \( R(\tilde{F}, \pi_0) = R(\pi_0) \) for all \( \tilde{F} \in C_{a,g} \). Then \( \pi \) can be chosen such that \( \pi(\sigma | w) = \sum_{i \in R(\tilde{F}, \pi_0) \text{ for all } \tilde{F} \in C_{a,g}} \pi_0(\sigma_i | w) \) for all \( w \in [l, h] \), as required.

If \( \pi_0 \) is sender-optimal, claim 6.4 implies that \( U(\pi) = \int_{[l, h]} \pi(\sigma | w) f_s(w) dw = \int_{[l, h]} \left( \sum_{i : i \in R(\tilde{F}, \pi_0) \text{ for all } \tilde{F} \in C_{a,g}} \pi_0(\sigma_i | w) \right) f_s(w) dw \), as required.

Lemma 7. Suppose that \( \pi \) is sender-optimal in the class of all information structures with two realizations and that there exists a partition \( (A, B) \) of \( \Omega^r_1 \) such that \( \pi \) satisfies

\[
\begin{align*}
\pi(\sigma | w') &= \min \left\{ \frac{\int_{[w^*, w]} g(w) dw - f_{a_1} g(w) \pi(\sigma | w) (w^* - w) dw - f_{a_2} g(w) (w^* - w) dw}{(w^* - w) \frac{1}{1 - a}}, 1 \right\} \text{ for all } w' \in A, \\
\pi(\sigma | w') &< \min \left\{ \frac{\int_{[w^*, w]} g(w) dw - f_{a_1} g(w) \pi(\sigma | w) (w^* - w) dw - f_{a_2} g(w) (w^* - w) dw}{(w^* - w) \frac{1}{1 - a}}, 1 \right\} \text{ for all } w' \in B,
\end{align*}
\]

and \( \pi(\sigma | w') = 1 \) for all \( w' \in \Omega^r_2 \).

Then there exists a constant \( c_1 > 0 \) such that

1. \( \pi(\sigma | w) = \frac{c_1}{w^* - w} \) for all \( w \in [l, w^* - c_1] \cap A \)

2. \( \pi(\sigma | w) = 1 \) for all \( w \in \Omega^r_2 = [w^* - c_1, w^*) \cap A \)

where \( c_1 \) satisfies \( c_1 = \frac{\int_{[w^*, w]} g(w) dw - f_{A \cup \Omega^r_2} g(w) (w^* - w) dw}{\frac{1}{1 - a} + \int_{A \cup \Omega^r_2} g(w) dw} \).

Proof of lemma 7.
The formula for $c_1$ follows from re-arranging the expressions. We next show that there exists a constant $c_1$ satisfying the formula in the lemma.

**Claim 7.1.** There exists $c_1 > 0$ satisfying the formula in the lemma.

**Proof of claim 7.1**

Consider a mapping $c \mapsto L(c)$ given by

\[
L(c) = \frac{\int_{[w^*,s]}(w-w^*)g(w)dw - \int_B g(w)\pi(w|w^*)dw - \int_{\Omega^w} g(w)(w^*-w)dw}{\int_{[w^*,s]}(w-w^*)g(w)dw + f_{[w^*,s]}g(w)dw} - c.
\]

We have $L(0) = \frac{\int_B g(w)\pi(w|w^*)dw + \int_{[w^*,h]} g(w)dw}{\int_{[w^*,s]}(w-w^*)g(w)dw + f_{[w^*,s]}g(w)dw} \leq w^*$. This, in turn, implies that $L(0) \leq 0$ is equivalent to $\int_B g(w)\pi(w|w^*)dw + \int_{[w^*,h]} g(w)dw \leq w^*$. Therefore, we must have $L(0) > 0$. Moreover, $\mu_{F_a} \pi[w|\sigma] < w^*$ for all $F \in F_a$, which contradicts the hypothesis that $\pi$ is sender-optimal. Therefore, there exists $\pi$ such that $L(c) = 0$.

**Lemma 8.** If Assumption 7 holds and $\pi$ is sender-optimal in the class of all information structures with two realizations, then it satisfies

\[
\pi(\sigma|w') = \min \left\{ \frac{\int_{[w^*,h]}(w-w^*)g(w)dw - \int_{\Omega^w} g(w)\pi(w|w^*)dw}{(w-w^*)}, 1 \right\} \mu_{F_a} \text{-a.e.} \ w' \in \Omega^w \text{ and } \pi(\sigma|w) = 1 \mu_{F_a} \text{-a.e.} \ w \in [w^*,h].
\]

Suppose that (for all parameter values) there exists an information structure satisfying $\pi(\sigma|w') = Z_\pi(w')$ $\mu_{F_a} \text{-a.e.} \ w' \in \Omega^w$ and $\pi(\sigma|w) = 1 \mu_{F_a} \text{-a.e.} \ w \in [w^*,h]$. If Assumption 7 fails, then there exists $\pi_0$ that is sender-optimal in the class of all information structures with two realizations and satisfies the above condition.

**Proof of lemma 8.** Suppose that Assumption 7 holds and $\pi$ is sender-optimal in the class of all information structures with two realizations. The fact that $\pi(\sigma|w) = 1$ $\mu_{F_a} \text{-a.e.} \ w \in [w^*,h]$ follows from lemma 4. Since $\pi$ is sender-optimal in the class of all information structures with two realizations, by lemma 4 for all $w' \in \Omega^w$, we have $\pi(\sigma|w') \leq \min \left\{ \frac{\int_{[w^*,h]}(w-w^*)g(w)dw - \int_{\Omega^w} g(w)\pi(w|w^*)dw}{(w-w^*)}, 1 \right\}$. Observe that we can write this as $\pi(\sigma|w) \leq Z_\pi(w)$.

Suppose for the sake of contradiction that there exists $\pi$ that is sender-optimal in the class of all information structures with two realizations such that $\pi(\sigma|w') < Z_\pi(w')$ for some
subset \( B \subseteq \Omega_1^\pi \) of a strictly positive measure under \( \mu_{F_s} \).

Define \( A = \Omega_1^\pi \setminus B \). Note that, because \( \Omega_2^\pi = \{ w \in \Omega_1^\pi : \pi(\sigma|w) = 1 \} \), we have \( \Omega_2^\pi = \{ w \in A : \pi(\sigma|w) = 1 \} \). Note also that we then have \( A = \{ w \in \Omega_1^\pi : \pi(\sigma|w) = Z_\pi(w) \} \) and \( B = \{ w \in \Omega_1^\pi : \pi(\sigma|w) < Z_\pi(w) \} \).

Fix \( x \in B \) such that for all intervals \( I \) satisfying \( x \in \text{int}(I) \) (where \( \text{int} \) denotes the interior), we have \( \mu(I \cap B) > 0 \). Because, by the hypothesis, \( \mu_{F_s}(B) > 0 \), such \( x \) exists. Given \( \epsilon > 0 \) sufficiently small, fix an interval \( I_{x,\epsilon} = [x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}] \). Note that then \( \mu(I_{x,\epsilon} \cap B) > 0 \).

We define a new information structure \( \pi_1 \) with two realizations as follows. For all \( w \in I_{x,\epsilon} \cap B \) we let \( \pi_1(\sigma|w) = \pi(\sigma|w) + \eta \) for some \( \eta > 0 \). We let \( \pi_1(\sigma|w) = \pi(\sigma|w) \) for all \( w \in B \setminus I_{x,\epsilon} \). We choose \( \eta \) small enough such that \( \pi_1(\sigma|w) < 1 \) and \( E_{F_s,\pi_1}[\omega|\sigma] > w^* \) for all \( w \in B \). Note that this is feasible because \( \pi(\sigma|w) < 1 \) for all \( w \in B \) and because, by lemma 4, if \( \pi(\sigma|w) < Z_\pi(w) \), then \( E_{F_s,\pi_1}[\omega|\sigma] > w^* \), which implies that \( E_{F_s,\pi_1}[\omega|\sigma] > w^* \) for all \( w \in B \). To complete the construction of the information structure \( \pi_1 \), we require that \( \pi_1(\sigma|w) = Z_\pi(\sigma|w) \) for all \( w \in A \).

Observe that, because \( \pi(\sigma|w) = Z_\pi(\sigma|w) \) for all \( w \in A \) and \( \pi(\sigma|w) < Z_\pi(\sigma|w) \) for all \( w \in B \), by lemma 7, there exists a constant \( c \) such that \( \Omega_2^\pi = [w^* - c, w^*] \cap A \), \( \pi(\sigma|w) = \frac{c}{w^* - w} \) for all \( w \in A \setminus \Omega_2^\pi \) and \( \pi(\sigma|w) = 1 \) for all \( w \in \Omega_2^\pi \). Similarly, because \( \pi_1(\sigma|w) = Z_\pi(\sigma|w) \) for all \( w \in A \) and \( \pi_1(\sigma|w) < Z_\pi(\sigma|w) \) for all \( w \in B \), by lemma 7, there exists a constant \( c_1 \) such that \( \Omega_2^\pi = [w^* - c_1, w^*] \cap A \), \( \pi_1(\sigma|w) = \frac{c_1}{w^* - w} \) for all \( w \in A \setminus \Omega_2^\pi \) and \( \pi_1(\sigma|w) = 1 \) for all \( w \in \Omega_2^\pi \).

Thus to ensure that \( \pi_1(\sigma|w) = Z_\pi(\sigma|w) \) for all \( w \in A \), we choose the constant \( c_1 \) such that
\[
\frac{c_1}{w^* - w'} = \int_{[w^* - w']\cup [w^* - w]} g(w)dw - \int_B g(w)\pi_1(\sigma|w)(w^* - w)dw - c_1 \int_{A\setminus \Omega_2^\pi} g(w)dw - c_1 \int_{\Omega_2^\pi} g(w)(w^* - w)dw
\]
for all \( w' \in A \setminus \Omega_2^\pi \) and \( \pi(\sigma|w') = 1 \) for all \( w' \in A \cap \Omega_2^\pi \) where \( \Omega_2^\pi = [w^* - c_1, w^*] \cap \Omega_1^\pi \).

**Claim 8.1.** \( U(\pi_1) - U(\pi) = \eta \int_{I_{x,\epsilon} \cap B} f_s(w)dw - \int_{A\setminus \Omega_2^\pi} \frac{c}{w^* - w}f_s(w)dw - \int_{[w^* - c_1, w^*] \cap A} \left( \frac{c_1}{w^* - w} - 1 \right) f_s(w)dw \).

**Proof of claim 8.1.** The result is immediate. \( \blacksquare \)

Define the derivative of the sender’s payoff with respect to the transformation of the information structure from \( \pi \) to \( \pi_1 \) given \( x \in B \) as \( E_x^\pi = \lim_{\epsilon \to 0} \frac{U(\pi_1) - U(\pi)}{\eta \epsilon} \).

**Claim 8.2.** \( E_x^\pi = \frac{f_s(x)}{g(x)(w^* - x)} - \frac{1}{1+c_1} \int_{A\setminus \Omega_2^\pi} g(w)dw \int_{A\setminus \Omega_2^\pi} \frac{f_s(w)}{w^* - w}dw \).

**Proof of claim 8.2.** Follows from approximating the integrals, the fact that \( g \) and \( f_s \) are \( C^1 \) and claim \( 7 \).\( \blacksquare \)
Observe that, in order for the sender to not have a strictly improving deviation, we need that for all \( x \in B \) such that for all intervals \( I \) satisfying \( x \in \text{int}(I) \) we have \( \mu(I \cap B) > 0 \), either \( E^\pi_x = 0 \) or \( E^\pi_x > 0 \) and \( \pi(\sigma|x) = 1 \) (note that if \( \pi(\sigma|x) = 0 \), then \( x \not\in B \)). Thus, because \( E^\pi_x = \frac{f(x)}{g(x)(w^*-x)} - \frac{1}{1-w_a + f_a \mu_2 g(w)dw} \int_{A \setminus \Omega^2_2} f(w)w^{*-w}dw \) by claim 8.2, we need that
\[
\frac{f(x)}{g(x)(w^*-x)} - \frac{1}{1-w_a + f_a \mu_2 g(w)dw} \int_{A \setminus \Omega^2_2} f(w)w^{*-w}dw = 0 \quad \text{for almost all } x \in B \text{ satisfying } \pi(\sigma|x) < 1.
\]

However Assumption 1 implies that almost everywhere on \( \Omega^2_1 \), \( \frac{f(x)}{g(x)(w^*-x)} - \frac{1}{1-w_a + f_a \mu_2 g(w)dw} \int_{A \setminus \Omega^2_2} f(w)w^{*-w}dw = 0 \) fails. Thus if \( \pi(\sigma|w) \leq Z^\pi(w) \) holds strictly on a set of a strictly positive measure, then the sender has a strictly improving deviation, which contradicts the optimality of \( \pi \). Therefore, the inequality \( \pi(\sigma|w) \leq Z^\pi(w) \) must be satisfied with equality almost everywhere on \( \Omega^2_1 \).

We now prove the second part of the lemma. Fix the sender’s prior \( f_s (\pi, f_s) \) and \( g \) such that for each \( f_s^n \) Assumption 1 is satisfied. The first part of the lemma and the assumption that for all parameter values there exists an information structure satisfying \( \pi(\sigma|w') = Z^\pi(w') \mu_{F^s} \text{-a.e.} w' \in \Omega^s_1 \) and \( \pi(\sigma|w) = 1 \mu_{F^s} \text{-a.e.} w \in [w^*, h] \) imply that for each \( f^n_s \) there exists a sender-optimal information structure \( \pi_n \) with two realizations satisfying \( \pi_n(\sigma|w) = Z_n^\pi(w) \) for almost all \( w \in \Omega^s_1 \). Because \( \{\pi_n\} \) is an infinite sequence in a compact set, it has a convergent subsequence. Without loss of generality, assume that this is the sequence itself, and let \( \pi^* \) denote the information structure that it converges to. Observe that, because \( \pi_n(\sigma|w) = Z_n^\pi(w) \) for almost all \( w \in \Omega^s_1 \) for all \( n \), we have \( \pi^*(\sigma|w) = Z_n^\pi(w) \) for almost all \( w \in \Omega^s_1 \).

Let \( U(\pi, f_s) = \int_{[l,h]} \pi(\sigma|w)f_s(w)dw \) denote the sender’s payoff from the information structure \( \pi \) given that the sender’s prior is \( f_s \). Suppose for the sake of contradiction that there exists an information structure \( \tilde{\pi} \) with two realizations that is sender-optimal given \( f_s \) such that \( \tilde{\pi}(\sigma|w) < Z^\pi \) for a subset \( B \subseteq \Omega^s_1 \) of a strictly positive measure and \( U(\tilde{\pi}, f_s) > U(\pi^*, f_s) \).

Observe that the fact that \( f_s \mapsto U(\pi, f_s) \) is continuous implies that for all \( \delta > 0 \) there exists \( N \) such that for all \( n > N \), \( |U(\tilde{\pi}, f_s) - U(\pi^*, f_s^n)| < \delta \) and \( |U(\pi^*, f_s) - U(\pi^*, f_s^n)| < \delta \). Moreover, because \( \{\pi_n\} \) converges to \( \pi^* \) and \( \pi \mapsto U(\pi, f_s) \) is continuous, for all \( \delta > 0 \) there exists \( N \) such that for all \( n > N \), \( |U(\pi^*, f_s^n) - U(\pi_n, f_s^n)| < \delta \).

Note that, because \( U(\tilde{\pi}, f_s) > U(\pi^*, f_s) \) and for all \( \delta > 0 \) there exists \( N \) such that for all \( n > N \), \( |U(\pi^*, f_s) - U(\pi^*, f_s^n)| < \delta \), there exists \( N_1 \) and \( \delta_1 > 0 \) such that for all \( n > N_1 \), we have \( U(\tilde{\pi}, f_s) - U(\pi^*, f_s^n) > \delta_1 \).
Fix $\delta > 0$. Then there exists $N_2$ such that for all $n > N_2$, we have $U(\tilde{\pi}, f_s) - U(\tilde{\pi}, f^n_s) + U(\pi_n, f^n_s) - U(\pi^*, f^n_s) = |U(\tilde{\pi}, f_s) - U(\tilde{\pi}, f^n_s)| + |U(\pi_n, f^n_s) - U(\pi^*, f^n_s)| < 2\delta$. This implies that $U(\tilde{\pi}, f^n_s) - U(\pi_n, f^n_s) > U(\tilde{\pi}, f_s) - U(\pi^*, f^n_s) - 2\delta$. Therefore, because $U(\tilde{\pi}, f_s) - U(\pi^*, f^n_s) > \delta_1$ for all $n > N_1$, we have $U(\tilde{\pi}, f^n_s) - U(\pi_n, f^n_s) > \delta_1 - 2\delta$ for all $n > \max\{N_2, N_1\}$.

This implies that there exists $\delta_0 > 0$ and $N$ such that for all $n > N$, $U(\tilde{\pi}, f^n_s) > U(\pi_n, f^n_s) + \delta_0$. However, this contradicts the hypothesis that the information structure $\pi_n$ is sender-optimal given that the sender’s prior is $f^n_s$.

Lemma 9. If $\pi$ is sender-optimal in the class of all information structures with two realizations and $\pi$ satisfies the condition in lemma 8 then there exists a threshold $t$ such that $\Omega^\pi_1 = \Omega_1 \mu_{f^s} - a.e.$ for some $\Omega_1 \in \Omega(t)$.

Either the threshold $t \geq \tilde{t}$ satisfies $t = \frac{\int_{\Omega^\pi_1 \cap [l, w^* - \epsilon] \frac{f_s(w)}{w} \, dw}{\int_{\Omega_1 \cap [l, w^* - \epsilon]} g(w) \, dw}$, where $c$ is a constant such that $\pi(\sigma|w) = \min \left\{ \frac{c}{w - w^*}, 1 \right\} \mu_{f_s} - a.e. \ w \in \Omega^\pi_1$, or $t > \frac{\int_{\Omega^\pi_1 \cap [l, w^* - \epsilon]} \frac{f_s(w)}{w} \, dw}{\int_{\Omega_1 \cap [l, w^* - \epsilon]} g(w) \, dw}$ and $t = \tilde{t}$.

Proof of lemma 9. Let $\Omega_1 = \Omega^\pi_1$. Given $x \in [l, w^*)$ and $\epsilon > 0$ satisfying $x - \frac{\epsilon}{2} \geq l$, $x + \frac{\epsilon}{2} < w^*$, define $I_{x, \epsilon} = [x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}]$. Consider an information structure $\pi_1$ obtained by adding an interval $I_{x, \epsilon}$ to the support of the information structure $\pi$ below the threshold $w^*$. That is, $\pi_1$ is an information structure with two signal realizations such that $\Omega^\pi_1 = \Omega_1 \cup I_{x, \epsilon}$ and $\mu(I_{x, \epsilon} \cap \Omega^\pi_1) < \epsilon$.

Because $\pi$ satisfies the condition in lemma 8 by lemmas 7 and 8, there exist constants $c$ and $c_1$ such that $\pi(\sigma|w) = \min \left\{ \frac{c}{w - w^*}, 1 \right\} \mu_{f_s} - a.e. \ w \in \Omega^\pi_1 = \Omega_1 = \pi_1(\sigma|w) = \min \left\{ \frac{c_1}{w^* - w}, 1 \right\} \mu_{f_s} - a.e. \ w \in \Omega^\pi_1 = \Omega_1 \cup I_{x, \epsilon}$.

Claim 9.1. If $\frac{f_s(x)}{w^* - x} \frac{1}{g(x)} > \frac{\int_{\Omega_1 \cap [l, w^* - \epsilon]} \frac{f_s(w)}{w} \, dw}{\int_{\Omega_1 \cap [l, w^* - \epsilon]} g(w) \, dw}$, then $U(\pi_1) > U(\pi)$. If $U(\pi_1) \geq U(\pi)$, then $
abla \frac{f_s(x)}{w^* - x} \frac{1}{g(x)} \geq \frac{\int_{\Omega_1 \cap [l, w^* - \epsilon]} \frac{f_s(w)}{w} \, dw}{\int_{\Omega_1 \cap [l, w^* - \epsilon]} g(w) \, dw}$.

Proof of claim 9.1. Follows from approximating the integrals.

Claim 9.2. If $w_1 \notin \Omega^\pi_1$, then $w_2 \notin \Omega^\pi_1 \mu_{f_s} - a.e. \ w_2 \in [l, w^*)$ such that $\frac{f_s(w_1)}{w^* - w_1} \frac{1}{g(w_1)} > \frac{f_s(w_2)}{w^* - w_2} \frac{1}{g(w_2)}$.

Proof of claim 9.2. If $w_1 \in [l, w^*)$ is such that $w_1 \notin \Omega^\pi_1$ and $\pi$ is a sender-optimal information structure, then adding an interval $I_{w_1, \epsilon}$ to the support does not strictly benefit the sender. That is, then $U(\pi_1) \leq U(\pi)$, where $\pi_1$ is an information structure obtained by adding an interval $I_{w_1, \epsilon}$ to $\Omega^\pi_1$. By claim 9.1, this implies that $\frac{f_s(w_1)}{w^* - w_1} \frac{1}{g(w_1)} \leq \frac{\int_{\Omega_1 \cap [l, w^* - \epsilon]} \frac{f_s(w)}{w} \, dw}{\int_{\Omega_1 \cap [l, w^* - \epsilon]} g(w) \, dw}$. 

37
Then for any \( w_2 \in [l, w^*) \) such that \( \frac{f_s(w_2)}{w^*-w_2} g(w_2) > \frac{f_s(w_1)}{w^*-w_1} g(w_1) \), we have \( \frac{f_{\Omega_1 \cap \{l, w^*-c\}} f_s(w)}{g(w)} > \frac{f_{\Omega_1 \cap \{l, w^*-c\}} f_s(w_1)}{g(w_1)} \). By claim \[ \boxed{1} \] this implies that \( U(\pi_2) < U(\pi) \), where \( \pi_2 \) is an information structure obtained by adding an interval \( I_{w_2, c} \) to \( \Omega_1^\pi \). Therefore, \( w_2 \not\in \Omega_1^\pi \mu_{F_\pi}\)-a.e. \( w_2 \in [l, w^*) \) such that \( \frac{f_s(w_1)}{w^*-w_1} g(w_1) > \frac{f_s(w_2)}{w^*-w_2} g(w_2) \), as required.

Thus claim \[ \boxed{2} \] implies that for \( \mu_{F_\pi}\)-a.e. \( w_1, w_2 \in [l, w^*) \), if \( w_2 \in \Omega_1^\pi \) and \( \frac{f_s(w_1)}{w^*-w_1} g(w_1) > \frac{f_s(w_2)}{w^*-w_2} g(w_2) \), then \( w_1 \in \Omega_1^\pi \). Therefore, \( \Omega_1^\pi = \Omega_1 \mu_{F_\pi}\)-a.e. for some \( \Omega_1 \in \Omega(t) \), as required. Observing that the threshold \( t \geq t \) satisfies either \( t = \frac{\int_{\Omega_1^\pi \cap \{l, w^*-c\}} f_s(w) g(w) dw}{\frac{a}{1-a} + \int_{\Omega_1^\pi \cap \{l, w^*-c\}} g(w) dw} \) or \( t > \frac{\int_{\Omega_1^\pi \cap \{l, w^*-c\}} f_s(w) g(w) dw}{\frac{a}{1-a} + \int_{\Omega_1^\pi \cap \{l, w^*-c\}} g(w) dw} \) and \( t = t \) concludes the proof.

**Lemma 10.** \( \Omega_1 \mapsto \hat{c}(\Omega_1) \) is a well-defined function. If Assumption \[ \boxed{1} \] is satisfied, then \( t \mapsto c(t) \) is a continuous and strictly increasing function for \( t \geq t \).

**Proof of lemma 10.** Let \( \hat{x}(c, \Omega_1) = \int_{[w^*, l]} (w - w^*) g(w) dw - \int_{\Omega_1 \cap \{w^*-c, w^*\}} g(w)(w^*-w) dw - c \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\}} g(w) dw \right) \), so that \( \Omega_1 \mapsto \hat{c}(\Omega_1) \) is the mapping satisfying \( \hat{x}(c, \Omega_1) = 0 \). We have \( \hat{x}(c + \eta, \Omega_1) - \hat{x}(c, \Omega_1) = -\eta \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\} - \eta} g(w) dw \right) + O(\eta^2) \) because \( g \) is \( C^1 \). Then \( \lim_{\eta \to 0} \frac{\hat{x}(c + \eta, \Omega_1) - \hat{x}(c, \Omega_1)}{\eta} = -\lim_{\eta \to 0} \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\} - \eta} g(w) dw \right) = \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\}} g(w) dw \right) < 0 \). We have \( \hat{x}(0, \Omega_1) = \int_{[w^*, l]} (w - w^*) g(w) dw > 0 \) and \( \hat{x}(\infty, \Omega_1) = -\infty \). Because \( c \mapsto \hat{x}(c, \Omega_1) \) is continuous and strictly decreasing, there exists a unique \( c \in (0, \infty) \) such that \( \hat{x}(c, \Omega_1) = 0 \), which implies that \( \Omega_1 \mapsto \hat{c}(\Omega_1) \) is a well-defined function, as required.

Recall that \( x(c, t) = \int_{[w^*, l]} (w - w^*) g(w) dw - \int_{\Omega_1(t) \cap \{w^*-c, w^*\}} g(w)(w^*-w) dw - c \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\}} g(w) dw \right) \) and \( t \mapsto c(t) \) is the mapping satisfying \( \hat{x}(c, t) = 0 \).

We have \( x(c + \eta, t) - x(c, t) = -\eta \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\} - \eta} g(w) dw \right) + O(\eta^2) \) because \( g \) is \( C^1 \). Then \( \lim_{\eta \to 0} \frac{x(c + \eta, t) - x(c, t)}{\eta} = -\lim_{\eta \to 0} \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\} - \eta} g(w) dw \right) = \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap \{l, w^*-c\}} g(w) dw \right) < 0 \).

Observe that, because \( \mu \left( \left\{ w \in [l, w^*) : \frac{f_s(w)}{g(w)(w^*-w)} = c_0 \right\} \right) = 0 \) for all \( c_0 > 0 \) by Assumption \[ \boxed{1} \] and \( f_s \) and \( g \) are continuous, we have \( \Omega_1(t + \eta) \cap \{l, w^*-c\} \subset \Omega_1(t) \cap \{l, w^*-c\} \) for all \( t \geq t \) and for all \( \eta > 0 \) sufficiently small. Then \( x(c, t + \eta) - x(c, t) = \int_{\Omega_1(t) \cap \Omega_1(t + \eta)} g(w)(w^*-w) dw + c \int_{\Omega_1(t) \cap \Omega_1(t + \eta)} g(w) dw \).

Because \( t \mapsto c(t) \) is the mapping satisfying \( x(c, t) = 0 \), the fact that
\[
\lim_{\eta \to 0} \frac{x(c, t + \eta) - x(c, t)}{\eta} < 0 \text{ and } x(c, t + \eta) - x(c, t) > 0 \text{ for all } \eta > 0 \text{ by the arguments above implies that } t \mapsto c(t) \text{ is a strictly increasing function, as required.}
\]

Observe that \((c, t) \mapsto x(c, t)\) is continuous because 
\[
\mu \left( \left\{ w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} = c_0 \right\} \right) = 0 \text{ for all } c_0 > 0 \text{ by Assumption 1 and } f_s \text{ and } g \text{ are continuous. The continuity of } t \mapsto c(t) \text{ then follows from the fact that } (c, t) \mapsto x(c, t) \text{ is continuous.} \]

\[\Box\]

**Lemma 11.** Let \(T = \{ t : |\bar{y}(t)| > 1 \} \) and let \(y_0(t, \Omega_1) = t \left( \frac{a}{1-a} + \int_{\Omega_1 \cap [l, w^* - \hat{c}(\Omega_1)]} g(w)dw \right) - \int_{\Omega_1 \cap [l, w^* - \hat{c}(\Omega_1)]} \frac{f_s(w)}{w^* - w}dw. \) Then

1. \(T\) is at most countable
2. for any selection \(t \mapsto \hat{y}(t)\) from the correspondence \(t \mapsto \bar{y}(t), \) if \(t_1 \leq t_2, \) then \(\hat{y}(t_1) < \hat{y}(t_2)\)
3. \(\bar{y}(t)\) is an interval
4. if \(|\bar{y}(t)| = 1\) for all \(t \in (t_1, t_2), \) then \(t \mapsto \bar{y}(t)\) is continuous on \((t_1, t_2)\)
5. \(\inf y(t_0) = \lim_{t \uparrow t_0} y(t), \) \(\sup y(t_0) = \lim_{t \downarrow t_0} y(t)\)
6. \(y(t) = y_0(t, \Omega_1), \) \(\Omega_1 \in \{ \Omega(t, z) \}_{z \in \{w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} = y\}}\)

**Proof of lemma 11.** Note that \(|T| \leq \left| \left\{ c_0 : \mu \left( \left\{ w : \frac{f_s(w)}{g(w)(w^*-w)} = c_0 \right\} \right) > 0 \right\} \right|, \) which implies that \(T\) is at most countable.

Fix \(\Omega_1, \) \(\Omega_2\) such that \(\Omega_1 \subset \Omega_2\), \(\Omega_1 \in \Omega_1(t + \eta) \) and \(\Omega_2 \subset \Omega_1(1) \) for some \(\eta \geq 0\) and \(t \geq t. \) We will show that \(y_0(t + \eta, \Omega_1) - y_0(t, \Omega_2) > 0 \) for all \(\eta \geq 0.\)

A proof similar to the proof that \(t \mapsto c(t)\) is strictly increasing for \(t \geq t\) in lemma 10 can be used to show that \(\hat{c}(\Omega_2) > \hat{c}(\Omega_1). \) Then a proof similar to the proof that \(y(t + \eta) - y(t) > 0\) for all \(t \geq t\) and for all \(\eta > 0\) in the proof of the second part of lemma 12 can be used to establish that \(y_0(t + \eta, \Omega_1) - y_0(t, \Omega_2) > 0.\)

Let \(\Omega_0(t) = \left\{ w : \frac{f_s(w)}{g(w)(w^*-w)} > t \right\}. \) Observe that the argument above implies that \(\inf y(t) = y_0(t, \Omega_1(t)) \) and \(\sup y(t) = y_0(t, \Omega_0(t)). \) It can be shown that for any \(y' \in (y_0(t, \Omega_1(t)), y_0(t, \Omega_0(t))) \) there exists \(\Omega_1\) such that \(\Omega_0(t) \subseteq \Omega_1 \subseteq \Omega_1(t)\) and \(y_0(t, \Omega_1) = y'. \) Then \(y(t)\) is an interval, as required.

Suppose that \(|\bar{y}(t)| = 1\) for all \(t \in (t_1, t_2). \) Observe that then for all \(t \in (t_1, t_2)\) we have \(\Omega_1 = \Omega_2 \) a.e. for all \(\Omega_1, \) \(\Omega_2 \in \Omega_1(t). \) This implies that \(\Omega_1 = \Omega_1(t) \) a.e. for all \(\Omega_1 \in \Omega_1(t)\) and
\( \overline{\gamma}(t) = y(t) \) for all \( t \in (t_1, t_2) \). Moreover, \( t \mapsto \hat{c}(\Omega_1(t)) \) and \( t \mapsto \mu(\Omega_1(t)) \) are continuous on \( (t_1, t_2) \). Therefore, \( t \mapsto \overline{\gamma}(t) \) is continuous on \( (t_1, t_2) \), as required.

Observe that \( \lim_{t \uparrow t_0} \Omega_1(t) = \Omega_1(t_0) \) and \( \lim_{t \downarrow t_0} \Omega_1(t) = \Omega_0(t_0) \) a.e. Because \( \Omega_1 \mapsto \hat{c}(\Omega_1) \) is continuous, this implies that \( \inf \overline{\gamma}(t_0) = y_0(t_0, \Omega_1(t_0)) = \lim_{t \uparrow t_0} y(t) = \lim_{t \downarrow t_0} y(t) = \sup \overline{\gamma}(t_0) = y_0(t_0, \Omega_0(t_0)) = \lim_{t \uparrow t_0} y(t) = \lim_{t \downarrow t_0} y(t) \), as required.

To prove the last claim, define \( y_1(t) = y_0(t, \Omega_1(t)) \), \( y_2(t) = y_0(t, \Omega_0(t)) \). Fix \( t \geq 0 \) and \( \Omega_1 \in \overline{\Omega}_1(t) \). Note that \( \overline{\gamma}(t) = \overline{\gamma}(t) \) for all \( t < t \). Observe that this and the second claim in the lemma imply that \( y_0(t, \Omega_1) \in [y_1(t), y_2(t)] \). Let \( Z(t) = \left\{ w \in [l, w^*] : \frac{f_s(w)}{g(w)(w^*-w)} = t \right\} \) and \( \underline{Z}(t) = \inf Z(t) \), \( \overline{Z}(t) = \sup Z(t) \). Observe that \( y_1(t) = y_0(t, \Omega(t, \underline{Z}(t))) \) and \( y_2(t) = y_0(t, \Omega(t, \overline{Z}(t))) \). Because \( z \mapsto y_0(t, \Omega(t, z)) \) is continuous, the Intermediate Value Theorem implies that there exists \( z \in Z(t) \) such that \( y_0(t, \Omega(t, z)) = y_0(t, \Omega_1) \).

Lemma 12. If \( \pi \) is sender-optimal in the class of all information structures with two realizations and \( \pi \) satisfies the condition in lemma \[ \text{8} \] then there exists a threshold \( t \geq 0 \) such that \( \Omega_1^\pi = \Omega_1 \mu_{F_s} \text{-a.e.} \) for some \( \Omega_1 \in \overline{\Omega}_1(t) \) and either \( t = j \) for some \( j \in \overline{J}(t) \) or \( t = j \) and \( t > j \) for all \( j \in \overline{J}(t') \) for all \( t' \). Moreover, \( \Omega_1 \) can be chosen such that \( \Omega_1 = \Omega(t, z) \) for some \( z \in [l, w^*] \). If Assumption \[ \text{7} \] is satisfied, then there exists a unique threshold \( t \geq 0 \) satisfying either \( t = J(t) \) or \( t = j \) and \( t > J(t') \) for all \( t' \). Finally, \( y(t) - y(t) > 0 \) for all \( t \geq 0 \) and for all \( \eta > 0 \).

Proof of lemma 12. Because \( \pi \) satisfies the condition in lemma \[ \text{8} \] lemma \[ \text{9} \] applies. Then lemma \[ \text{9} \] implies that either \( t \geq 0 \) satisfies \( = \frac{\int_{\Omega_1^\pi \cap [l, w^*-c)} f_s(w) \, dw}{\int_{\Omega_1^\pi \cap [l, w^*-c)} g(w) \, dw} \), where \( c \) is a constant such that \( \pi(\sigma|w) = \frac{c}{w^*-w} \mu_{F_s} \) a.e. \( w \in [l, w^* - c) \cap \Omega_1^\pi \), and \( \pi(\sigma|w) = 1 \mu_{F_s} \) a.e. \( w \in [l, w^*-c, w^*) \setminus \Omega_1^\pi \), or \( t > \frac{\int_{\Omega_1^\pi \cap [l, w^*-c)} f_s(w) \, dw}{\int_{\Omega_1^\pi \cap [l, w^*-c)} g(w) \, dw} \) and \( t = t \). Note that, because \( \pi \) satisfies the condition in lemma \[ \text{8} \] lemma \[ \text{7} \] implies that \( c = \hat{c}(\Omega_1^\pi) \). Then, because \( \Omega_1^\pi = \Omega_1 \) for some \( \Omega_1 \in \overline{\Omega}_1(t) \) by lemma \[ \text{9} \] we have \( c = \hat{c}(\Omega_1) \) for some \( \Omega_1 \in \overline{\Omega}_1(t) \). Because \( \overline{J}(t) = \frac{\int_{\Omega_1^\pi \cap [l, w^*-c(\Omega_1))] f_s(w) \, dw}{\int_{\Omega_1^\pi \cap [l, w^*-c(\Omega_1))] g(w) \, dw} \), \( \Omega_1 \in \overline{\Omega}_1(t) \), either \( t \) satisfies \( t = j \) for some \( j \in \overline{J}(t) \) or \( t > j \) for all \( j \in \overline{J}(t') \) for all \( t' \) and \( t = t \). Thus a necessary condition for \( \pi \) to be sender-optimal is that there is a threshold \( t \geq 0 \) such that \( \Omega_1^\pi = \Omega_1 \mu_{F_s} \) a.e. for some \( \Omega_1 \in \overline{\Omega}_1(t) \) and either \( t = j \) for some \( j \in \overline{J}(t) \) or \( t > j \) for all \( j \in \overline{J}(t') \) for all \( t' \) and \( t = t \).

I next show that a threshold \( t \geq 0 \) satisfying the above properties exists. Note that, because \( \overline{\gamma}(t) = t \left( \frac{a}{1-a} + \int_{\Omega_1^\pi \cap [l, w^*-c(\Omega_1))] g(w) \, dw} - \int_{\Omega_1^\pi \cap [l, w^*-c(\Omega_1))] \frac{f_s(w)}{w^*-w} \, dw \right) \), \( \Omega_1 \in \overline{\Omega}_1(t) \), we have that \( t = j \) for some \( j \in \overline{J}(t) \) is equivalent to \( 0 \in \overline{\gamma}(t) \). Observe that it follows from lemma \[ \text{11} \] that \( t \mapsto \overline{\gamma}(t) \) is a convex-valued and upper hemicontinuous correspondence,
which implies that the Intermediate Value Theorem for correspondences (see Theorem 9.39 on page 145 in [Moore]1999) applies to it. Next, observe that, because $y' \leq 0$ for all $y' \in \overline{y}(0)$ and $\lim_{t \to \infty} \overline{y}(t) = \{\infty\}$, the Intermediate Value Theorem for correspondences implies that there exists $t^* \in [0, \infty)$ such that $0 \in \overline{y}(t^*)$. This implies that there exists $t^* \geq 0$ such that $\Omega_1^t = \Omega_1 \mu_{F_1}$-a.e. for some $\Omega_1 \in \overline{\Omega}_1(t^*)$ and $t^* = j$ for some $j \in \overline{T}(t^*)$.

Moreover, because the last claim of lemma [11] implies that $\overline{T}(t) = \frac{\int_{\Omega_1 \cap [l, w^* - c(t_j)]} f_s(w) \, dw}{\int_{\Omega_1 \cap [l, w^* - c(t_j)]} \int_{\Omega_1 \cap [l, w^* - c(t_j)]} g(w) \, dw}$, $\Omega_1 \in \{\Omega(z, t)\}_{z \in \{w \in [l, w^*] : \tilde{g}(w)(\tilde{w} = t) = 1\}}$, $\Omega_1$ can be chosen such that $\Omega_1 = \Omega(t, z)$ for some $z \in [l, w^*)$, as required. Note that if $t^* < t$ for all fixed points $t^*$, then $j < t$ for all $j \in \overline{T}(t')$ for all $t'$. To conclude the proof, if $t^* \geq t$ for some fixed point $t^*$, set $t = t^*$, otherwise set $t = t$. 

In the rest of the proof I suppose that Assumption [1] is satisfied. Observe that, because $y(t) = t \left(\frac{a}{1-a} + \int_{\Omega_1(t) \cap [l, w^* - c(t_j)]} f_s(w) \, dw\right) - \int_{\Omega_1(t) \cap [l, w^* - c(t_j)]} \frac{f_s(w)}{w^*-w} \, dw$, $t = J(t)$ is equivalent to $y(t) = 0$. Because, provided that Assumption [1] is satisfied, $t \mapsto c(t)$ is a strictly increasing function for $t \geq t$ by lemma [10], we have $c(t) < c(t + \eta)$ for all $\eta > 0$ sufficiently small.

Define $S^1_{\eta} = \left(\Omega_1(t) \cap [w^* - c(t + \eta), w^* - c(t)]\right) \cup \left(\Omega_1(t) \cap [w^* - c(t + \eta), w^* - c(t)]\right) \cup \left(\Omega_1(t) \cap [l, w^* - c(t)]\right) \cap S^1_{\eta}$. Then $y(t + \eta) - y(t) = \eta \left(\frac{a}{1-a} + \int_{\Omega_1(t+\eta) \cap [l, w^* - c(t+\eta)]} f_s(w) \, dw\right) - t \int_{S^1_{\eta}} g(w) \, dw + \int_{S^1_{\eta}} \frac{f_s(w)}{w^*-w} \, dw$.

We claim that $\int_{S^1_{\eta}} \frac{f_s(w)}{w^*-w} \, dw - t \int_{S^1_{\eta}} g(w) \, dw \geq 0$ for all $\eta > 0$. This is equivalent to $t \leq \frac{\int_{S^1_{\eta}} \frac{f_s(w)}{g(w)(w^*-w)} \, dw}{\int_{S^1_{\eta}} g(w) \, dw}$. Observe that the definition of $\Omega_1(t)$ implies that for all $w \in \Omega_1(t)$ we have $\frac{f_s(w)}{g(w)(w^*-w)} \geq t$. Because $S^1_{\eta} \subseteq \Omega_1(t)$, this implies that for all $w \in S^1_{\eta}$ we have $\frac{f_s(w)}{g(w)(w^*-w)} \geq t$. Because $\int_{S^1_{\eta}} \frac{f_s(w)}{g(w)(w^*-w)} \, dw \geq \inf_{w \in S^1_{\eta}} \frac{f_s(w)}{g(w)(w^*-w)} \geq t$, the inequality $t \leq \frac{\int_{S^1_{\eta}} \frac{f_s(w)}{g(w)(w^*-w)} \, dw}{\int_{S^1_{\eta}} g(w) \, dw}$ is satisfied.

Because $\frac{a}{1-a} + \int_{\Omega_1(t+\eta) \cap [l, w^* - c(t+\eta)]} f_s(w) \, dw > 0$, this implies that $y(t + \eta) - y(t) > 0$ for all $\eta > 0$. Consider a point $t^*$ such that $t^* = J(t^*)$. By the argument above, $t^* = J(t^*)$ is equivalent to $y(t^*) = 0$. The fact that $y(t + \eta) - y(t) > 0$ for all $t \geq t$ and for all $\eta > 0$ implies that the function $t \mapsto y(t)$ can intersect the zero function at $t \geq t$ at most one point.

The argument in the first part of the proof implies that there exists $t^*$ such that $y(t^*) = 0$. If there exists a fixed point $t^*$ such that $t^* \geq \bar{t}$, set $t = t^*$, and if $t^* < \bar{t}$ for all fixed points $t^*$, set $t = \bar{t}$. Note that if $t^* < \bar{t}$ for all fixed points $t^*$, then $\bar{t} > J(t')$ for all $t'$. ■
Letting $t = \min_{w \in [l, w^*]} \frac{f_s(w)}{g(w)(w^* - w)}$, I define a class of distributions $S^{tcz}$ over the receiver's actions as follows: given constants $t \geq t$, $c, z \geq 0$, the probability of action 1 in state $\omega$ is given by

$$S^{tcz}(\omega) = \begin{cases} 
1 & \text{for } \omega \in [w^*, h] \\
\min \left\{ \frac{c}{w^* - \omega}, 1 \right\} & \text{for } \omega \in \Omega(t, z) \\
0 & \text{for } \omega \in [l, w^*) \setminus \Omega(t, z)
\end{cases}$$

where

$$\Omega(t, z) = \left\{ w \in [l, w^*) : \frac{f_s(w)}{g(w)(w^* - w)} > t \right\} \cup \left\{ w \in [l, w^*) : \frac{f_s(w)}{g(w)(w^* - w)} = t \text{ and } w \geq z \right\}$$

**Proposition 6.** There exist $t \geq t$, $c, z \geq 0$ and a sender-optimal information structure that induces a distribution $s$ over the receiver’s actions satisfying $s = S^{tcz}$.

A sender-optimal information structure inducing the distribution $s$ is given by $\pi(\sigma|\omega) = s(\omega)$, $\pi(\sigma_0|\omega) = 1 - s(\omega)$ for all $\omega \in [l, h]$.

**Proof of Proposition 6 and Theorem 1.** We will prove the following statement:

There exist $t \geq t$, $c, z \geq 0$ and a sender-optimal information structure that induces a distribution $s$ over the receiver’s actions satisfying $s = S^{tcz}$. A sender-optimal information structure inducing the distribution $s$ is given by $\pi(\sigma|\omega) = s(\omega)$, $\pi(\sigma_0|\omega) = 1 - s(\omega)$ for all $\omega \in [l, h]$. We have $\Omega_1^* = \Omega_1 \mu_{F^*}$-a.e. for some $\Omega_1 \subseteq \Omega(t)$, $c = \hat{c}(\Omega_1)$ for some $\Omega_1 \subseteq \Omega_1(t)$ and either $t = j$ for some $j \in \mathcal{J}(t)$ or $t = \hat{t}$ and $\hat{t} > j$ for all $j \in \mathcal{J}(t')$ for all $t'$. Moreover, $\Omega_1$ can be chosen such that $\Omega_1 = \Omega(t, z)$ for some $z \in [l, w^*)$.

If Assumption 1 is satisfied, then there exist unique $t \geq t$, $c \geq 0$ such that any sender-optimal information structure induces a distribution $s$ satisfying $s = S^{tcz} \mu_{F^*}$-a.e. Moreover, $\Omega_1^* = \Omega_1(t) \mu_{F^*}$-a.e., $c = c(t)$ and either $t = J(t)$ or $t = \hat{t}$ and $\hat{t} > J(t')$ for all $t'$.

Lemma 6 shows that there exists a sender-optimal signal with two realizations.

Lemma 7 shows that if an information structure $\pi$ with two realizations satisfies $\pi(w) \leq Z_\pi(w)$ for all $w \in [l, w^*)$ and satisfies $\pi(\sigma|w) = Z_\pi(w)$ for all $w$ in some $A \subseteq [l, w^*)$, then $\pi(\sigma|w) = \min \left\{ \frac{c}{w^* - \omega}, 1 \right\}$ for $\mu_{F^*}$-a.e. $w \in \Omega_1^* \cap A$ and $\pi(\sigma|w) = 1$ for $\mu_{F^*}$-a.e. $w \in [w^*, h]$, where $c$ is given by $c = \frac{\int_{[w^*, h]}^{w^*}(w-w)g(w)dw - \int_{[\hat{t}, w^*)}^{w^*}(w-w)g(w)dw}{\int_{[\hat{t}, w^*)}^{w^*}g(w)dw}$.
Lemma 8 shows that if Assumption 1 is satisfied and \( \pi \) is a sender-optimal information structure with two realizations, then \( A = \Omega^\pi_1 \mu_{F_1} \)-a.e., while if Assumption 1 fails and (for all parameter values) there exists an information structure satisfying \( \pi(\sigma|w') = Z_\pi(w') \mu_{F_1} \)-a.e. \( w' \in \Omega^r_1 \) and \( \pi(\sigma|w) = 1 \mu_{F_1} \)-a.e. \( w \in [w^*, h] \), then there exists a sender-optimal information structure with two realizations satisfying \( A = \Omega^\pi_1 \mu_{F_1} \)-a.e.

The proof of lemma 12 shows that if \( \pi \) satisfies the condition in lemma 8, then there exists unique \( t \geq t_0 \) such that \( \pi(\sigma|w) = Z_\pi(w') \mu_{F_1} \)-a.e. \( w' \in \Omega^r_1 \) and \( \pi(\sigma|w) = 1 \mu_{F_1} \)-a.e. \( w \in [w^*, h] \), then there exists a sender-optimal information structure with two realizations satisfying \( A = \Omega^\pi_1 \mu_{F_1} \)-a.e.

Now suppose that Assumption 1 is satisfied. Observe that then any selection \( t \mapsto \Omega(t) \) from \( t \mapsto \Omega(t) \) satisfies \( \Omega(t) = \Omega_1(t) \) a.e. This is because \( \Omega(t) = \left\{ w : \frac{f_s(w)}{g(w)(w^*_s - w)} > t \right\} \cup \Omega_1 \), \( \Omega_1(t) = \left\{ w : \frac{f_s(w)}{g(w)(w^*_s - w)} > t \right\} \cup \Omega_2 \) for some \( \Omega_1 \), \( \Omega_2 \subseteq \left\{ w : \frac{f_s(w)}{g(w)(w^*_s - w)} = t \right\} \), and Assumption 1 implies that \( \mu(\Omega_1') = \mu(\Omega_2') = 0 \). Then the correspondence \( t \mapsto \Omega(t) \) is a function equal to \( t \mapsto J(t) \). Note that \( c(t) = \tilde{c}(\Omega_1(t)) \). Then the result in the paragraph above implies that if \( \pi \) satisfies the condition in lemma 8, then there exists a sender-optimal information structure \( \pi \) with two realizations such that there is a threshold \( t \geq t_0 \) satisfying \( \Omega^\pi_1 = \Omega_1(t) \mu_{F_1} \)-a.e., \( c = c(t) \) and either \( t = J(t) \) or \( t = t_0 \) and \( t > J(t') \) for all \( t' \). By lemma 12 if Assumption 1 is satisfied, then the threshold \( t \geq t_0 \) is unique. Observe that, because Assumption 1 is satisfied, by lemma 8 any sender-optimal information structure \( \pi \) with two realizations satisfies the condition in lemma 8. Then there exist unique \( t \geq t_0 \), \( c \geq 0 \) such that any optimal sender-optimal information structure induces a distribution \( s = S^{tez} \mu_{F_1} \)-a.e., \( \Omega^\pi_1 = \Omega_1(t) \mu_{F_1} \)-a.e., \( c = c(t) \) and either \( t = J(t) \) or \( t = t_0 \) and \( t > J(t') \) for all \( t' \), as required.

We next show that, provided that Assumption 1 is satisfied, if a sender-optimal information structure \( \pi_0 \) induces a distribution \( s_0 \) over the receiver’s actions, where the probability of action 1 in state \( w \) is given by \( s_0(w) = \pi(\sigma|w) \mu_{F_1} \)-a.e. \( w \in [l, h] \).

Lemma 6 shows that for any sender-optimal information structure \( \pi_0 \) the sender’s payoff from \( \pi_0 \) is \( \int_{[l, h]} \left( \sum_{i \in R(\tilde{f}, \pi_0)} \pi_0(\sigma_i|w) \right) f_s(w) dw \) and that there exists a sender-optimal information structure \( \pi_1 \) with two realizations satisfying \( \pi_1(\sigma|w) = \sum_{i \in R(\tilde{f}, \pi_0)} \pi_0(\sigma_i|w) \) for all \( w \in [l, h] \).
Note that, because the sender’s payoff from $\pi_0$ is
$$\int_{[l,h]} \left( \sum_{i \in R(\tilde{F},\pi_0)} \pi_0(\sigma_i | w) \right) f_s(w) dw,$$
we have $s_0(w) = \sum_{i \in R(\tilde{F},\pi_0)} \pi_0(\sigma_i | w)$. Then the fact that $\pi_1(\sigma | w) = \sum_{i \in R(\tilde{F},\pi_0)} \pi_0(\sigma_i | w)$ for all $w \in [l,h]$ implies that $s_0(w) = \pi_1(\sigma | w)$ for all $w \in [l,h]$.

The fact that, provided that Assumption $[1]$ is satisfied, the optimal information structure with two realizations in unique $\mu_{F_z}$-a.e. by the arguments above implies that $\pi_1(\sigma | w) = \pi(\sigma | w)$ $\mu_{F_z}$-a.e. $w \in [l,h]$. Therefore, $s_0(w) = \pi(\sigma | w)$ $\mu_{F_z}$-a.e. $w \in [l,h]$, as required.

**Proof of Proposition $[3]$**. Suppose that $\mu_G([l,w^*]) = 0$. Then $g(w) = 0$ a.e. $w \in [l,w^*)$, so that $\frac{f_s(w)}{g(w)(w^*-w)} = \infty$ a.e. $w \in [l,w^*)$. This implies that
$$\{w \in [l,w^*) : \frac{f_s(w)}{g(w)(w^*-w)} > t\} = [l,w^*)$$
for all $t < \infty$. Thus, because $f_s$ has full support, the proof of Proposition $[5]$ and Theorem $[1]$ implies that an optimal information structure induces the probability of action 1 in state $w$ given by $s(w) = 1$ a.e. $w \in [w^*,h]$ and $s(w) = \min \{\frac{c}{w^*-w}, 1\}$ a.e. $w \in [l,w^*)$, where $c$ satisfies $c = \frac{\int_{[w^*,h]} g(w)(w^*-w) dw - \int_{[w^*,w^*-c]} g(w)(w^*-w)dw}{\frac{1}{1-a} + \int_{[l,w^*-c]} g(w) dw}$. Observe that $s$ is independent of $F_s$, as required.

Next suppose that a sender-optimal information structure with two realizations induces a distribution over the receiver’s actions that is independent of $F_s$. Suppose for the sake of contradiction that $\mu_G([l,w^*)) > 0$. Let $A \subseteq [l,w^*)$ denote the set satisfying $\mu_G(A) = \mu_G([l,w^*))$ (note that $A$ is unique up to sets of measure zero).

For simplicity, I will provide a proof allowing for discontinuous density functions $f_s$. Fix $\eta > 1$, $A_1 \subseteq A$ such that $\mu(A_1) = \frac{1}{\eta}$ and the sender’s prior $F_s^\eta$ for $\eta \in (0,\eta]$ such that $f_s^\eta(w) = \eta$ for $w \in A_1$ and $f_s^\eta(w) = \epsilon_\eta$ for $w \in [l,h] \setminus A_1$ where $\epsilon_\eta$ is given by $\epsilon_\eta = \frac{\eta - \eta}{\frac{\eta}{\eta - 1}}$ (which ensures that $f_s^\eta$ is a probability density function). Let $A_2 = A \setminus A_1$.

Let $\pi^\eta$ denote an information structure that is sender-optimal when the sender’s prior is $F_s^\eta$ and satisfies the condition in lemma $[8]$ and let $\Omega^\eta = \Omega^\pi_{\eta} = \{w \in [l,w^*) : \pi^\eta(\sigma | w) > 0\}$. Let $A_\pi = \{w \in [l,w^*) : g(w) = 0\}$. Observe that, because for all $w \in A_\pi$ we have $g(w) = 0$ and $f_s^\pi(w) = 0$, adding a subset of $A_\pi$ to the support of the signal realization recommending action 1 under $\pi^\pi$ does not change the sender’s payoff. Then, without loss of generality, we may assume that $A_\pi \subseteq \Omega^\pi$.

Note that, because $f_s^\eta(w) = 0$ for $w \in [l,h] \setminus A_1$ and $\mu_G(A') > 0$ for all $A' \subseteq A$ such that $\mu(A') > 0$, we must have $\pi^\eta(\sigma | w) = 0$ a.e. $w \in A \setminus A_1$. Let $A_0 = \Omega^\pi \cap A_1$. It can be shown
that $\mu(A_0) > 0$. Then, because $A_0 \subseteq \Omega^\pi$ by the argument above, we have $\Omega^\pi = A_0 \cup A_0$.

We claim that there exists $\tilde{\varepsilon} > 0$ such that for all $\varepsilon \in (0, \tilde{\varepsilon})$, there is a set $A_3^\varepsilon \subseteq A \setminus A_1 = A_2$ satisfying $\mu(A_3^\varepsilon) > 0$ and $\mu(A_3^\varepsilon \cap \Omega^\pi) = 0$. Suppose for the sake of contradiction that for all $\tilde{\varepsilon} > 0$ there exists $\varepsilon \in (0, \tilde{\varepsilon})$ such that if $\varepsilon = \varepsilon$, then $\mu(A_2 \setminus \Omega^\pi) = 0$. Fix $\tilde{\varepsilon} > 0$ and $\varepsilon \in (0, \tilde{\varepsilon})$ such that if $\varepsilon = \varepsilon$, then $\mu(A_2 \setminus \Omega^\pi) = 0$. Fix $\eta \in (0, \tilde{\eta})$ such that $\varepsilon = \varepsilon$. We will show that the sender’s payoff given $F_\pi^\eta$ is strictly higher under the information structure $\pi^\varepsilon$ than under $\pi^\eta$, which would lead to a contradiction.

Because $\pi$ satisfies the condition in lemma $\mathbf{8}$ the proof of Proposition $\mathbf{9}$ and Theorem $\mathbf{1}$ implies that $\pi^\eta(\sigma|w) = \frac{\pi^\varepsilon(\lambda|w)}{w - w^*}$ for $w \in \Omega^\pi \cap [l, w^* - \hat{c}(\Omega^\eta)]$. Note that, because $\mu(A_2 \setminus \Omega^\eta) = 0$, we have $A_0 \cup A_2 \cup A_2 \subseteq \Omega^\eta$ and $\hat{c}(\Omega^\eta) \leq \hat{c}(A_0 \cup A_2 \cup A_2) = c$. Observe that $c$ is the constant of proportionality for the information structure resulting from adding $A_2$ to the support of the signal realization recommending action $1$ under $\pi^\varepsilon$. $\hat{c}(\Omega^\eta) \leq c$ then follows from the fact that $\Omega^\pi = A_0 \cup A_0$ and $A_0 \cup A_0 \cup A_2 \subseteq \Omega^\eta$.

Thus, because $\Omega^\pi = A_0 \cup A_0$, we have $\Omega^\pi \cup A_2 = A_0 \cup A_0 \cup A_2 \subseteq \Omega^\eta$, so that $\Omega^\eta$ is obtained by adding $A_2$ and possibly other sets to the support of the signal realization recommending action $1$ under $\pi^\varepsilon$. The gain in the sender’s payoff from adding $A_2$ and possibly other sets to the support is at most $\int_{l, w^* - \hat{c}(\Omega^\eta)} f_\eta^\pi(w) dw = \varepsilon \mu([l, w^* - \hat{c}(\Omega^\eta)] \setminus A_0)$. Observe that the loss in the sender’s payoff from adding $A_2$ and possibly other sets to the support is greater than the loss in the sender’s payoff from adding only $A_2$ to the support. The loss in the sender’s payoff from adding $A_2$ to the support is at least $\Delta U = \int_{l, w^* - \hat{c}(\Omega^\eta)} f_\eta^\pi(w) dw = \frac{\pi^\varepsilon(\lambda|w)}{w - w^*} \eta \mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0) = \Delta U_0$.

We claim that for $\varepsilon \eta \in (0, \tilde{\eta})$ where $\tilde{\eta} = \frac{\Delta U_0}{\mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0)}$ we have $\varepsilon < \frac{\pi^\varepsilon(\lambda|w)}{w - w^*} \mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0)$. Note that the definition of $\Omega \mapsto \hat{c}(\Omega)$ and the fact that $\mu_G(A_2) > 0$ imply that $\hat{c}(\Omega^\eta) > 0$. It can be shown that $\mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0) > 0$. Observe also that $\mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0)$ does not depend on $\eta$. Then the claim follows from the observation that $\eta \mapsto \frac{c - \hat{c}(\Omega^\eta)}{w - l} \eta \mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0)$ is strictly increasing, while $\eta \mapsto \varepsilon \eta$ is strictly decreasing, that $\eta \mapsto \frac{c - \hat{c}(\Omega^\eta)}{w - l} \eta \mu([l, w^* - \hat{c}(\Omega^\eta)] \cap A_0)$ and $\eta \mapsto \varepsilon \eta$ are continuous, and that $\varepsilon \eta < \frac{\Delta U_0}{\mu([l, w^* - w^] \cap A_0)}$ holds at $\eta = \tilde{\eta}$ because $\varepsilon \eta = 0$ and $\tilde{\eta} > 0$.

Then the gain from adding $A_2$ and possibly other sets to the support is strictly less than the loss. Therefore, the sender’s payoff given $F_\pi^\eta$ is strictly higher under $\pi^\varepsilon$ than under $\pi^\eta$, which is a contradiction. Thus there exists $\tilde{\varepsilon} > 0$ such that for all $\varepsilon \eta \in (0, \tilde{\varepsilon})$, there is a set $A_3^\varepsilon \subseteq A \setminus A_1 = A_2$ satisfying $\mu(A_3^\varepsilon) > 0$ and $\mu(A_3^\varepsilon \cap \Omega^\eta) = 0$.

Fix $A_1^\varepsilon \subseteq A_3^\varepsilon$ such that $\mu(A_1^\varepsilon) > 0$ and consider the sender’s prior $F_\eta^\varepsilon$ with full support
constructed as above with \( A'_1 \) instead of \( A_1 \). Then the above argument implies that there exists \( A'_0 \subseteq A'_1 \) such that \( \mu(A'_0) > 0 \) and under the sender’s prior \( F_s^{\eta'} \) we have \( \mu(A'_0 \cap \Omega^{\eta'}) > 0 \) for some \( \eta' \in (0,\bar{\eta}) \). Because \( A'_0 \subseteq A_3^{\eta} \) and \( \mu(A_3^{\eta} \cap \Omega^{\eta'}) = 0 \), this contradicts the hypothesis that the distribution over the receiver’s actions induced by a sender-optimal information structure is independent of the sender’s prior.

\[\square\]

**Proof of Proposition 4.** Let us write the fixed point equation pinning down the values of the threshold \( t > t' \) as \( t^* = J(t^*) \). Observe that \( \lim_{a \to 1} J(t) = 0 \). Then \( \lim_{a \to 1} t^* = 0 \).

The fact that \( g \) is \( C^1 \) and \([l,h]\) is a compact set implies that \( g \) attains a maximum \( M \) on \([l,h]\), which implies that \( g(w) \leq M \) for all \( w \in [l,w^*] \). It follows that if there exists \( m > 0 \) such that \( f_s(w) \geq m \) for all \( w \in [l,w^*] \), then \( t = \min_{w \in [l,w^*]} \frac{f_s(w)}{g(w)(w^*-w)} > 0 \).

Thus, because \( \lim_{a \to 1} t^* = 0 \), there exists \( \bar{a} \in (0,1) \) such that for all \( a \in (\bar{a},1) \) we have \( t^* < t' \). Therefore, for all \( a \in (\bar{a},1) \) the threshold is \( t = t' \), so that \( \Omega_1(t) = [l,w^*) \), as required.

\[\square\]

**Proof of Proposition 1.** Let us write \( y_a(t) = t \left( \frac{a}{1-a} + \int_{\Omega_1(t) \cap [l,w^*-c(t)]} g(w)dw \right) - \int_{\Omega_1(t) \cap [l,w^*-c(t)]} \frac{f_s(w)}{w^*-w}dw \).

**Claim 4.1.** \( y_a'(t) - y_a(t) < 0 \) for all \( t \geq t' \).

**Proof of claim 4.1.** \( y_a'(t) - y_a(t) = t \left( \frac{a'}{1-a'} - \frac{a}{1-a} \right) < 0 \) because \( t > 0 \) and \( a' < a \).

**Claim 4.2.** \( t_a' - t_a > 0 \).

**Proof of claim 4.2.** Because \( \mu(\Omega_1(t_a)) < \mu([l,w^*)) \), we have \( t_a > t' \). Then, because \( t = J(t) \) is equivalent to \( y(t) = 0 \) and, by lemma 12, we have \( y(t+\eta) > y(t) \) for all \( t \geq t' \) and for all \( \eta \), to show that \( t_a' - t_a > 0 \) it is enough to show that \( y_a'(t) - y_a(t) < 0 \) for all \( t \geq t' \). Claim 4.1 implies that this is satisfied.

The fact that \( t_a' - t_a > 0 \) by claim 4.2 implies that \( \Omega_1(t_a') \subseteq \Omega_1(t_a) \) \( \mu_{F_s} \)-a.e. Finally, the fact that \( c(t_a') > c(t_a) \) follows from the fact that, provided that Assumption 11 is satisfied, \( t \mapsto c(t) \) is strictly increasing by lemma 10.

\[\square\]

**Proof of Proposition 2.** We will prove the following result.

Suppose that the receiver’s prior is \( F \in C_{a',g} \subseteq C_{a,g} \) for \( 0 < a' < a < 1 \). Then

1. If \( V(a') = 0 \), then \( V(a) > 0 \).
2. Suppose that the optimal signal recommends approval in all states with a strictly positive probability under \( a \) and \( a' \). Then \( V(a) \geq V(a') \). Moreover, if the sender’s knowledge is detail-free, then \( V(a) = V(a') \), and if the reference prior \( g \) has full support, then \( V(a) > V(a') \).

3. There exist \( g, F_s \) such that \( 0 < V(a) < V(a') \).

Let \( d \in \{0,1\} \) denote the actions available to the receiver. Let us write the receiver’s utility function as \( u(d, w) = 1_{d=1}(w - w^*) \). Let \( \pi_a \) denote the sender-optimal signal with two realizations when the receiver’s set of priors is \( C_{a,g} \). Then the expected payoff of a receiver with belief \( F \) is \( V(a) = \int_{[l,h]} \pi_a(\sigma|w)(w-w^*)dF(w) \).

Observe that \( V(a) = \int_{[l,h]} \pi_a(\sigma|w)wdF(w) - w^* \int_{[l,h]} \pi_a(\sigma|w)dF(w) = \int_{[l,h]} \pi_a(\sigma|w)dF(w) \left( \int_{[l,h]} \pi_a(\sigma|w)wdF(w) - w^* \right) = \int_{[l,h]} \pi_a(\sigma|w)dF(w)(E_{F,\pi_a}|w|\sigma - w^*) \), so that \( V(a) = \int_{[l,h]} \pi_a(\sigma|w)dF(w)(E_{F,\pi_a}|w|\sigma - w^*) \). This implies that \( V(a') = 0 \) if and only if \( E_{F,\pi_a}|w|\sigma - w^* = 0 \). Then, because \( \int_{[l,h]} \pi_a(\sigma|w)dF(\omega) > 0 \), to prove the first part of the Proposition it is enough to show that if \( E_{F,\pi_a}|w|\sigma = w^* \), then \( E_{F,\pi_a}|w|\sigma > w^* \).

Observe that \( \mu_F = \int_{[l,h]} \tau_\omega(a') \delta_\omega + (1 - a') \mu_G d\omega \) for some \( \{\tau_\omega\}_r \) such that \( \int_{[l,h]} \tau_\omega d\omega = 1 \). Then \( E_{F,\pi_a}|w|\sigma = \int_{[l,h]} \tau_\omega(a' \omega \pi_a |\omega\| + (1 - a') \omega \pi_a |\omega\|)d\omega \). Let \( \mu_F = a' \delta_{w^*} + (1 - a') \mu_G \) and note that \( E_{F_w,\pi_a}|w|\sigma = \frac{a' \omega \pi_a |\omega\| + (1 - a') \omega \pi_a |\omega\| g(w)dw}{a' \omega \pi_a |\omega\| + (1 - a') \omega \pi_a |\omega\| g(w)dw} \).

Because \( E_{F_w,\pi_a}|w|\sigma = w^* \) for some \( \omega \in [l, h] \) and \( a > a' \), we have \( E_{F_w,\pi_a}|w|\sigma > w^* \) for all \( \omega \in [l, h] \). Then the above results imply that \( E_{F,\pi_a}|w|\sigma > w^* \), as required.

Observe that \( V(a) = \int_{[w^*, l]} (w - w^*)dF(w) - \int_{[l, w^*]} (w - w^*) \pi_a(\sigma|w)dF(w) \). Then \( V(a) - V(a') = \int_{[l, w^*]} (w^* - w) \pi_{a'}(\sigma|w) - \pi_a(\sigma|w)dF(w) \).

We next consider the case in which \( \Omega_1^{\pi_a} = \Omega_1^{\pi_{a'}} = [l, w^*] \). Observe that \( \pi_a(\sigma|w) = \min\{\frac{a \omega}{w^* - w}, 1\} \) for some constant \( c_a \) if \( w \in [l, w^*] \) and \( \pi_a(\sigma|w) = 1 \) if \( w \in [w^*, h] \). Moreover, \( c_a < c_{a'} \). Then we have \( \pi_a(\sigma|w) < \pi_{a'}(\sigma|w) \) for \( w \in [l, w^* - c_a] \) and \( \pi_a(\sigma|w) \leq \pi_{a'}(\sigma|w) \) for \( w \in [w^* - c_a, w^*] \). Then \( V(a) - V(a') \geq 0 \). Moreover, \( V(a) - V(a') > 0 \) if \( \mu_F([l, w^* - c_a]) > 0 \) and \( V(a) - V(a') = 0 \) if \( \mu_F([l, w^* - c_a]) = 0 \). Therefore, if \( \mu_F([l, w^*]) = 0 \), then \( V(a) = V(a') \), while if \( g \) has full support, then \( V(a) > V(a') \), as required.

We now provide an example of the parameters under which \( V(a) < V(a') \). For simplicity, suppose that \( \mu_G = (1 - \kappa)\mu_G_0 + \kappa \delta_h \) for some \( \kappa \in (0,1) \) and \( \mu_G_0 \) that is uniform on \( [w_0, w_1] \) for some \( l < w_0 < w_1 < w^* \).
Claim 6.1. There exist parameters such that $\Omega_1^{\pi a} = [l, w^*)$ and $[w_0, w_1] \cap \Omega_1^{\pi a'} = \emptyset$.

Proof of claim 6.1 First observe that, because $\mu_G([l, w_0]) = 0$, we have $[l, w_0] \subseteq \Omega_1^{\pi a}$, $\Omega_1^{\pi a'}$. Fix $a' \in (0, 1)$. Consider $F_*$ such that $f_*(w) = \eta_1$ for $w \in [w_0, w_1]$ and $f_*(w) = \eta_0$ for $w \in [l, h] \setminus [w_0, w_1]$. Then for $\eta_1$ sufficiently small we have $[w_0, w_1] \cap \Omega_1^{\pi a'} = \emptyset$. This is because 

$$\lim_{\eta_1 \to 0} \frac{f_*(w)}{g(w)} \cdot (w^* - w) = 0$$

for $w \in [w_0, w_1]$ but 

$$\lim_{\eta_1 \to 0} \int_{[l, w^*) \cap [l, w^* - c(t))] \frac{f_*(w)}{g(w)} dw > 0$$

because $\mu(\Omega_1(t) \cap [l, w_0)) > 0$ and $f_*(w) \geq \eta_0$ for all $w \in [l, w_0]$.

Observe next that $l < w^* - c(t)$ for $t$ such that $\Omega_1(t) = \Omega_1^{\pi a}$ for $a > a'$ sufficiently large. This is because 

$$\lim_{a \to 1} \frac{f_*(w)}{g(w)} = \frac{1}{\pi} \int_{[l, w^* - c(t))] \frac{f_*(w)}{g(w)} dw = 0$$

and $l < w^*$.

The fact that $l < w^* - c(t)$ for $t$ such that $\Omega_1(t) = \Omega_1^{\pi a}$ for $a > a'$ sufficiently large and that $[l, w_0] \subseteq \Omega_1(t)$ implies that $\mu(\Omega_1(t) \cap [l, w^* - c(t))] > 0$ for $a > a'$ sufficiently large. Because 

$$\lim_{a \to 1} \frac{a}{1-a} = \infty,$$

the fact that $\mu(\Omega_1(t) \cap [l, w^* - c(t))] > 0$ for $a > a'$ sufficiently large implies that 

$$\lim_{a \to 1} t = \lim_{a \to 1} \frac{f_*(w)}{g(w)(w^* - w)} = \frac{\eta_1}{g(w)(w^* - w)} > 0$$

for $w \in [w_0, w_1]$, this implies that as $a \to 1$, $\Omega_1^{\pi a} \to [l, w^*)$.

Consider $\mu_F = (1 - a)(1 - \kappa)\mu_G + (1 - a)\kappa\delta_h + ah$. Because $V(a) = \int_{[w^*, h]} (w - w^*) F(w) - \int_{[l, w^*)} (w^* - w) \pi_0 (\sigma|w) dF(w)$, we have $V(a') = \int_{[w^*, h]} (w - w^*) F(w)$ because $[w_0, w_1] \cap \Omega_1^{\pi a'} = \emptyset$ and $V(a) = \int_{[w^*, h]} (w - w^*) F(w) - \int_{[l, w^*)} (w^* - w) \pi_0 (\sigma|w) dF(w)$. Note that $\int_{[l, w^*)} (w^* - w) \pi_0 (\sigma|w) dF(w) > 0$ because $\Omega_1^{\pi a} = [l, w^*)$. This implies that $V(a') > V(a)$, as required.

Proof of Proposition 5. We first characterize the optimal information structure when $a = 0$. We will show that in this case the optimal information structure is given by 

$$\pi(\sigma|w) = 1 \text{ for } w \in \Omega_1^* \cup [w^*, h], \pi(\sigma|w) = 0 \text{ for } w \in [l, w^*) \setminus \Omega_1^*,$$

where $\Omega_1^* = \Omega_1(t)$ for some $t \geq 0$ such that $E_{G, \pi}[\omega|\sigma] = w^*$.

Because $a = 0$, the revelation principle applies. Thus, because the receiver has two actions, there exists a sender-optimal information structure with two realizations, $\sigma$ and $\sigma_0$. Then the sender’s problem is 

$$\max_{\pi(\sigma|w) \in [0, 1]} \int_{[l, h]} f_s(w) \pi(\sigma|w) dw$$

subject to the constraint that $E_{G, \pi}[\omega|\sigma] \geq w^*$. The constraint that $E_{G, \pi}[\omega|\sigma] \geq w^*$ is equivalent to 

$$\int_{[l, h]} w g(w) \pi(\sigma|w) dw \geq w^*.$$
We can write the Lagrangian for the constrained optimization problem as
\[ \int_{[l,h]} f_s(w)\pi(\sigma|w)dw - \lambda \int_{[l,h]}(w^* - w)g(w)\pi(\sigma|w)dw. \]
Equivalently, \( \int_{[l,h]}(f_s(w) - \lambda(w^* - w))g(w)\pi(\sigma|w)dw. \)

Then the solution is \( \pi(\sigma|w) = 1 \) if \( f_s(w) \geq \lambda(w^* - w) \) and \( \pi(\sigma|w) = 0 \) if \( f_s(w) < \lambda(w^* - w) \) where \( \lambda > 0 \) is such that \( E_{G,\pi}[\omega|\sigma] = w^* \). This implies that \( \pi(\sigma|w) = 1 \) for all \( w \in [w^*,h] \). Moreover, for \( w \in [l,w^*) \), we have \( \pi(\sigma|w) = 1 \) if \( \frac{f_s(w)}{g(w)(w^*-w)} \geq \lambda \) and \( \pi(\sigma|w) = 0 \) if \( \frac{f_s(w)}{g(w)(w^*-w)} \leq \lambda \) where \( \lambda \) is such that \( E_{G,\pi}[\omega|\sigma] = w^* \).

Next, we prove the second part of the Proposition. Let \( \pi_a \) denote the sender-optimal signal with two realizations when the set of priors is \( C_{a,g} \). Without loss of generality, suppose that the limits \( \lim_{a \to 0} \int_{l}^{h} |\pi_a(\sigma|w) - \pi(\sigma|w)|dF_s \) and \( \tilde{\pi} = \lim_{a \to 0} \pi_a(\sigma|w) \) exist. Suppose for the sake of contradiction that \( \lim_{a \to 0} \int_{l}^{h} |\pi_a(\sigma|w) - \pi(\sigma|w)|dF_s \neq 0 \).

Note that the sender’s payoff from \( \pi_a \) is \( U_s(\pi_a) = \int_{l}^{h} \pi_a(\sigma|w)dF_s(w) \). Suppose first that \( \lim_{a \to 0} U_s(\pi_a) = U_s(\pi) \). Note that we must have \( E_{G,\pi_a}[w|\sigma] \geq w^* \) for all \( a > 0 \). This implies that \( \lim_{a \to 0} E_{G,\pi_a}[w|\sigma] = E_{G,\tilde{\pi}}[w|\sigma] \geq w^* \), so under the limit signal \( \tilde{\pi} \) the receiver with prior \( G \) takes action 1 after the signal realization \( \sigma \). Then \( \tilde{\pi} \) is a signal with two realizations such that \( \int_{l}^{h} |\pi_a(\sigma|w) - \tilde{\pi}(\sigma|w)|dF_s \neq 0 \) and \( U_s(\tilde{\pi}) = U_s(\pi) \) given that the receiver’s prior is \( G \). This contradicts the fact that, because of Assumption \[ \exists \pi \] is unique up to sets of measure zero under \( F_s \).

Next, suppose that \( \lim_{a \to 0} U_s(\pi_a) \neq U_s(\pi) \). Because \( U_s(\pi_a) < U_s(\pi) \) for all \( a > 0 \), this implies that \( \lim_{a \to 0} U_s(\pi_a) < U_s(\pi) \). We will show that there exists a sequence of signals \( \{\pi_a^0\} \) with two realizations such that \( E_{F,\pi_a^0}[w|\sigma] \geq w^* \) for all \( F \in C_{a,g} \) and \( \lim_{a \to 0} U_s(\pi_a^0) = U_s(\pi) \). Because \( \lim_{a \to 0} U_s(\pi_a^0) < U_s(\pi) \), this would establish that \( \{\pi_a\}_{a \in (0,1)} \) was not a collection of sender-optimal signals, which is a contradiction.

Let \( \pi_a^0(\sigma|w) = \pi(\sigma|w) - \epsilon_a \) if \( \pi(\sigma|w) = 1 \) and \( w < w^* \), \( \pi_a^0(\sigma|w) = \pi(\sigma|w) = 1 \) if \( w \geq w^* \) and \( \pi_a^0(\sigma|w) = 0 \) if \( \pi(\sigma|w) = 0 \). Choose the minimal \( \epsilon_a > 0 \) such that \( E_{F,\pi_a^0}[w|\sigma] \geq w^* \) for all \( F \in C_{a,g} \). Note that this is feasible because if \( \epsilon_a = 1 \), then \( E_{F,\pi_a^0}[w|\sigma] > w^* \) for all \( F \in C_{a,g} \). Observe that \( \lim_{a \to 0} \epsilon_a = 0 \). This implies that \( \lim_{a \to 0} U_s(\pi_a^0) = U_s(\pi) \), as required.

\[ \blacksquare \]

References


